Multi-Stream Write SSD
Increasing SSD Performance and Lifetime with Multi-Stream Write Technology

Changho Choi, PhD

Principal Engineer

changho.c@samsung.com

Samsung Semiconductor

Disclaimer

This presentation is intended to provide information concerning SSD and memory industry. We do our best to make sure that information presented is accurate and fully up-to-date. However, the presentation may be subject to technical inaccuracies, information that is not up-to-date or typographical errors. As a consequence, Samsung does not in any way guarantee the accuracy or completeness of information provided on this presentation.

The information in this presentation or accompanying oral statements may include forward-looking statements. These forward-looking statements include all matters that are not historical facts, statements regarding the Samsung Electronics' intentions, beliefs or current expectations concerning, among other things, market prospects, growth, strategies, and the industry in which Samsung operates. By their nature, forward-looking statements involve risks and uncertainties, because they relate to events and depend on circumstances that may or may not occur in the future. Samsung cautions you that forward looking statements are not guarantees of future performance and that the actual developments of Samsung, the market, or industry in which Samsung operates may differ materially from those made or suggested by the forward-looking statements contained in this presentation or in the accompanying oral statements. In addition, even if the information contained herein or the oral statements are shown to be accurate, those developments may not be indicative developments in future periods.

Agenda

- NAND flash characteristics

- Multi-Stream Write

  - Multi-Stream Write concept
  - Multi-Stream Write system architecture
  - Multi-Stream Write operation

- Performance benefit

- Standards
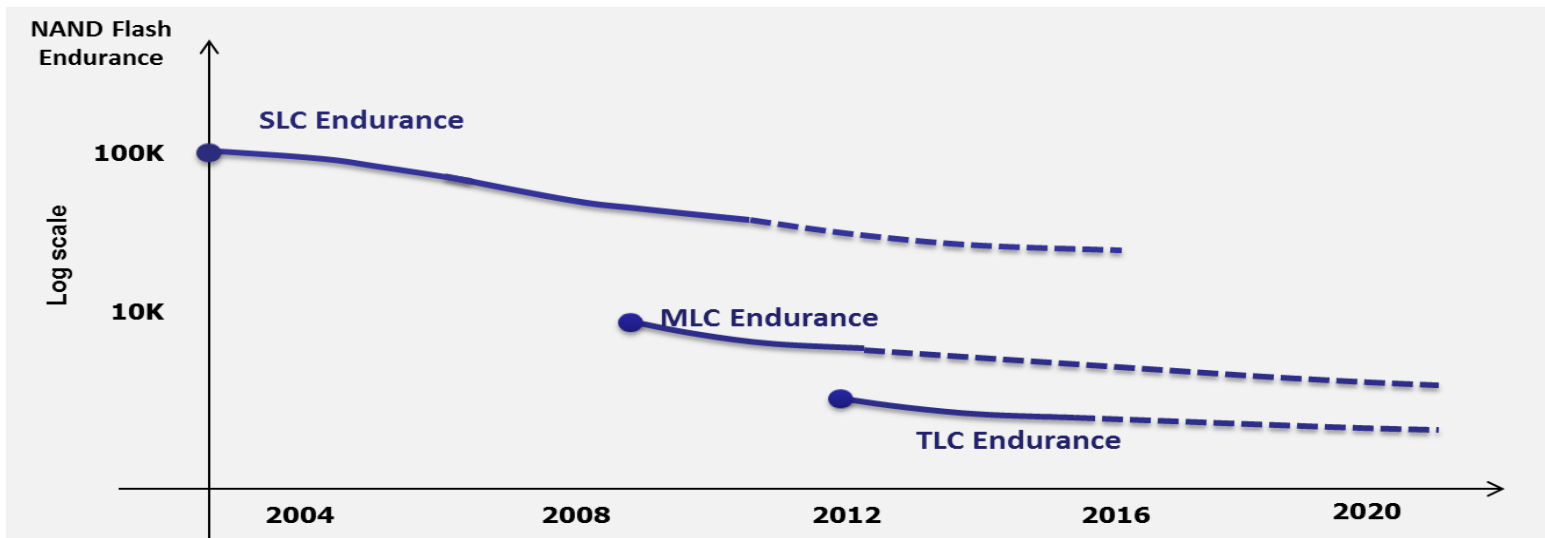
- Summary

- Q&A

NAND Flash Characteristics

- Operation: Read/Program/Erase

- Operation unit
  - Read/Program: Page
  - Erase: block (= multiple pages)

