NVMe® Technology
Powering the Connected Universe

Amber Huffman
Fellow & Chief Technologist of IP Engineering Group, Intel Corporation
President, NVM Express, Inc.
Agenda

Fixing the Memory & Storage Hierarchy

Refactoring for the Next Decade of Growth

NVMe® Architecture Advancements

What’s Next: Computational Storage
Agenda

- Fixing the Memory & Storage Hierarchy
- Refactoring for the Next Decade of Growth
- NVMe® Architecture Advancements
- What’s Next: Computational Storage
Exponential advances in all levels of memory hierarchy are needed to match the ever increasing compute demand.

While compute has grown at exponential rate, memory performance only grew at linear rate.
Memory and Storage Hierarchy Gaps

- **Memory**
  - Compute Cache: 10s MB, ~1ns
  - In-Package Memory: 1s GB, ~10ns
  - Memory: 10s GB, <100ns
  - Secondary Storage: 100s GB, <1usec
  - Tertiary Storage: 1s TB, <10µsecs

- **Storage**
  - Capacity Gap
    - Memory: 100s GB, <1usec
    - Secondary Storage: 1s TB, <10µsecs
    - Tertiary Storage: 10s TB, <100µsecs
  - Storage Performance Gap
    - Memory: 10s GB, <100ns
    - Secondary Storage: 100s GB, <1usec
  - Cost-Performance Gap
    - Memory: 10s MB, ~1ns
    - Secondary Storage: 1s GB, ~10ns
Memory and Storage Hierarchy Gaps

For illustrative purposes only
Types of Memory

- **DRAM**: One DRAM memory cell = 1 bit
- **3D XPoint**: One 3D XPoint memory cell = 1 bit
- **Intel 3D NAND**: One 3D NAND memory cell = 1-4 bits
Types of Memory Compared

- **DRAM**: One DRAM memory cell = 1 bit
- **3D XPoint**: One 3D XPoint memory cell = 1 bit
- **Intel 3D NAND**: One 3D NAND memory cell = 1-4 bits
3D NAND Roadmap

- **32L TLC**
  - Gen 1
- **64L QLC**
  - Gen 2
- **96L QLC**
  - Gen 3
- **128L**
  - Industry Target

- 2016
- 2017
- 2019
- 2020
3D NAND Roadmap

- **32L TLC Gen 1**
  - Bit Density Increase: 133%

- **64L QLC Gen 2**
  - Bit Density Increase: ~50%

- **96L QLC Gen 3**
  - Bit Density Increase: ~50%

- **144L QLC Gen 4**
  - Bit Density Increase: ~50%

- **2016**
- **2017**
- **2019**
- **2020**
3D XPoint Memory Roadmap

1st Gen
2-Deck

2nd Gen
4-Deck

Multiple Millions of IOPS on 2nd Generation Intel® Optane™ SSD

*Target Production
Media Innovation Realized in Product

Innovation Powered by nvm EXPRESS

Fast
Simple
Scalable

Intel Optane™ SSD
Intel Optane™ Memory H10
Optane + 3D QLC NAND
Intel 3D NAND SSD
Memory and Storage Hierarchy Gaps

For illustrative purposes only

Courtesy 2020 Intel Architecture Day
Memory and Storage Hierarchy Gaps Solutions

For illustrative purposes only

Innovation Powered by

Courtesy 2020 Intel Architecture Day
NVMe® Technology Powers the Connected Universe

<table>
<thead>
<tr>
<th>Units (Ku)</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
<th>2020*</th>
<th>2021*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enterprise</td>
<td>364</td>
<td>749</td>
<td>1,069</td>
<td>2,045</td>
<td>4,067</td>
<td>5,554</td>
</tr>
<tr>
<td>Cloud</td>
<td>2,051</td>
<td>3,861</td>
<td>10,369</td>
<td>12,276</td>
<td>18,982</td>
<td>21,999</td>
</tr>
<tr>
<td>Client</td>
<td>33,128</td>
<td>48,951</td>
<td>82,587</td>
<td>143,236</td>
<td>202,348</td>
<td>258,791</td>
</tr>
</tbody>
</table>

* Projections provided by Forward Insights Q2’20

- NVMe technology grew from 3 Petabytes to 29 PB shipped per year from 2016 to 2019
- For 2020, the projection is 54 PB
- NVMe technology demand projected to remain strong in a post COVID world

![Average Capacity (GB)](chart.png)
Agenda

- Fixing the Memory & Storage Hierarchy
- Refactoring for the Next Decade of Growth
- NVMe® Architecture Advancements
- What’s Next: Computational Storage
The Evolution of NVMe® Technology

Deployments

2010

2020

UNIFY PCIE* SSDs

SCALE OVER FABRICS

ENABLE INNOVATION
Driving Simplicity in a World of Complexity

TOO COMPLEX

Zoned Namespaces
Original NVMe
More expansion

NVMe/New Technology
NVMe/RDMA

TOO DEPENDENT ON HUMAN GLUE

TOO COMPLEX TOO DEPENDENT ON HUMAN GLUE
Driving Simplicity in a World of Complexity

