NVMe HDDs
The evolution of the HDD
NVMe HDD
The Evolution of the HDD

Jason Adrian, Sr. Director, Azure Platform Architecture, Microsoft
Mohamad El-Batal, Technologist, CTO Office at Seagate Technology
Reshaping the HDDs of Tomorrow

- Storage workstream has discussed NVMe™ HDDs for 3+ years
- NVMe™ HDD workstream was established - Q4 2020
- NVMe™ TP 4088 Rotational Media ratified – Q2 2021
- NVMe™ HDD specification will be contributed to OCP Q4 2021
- Incorporates feedback from:
  - NVMe™ HDD consumers
  - HDD vendors
  - PCIe switch vendors
  - Interconnect vendors
The Future - Simplified

Today
SAS/SATA HDD

Tomorrow
NVMe HDDs

OPEN POSSIBILITIES.
The Future - Simplified

- Eliminates proprietary SAS/SATA drivers
- Common NVMe™ interface for HDD and SSD
- Native support for multi-actuator
- Opens opportunities for NVMeOF™, computation storage applications, etc
Requirements Alignment

**Complete**
- Form Factor
- Power Requirements
- Single and dual port support
- Connectors
- Link Speeds
- SRIS support required (no clock)

**Ongoing**
- Feature requirements
- Security requirements
# Success Factors for NVMe HDD

<table>
<thead>
<tr>
<th>Factor</th>
<th>Key Metric</th>
</tr>
</thead>
</table>
| Technology Cost / Power Comparison  | • Does PCIe-based infrastructure for NVMe HDD have same or less cost than SAS-based infrastructure?  
                                         • Does PCIe-based infrastructure for NVMe HDD have same or less power than SAS-based infrastructure?  |
| Multi-Vendor Support                | • Are there 2 or more suppliers for key components and interconnects?       |
| Technology Readiness                | • Is the technology ready to be adopted or are there still technical challenges to be solved? |
The NVMe HDD is Born ...

- SOC Native NVMe port w/ Tri-Mode (SAS, SATA & NVMe) Transceivers
- Proven 3rd gen design leveraged from SSD SOC HW-IP block & FW
- EDUs available to key customers Sept-2022 with single port/lane
- CDUs will be available in Mid-2024 in Single and Dual-Port SKUs
EDU Demo in OCP Experience Center

- Phase-1 IOMs supports PCIe-Gen3
- Future Phase-2 IOMs Sept-2022 to support Gen4 & NVMe-oF x16 RNIC

Two SFF8644 Copper Cables
PCIe Switch/Redriver Host HBA
2U12-3.5” NVMe-HDD POC JBOD

OPEN POSSIBILITIES.
OCP Requirement-Based Design

- EDU units use SFF8639 conn to provide all PCIe-Gen4 connections
- CDU units will validate SRIS clockless PCIe-Gen4 capable Design
- CDUs to use OCP requirement-compliant connectors:
  - SATA Connector for Single-Port (Same HW SKU as SATA HDD)
  - SAS Connector for Dual-Port (Same HW SKU as SAS HDD)
**HDD SKU Reduction Focused**

- Significant manufacturing benefits w/ NVMe-HDDs using SAS or SATA connectors
- The same HW SKUs can be produced for
  - SATA Connector for (SATA & NVMe) Single-Port/Single-Actuator
  - SAS Connector for (SAS & NVMe) Dual-Port/Multi-Actuator
- FW determining the final interface type and test bench for both HW SKUs

<table>
<thead>
<tr>
<th>Market</th>
<th>Current SKUs</th>
<th>Near-Term Transition</th>
<th>Long-Term Convergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperscalers</td>
<td>SATA(Single-Port)</td>
<td>SATA(Single-Port) NVMe(Single-Port)</td>
<td>SATA/NVMe(Single-Port)</td>
</tr>
<tr>
<td>Both</td>
<td>SAS(Dual-Port)</td>
<td>SAS(Single-Port)</td>
<td>SAS/NVMe(Dual-Port)</td>
</tr>
<tr>
<td>Enterprise</td>
<td>SAS(Dual-Port)</td>
<td>SAS(Dual-Port) NVMe(Dual-Port)</td>
<td></td>
</tr>
</tbody>
</table>

**OPEN POSSIBILITIES.**

STORAGE
Optimizing Connectivity & TCO

• Best – Direct connect to CPU
• Better - Replacing SAS native devices with NVMe™ native devices.
  Example: PCIe Switch or Trimode IOCs

<table>
<thead>
<tr>
<th>Broadcom</th>
<th>Microchip</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCIe switches with full x1</td>
<td>IOCs with 32 x1’s (native NVMe)</td>
</tr>
<tr>
<td>bifurcation</td>
<td>Optional HW acceleration</td>
</tr>
<tr>
<td></td>
<td>Switches support x1</td>
</tr>
</tbody>
</table>

Example:
- Broadcom PCIe Switch or Trimode IOCs
- Microchip SmartIOC 2200
  16x32
- Microchip Switchtec™
  PFX Fanout PCIe Switches

STORAGE
What is needed from the industry?

**CPU/DPU/ASIC**
- Ensure x1 PCIe bifurcation

**PCIe Switches**
- Create cost and power optimized switches
- Gen 3 or 4 PCIe x1 downlinks to the NVMe HDD
- Gen 4 or 5 PCIe uplinks, x8 or x16 to the host
Call to Action

• Check out the NVMe™ HDD demo in the experience center
• NVMe™ HDD specification contribution is expected to be contributed Q4 2021
• Optimized PCIe switches and CPU’s can enable new architectures
• Bring ideas to the OCP Storage working group or NVMe™ HDD workstream

Storage Wiki: https://www.opencompute.org/wiki/Storage
NVMe HDD Wiki: https://www.opencompute.org/wiki/Storage/NVMeHDD
Thank you!