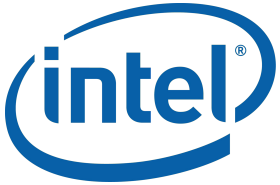
NVMe™ Annual Update

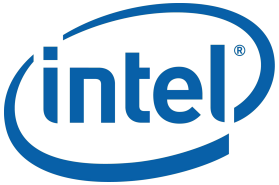
Sponsored by NVM Express™ organization, the owner of NVMe™, NVMe-oF™ and NVMe-MI™ standards

Speakers

Peter Onufryk



Nick Adams



nvm EXPRESS®

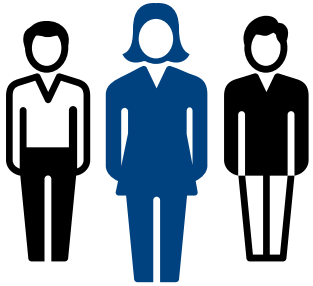
NVMe™ State of the Union

Peter Onufryk

intel

NVM Express™, Inc.
120+ Companies Defining NVMe™ Together

Board of Directors
13 elected companies, stewards of the technology & driving processes
Chair: Amber Huffman



Marketing Workgroup
NVMexpress.org, webcasts, tradeshows, social media, and press
Co-Chairs: Jonmichael Hands and Cameron Brett

Technical Workgroup
NVMe™ Base and NVMe Over Fabrics
Chair: Peter Onufryk

Management Intf. Workgroup
NVMe Management
Chair: Peter Onufryk
Vice Chair: Austin Bolen

Interop (ICC) Workgroup
Interop & Conformance Testing in collaboration with UNH-IOL
Chair: Ryan Holmqvist

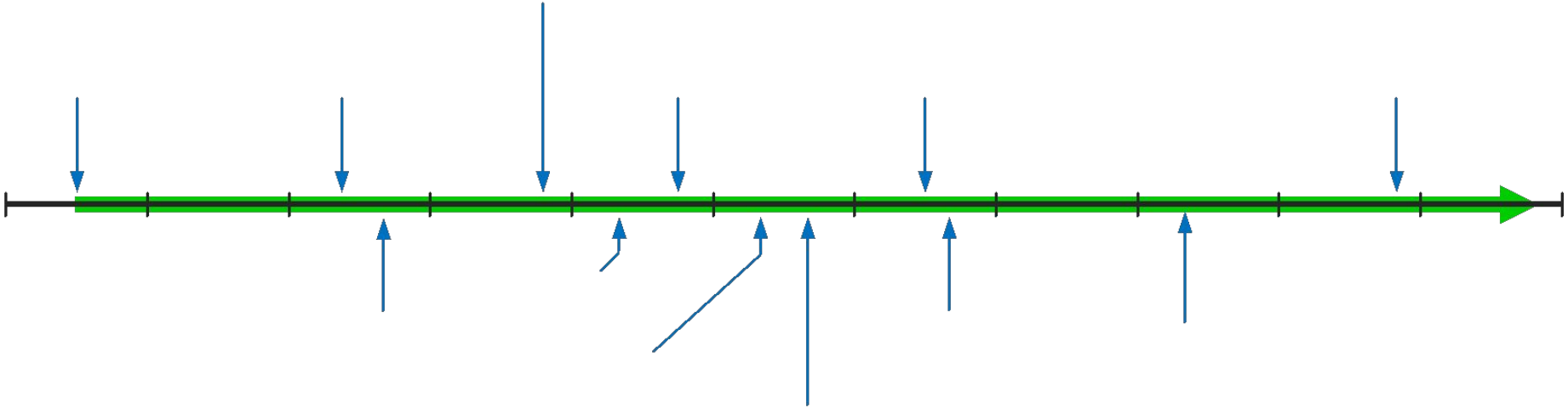
facebook Microsoft intel cisco DELL EMC SEAGATE Toshiba Memory