- Out-of-place update: in-place update(=overwrite) NOT allowed
  - Invalidate overwritten data

- Page MUST be erased before programming(writes)
  - Program/Erase (P/E) cycles
  - Need garbage collection operation

Efficient data placement increases performance
with reduced garbage collection overhead

NAND Flash Characteristics (Cont'd)

- Limited number of Program/Erase cycles

NAND Flash Endurance (Log scale)

- SLC Endurance: 100K
- MLC Endurance: ~10K
- TLC Endurance

Years: 2004, 2008, 2012, 2016, 2020

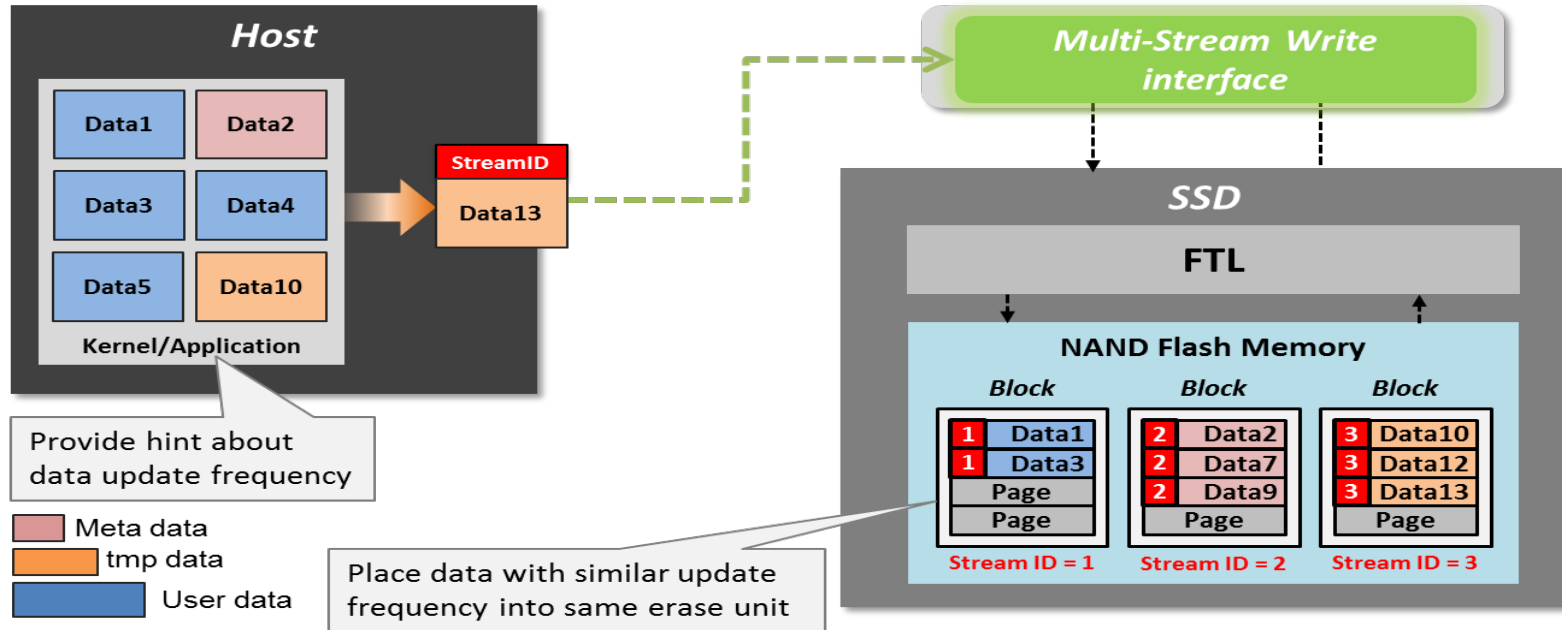Efficient data placement increases lifetime(endurance) of SSD

