Abstract

- **Author’s Name:** Kyle Zerner
- **Job Title:** Senior Engineer, Systems Design Engineering
- **Phone:**
- **Email:**
- **Organization**: SanDisk Flash Products Group

- **Author Bio:** Kyle Zerner joined Western Digital/SanDisk in 2022 after graduating from the University of Maryland. He focuses on systems validation, prototyping, and scripting for interface characterization, optimizing system-level designs and ensuring robust validation processes to advance storage technology.
- **Title**: Adaptive PCIe Link Speed Management Power Saving benchmarking for Gen5 client storage devices
- **Target Audience**:
  - Gen5 client customers and ecosystem partners

As PCIe PHY technology advances to support increasing bandwidth demands, the importance of power efficiency has become paramount for SSDs operating at PCIe Gen5 speeds. The challenge no longer lies solely in NAND efficiency; SSDs must now also contend with the high power requirements necessary to sustain high PCIe link speeds, even when workloads do not fully exploit the available bandwidth. Many real-world storage scenarios exhibit dynamic bandwidth needs and low queue depths, resulting in significant, yet unused, power-saving potential. To mitigate this inefficiency, adaptive PCIe link speed management emerges as an effective solution, dynamically regulating link speed according to real-time host demands and system power policies, thereby enhancing energy efficiency with minimal compromise to performance.

This presentation will provide an in-depth exploration of the mechanisms that enable adaptive PCIe link speed management in Gen5 SSDs, detailing how real-time workload monitoring and link speed adaptation contribute to improved power consumption and enhanced power efficiency. This is aided by power policy hints to reflect host power savings policies.

Attendees will gain insights into the trade-offs between power efficiency and performance, the latency impact of link speed changes, and how simple speed change algorithms balance responsiveness with energy savings. Additionally, the session will explore real-world workload profiling, demonstrating scenarios where this feature yields tangible benefits in power-constrained environments.

SANDISK™

FMS the Future of Memory and Storage

the Future of Memory and Storage

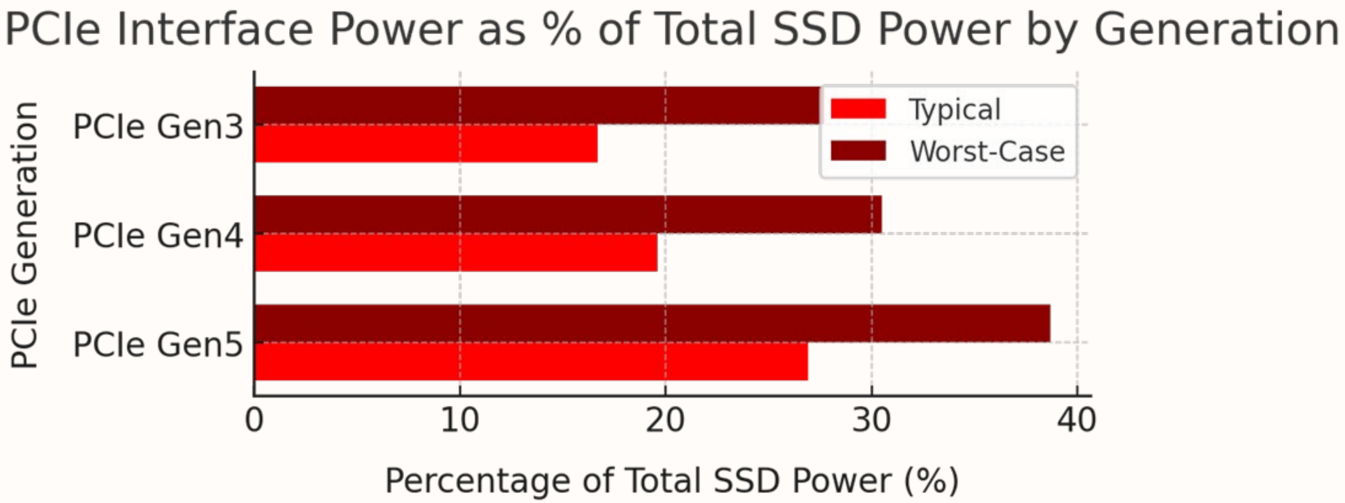
Adaptive Link Speed Management Benchmarking