Micron MARVELL SAMSUNG Microsemi NetApp Western Digital

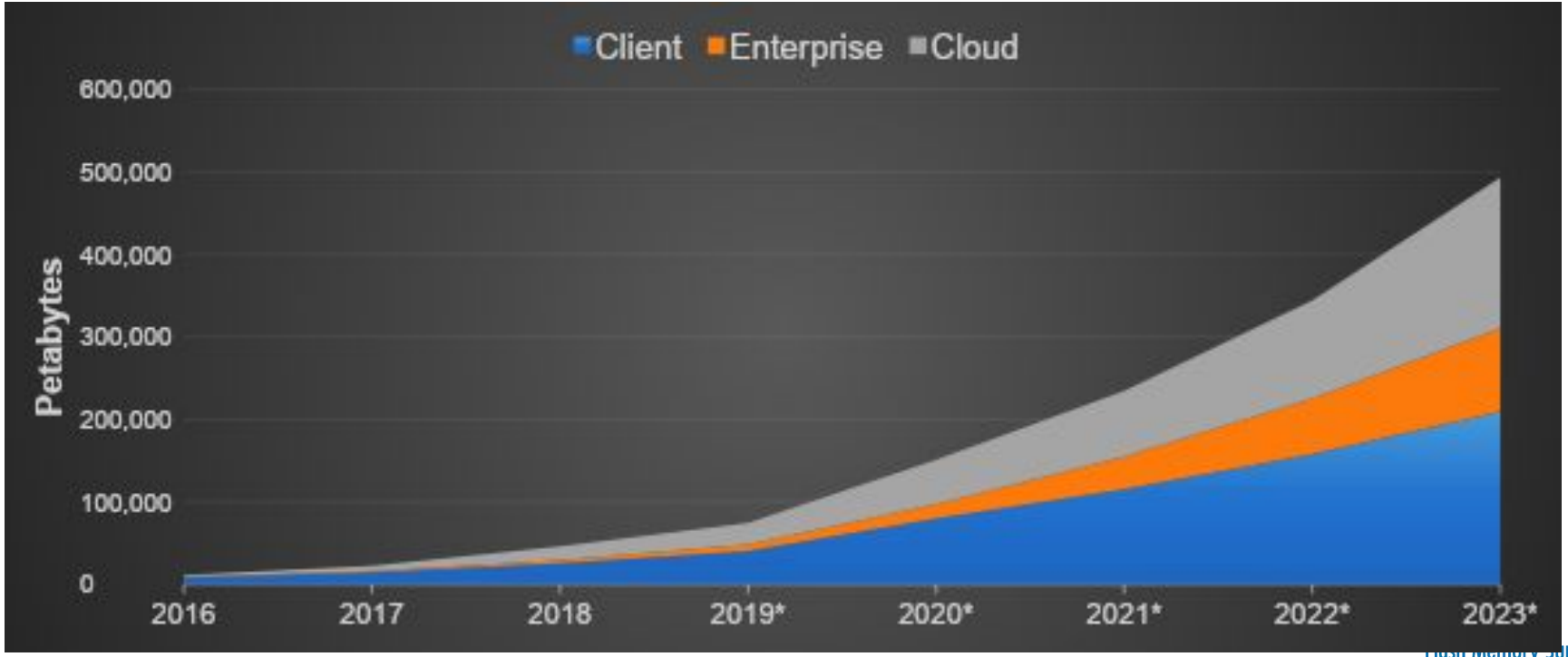
Flash Memory Summit

nvm EXPRESS®

Ten Years of NVMe™



Strong Growth Across Segments



* Projections provided by Forward Insights Q2'19

NVMe™ is The New Language of Storage

NVMe™ SSDs	23 Companies Shipping 96 Models
NVMe Severs	13 Companies Shipping 93 Models
NVMe AFAs	11 Companies Shipping 21 Models
NVMe Appliances	8 Companies Shipping 21 Models
NVMe-oF HBAs/NICs/RNICs	5 Companies Shipping 53 Models
NVMe-oF Accelerated Adapters	6 Companies Shipping

* Data provided by G2M Research

With Millions of Units Shipping

K Units	2016	2017	2018	2019*	2020*	2021*
Enterprise	364	749	1,048	2,774	5,740	11,192
Cloud	2,051	3,861	10,231	17,338	25,891	31,050
Client	33,128	50,385	82,613	111,888	187,689	243,889

* Projections provided by Forward Insights Q2'19

2019 NVMe™ Deliverables

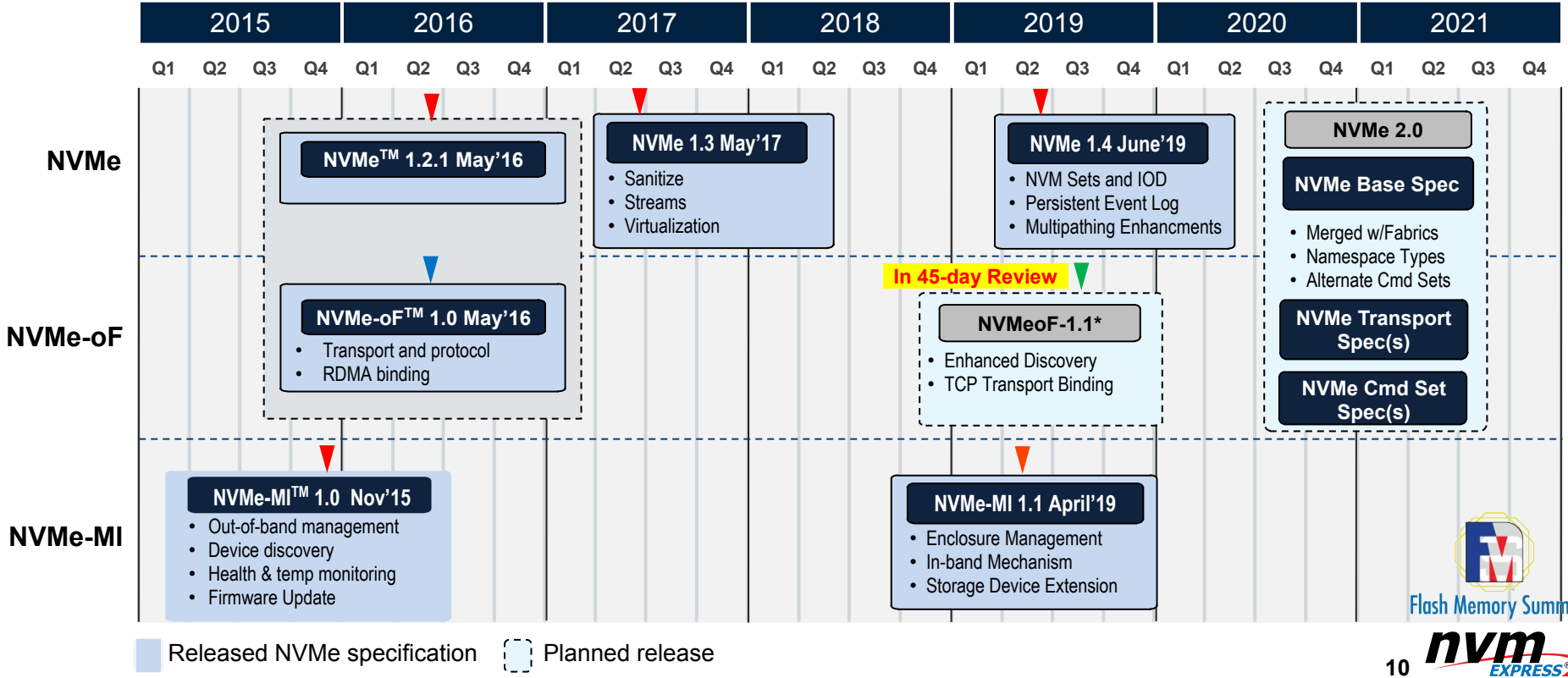
NVMe™ Base Specification	
NVMe 1.4	6/10/2019
NVMe 1.3	5/1/2017
NVMe 1.2	11/3/2014
NVMe 1.1	10/11/2012
NVMe 1.0	5/14/2008

