Standards for improving SSD performance and endurance

Bill Martin
Samsung Semiconductor, Inc.

What is being standardized?

☐ An interface to provide better collaboration between SSD and storage systems

    ☐ Stream operation

        ☐ Stores data with similar lifetime in associated physical locations

    ☐ Background operation control

        ☐ Advanced garbage collection

    ☐ Read Determinism

☐ Features to help limit long tail effect

    ☐ Rebuild Assist for SSDs

    ☐ Depopulation

☐ Object Storage

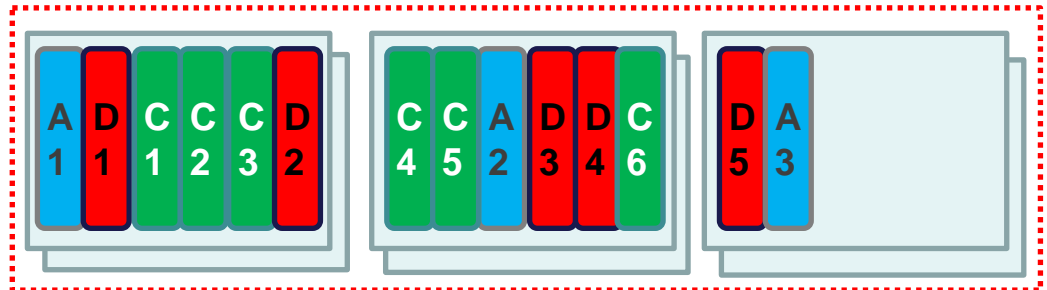   ☐ IP Based Drive management

    ☐ Key Value Storage API

2

Write Streams

- Development in NVMe, SCSI, and SATA
- Allows host to associate each write operation with a stream
- Device places all data associated with a stream in physically associated locations
- All data associated with a stream is expected to be invalidated at the same time (e.g., trimmed, unmapped)
- Why stream operation?
  - When different lifetime data is intermixed
    - Garbage Collection overhead increases
    - Write Amplification Factor increases
- Improves system performance
- Improves device endurance
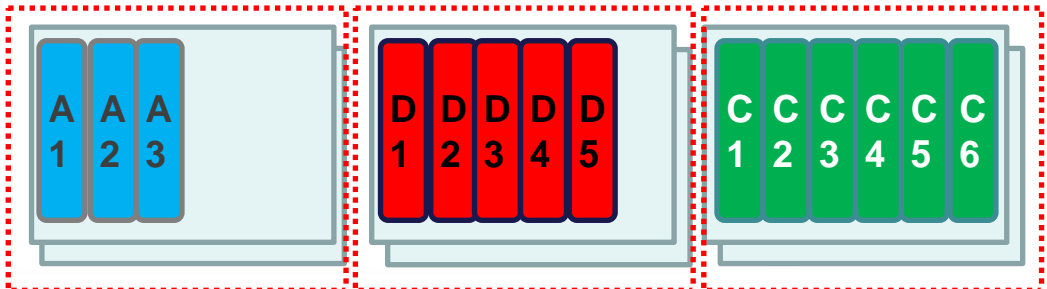
3

Stream comparison

Virtual Machine A

Database

Virtual Machine C
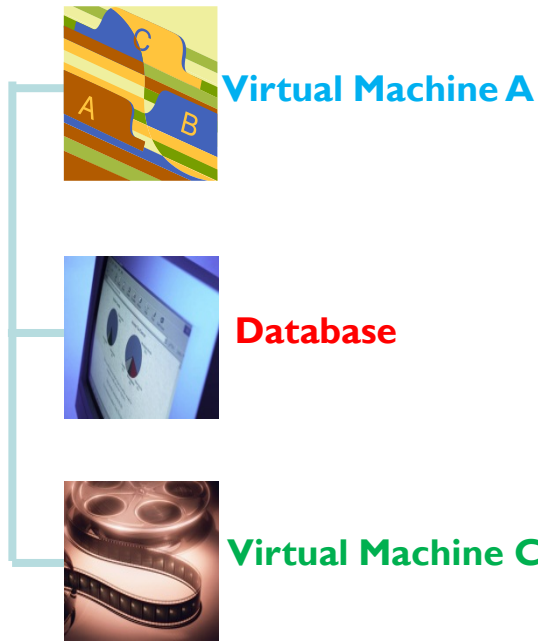
Non Stream
Data is written in order writes are processed