Multi-Stream Write

- Provide better endurance, improved performance, and consistent latency
  - Allow host to associate each write operation with a stream
  - All data associated with a stream is expected to be invalidated at the same time (e.g., updated, trimmed, unmapped, deallocated)
  - Align NAND block allocation based on application data characteristics(e.g., update frequency)
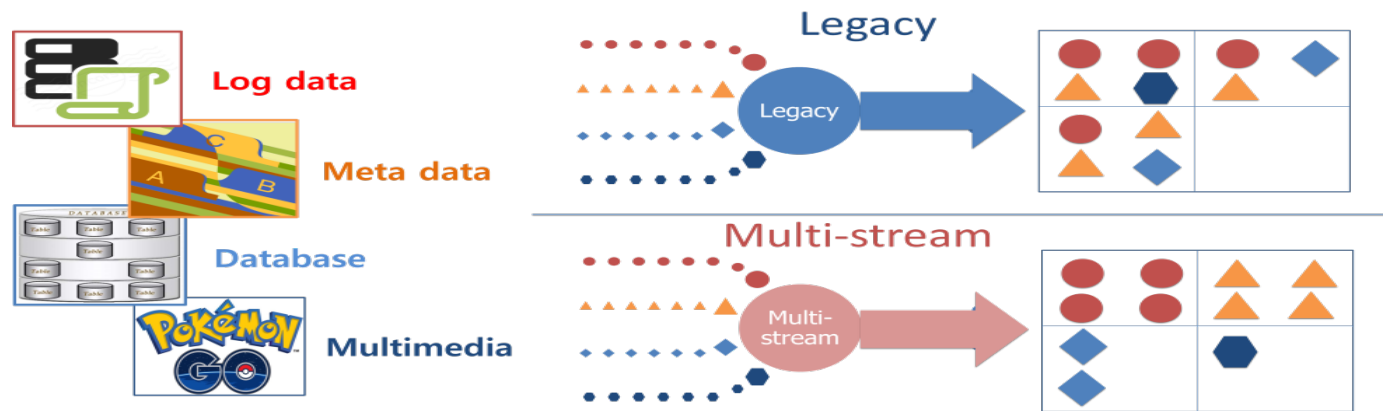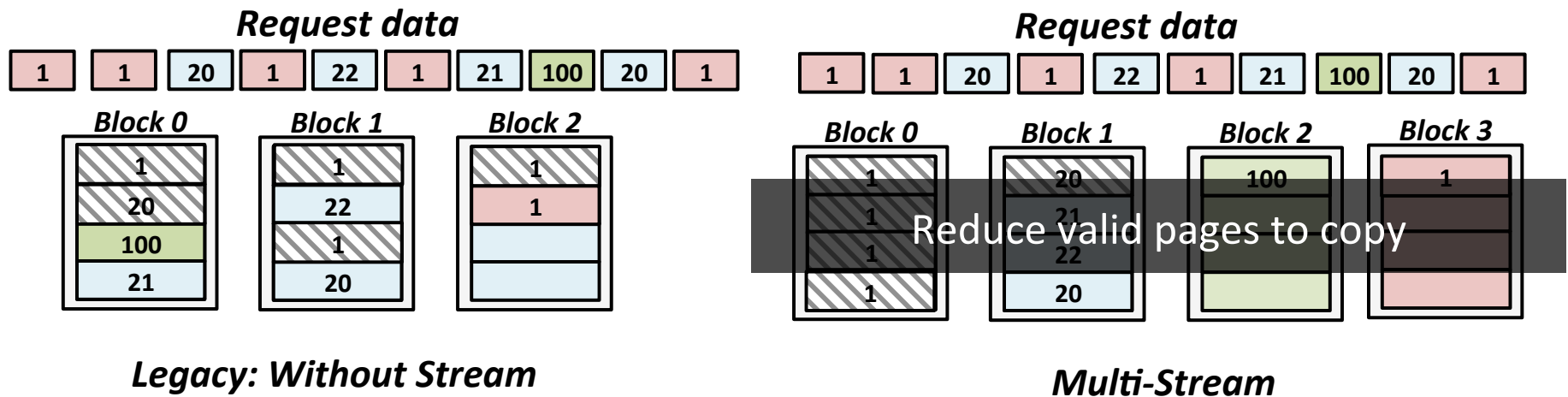
Multi-Stream Write Operation

- Mapping data with different update frequency to different streams

Operation Example

- Efficient data placement with Multi-Stream Write
  - Reduce GC overheads -> better performance and lifetime!

Request data

| 1 | 1 | 20 | 1 | 22 | 1 | 21 | 100 | 20 | 1 |

Request data

| 1 | 1 | 20 | 1 | 22 | 1 | 21 | 100 | 20 | 1 |

Block 0

| 1 |
| 20 |
| 100 |
| 21 |

Block 1

| 1 |
| 22 |
| 1 |
| 20 |

Block 2

| 1 |
| 1 |
| |
| |

Block 0

| 1 |
| 1 |
| 1 |
| 1 |

Block 1

| 20 |
| 21 |
| 22 |
| 20 |

Block 2

| 100 |
| |
| |
| |

Block 3

| 1 |
| |
| |
| |

Reduce valid pages to copy

Legacy: Without Stream

Multi-Stream

For effective multi-streaming,
proper mapping of data to streams is important!