- Back to core values... Fast, Simple, Scalable
- Foster areas of innovation AND avoid impact to broadly deployed solutions
- Create an extensible infrastructure that will take us through the next decade of growth
Specification Families

- The core of NVMe and NVMe over Fabrics integrated into a base specification
- Modular command set specifications (Block, Zoned Namespaces, Key Value, etc)
- Modular transport layer specifications (PCI Express®, RDMA, TCP)
- Maintain Management Interface as separate modular specification
Agenda

- Fixing the Memory & Storage Hierarchy
- Refactoring for the Next Decade of Growth
- NVMe Architecture® Advancements
- What’s Next: Computational Storage
NVM Express Technology Specification Roadmap

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NVMe Spec**
- NVMe 1.2.1 May'16
  - Sanitize
  - Streams
  - Virtualization

**NVMe-oF Spec**
- NVMe-oF 1.0 May'16
  - Transport and protocol
  - RDMA binding

**NVMe-MI Spec**
- NVMe-MI 1.0 Nov'15
  - Out-of-band management
  - Device discovery
  - Health & temp monitoring
  - Firmware Update

**NVMe 1.3 May’17**
- IO Determinism (NVM Sets)
- Persistent Event Log, Rebuild Assist
- Persistent Memory Region (PMR)
- Asymmetric Namespace Access (ANA)

**NVMe 1.4 June’19**
- Enhanced Discovery
- TCP Transport Binding

**NVMe 2.0**
- Merged w/Fabrics
- Namespace Types
- Alternate Cmd Sets

**NVMe-oF 1.1 July’19**
- Enhanced Discovery
- TCP Transport Binding

**NVMe-MI 1.1 May’19**
- Enclosure Management
- In-band Mechanism
- Storage Device Extension

Released NVMe specification

 WNed release
Namespace Types
Enable Alternate Command Sets

- Discover Command Set supported for a Namespace
- Command Sets: Block I/O, Key Value, Zoned, < future >
- Extensible approach for future innovation
Zoned Namespaces

- NVMe® technology evolving to address underlying media changes with larger Erase Blocks and more
- Zoned Namespaces require Logical Blocks to be written sequentially in a Zone
- Reduces write amplification and overprovisioning
Endurance Groups

- Flexible capacity management model
- Create appropriate groupings based on access pattern, media type(s), and more
Domains and Partitions

- Supports subdividing LARGE scale solutions – enabling partial operations & maintenance flows
Building Highways for Client, Cloud, & Enterprise

Use cases for NVMe® architecture continue to expand across Client, Cloud, and Enterprise

- Enhancements in Management, Telemetry and large storage systems meet the growing needs

<table>
<thead>
<tr>
<th>INFRASTRUCTURE ENHANCEMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Copy Command</td>
</tr>
<tr>
<td>Command Group Control Feature</td>
</tr>
<tr>
<td>Controller Memory Buffer Write Elasticity Status</td>
</tr>
<tr>
<td>Namespace Attachment Limits</td>
</tr>
<tr>
<td>Multiple Controller Firmware Update</td>
</tr>
<tr>
<td>Telemetry Enhancements</td>
</tr>
<tr>
<td>Command and Effect Log Enhancements</td>
</tr>
<tr>
<td>Non-”Maximum Data Transmit Size” Command Size Limits</td>
</tr>
</tbody>
</table>
The Complex Database Universe

- Data Warehouses (Presto, SparkSQL, AWS Redshift AQUA, ...) store **LOTS** of data
- Data stored in **LOTS** of (arbitrary) formats
- Data stored compressed and encrypted
- Formats and data constantly evolving

**Compressed, Encrypted, Arbitrary Format**

**Stored Table**

<table>
<thead>
<tr>
<th>NAME</th>
<th>STATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bruce Banner</td>
<td>NH</td>
</tr>
<tr>
<td>Bruce Wayne</td>
<td>NY</td>
</tr>
<tr>
<td>Diana Prince</td>
<td>NJ</td>
</tr>
<tr>
<td>Jean Grey</td>
<td>NH</td>
</tr>
<tr>
<td>Jessica Jones</td>
<td>NY</td>
</tr>
<tr>
<td>Peter Parker</td>
<td>NY</td>
</tr>
<tr>
<td>Selina Kyle</td>
<td>CA</td>
</tr>
<tr>
<td>Tony Stark</td>
<td>CA</td>
</tr>
</tbody>
</table>
Finding the Needle in the Haystack

**FILTER** Select Name where State="NY"

Compressed, Encrypted, Arbitrary Format
Programs as Computational Storage Offloads

- Programs invoked and used in standard way
- Programs in hardware agnostic bytecode and downloaded from host for later execution
- Device may offer fixed function programs
- Programs operate on data in on-device memory

Saving Power
Increasing Performance
The Evolution of NVMe® Technology

UNIFY PCIE* SSDs

SCALE OVER FABRICS

ENABLING INNOVATION

2010 2020