| A1 | D1 | C1 | C2 | C3 | D2 |

| C4 | C5 | A2 | D3 | D4 | C6 |

| D5 | A3 |

Stream
Data is grouped according to stream

| A1 | A2 | A3 |

| D1 | D2 | D3 | D4 | D5 |

| C1 | C2 | C3 | C4 | C5 | C6 |

Stream improvement

Greater than 9x performance and 2x SSD endurance

Write Throughput

Stream

Legacy

>9x Performance

WAF

>2x SSD Endurance
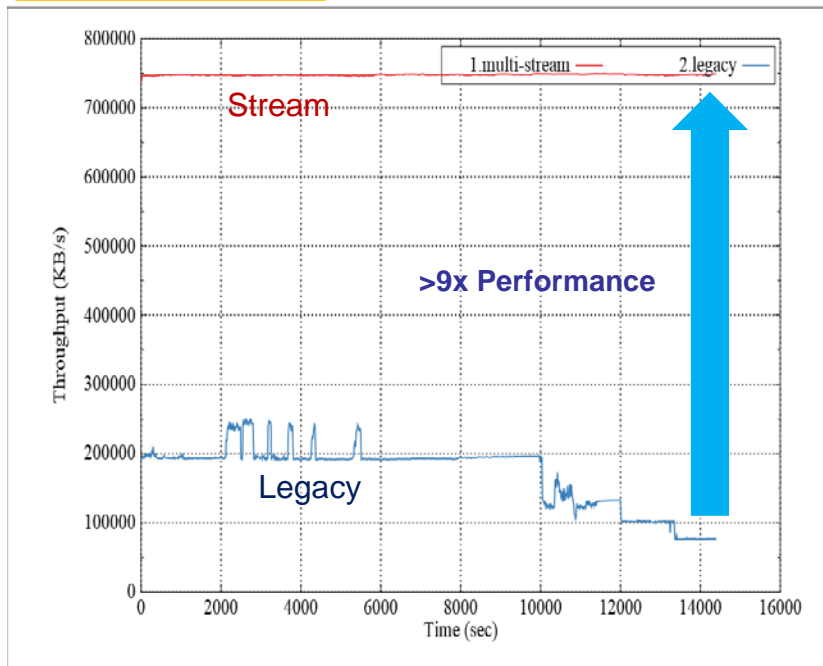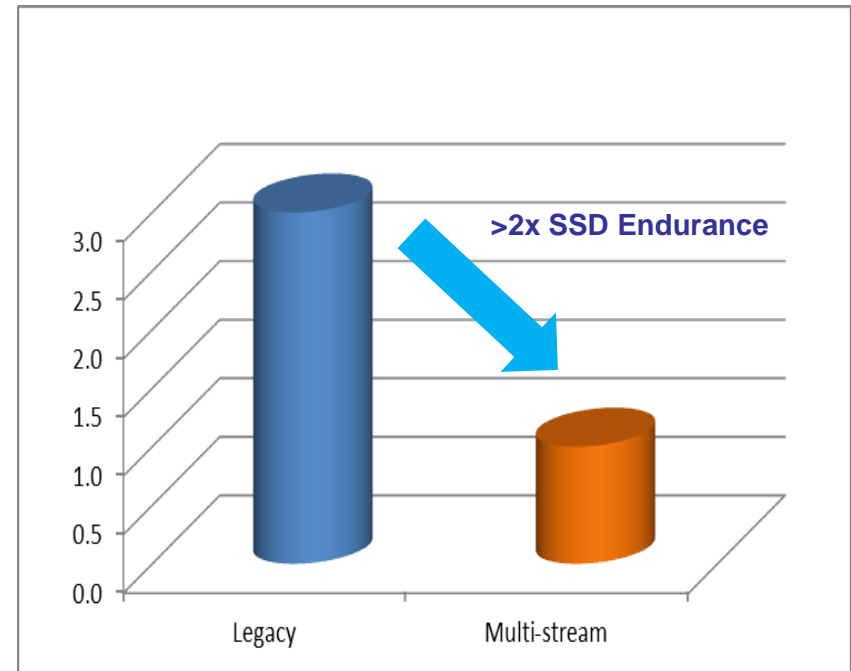
FIO 100% 128K writes with four different Streams of data

NVMe standardization for Streams

- Directives Technical Proposal
  - Extensible feature to provide host directives to device
    - Directive type
    - Directive identifier
  - Identify Directive
  - Streams Directive

6

Directives Commands

- New Commands
  - Directive Send
    - Configure specified directive
  - Directive Receive
    - Return parameters for a specified directive
- Existing I/O Commands
  - Write
  - In future other I/O commands may be added

