Endurance Group Management: Host control of SSD Media Organization

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

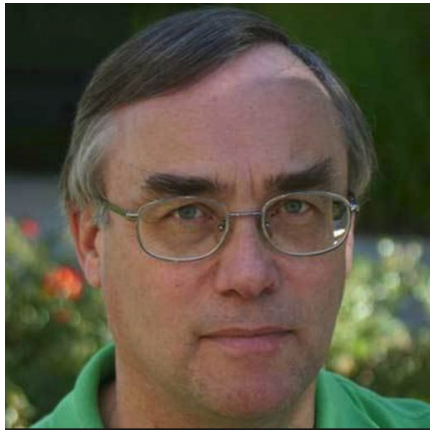
Abstract

SSD customers can have different requirements for the organization of the media in a drive: one large pool of capacity, separate sub-drives with performance isolation (IO determinism), or one large pool plus a small pool capable of higher-performance writes.

By allowing the host to configure a drive's media in the field, a single SSD model can satisfy very different use cases. NVMe™ Endurance Group Management provides a mechanism for media to be configured into Endurance Groups and NVM Sets.

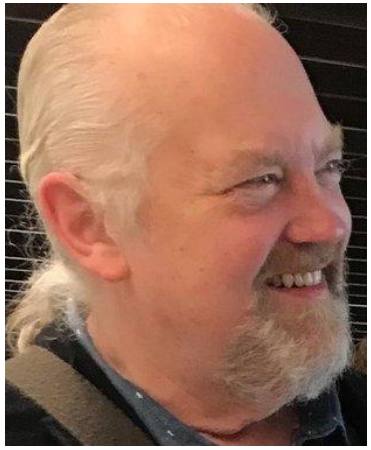
This presentation will explain various use cases and how the mechanism is used to configure not just SSD media but also storage array components.

Speakers



Paul Suhler

Micron



Mark Carlson

Toshiba Memory

Agenda

NVMe™ Capacity Entities

SSD Organizations (Use Cases)

Management Methods

Future Work

TP 4052 Endurance Group / NVM Set Management

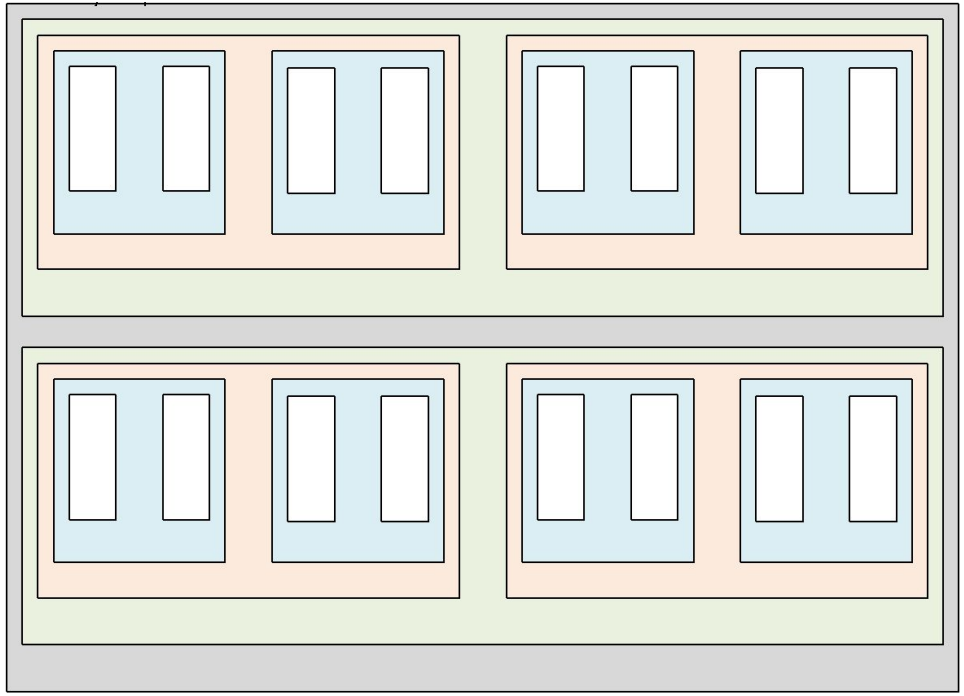
Use cases:

