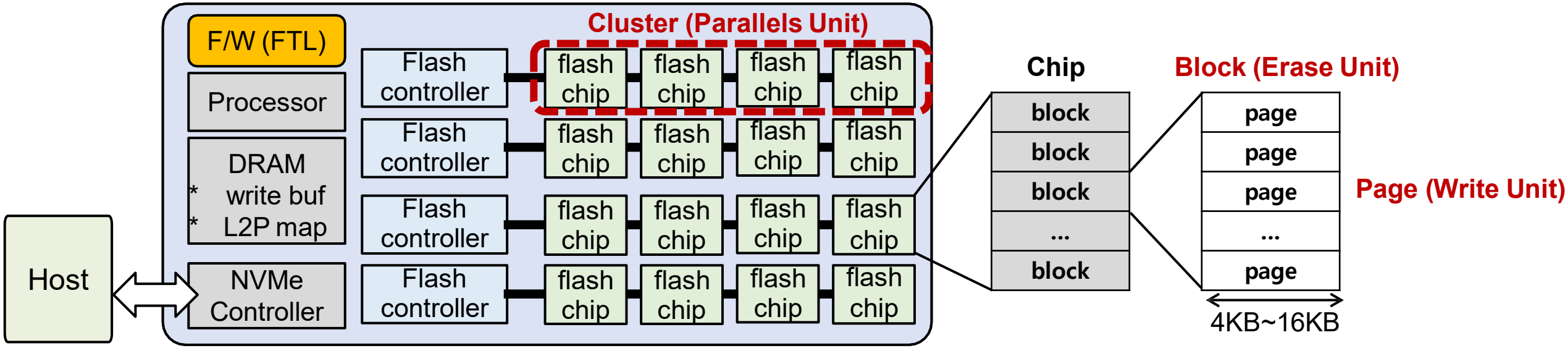
Improving the Reliability of Next Generation SSDs using WOM-v Codes

Shehbaz Jaffer, Kaveh Mahdaviani and Bianca Schroeder

Awarded Best Paper

FAST 22

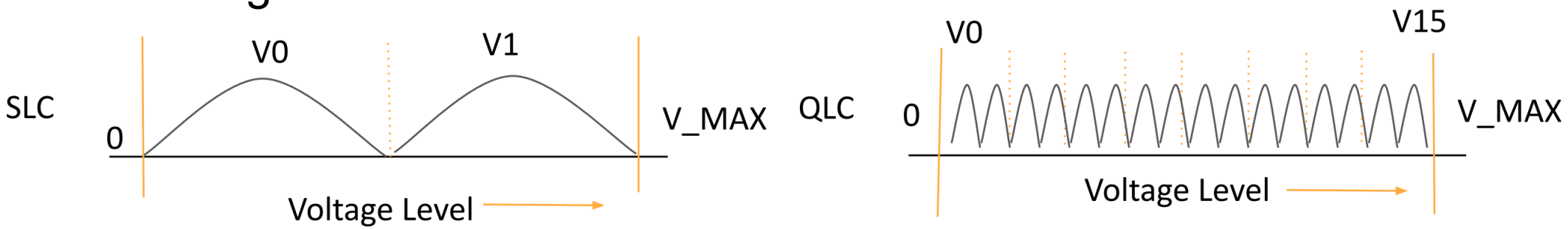
Review: SSD Architecture



- Reading and writing in units of pages
- Erase in units of blocks
- Parallel operation in units of different blocks in the same cluster

SSD Storage Principle

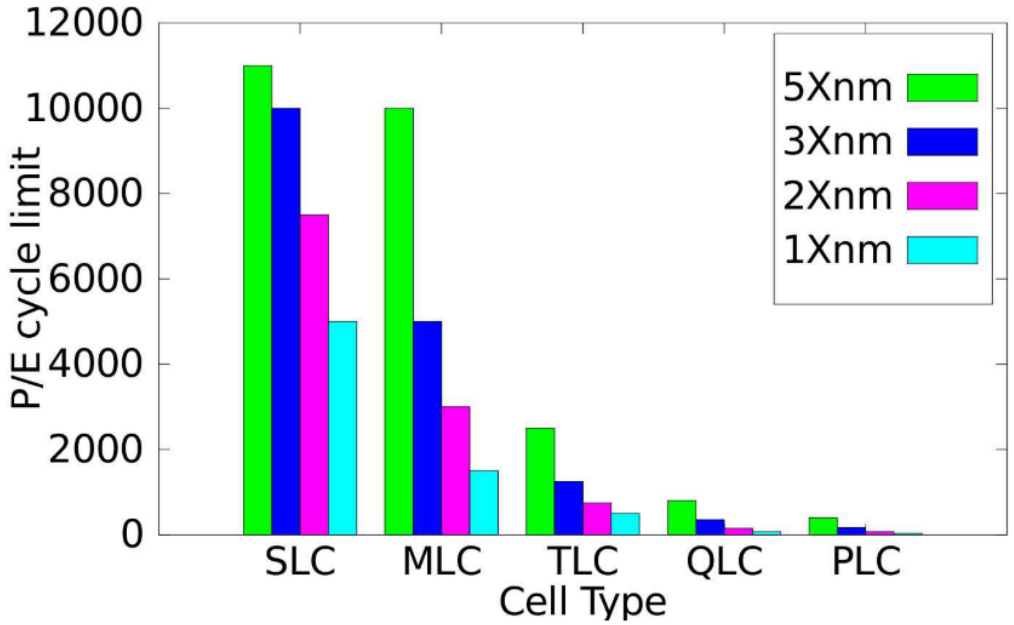
- NAND Flash storage method:
  - Based on MOSFET: Determine whether the stored bit is 0 or 1 based on the voltage