Identify Directive

❏ Directive Receive

    ❏ Returns bitmaps of all directives

        ❏ Supported

        ❏ Enabled

❏ Directive Send

    ❏ Enable/Disable a specified directive

Streams Directive Overview

- Identifier scope
  - Namespace associated with a specific host
- Resource allocation
  - Per namespace associated with a specific host
  - Per NVM Subsystem

Streams Directive Receive

- Reports properties
  - NVM Subsystem
    - Maximum streams
    - NVM Subsystem streams available
    - NVM Subsystem streams open
  - Namespace
    - Stream Write Size
    - Stream Granularity Size
    - Allocated Stream Resources
    - Streams open
- Status of open streams
- Request Stream Resources for a namespace associated with a specific host

10

Streams Directive Send

☐ Release stream identifier for a namespace associated with a specific host

☐ Release stream resources for a namespace associated with a specific host

Streams Write command

❑ Specifies an identifier

  ❑ Associated with a namespace associated with a host

  ❑ If identifier specified is zero, then operates as a normal write

❑ Identifier implicitly opens specified stream

❑ If necessary stream identifier implicitly released

  ❑ For host that wants tight control over identifier use – Release Identifier

  ❑ For host that does not wand control over identifiers – allow controller to implicitly release identifiers

12

SDC 16

Advanced Background Operation (ABO) Control

- Development in NVMe, SCSI, and SATA
- Why ABO control?
  - IO performance is degrade when ABO occurs at the same time as IO
  - Avoids overlap of IO and ABO
- What does ABO provide?
  - Notification when ABO is imminent
  - Mechanism for host to specify the time to spend on ABO
  - Mechanism for host to specify an amount of resources to free during ABO
- Provides predictable and consistent performance
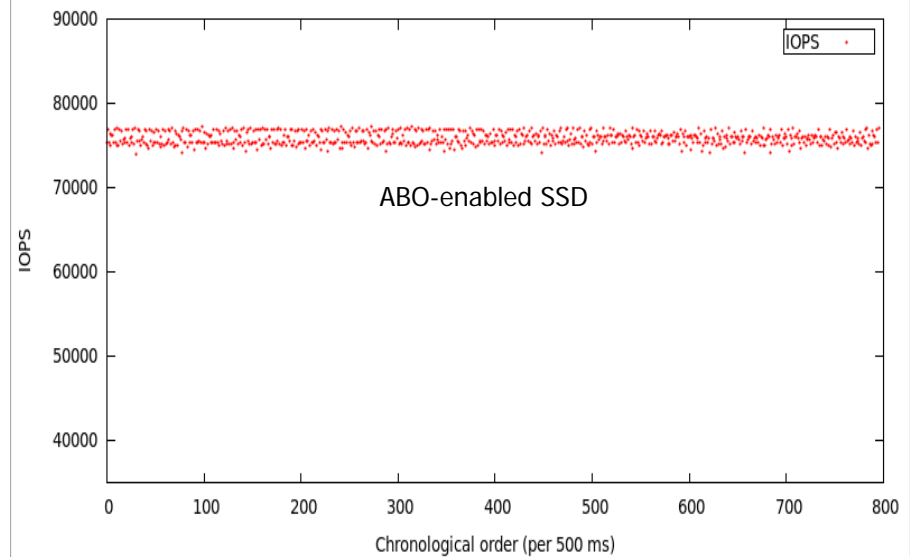
13

ABO performance improvement
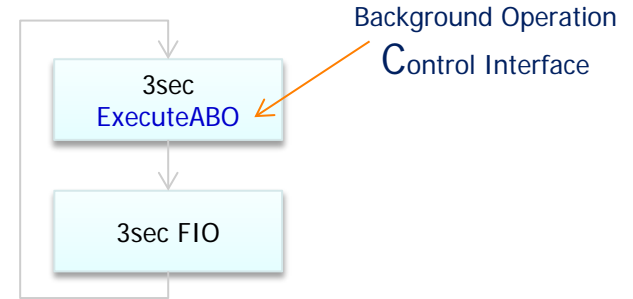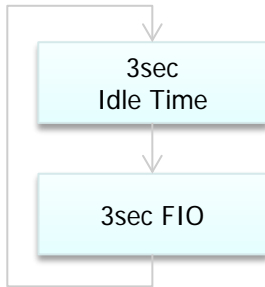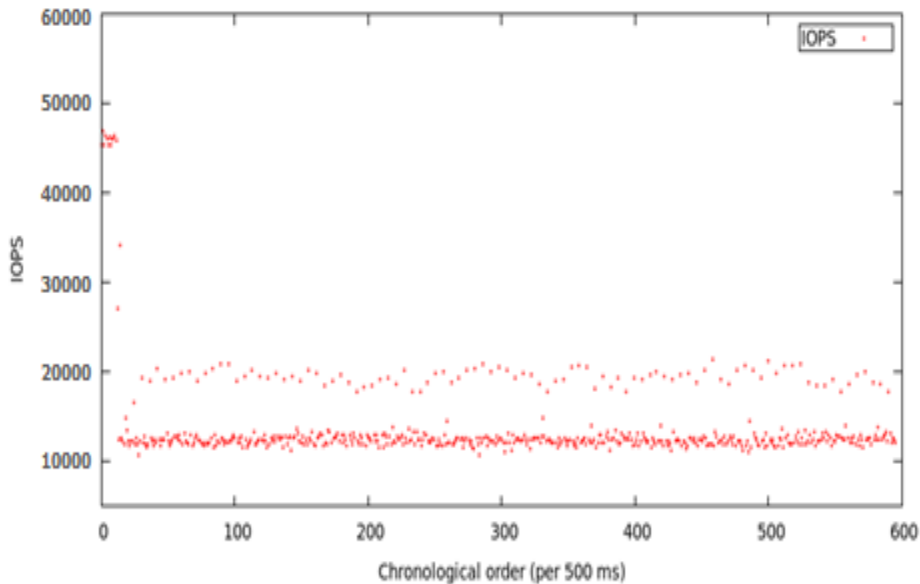
FIO 100% 4K random writes