NVMe Over Fabrics Specification	
NVMe-oF™ 1.1	45-day Review
NVMe-oF 1.0	6/5/2016

NVMe Management Interface Specification	
NVMe-MI™ 1.1	4/29/2019
NVMe-MI 1.0	11/17/2015

NVMe Plugfest	
Plugfest #11	6/24/2019
Plugfest #10	11/12/2018
…	
Plugfest #1	5/13/2013

NVMe™ Specification Roadmap



Three New Specifications for 2019

NVMe™ 1.4

NVM Sets and **IO Determinism** enable better performance, isolation, and QoS for hyperscale data centers. **Persistent event log** provides robust drive history for issue triage and debug. **Multipathing** provides optimal path for a namespace in multi-controller topologies

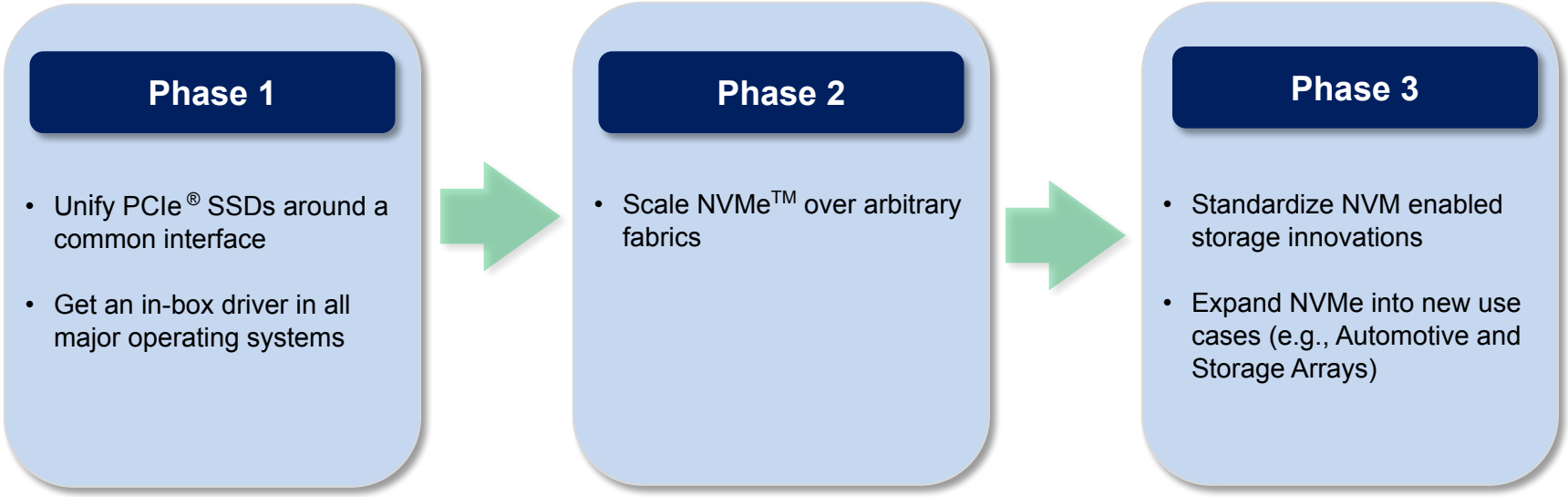
NVMe-oF™ 1.1

Enhanced Discovery for hosts to discover new NVMe devices. **TCP Transport Binding** NVMe/TCP enables efficient end-to-end NVMe operations with standard IP network with excellent performance and latency characteristics

