Building efficient RAID-5 systems across SSDs at the FTL Layer

Xueshi Yang

CEO & Co-founder

Shannon Systems

Outlines

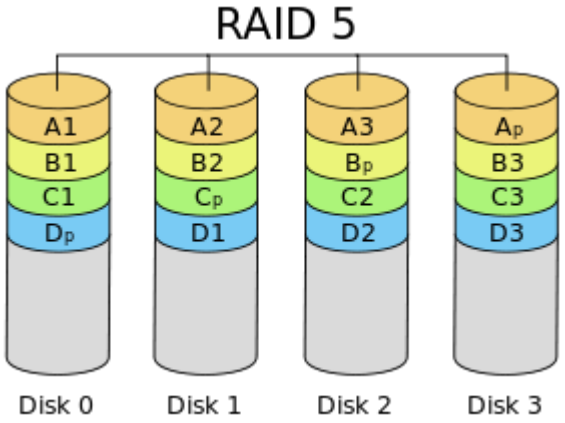
- Motivation
- Conventional RAID5 architecture drawbacks
- Building efficient RAID5 systems across SSDs at the FTL layer
- Conclusions

Motivation

- SSDs provides sufficient IOPS, but RAID-5 is still commonly employed for
  - Higher capacity (e.g., 8 to 12 drives)
  - Increased bandwidth
  - Improved reliability
- Issues:
  - RAID-5 write hole
  - Performance loss due to partial-stripe writes
  - Performance degrades due to performance fluctuation between SSDs

Conventional RAID-5

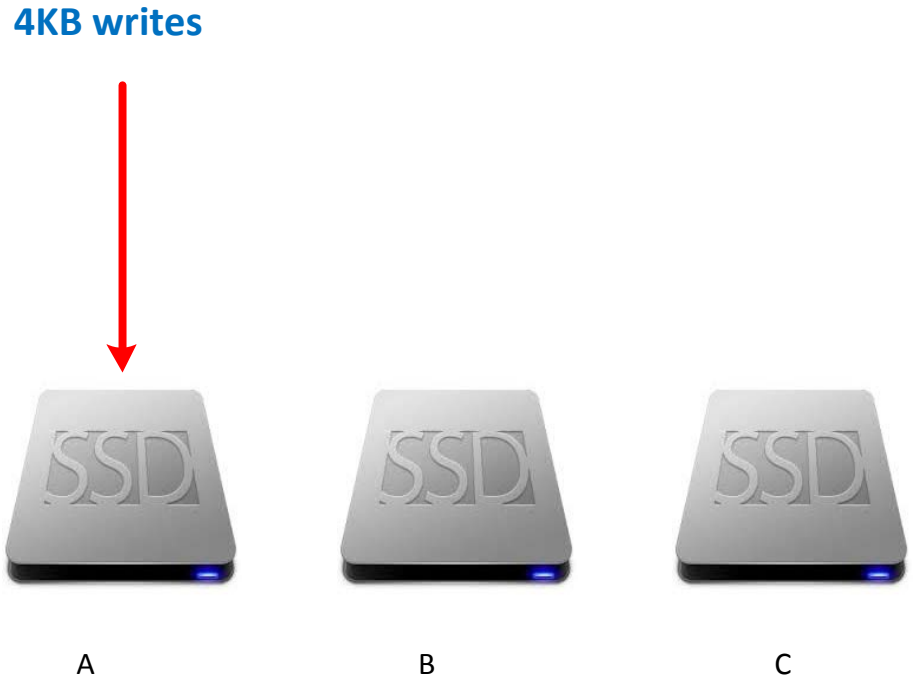
- Rotational parity, resilient to one SSD failure



- Without parity cache, random writes result in
  - Write amplification >>2
  - Faster wear out
  - Write holes