FIO Performance Measurement System

- Hardware
  - Quad Core Intel i7-4790 CPU 3.60GHz
  - 16GB memory
- Software
  - Ubuntu 14.04 LTS, v4.0.3 Kernel with Multi-Stream Write patch
  - FIO 2.2.5 with Multi-Stream Write patch
- Device
  - Multi-Stream Write enabled NVMe 960GB M.2 SSD
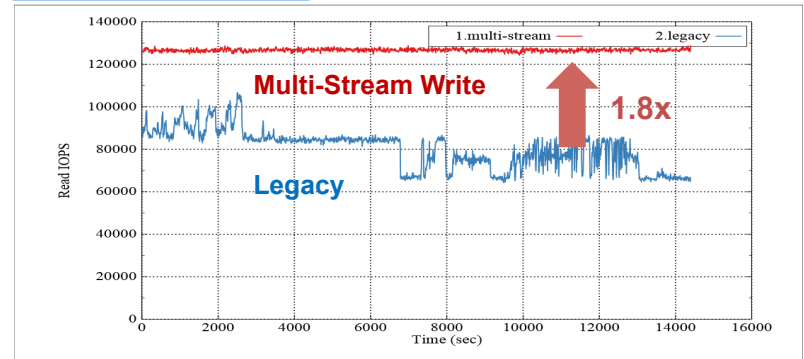
Performance Measurement Configuration

- Four sequential write jobs and six random read jobs
  - Different data lifetime: 1x, 10x, 33x, 55x
- Precondition
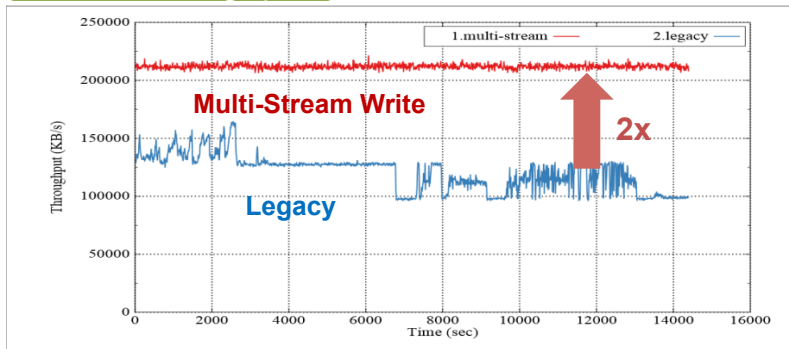 - 2 hours with four-write jobs

Four Streams – Read/Write(70%/30%)

- **Reads**
  - Jobs: 6
  - Block size: 4k
  - Iodepth: 64

- **Writes**
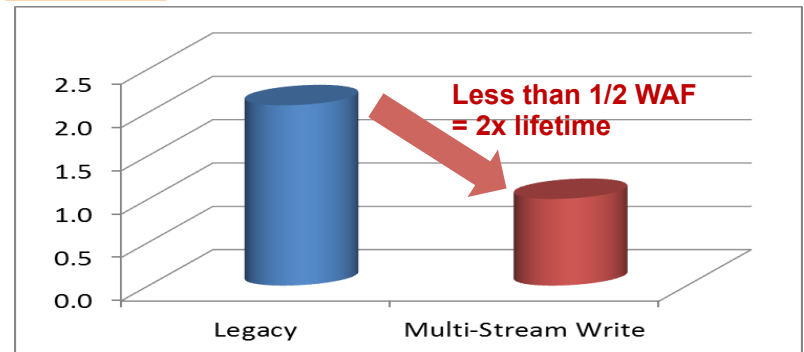 - Jobs: 4
  - Block size: 128k
  - Iodepth: 1

Read IOPS

Multi-Stream Write

1.8x

Legacy

Write Throughput

Multi-Stream Write

2x

Legacy

WAF

Less than 1/2 WAF = 2x lifetime

Legacy   Multi-Stream Write

$$WAF = \frac{Amount\ of\ data\ written\ to\ NAND}{Amount\ of\ data\ written\ by\ host}$$