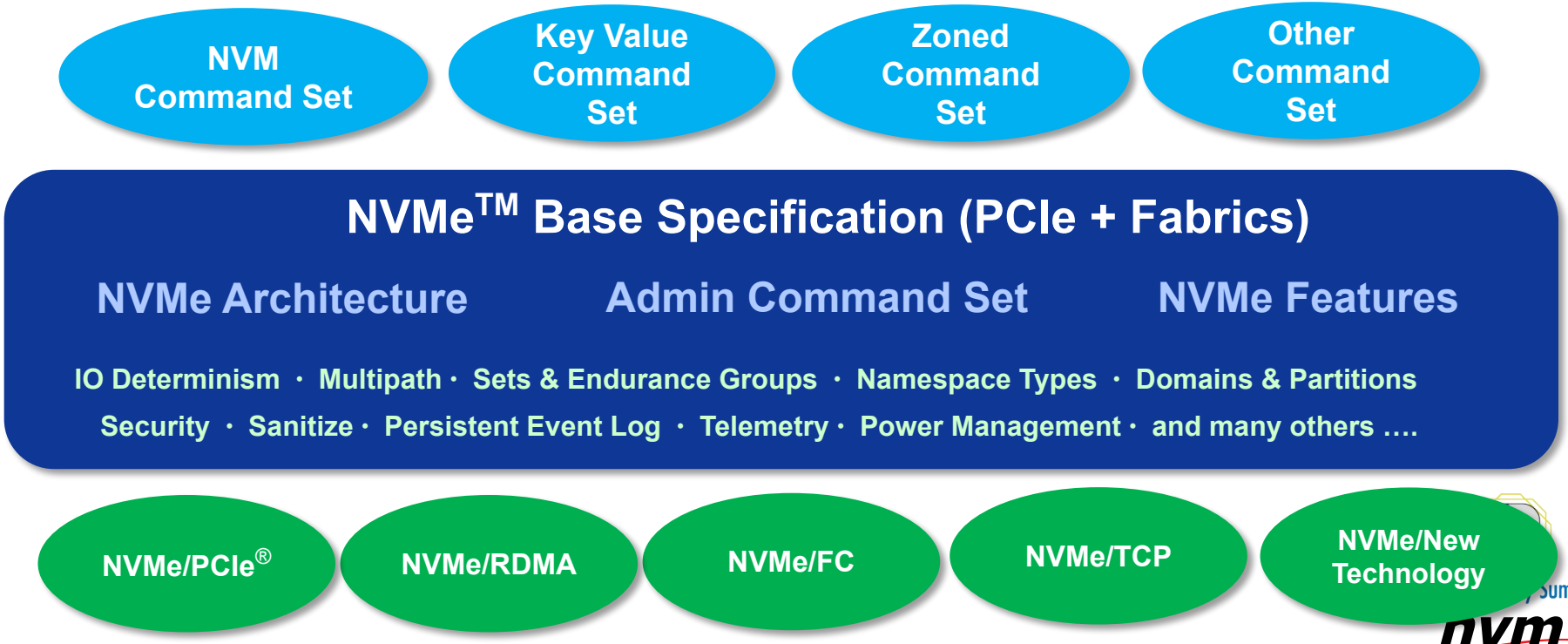
NVMe-MI™ 1.1

Enclosure Management enhances NVMe-MI for storage arrays for slot control, LED, and fans. **In-band Mechanism** opens up the NVMe-MI command set to standard NVMe driver (VPD, FRU). **Storage Device Extension** extends NVMe-MI to carrier cards and multiple controller devices