Random writes in RAID-5

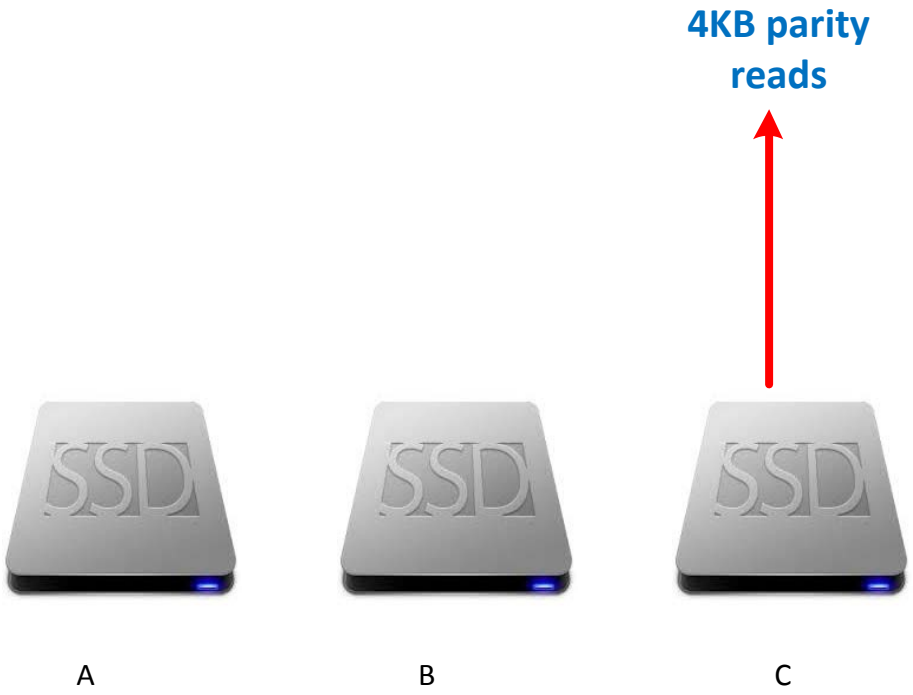
- 4K random write steps -1



Write amplification occurs when Drive-A is full

Random writes in RAID-5

- 4K random write steps -2



Random writes in RAID-5

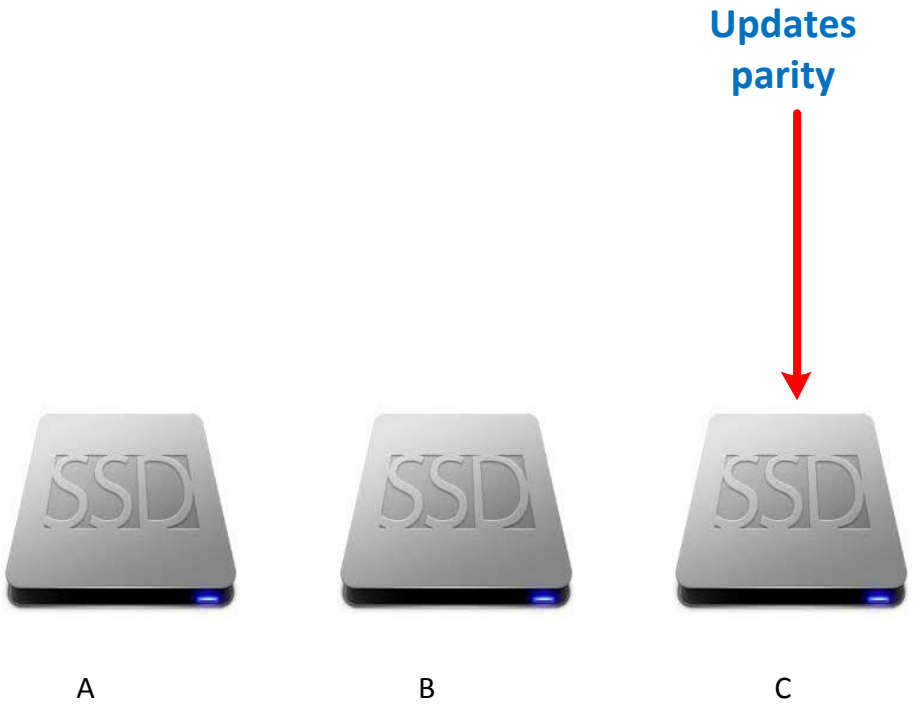
- 4K random write steps -3

Computes updated parity

A B C

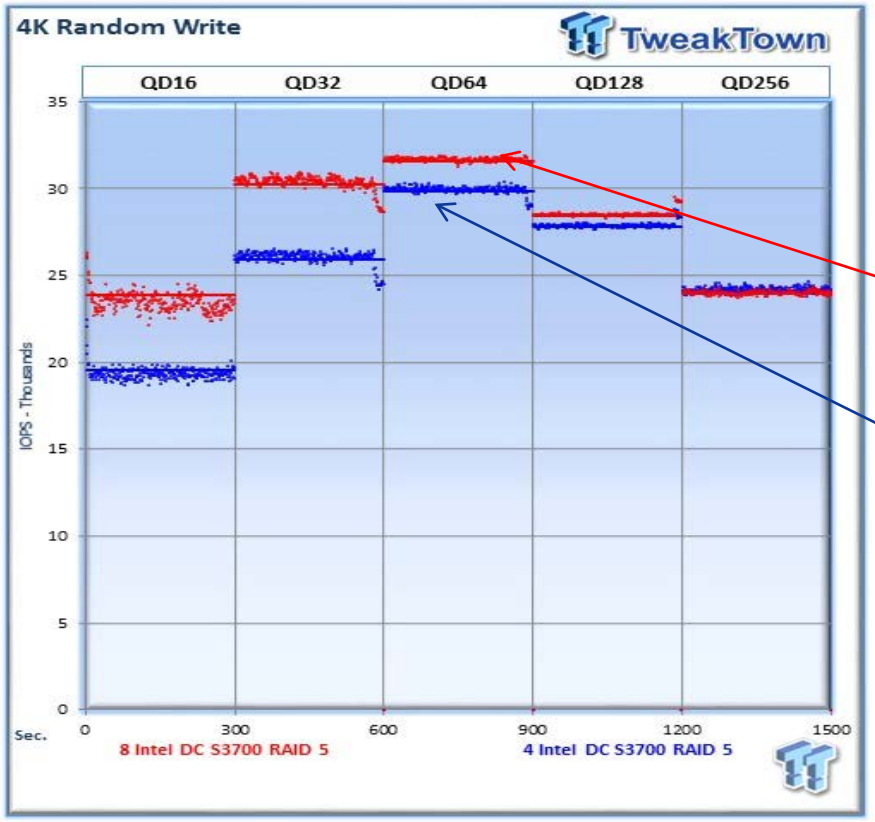
Random writes in RAID-5

