

CR5M: A Mirroring-Powered Channel-RAID5 Architecture for An SSD

Yu Wang¹, Wei Wang², Tao Xie², Wen Pan¹, Yanyan Gao¹,
Yiming Ouyang¹,

¹Hefei University of Technology

²San Diego State University

The 30th International Conference on Massive Storage Systems and Technology (MSST
2014), California, 2014



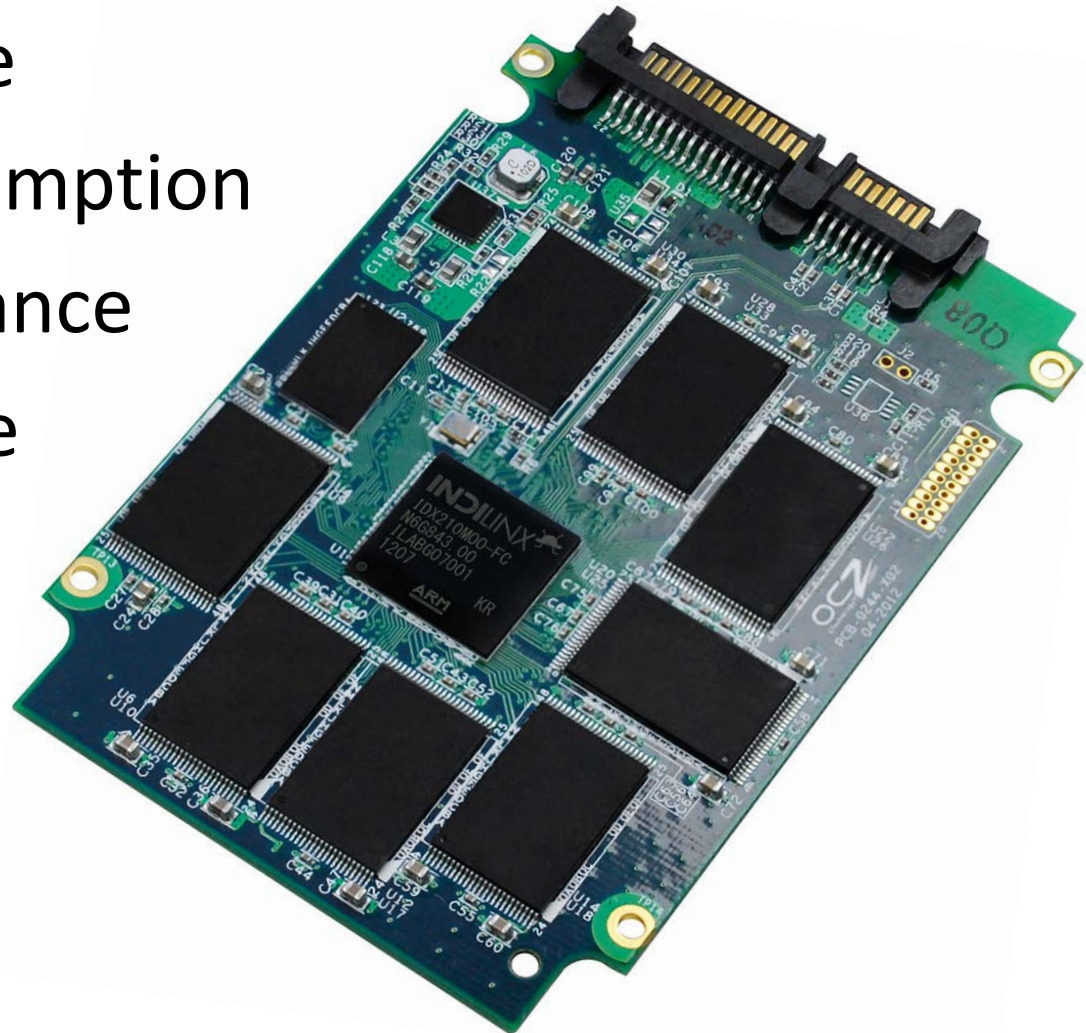
Outline

- Introduction
- Design and implementation
- Experimental results
- Conclusions



Introduction of NAND Flash

- High performance
- Low-power consumption
- High shock resistance
- Small physical size