Presenter: Kyle Zerner
Contributors: Ritesh Deshmukh

SANDISK™

Background

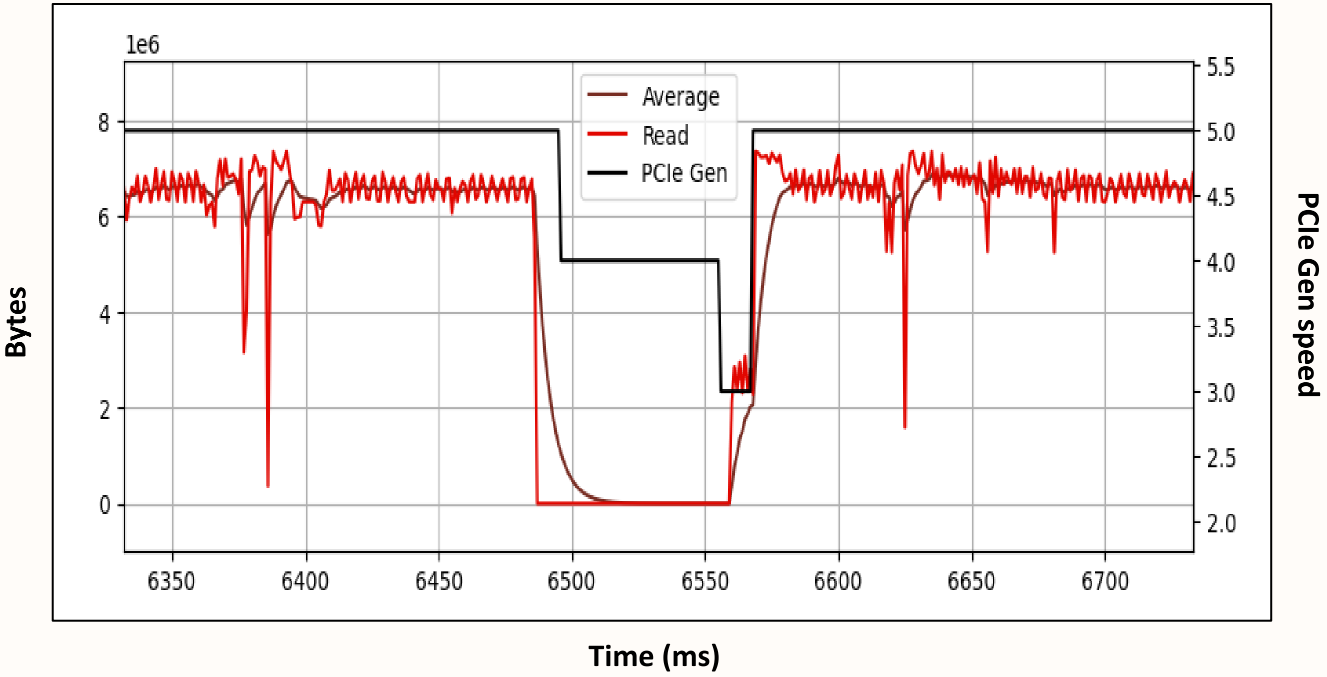
- PCIe Gen5 Link Power Consumption accounts for 20%~50% of the SSD power consumption, depending on thermal state and performance.
- PCIe Gen5 is overkill for all but data-intensive operations, wasting power
  - Running a PCIe PHY at Gen5 speeds takes ~1.5W on average, ~3W peak