- 4K random write steps -4



Performance issues of RAID-5

- Random write performance does not scale linearly to # of SSDs



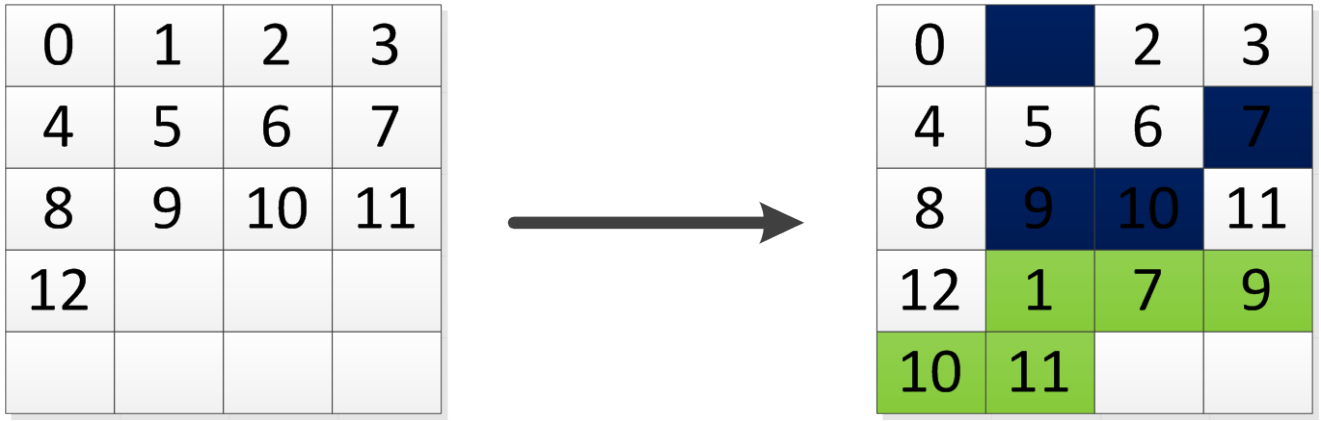
8 Drives (RAID-5)

vs

4 Drives (RAID-5)

Dynamic mapping of LBA and PBA in SSDs

- Conventional RAID controller treats SSD just as another type of HDD
- SSDs are fundamentally different from HDDs viz. dynamic association of LBAs to PBAs