The Evolution of NVMe™

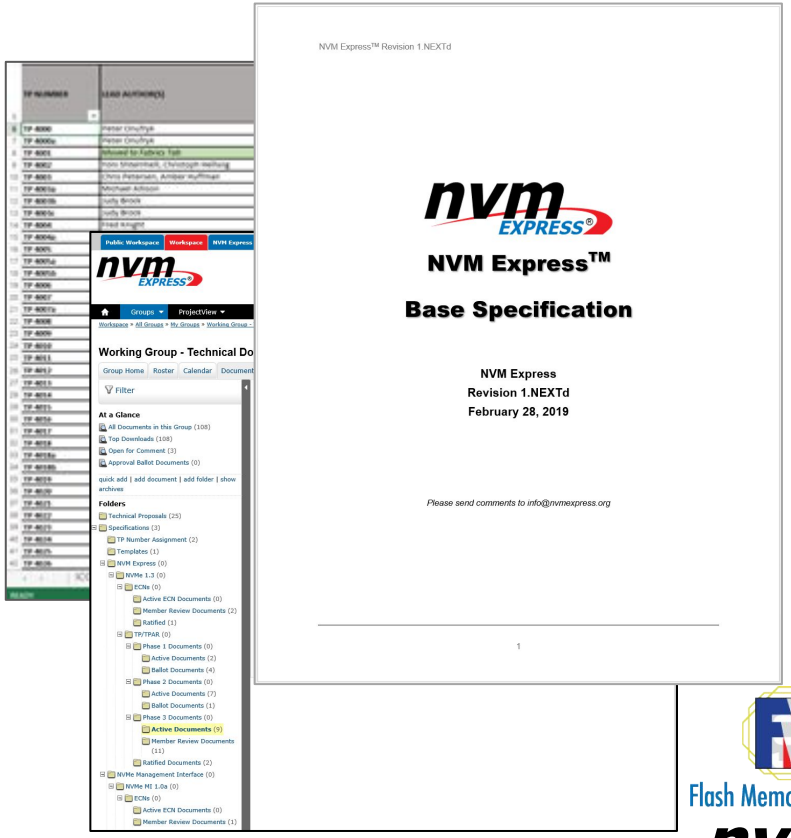


NVMe™ Continues to Drive Simplicity in A World of Complexity



Increasing the Rate of Innovation Together with Greater Quality

- Formalized sub-teams with publicly published calendars and minutes
- Technical proposal phases with clear entries and exits
- Document repository with revision history
- Integrated draft specification always up to date
- Weekly electronic ballots



Summary

NVMe™ has unified client, cloud, and enterprise storage around a common command set and interface

The growth in NVMe adoption continues to accelerate

The NVMe organization has put in place processes and initiatives to support the increased rate of innovation enabled by NVM and new use cases

NVMe remains true to its core principles of simplicity and efficiency as it enters is second decade

NVMe™ Base Spec 1.4 Features Overview

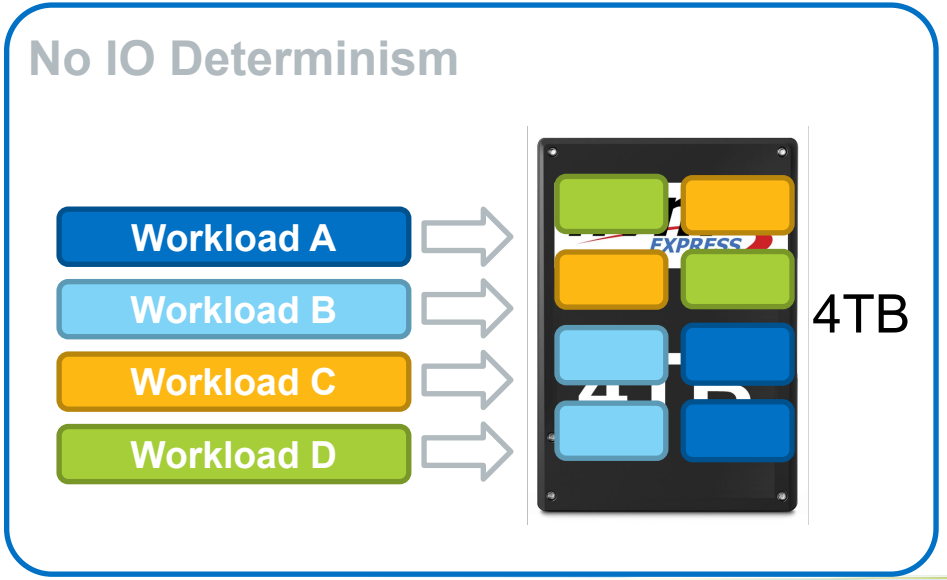
Nick Adams

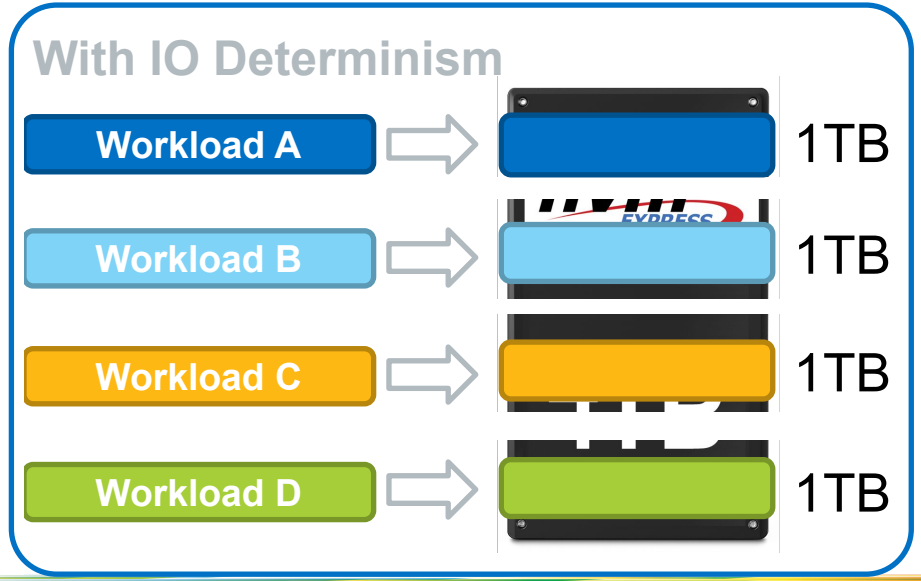
NVMe™ 1.4 Technical Highlights

IO Determinism

- Service isolation region
- Increase Read IOPs and reduce tail latency
- Provides strict QoS profile
- Significantly improves P99 and P9999 for a well-behaved host

IO Determinism & NVM Sets work together to provide improved QoS!