- More flexible IO Determinism
  - SSD vendors currently ship static configurations
- Address the need to divide work between host and drive
- Enable Endurance Groups for Storage Systems
  - Capacity management

Enable one SKU to be configured by customer for their use case

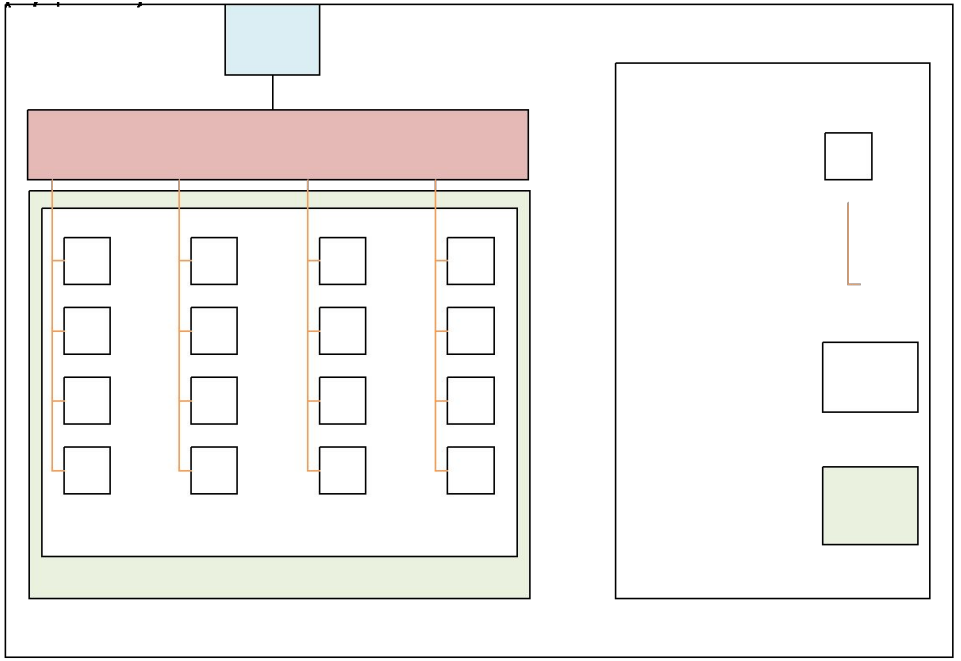
NVMe™ Capacity Entity Hierarchy

- Namespaces – Contain an array of logical blocks
- NVM Sets – Contain namespaces
- Endurance Groups – Contain NVM Sets
- Domains – Contain Endurance Groups, controllers, etc.



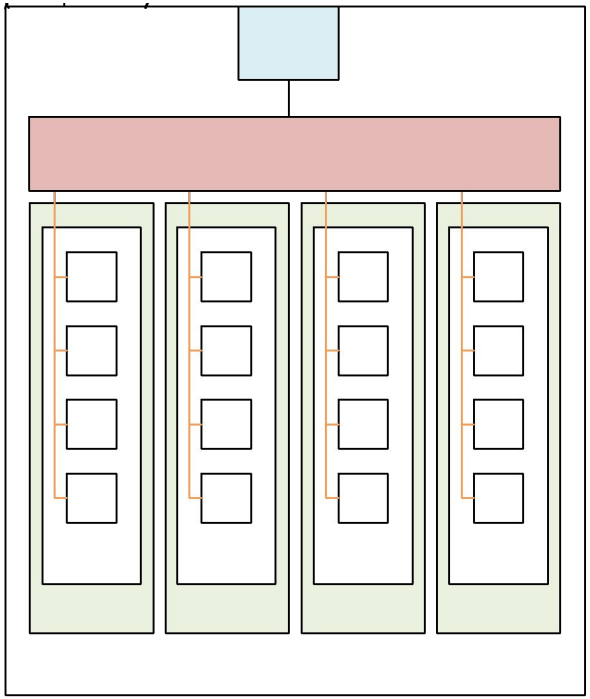
Original NVMe™ SSD Organization

- Media Units (e.g., dies) are connected to the controller by channels.
- Endurance is managed across all Media Units.