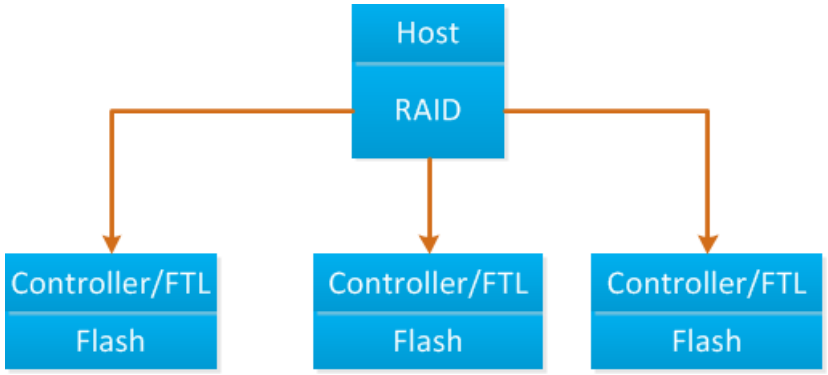


The invariants in SSDs

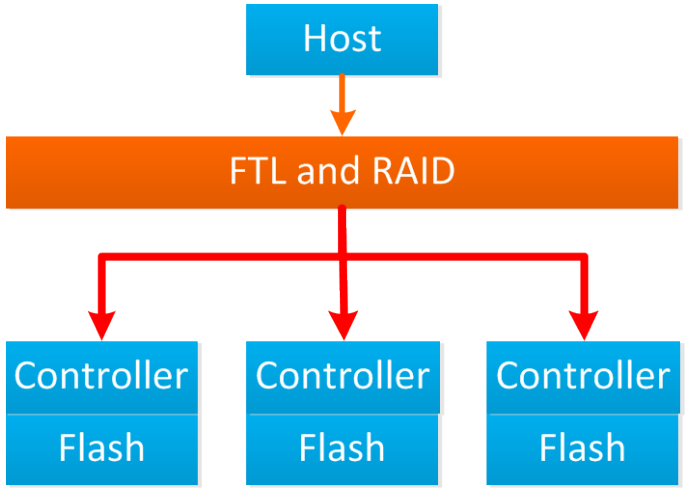
- In SSDs, the invariants are Physical Block Addresses (PBAs)
- RAID could be built upon the invariants of PBAs
- Off-loading FTL from SSDs facilitates RAID construction on PBAs

Off-loading FTL from SSDs

- Off-loading FTL from SSDs
- Unified FTL and RAID layer solves the issues with conventional RAID.