Introduction of NAND Flash

- Increasing Flash Capacity Density
 - Smaller geometries of flash memory cell
45nm~20nm
 - More bits each cell store
SLC~TLC
- Decreasing Endurance and Reliability
 - SLC ~100k P/E cycles
 - MLC ~10k P/E cycles





Introduction of NAND Flash

- Flash Memory Errors
 - Transient (or soft) Errors
 - Permanent (or hard) Errors
- ECC (Error Correction Code)

Per 256 to 512 bytes, ECC typically can

- Detect two bit errors
- Correct one bit error

Errors beyond that range may be unrecoverable.





Introduction of NAND Flash

ECC are incapable of correcting these errors:

- Word line errors
- Block or die errors
- Multiple-bit transient errors





RAID

RAID has successfully been implemented in

- HDD arrays
- SSD arrays

Im and Shin proposed a Delayed Partial Parity Scheme for Reliable and High-Performance Flash Memory SSD (MSST2010)

Kadav et al. presented Diff-RAID, a new RAID variant that distributes parity unevenly across SSDs to create age disparities within arrays (ACM Transactions on Storage 2010)

- HDD+SSD hybrid arrays



Channel-RAID (CR) - Requirement

- Cases where only one SSD can reliability is critical.

Such as:

- a) Wireless Healthcare System
- b) Mobile Military Application



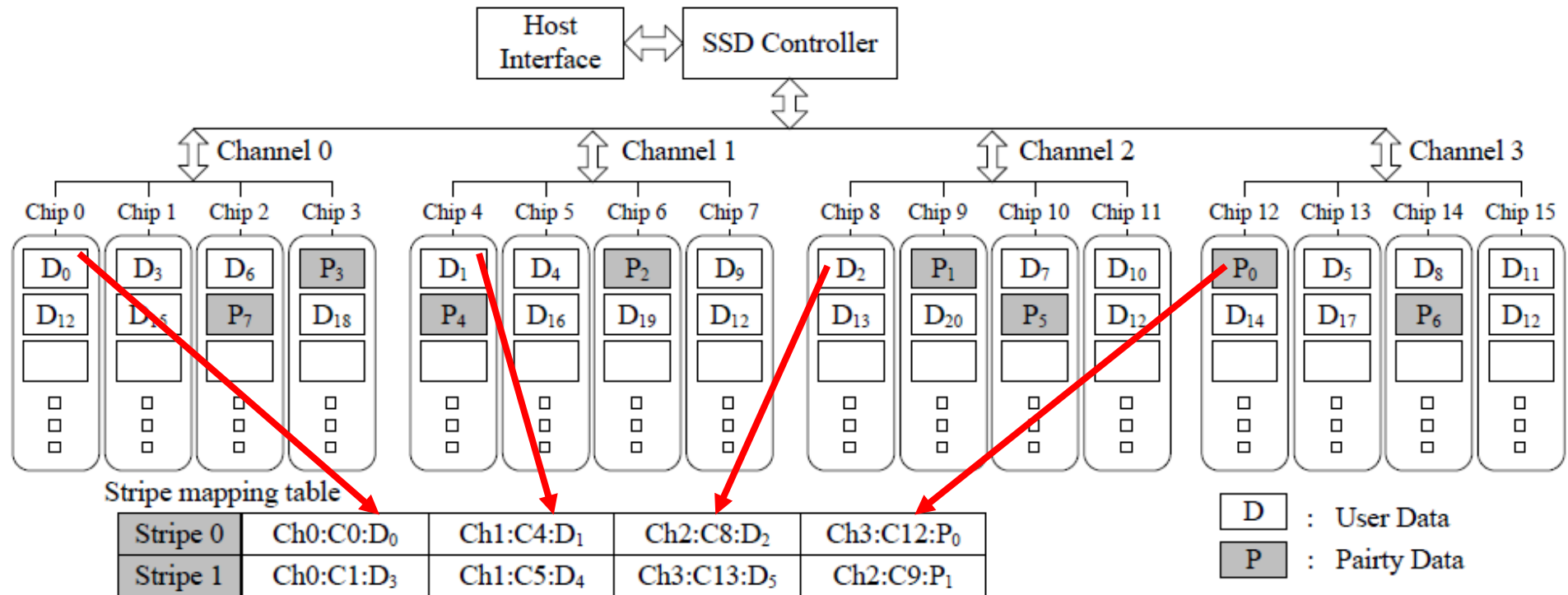


Channel-RAID (CR) - Feasibility

- The multi-channel structure provides an opportunity to implement RAID into a single SSD
 - CR1 (Channel-RAID1)
 - CR4 (Channel-RAID4)
 - CR5 (Channel-RAID5)



CR5 (Channel-RAID5)



- Striping size is adjusted to (N-1) page size
 N means the number of channels.



CR5 (Channel-RAID5)

- Full-Stripe Write: no extra read operation
- Partial-Stripe Write
 - RMW (Read-Modify-Write): reads the old data of the updates and its associated parity.
 - RCW (Read-Reconstruct-Write): reads the rest part of the stripe (i.e., the data that are not going to be updated).

The method whose pre-read operation number is less will be selected.





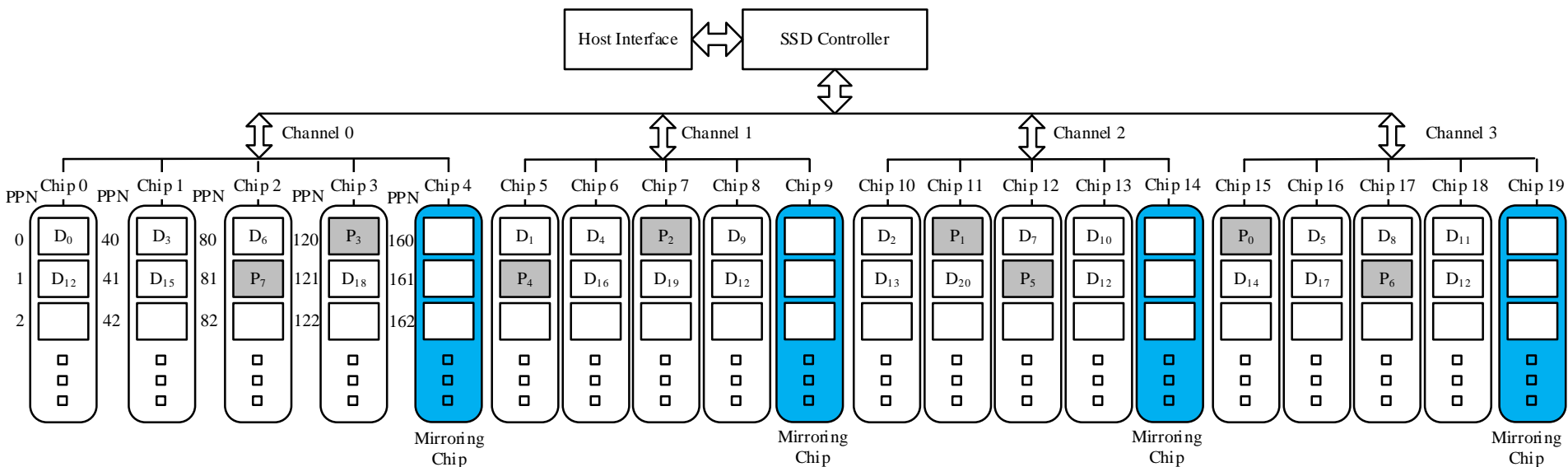
Limitations of CR5 SSD

- Decreased Lifetime
- Degraded Performance
- Vulnerability



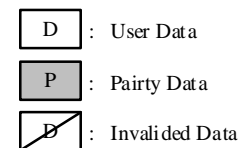
CR5M (Mirroring-Powered Channel-RAID5)

- The key feature: an extra chip is introduced to each channel serve as a mirroring chip.