IO Determinism Use Case

- Need to create NVM Sets according to their capacity requirements (e.g. 1TB sets)
  - Typically once at the beginning of drive life
  - Supported Media Unit configurations are available indicating NVM Sets formed from Media Units along channels for isolation
  - Drive may only support two configurations (e.g. ½ TB and 1 TB sets) for this market



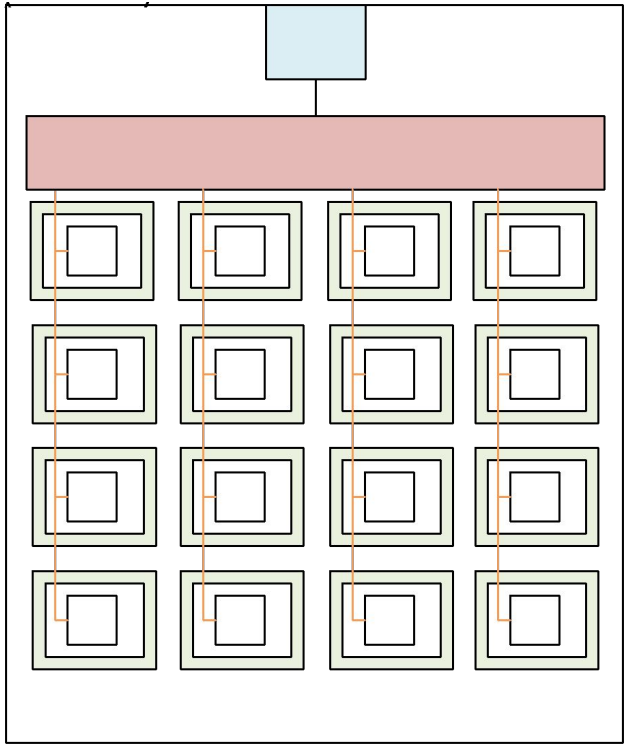
Four separate, isolated NVM Sets

Host Managed Media Users

- Need to closely manage placement of data and accommodate append behavior
  - Big concerns about Write Amplification and managing wear

1 EG / 1 NVM Set / 1 MU

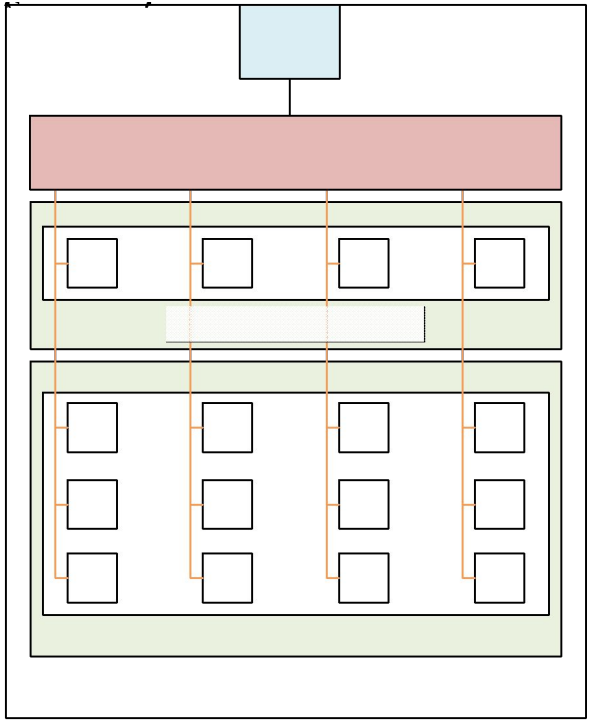
- No predictable latency
- Raw UBER
- 1 Namespace / MU