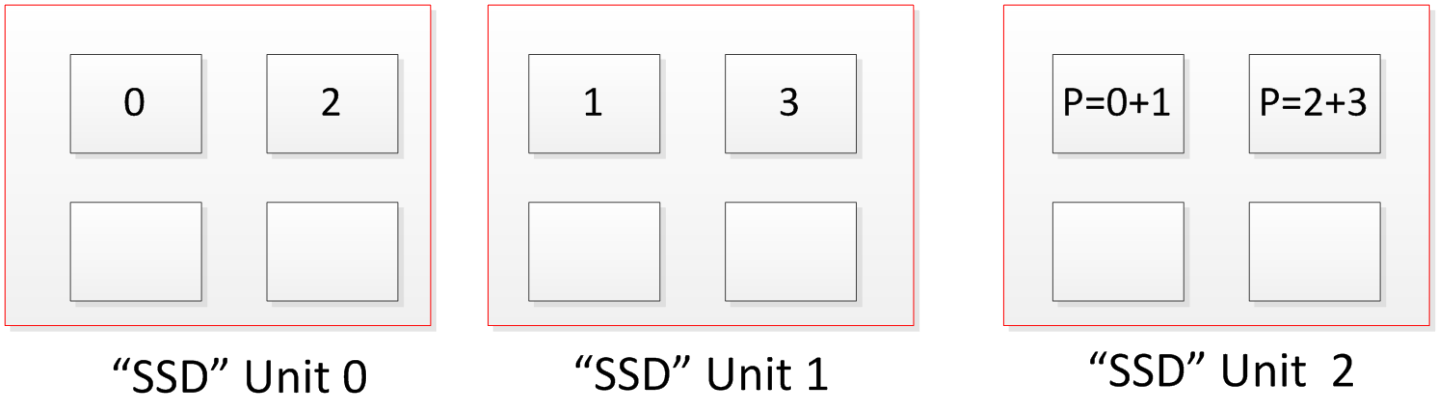
Conventional RAID5 of SSDs



The new RAID5 architecture

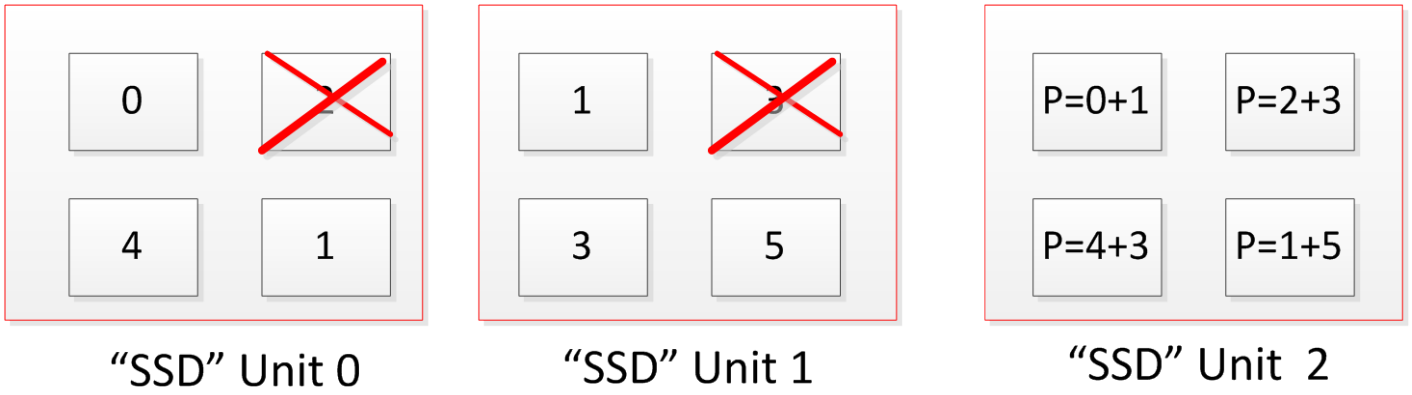
RAID construction at FTL (1)

- Initial state
  - Numbers are LBAs
  - Blocks are PBAs



RAID construction at FTL (2)

- Next state
  - Numbers are LBAs
  - Blocks are PBAs



Advantages

- Joint FTL and RAID-5 layer provide better system efficiency:
  - Global GC and wear leveling
  - Random small writes WAF can be less than 2
  - No more read-modify-write
  - No partial parity stripes write performance loss
  - Write hole can be gracefully mitigated by partial parity writes or filling dummy data.

Hardware accelerated parity calculation

- FTL and RAID are co-located
- FTL are primarily based on software
- Parity computation can be hardware accelerated by seeding raw data to the hardware controller.
  - Parity computation are carried out in hardware
  - Parity are subsequently written out to the parity “SSD”

Shannon Direct-IO™ SFF-8639 PCIe Flash

- Natively PCIe, supports hot plug
- Built in cross-drive RAID with hardware acceleration
- Superior sustained 4KB random write performance

About Shannon Systems

- Shannon Systems is the leading provider of PCIe Flash and associated flash systems in the China market.

- Come to visit us at booth #700-702
  - SFF-8639 PCIe Flash Drive
  - 6.4TB PCIe Flash