  - MOSFET divides voltage range in 1.8~3.3V and discharges over time
  - Voltage can only be increased (Program) and reset to zero (Erase)

- The Problem: Fine-grained voltage control leads to reduced lifetime and performance

Motivation

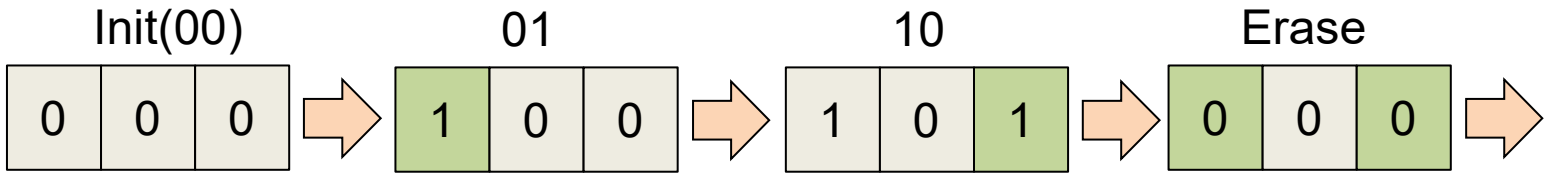


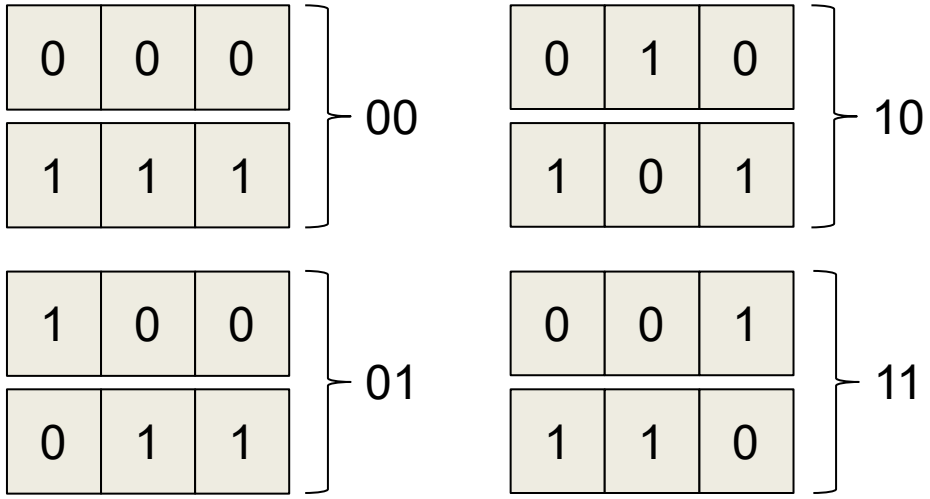
- QLC has entered the data center, PLC is on the way
- As process becomes more advanced, the P/E cycle limit is getting smaller

➢ Reduced Program and Erase (P/E) cycles by Overwrite between Erase: Write Once Memory Code