Cassandra

- Free open-source distributed NoSQL DB
- Provide high availability with no single point of failure
- Support clusters across multiple data centers
- Scalable
- Fault tolerant with automatic replication
- Support query language (CQL: Cassandra Query Language)
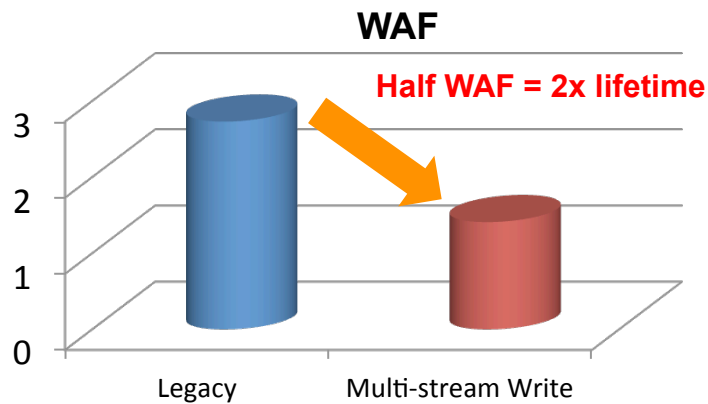

cassandra

Performance Measurement Configuration

- Hardware
  - Dell Precision T7810 Workstation
  - Intel Xeon E5-2630 CPU 2.40GHz
  - 64GB RAM
- Software
  - Ubuntu 16.04 LTS, v4.6.0 Kernel with Multi-Stream Write patch
  - Cassandra 3.5.0 w/ Multi-Stream Write Patch
- Benchmark
  - Cassandra built-in tool (cassandra-stress)
  - 50%/50% Read/Write
  - Total records: 1M
- Device
  - Multi-Stream Write enabled SAS 480GB SSD
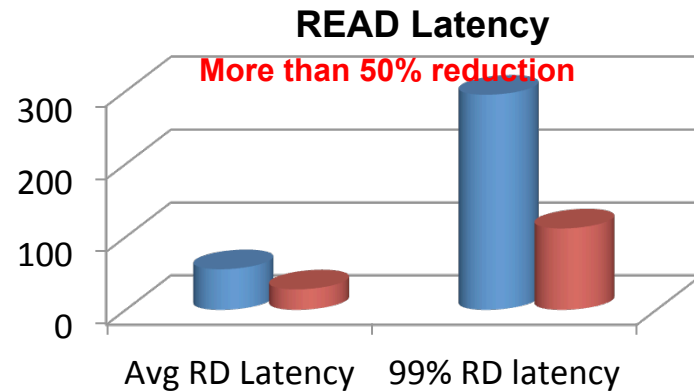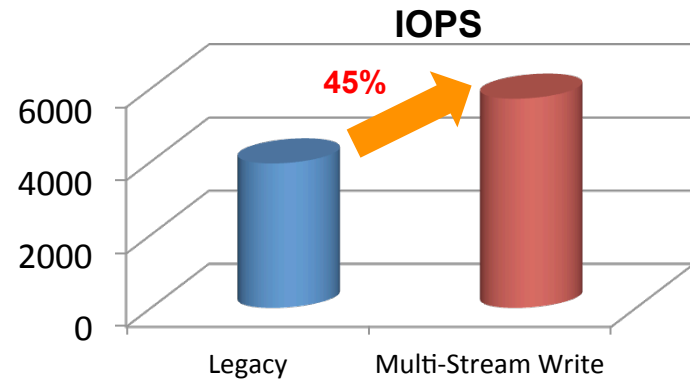 - 4 hour pre-conditioning with 100% write

Cassandra Performance

- 45% performance improvement
- 2x lifetime
- More than 50% READ latency reduction

IOPS

45%

6000
4000
2000
0

Legacy   Multi-Stream Write

WAF

Half WAF = 2x lifetime

3
2
1
0

Legacy   Multi-stream Write

$$WAF = \frac{Amount\ of\ data\ written\ to\ NAND}{Amount\ of\ data\ written\ by\ host}$$

READ Latency

More than 50% reduction

300
200
100
0

Avg RD Latency   99% RD latency

Standards

- SCSI/SAS: Completed in May, 2015
  - Standard spec: http://www.t10.org/cgi-bin/ac.pl?t=f&f=sbc4r10.pdf

- NVMe: standardization in final review stage

Summary

- With Multi-Stream Write, SSDs can be more efficiently used for
  - Consistent better performance
  - Better endurance (=better SSD lifetime)
- With Multi-Stream Write
  - FIO: more than 2x SSD lifetime in addition to the decent I/O performance enhancement
  - Cassandra: 2x SSD lifetime as well as 45% I/O performance improvement
- Multi-Stream Write collateral site
  - http://www.samsung.com/semiconductor/insights/article/25465/multistream