Interface Optimizations

Created new **mechanisms** for Hosts to **optimize** their use of NVMe™ devices

- IO Performance & Endurance Hints
  - Exposes preferred Size, Granularity and Alignment for both Write and Deallocate to the Host
- Endurance Groups
  - Enables drives to be configured for endurance management across one or more NVM Sets
- Namespace Granularity
  - Enables optimal Namespace Size and Capacity by the Host at Namespace creation time
- Submission Queue Associations
  - Enables the Host to associate an IO Submission Queue it created with a specific NVM Set
- Verify Command
  - Verifies ability to read data without sending that data across the bus to the host

Focused effort on optimizing the Host's ability to improve the performance and endurance of NVMe devices

Persistent Memory Regions

- Persistent Memory Region (PMR)
  - PCI memory space on the SSD exposed to the Host
  - May be used to store command data
  - Contents persist across power cycles, resets and disabling of the PMR

- **Usage Models for PMR**
  - **Logs for SW RAID, EC & Databases**
  - **Journals for File Systems**
  - **Metadata**
  - **Staging area for data pre-processing**
  - **Network transactions**

Enhanced Telemetry Capabilities

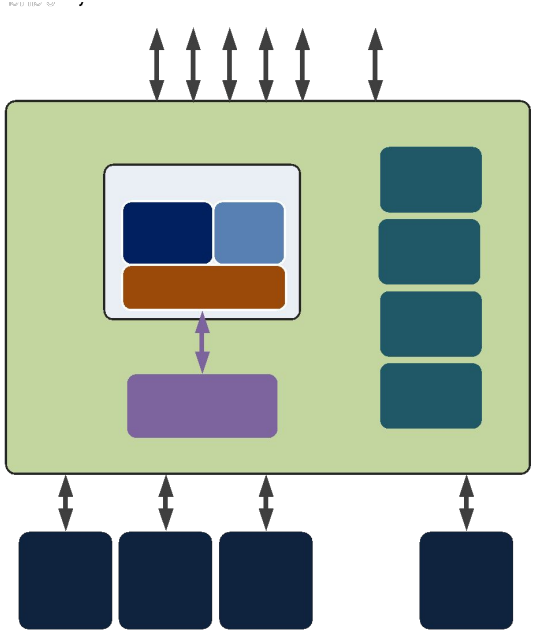
- The **Persistent** Event Log defines the features necessary to build a scaffolding that enables extensible debug infrastructure that is usable at scale
- Comprehensive set of events defined
  - Health Snapshot
  - Firmware Commits
  - Timestamp Changes
  - Power-on or Resets
  - Thermal Excursions
  - Vendor Specific
  - TCG-defined Events
  - Hardware Errors

Allows SSD customers to get consistent debug capabilities across vendors!

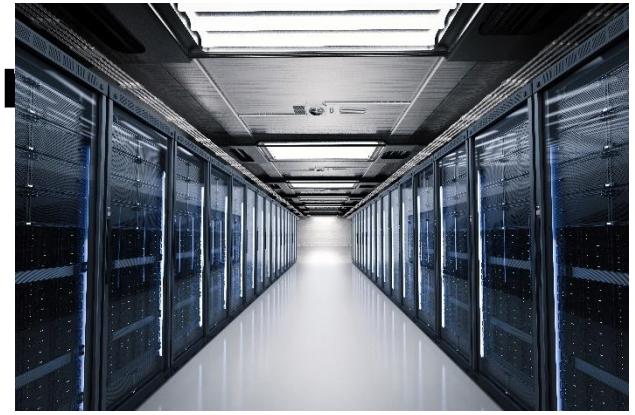
Allows SSD vendors an extensible framework for custom debug content!

Administrative Controller

- New controller type for Enclosure Management
- **Requires minimal functionality** while enabling innovative management solutions
  - NVMe-MI™ in-band support (NVMe-MI Send & Receive)
  - Namespace Management
  - NVM Subsystem Reset
  - Unique PCI interface allows loading a custom driver
- Key Functional Aspects
  - Not required to support IO Queues or Command Set(s)
  - Reduced set of required Admin Commands, Features & Log Pages to support