Binary-WOM Codes

- Work on bit-level, each bit can only be written in one direction

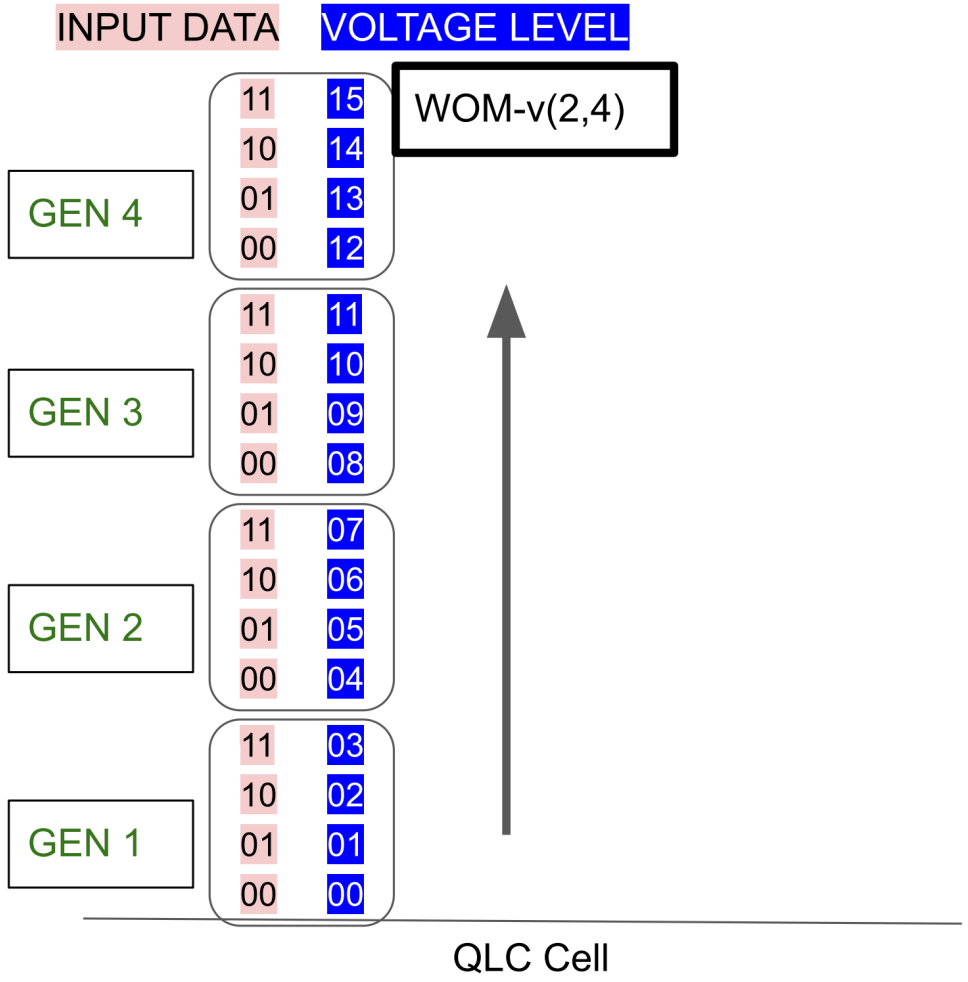




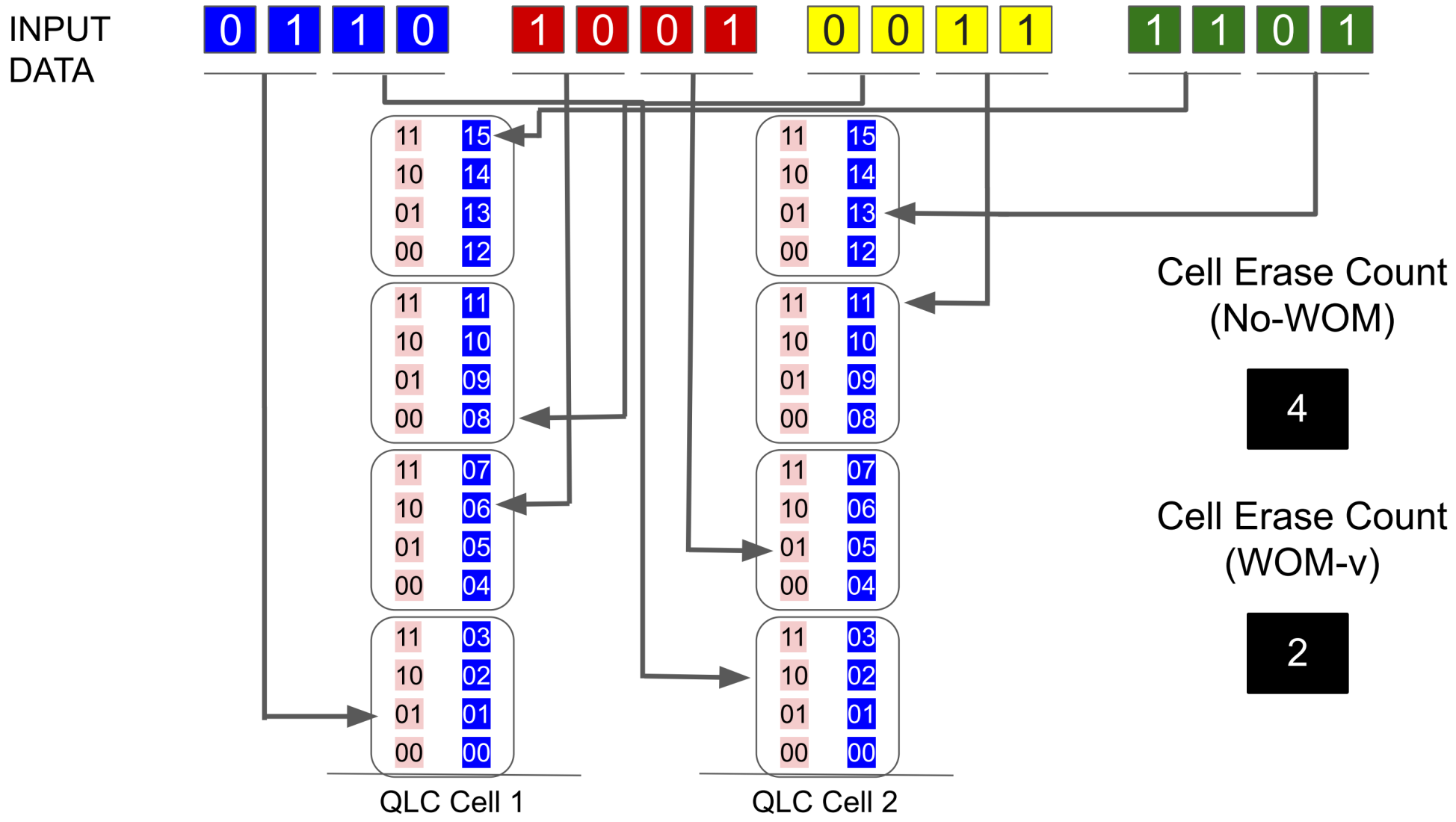
- Each group could used for twice before erase
- Each voltage in QLC can only be increased and cannot be modified by a single bit

WOM-v Codes

- Voltage-based WOM Codes
  - WOM-v (k,N) Codes set N to bit number per cell, and k to 1/2/3 for different configuration
  - Tradeoff between coding efficiency and erasing times