Overview of Feature

- Designed for Gen5 products
- Changes link speed according to average bandwidth
- Moving average calculated and stored at regular intervals
    - ○ Average calculation has tunable filter parameters
- Average compared against tunable speed change thresholds at longer intervals
    - ○ Thresholds scaled according to current queue depth

Averaging Filter Example



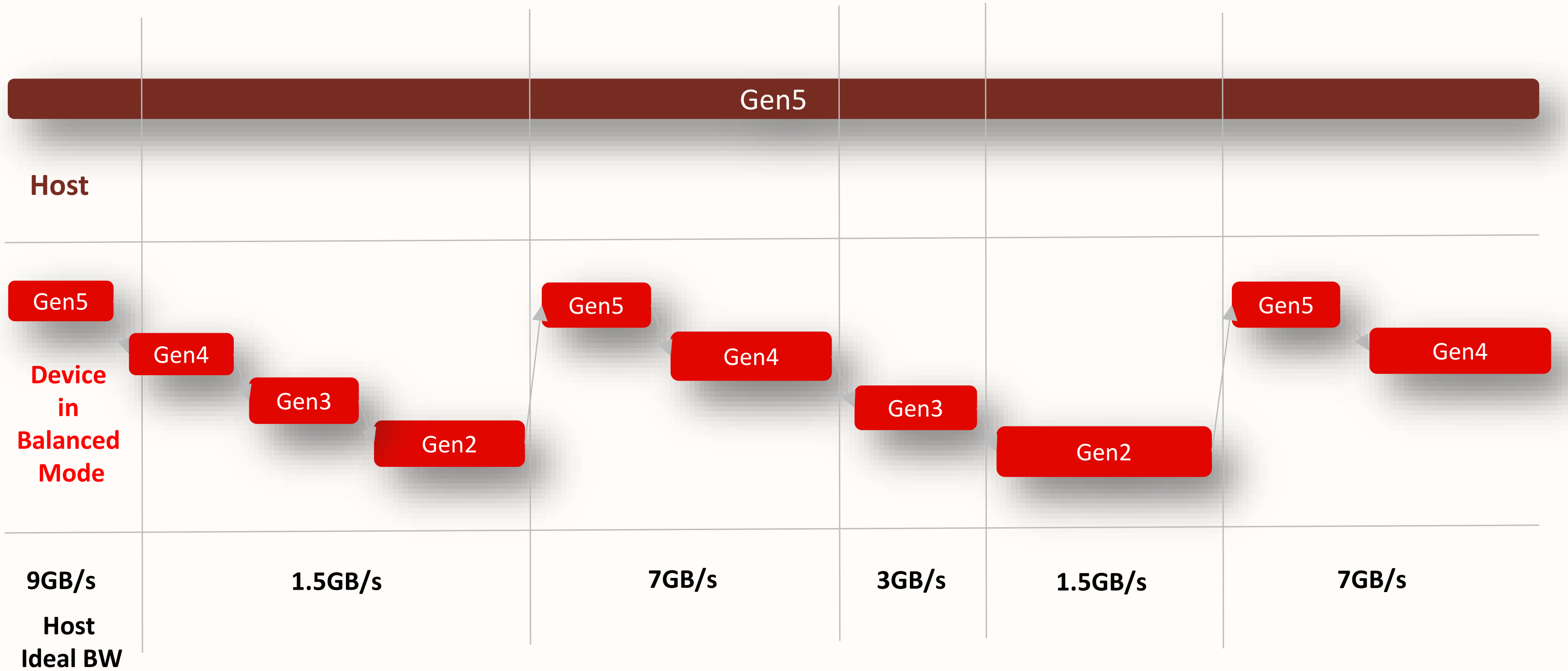
More Feature Details

- Respects host-limited link speed as maximum speed
  - Will never exceed last host-selected link speed
- Vendor specific NVMe feature w/ opportunity for standardization in the future
- Host can signal the desired link-speed management option to the device via NVMe admin command