Maturing the NVMe™ Infrastructu

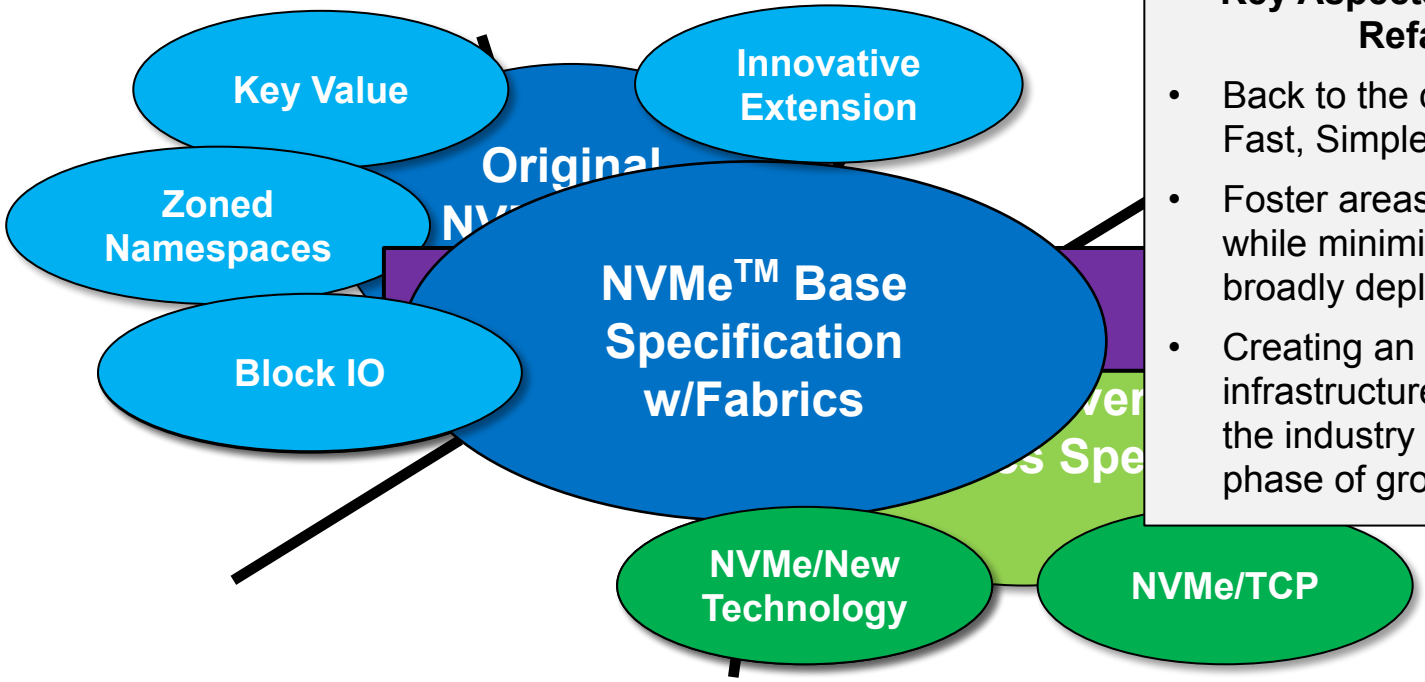


- Working together to make the NVMe™ infrastructure **robust** and **mature** for the industry.
  - Addressed industry needs across a variety of areas
    - **Rebuild Assist** – Improved Recovery Scenarios
    - **UUIDs for Vendor-specific Info** – Mechanism to ensure vendor-specific events don't collide
    - **Multi-host Shared Stream Write** – Improved multi-host functionality for Cloud & Enterprise
    - **Enhanced Command Retry** – Adds robustness in heavy load & other abnormal conditions
    - **Namespace Write Protect** – Enables finer granularity control over areas to Write Protect
  - Added clarifications and clean up to over 25 functional areas of the specification

Contributions to the infrastructure come from broad industry collaboration. This highlights NVMe's ability to come together as a community & deliver value for the entire industry!

Refactoring the NVMe™ Family of Specifications

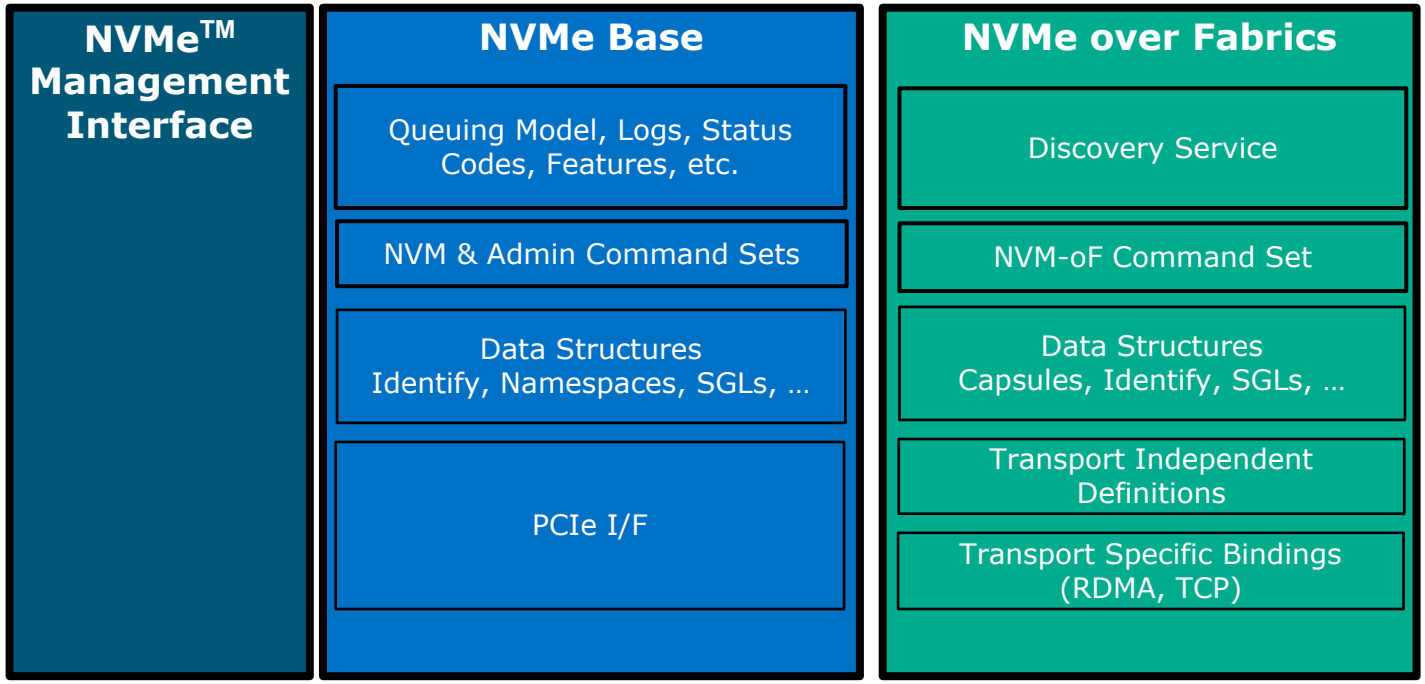
Driving Simplicity in a World of Complexity