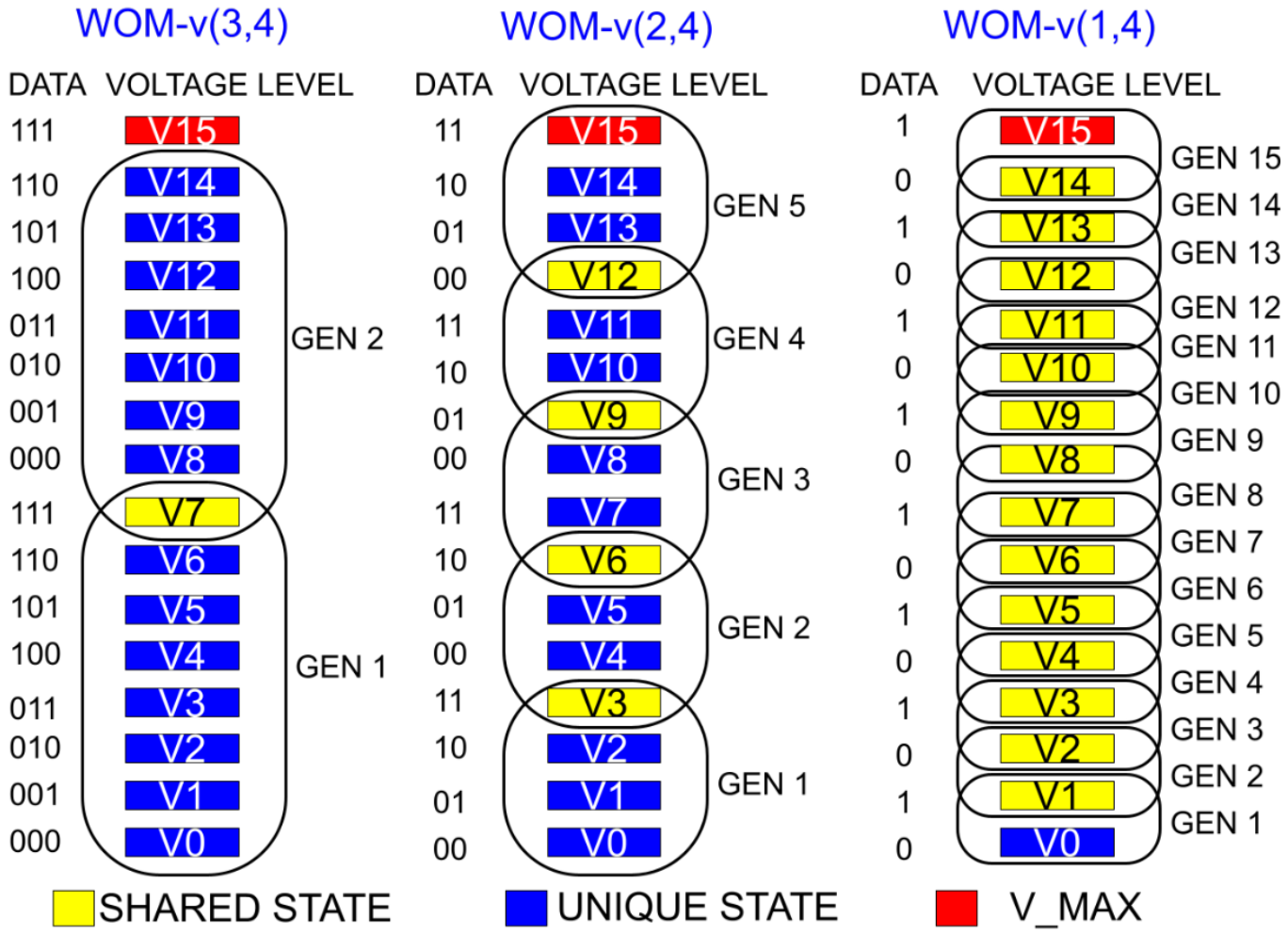


WOM-v Codes


INPUT DATA
0 1 1 0
1 0 0 1
0 0 1 1
1 1 0 1
11 15
10 14
01 13
00 12
11 11
10 10
01 09
00 08
11 07
10 06
01 05
00 04
11 03
10 02
01 01
00 00
QLC Cell 1
QLC Cell 2
Cell Erase Count (No-WOM)
4
Cell Erase Count (WOM-v)
2

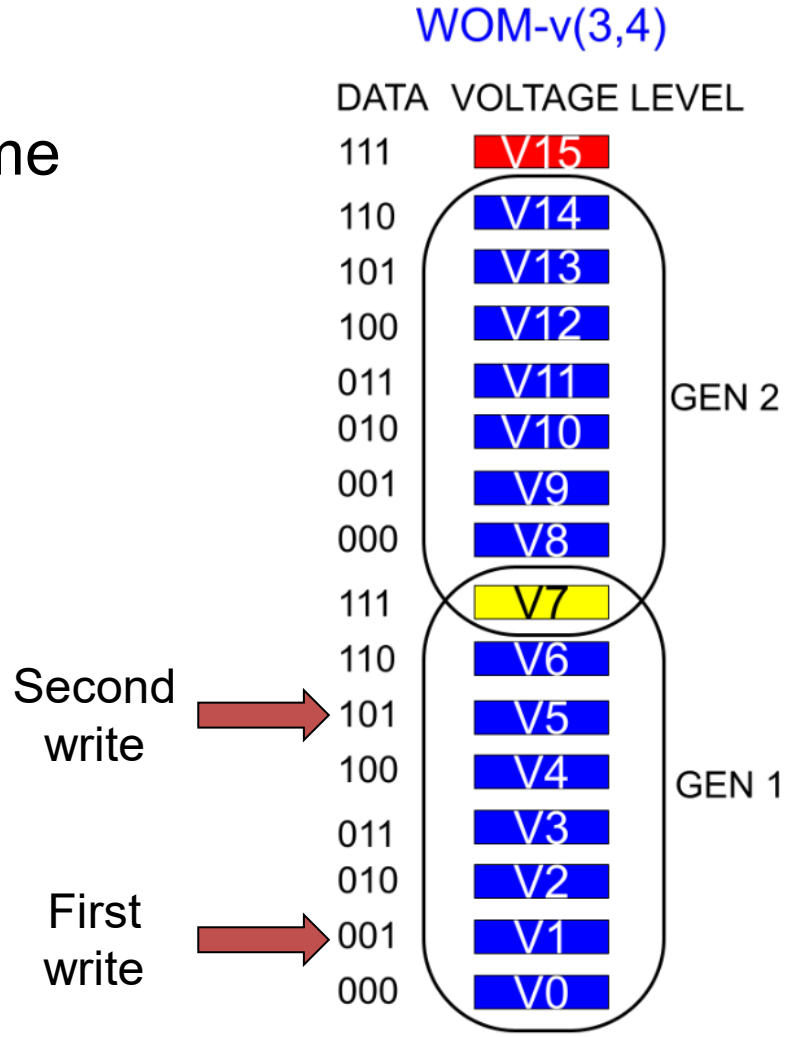
Optimize WOM-v Codes

- Share code words
  - In WOM-v(2,4) + 1 GEN
  - In WOM-v(1,4) + 7 GEN
- More generations