Get Log Page – Media Unit Status

Mixed-Mode NAND Operation

- NAND cells allow operation at a maximum number of bits per cell (e.g., QLC), as well as at a smaller number (e.g., TLC, SLC).
- Different Endurance Groups can have different bits per cell.
- One SSD can use some Media Units for a small amount of fast capacity and the remaining Media Units at a much higher density.



Storage Systems Users

- Need to create, resize and delete Endurance Groups within a Domain
  - No need to directly configure Media Units
  - Primarily tied to domains and partitions (TP 4009)
- Capacity Endurance Group Management
  - Capacity is drawn from the Domain
  - NVM Set is created as well
  - Deletion of Endurance Group also deletes NVM Set(s), Namespace(s)

Management Methods

- Two methods:
  - Direct Endurance Group Management
  - Capacity Endurance Group Management
- Direct Endurance Group Management is used for drives.
  - The operation will select from a fixed set of complete configurations; the selected configuration typically will be for the lifetime of the NVM subsystem.
  - This satisfies the requirements of hyperscalers.
    - Incrementally configuring endurance groups / NVM sets will not be supported for this method (not needed). Changing the configuration after the media has been used will not be supported in this method.

Endurance Group Management command

Operation (primarily) for SSDs:

- **Select Media Unit Configuration:** Selects one of the supported configurations.

Operations for Storage Arrays:

- **Create Endurance Group:** Creates an Endurance Group of a specified size from a Domain.
- **Create NVM Set:** Creates an NVM Set of a specified size from an Endurance Group.
- **Delete NVM Set :** Deletes a specified NVM Set and all its namespaces.
- **Delete Endurance Group:** Deletes a specified Endurance Group and all its contents.

Capacity Configuration Descriptor

Endurance Group Configuration Descriptor per endurance group

- Endurance Group Information
  - Capacity Adjustment Factor
  - Total Endurance Group Capacity
  - Spare Endurance Group Capacity
  - Endurance Estimate
- NVM Set Identifiers
- Channel Descriptors
  - Media Units on each channel

Media Unit Status Descriptor

- Media Unit Identifier
- Domain Identifier
- Endurance Group Identifier
- NVM Set Identifier
- Capacity Adjustment Factor
- Available Spare
- Percentage Used
- Number of Channels attached to this Media Unit
- Channel Identifier List

Capacity Endurance Group Management

- Capacity Endurance Group Management is for systems to dynamically create Endurance Groups and NVM Sets. The operation specifies a capacity for endurance groups and NVM sets without understanding of the underlying media units.

Endurance Group Management command

Operation (primarily) for SSDs:

- **Select Media Unit Configuration:** Selects one of the supported configurations.

Operations for Storage Arrays:

- **Create Endurance Group:** Creates an Endurance Group of a specified size from a Domain.
- **Create NVM Set:** Creates an NVM Set of a specified size from an Endurance Group.
- **Delete NVM Set :** Deletes a specified NVM Set and all its namespaces.
- **Delete Endurance Group:** Deletes a specified Endurance Group and all its contents.

Future Work

After we have experience with endurance group management, we will know whether and how to address:

- Indicating error correction provided by controller. Would allow storage system to rely on SSD ECC and not implement ECC across SSDs.
- Incrementally configuring Endurance Groups (rather than selecting a single SSD-wide configuration.
- Reconfiguring used Media Units to repurpose an SSD.
- Indicating levels of capacity organization below the Media Unit, e.g., planes or dies. Would it be useful?
- Indicating mapping of zones to Media Units. Would it provide any benefit?

Questions?


nvm
EXPRESS®
Architected for Performance