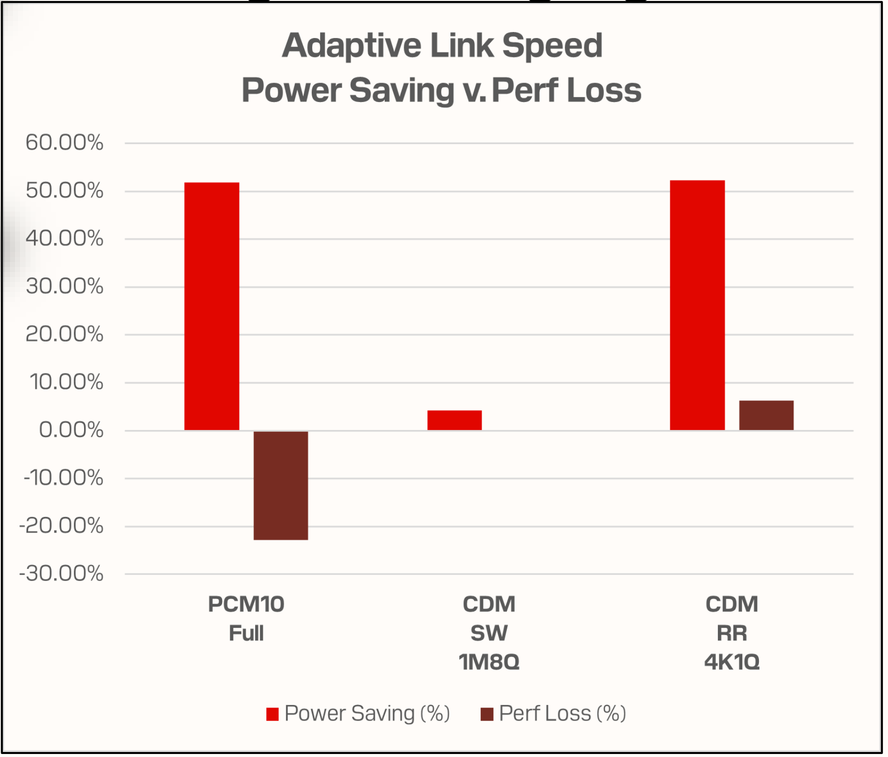
Modes of Operation:

Performance Mode (ALSM disabled)	Balanced Mode (ALSM enabled)
• **No adaptive link speed changes** • Module is effectively disabled • **No performance impact or extra power-saving**	• Link speed **always increases to maximum,** when required • Link speed **decreases one gen** at a time • Should have **minimized performance impact** w/ **reasonable power-saving** • Utilizes **early saturation detection** feature

Example of ALSM in action



Testing Result Highlights



- Observed ~50% power reduction in low perf WL with minimal perf loss in Gen5
- In dynamic WLs, observed ~20% score degradation vs ~50% power saving
- Little effect on sequential WLs, as the host traffic is steady