Optimize WOM-v Codes

- Same-generation transitions
  - Only new data has a lower voltage in the same generation then migrate to new generation
- Writeable times > generation number

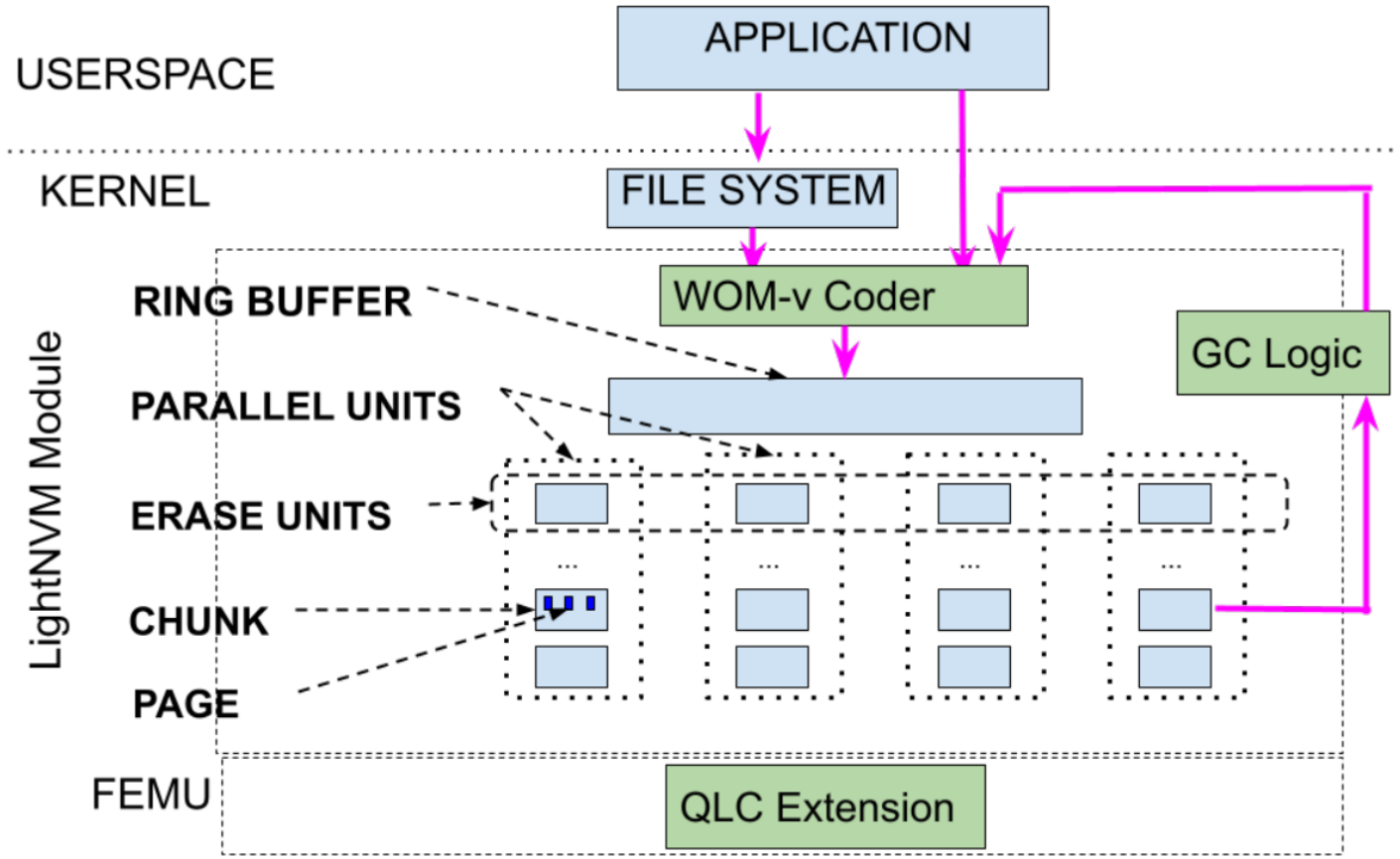


Optimize WOM-v Codes

- Use ECC to increase pages write between erase
  - The unit of write is pages. When a cell in the page reaches the maximum generation, it can't be overwritten
  - Observation I: The generation of each cell in each page is relatively average, and only a few cells will reach max generation first
  - Observation II: SSD itself uses ECC to solve bit errors, and ECC is over-provisioned in the early stage of SSD life
- Mark the cell as invalid, and read it through ECC
  - Faster recovery than random bit errors due to location determination

Deploy WOM-v Codes to QLC SSDs

- Use LightNVM Module to control SSD flash chips
  - One of the Open Channel SSD implementation in Linux Kernel