NVMe standardization of ABO Directive Receive

- Device Returns characteristics for ABO
  - Minimum fraction resources available
  - Current fraction resources available
  - Maximum fraction resources possible
  - Current status of ABO
    - ABO not in process
    - Host initiated ABO in process
    - Controller initiated ABO in process
- Host requests notification of ABO imminent

NVMe standardization of ABO Directive Send

❑ Host informs device of time to perform ABO

    ❑ Notification threshold

    ❑ Maximum time to perform ABO

    ❑ Target fraction of resources

16

Read Determinism

- Development in NVMe
- Allows data that is expected to be read in parallel to be placed in physical locations that avoid blocking
  - Read/Read blocking
  - Read/Write blocking
- Alternate mechanisms being discussed now
  - Provide request mechanism to storage for different read groups
  - Provide layout information to application and allow application to manage

17

Rebuild Assist for SSDs

- Development in SCSI
- Allows storage subsystem to easily determine LBAs that should be recovered from another source
- Extension of GET LBA STATUS
  - Returns information on LBA extents that are anticipated to return unrecovered errors
  - Storage subsystem is able to read LBA extents marked as returning unrecovered errors and re-write or mark as not available from this device
- Completed for SCSI September 2015

18

Depopulation

- Development in SCSI and SATA
- Data Preserving Depopulation
  - Return a reduced capacity drive to service with slow storage elements removed from the logical address space to address the long tail effect
  - Preserves data not associated with depopulated elements
- Repurposing Depopulation
  - Similar to Data Preserving Depopulation
  - Is intended for a storage device that will be taken out of service in one locality and re-deployed in another locality

Data Preserving Depopulation

- Report physical element status
- Remove physical element
- Remove LBA range from use
- Truncate logical capacity

Repurposing Depopulation

□ Has characteristics of Data Preserving Depopulation

  □ Does not retain data in elements that are not depopulated

□ Following depopulation the storage device may be initialized to a lower capacity

NVMe Standardization Status

❑ Directives (Identify and Streams)

  ❑ Technical proposal complete

  ❑ 30 day member review started 15 September

❑ Advanced Background Operation Directive

  ❑ Technical proposal almost complete

  ❑ Other alternatives being considered

❑ Read Determinism

  ❑ Just starting process

  ❑ 6 months or more to settle on mechanism

22

SCSI Standardization Status

- ABO and Streams standardization approved May 2015
  - Streams does not support implicit identifier assignment and release
  - ABO does not support Target fraction
  - Will be updated to match the NVMe model
- Rebuild Assist for SSDs
  - Approved 15 September 2016
- Depopulation
  - Under development in parallel in SCSI and SATA
  - Should complete in 2017

23

SATA Standardization Status

❏ ABO standardization completed April 2016

  ❏ Does not support target fraction

❏ Stream standardization

  ❏ Following NVMe standard

  ❏ Expect completion later this year

❏ Depopulation

  ❏ Being developed in parallel with SCSI development

  ❏ Should complete in 2017

Object Drive

- Development in SNIA
- Starting with IP Based Storage Management
  - Currently being developed in Object Drive TWG
  - Current document is revision 0 version 12
  - Supported by Swordfish 1.0
- Future directions
  - Key Value API
    - No proposals yet

Questions?

SD C 16