Key Aspects Driving the Refactor

- Back to the core values… Fast, Simple, Scalable
- Foster areas of innovation while minimizing impact to broadly deployed solutions
- Creating an extensible spec infrastructure that will take the industry through the next phase of growth for NVMe!

NVMe™ spec family wasn't structured for extensibility

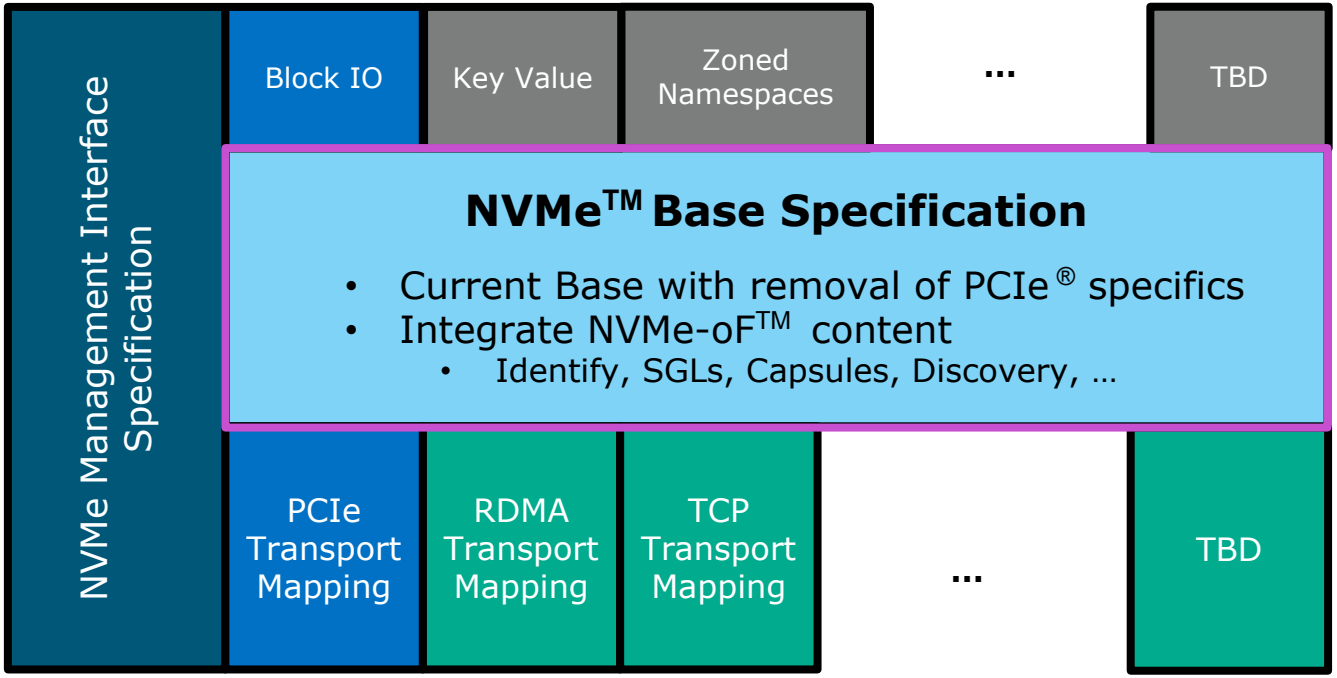


- Similar functions split between Base & Fabrics
- PCIe ® transport integrated into Base
- Command Sets not layered to enable extensibility

Need a new structure to enable innovation!

Optimizing the Specifications for Evolution

Transport Separation, Command Set Extensibility, Fabrics Base Integration



- Adds Fabrics concepts as core to NVMe
- Eliminates duplication in data structures
- Integration of NVMe and NVMe-oF base functions
- Separate command set specs
- Modular transport mapping layer, including PCIe

Questions?


nvm
EXPRESS®
Architected for Performance