Optimizations

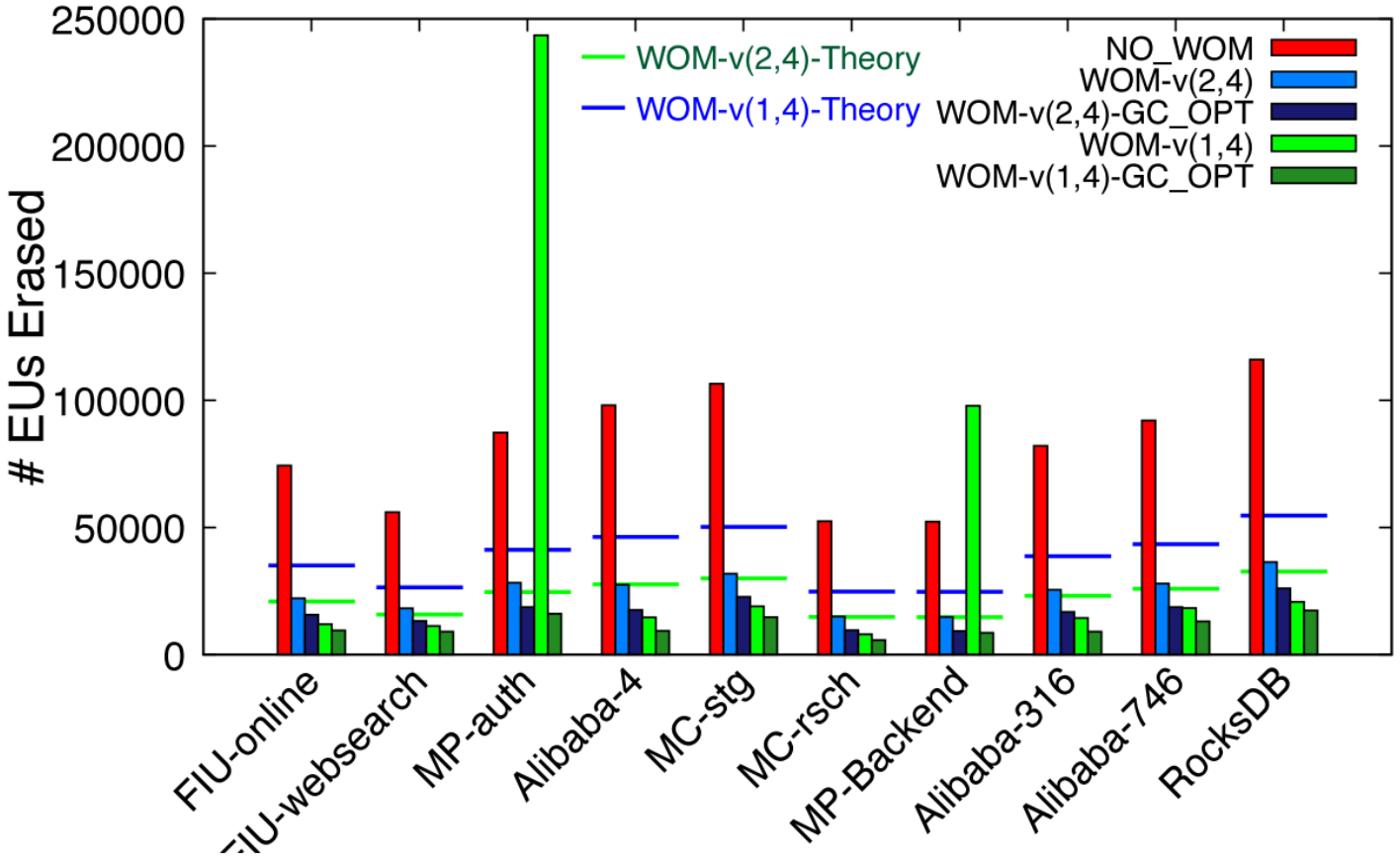
- No-Read Mode
  - Disable same-generation transition to reduce read-before-write performance overhead
  - Trade-off between P/E cycles and SSD write performance
- GC_OPT Mode
  - WOM-v codes allow many pages in EU to be rewritten, but the GC will copy them for reclaim space
  - During GC, the contents of valid pages are not copied to other locations, and not erase the blocks with valid pages

Evaluation

- Platforms
  - FEMU simulated SSDs
  - Add 445 LOC in LightNVM and 240 LOC in FEMU
- Traces
  - Block traces:

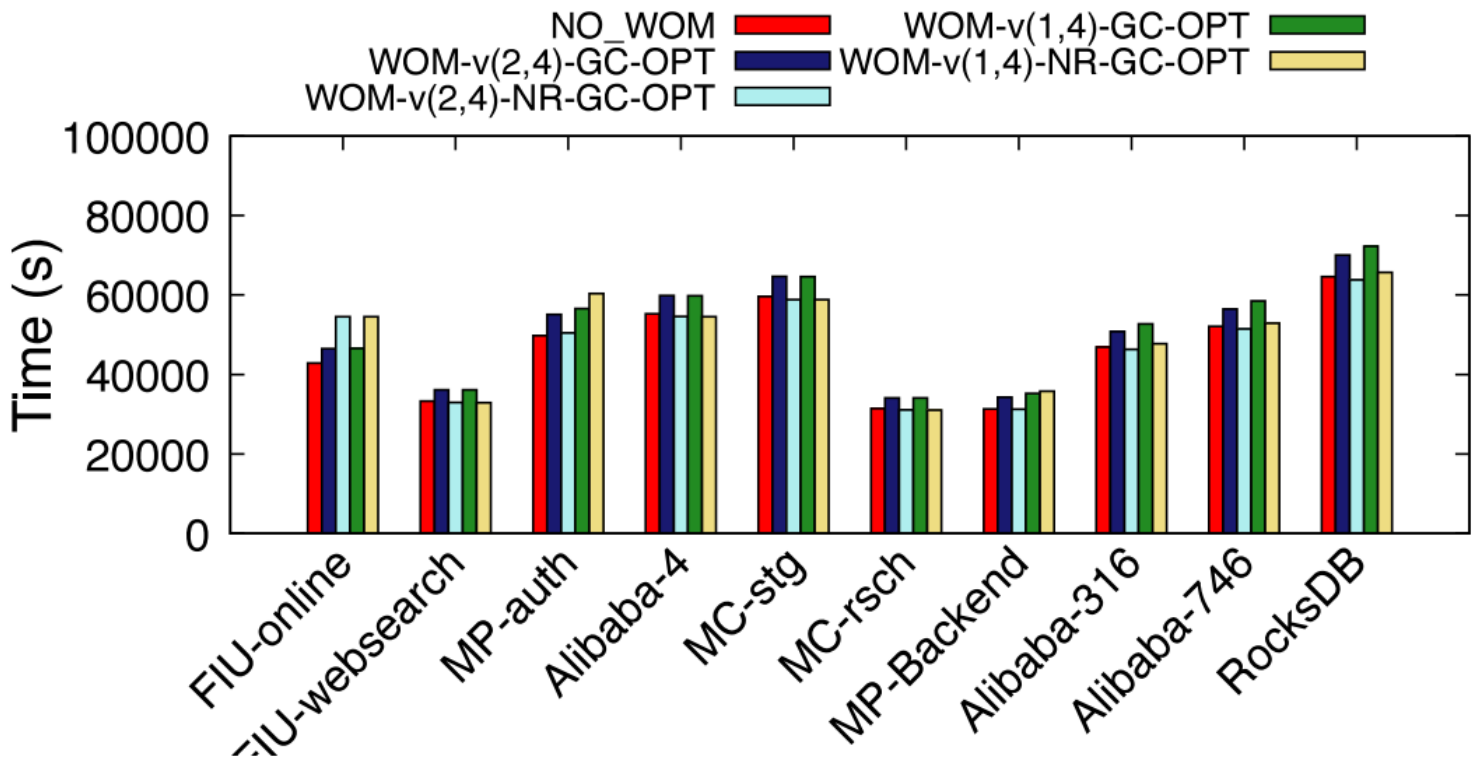
Source	# Traces	Medium	Year
Alibaba [18]	814	SSD	2020
RocksDB/YCSB Trace [25, 30]	1	SSD	2020
Microsoft Cambridge [22]	11	HDD	2008
Microsoft Production [15]	9	HDD	2008
FIU [16]	7	HDD	2010