SANDISK™

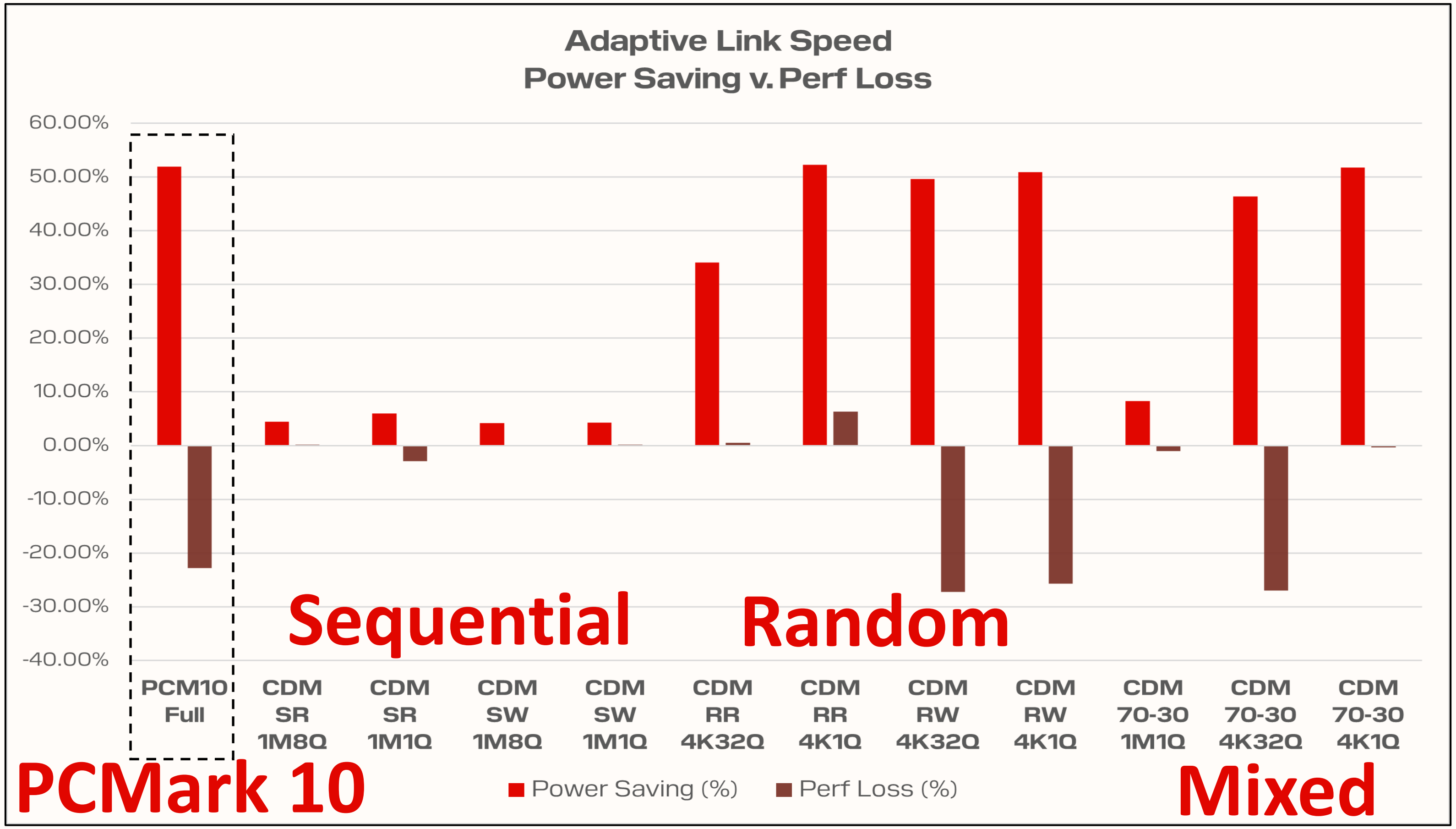
FMS the Future of Memory and Storage

Optimization Features

- **Queue-depth adjusted thresholds**: scale speed-change thresholds and saturation detection points depending on number of IO commands currently in queue
  - Lower queue-depth = lower thresholds
  - Higher queue-depth = higher thresholds

- **Early Saturation detection**: if bandwidth demand at or above saturation point for number of samples, raise speed without waiting for average
  - Saturation point also adjusted with queue-depth
  - Wait for tunable number of speed change periods before subsequent adjustments

Full Test Results



PCMark 10

Sequential

Random

Mixed

Conclusions

Significant power-savings opportunity with device-initiated PCIe link management

Most power-savings in request-limited, low performance workloads (random reads/writes)

Some performance loss in dynamic applications, but impressive power savings

Feature parameters allow for vendor-specific optimization and tuning

Thank you all for your time!

FMS the Future of Memory and Storage

SANDISK™