Erase Operation Reduction



4.4 - 11.1x reduction in erase cycles

Write Performance



< 8% write performance overhead

Read Performance

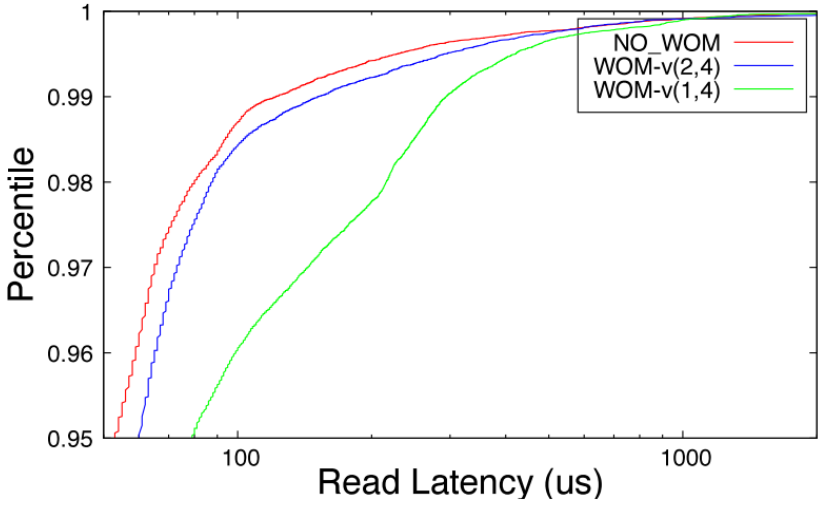


Figure 10: *MSR-Cambridge Web1*

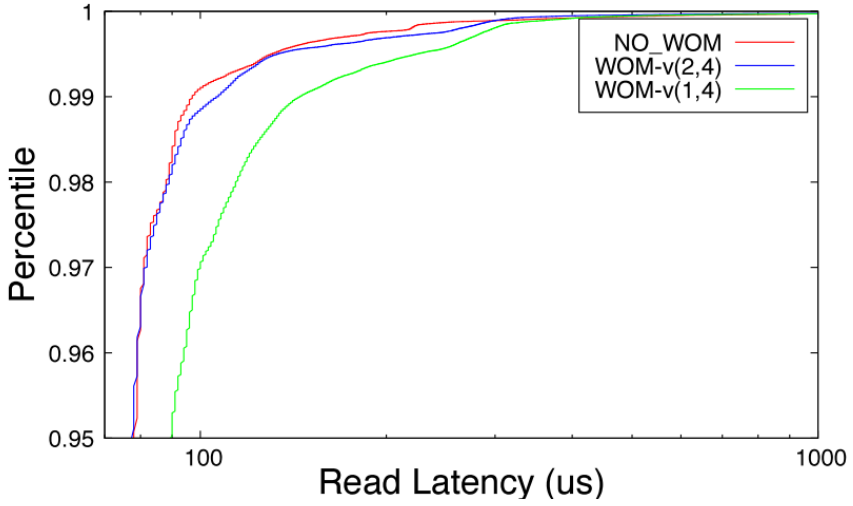


Figure 11: *MSR-Production DAPL*

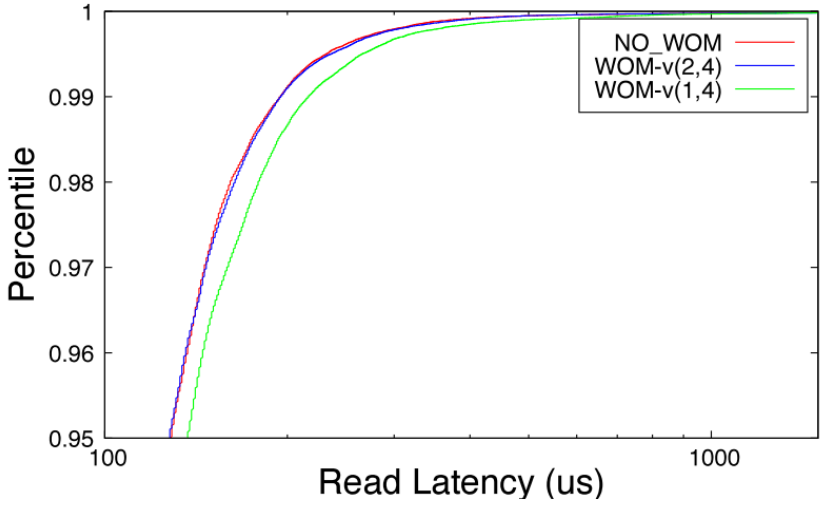
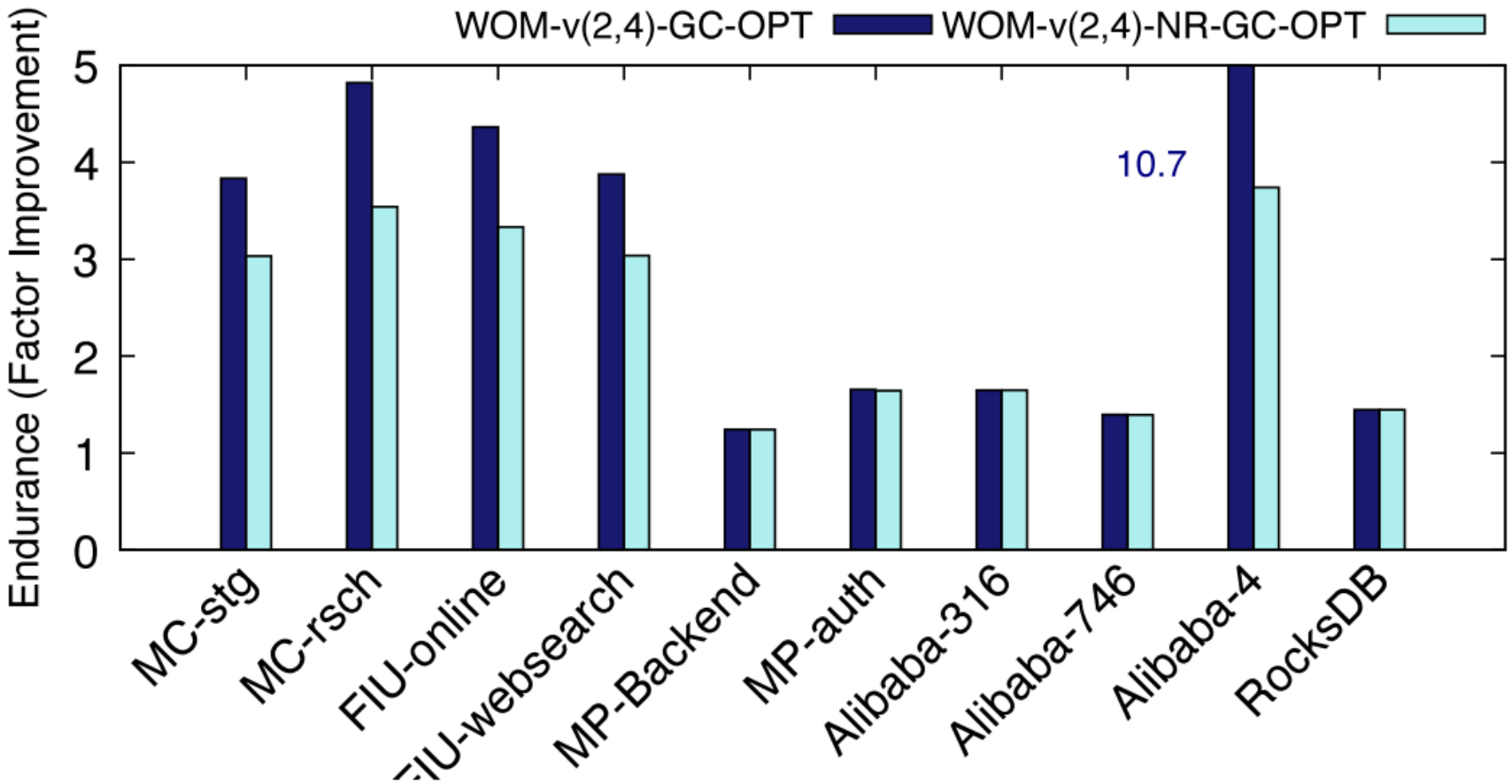


Figure 12: *Alibaba 3*

0.6-22% read performance overhead

Extend to MLC SSDs



- Lifespan: MLC(10K) > QLC(3K)
- MLC: 2-bit per cell
- QLC with WOM-v(2,4): 2-bit per cell

WOM-v(2,4) QLC endurance > MLC endurance

Conclusion

- The Problem: The trade-off between SSD life and capacity
- Main Idea: Map different voltage values to the same content to increase the number of write generations to the same cell
- Key Designs:
  - WOM-v with share code words for more write GEN on each cell
  - Same-generation transition to write more times in same GEN
  - ECC-based block erase reduction by mark cell with max GEN as invalid
- Result: 4.4-11.1x reduced erase operations with minimal performance overheads.