Stripe mapping table

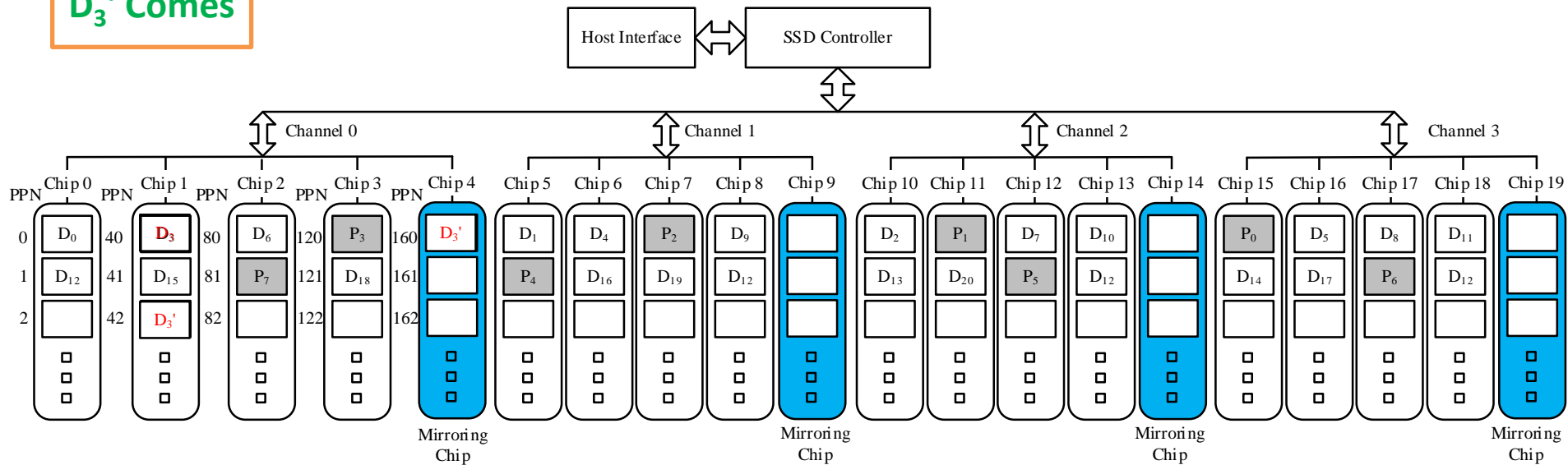
Stripe 0	Ch0:C0:D ₀	Ch1:C5:D ₁	Ch2:C10:D ₂	Ch3:C15:P ₀
Stripe 1	Ch0:C1:D ₃ '	Ch1:C6:D ₄	Ch3:C16:D ₅	Ch2:C11:P ₁



MW (Mirroring Write)

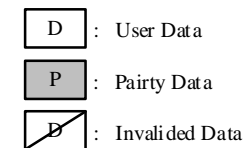
- MW concurrently writes both the original update and a copy of it onto its destination chip and the mirroring chip.

D₃' Comes



Stripe mapping table

Stripe 0	Ch0:C0:D ₀	Ch1:C5:D ₁	Ch2:C10:D ₂	Ch3:C15:P ₀
Stripe 1	Ch0:C1:D ₃ '	Ch1:C6:D ₄	Ch3:C16:D ₅	Ch2:C11:P ₁

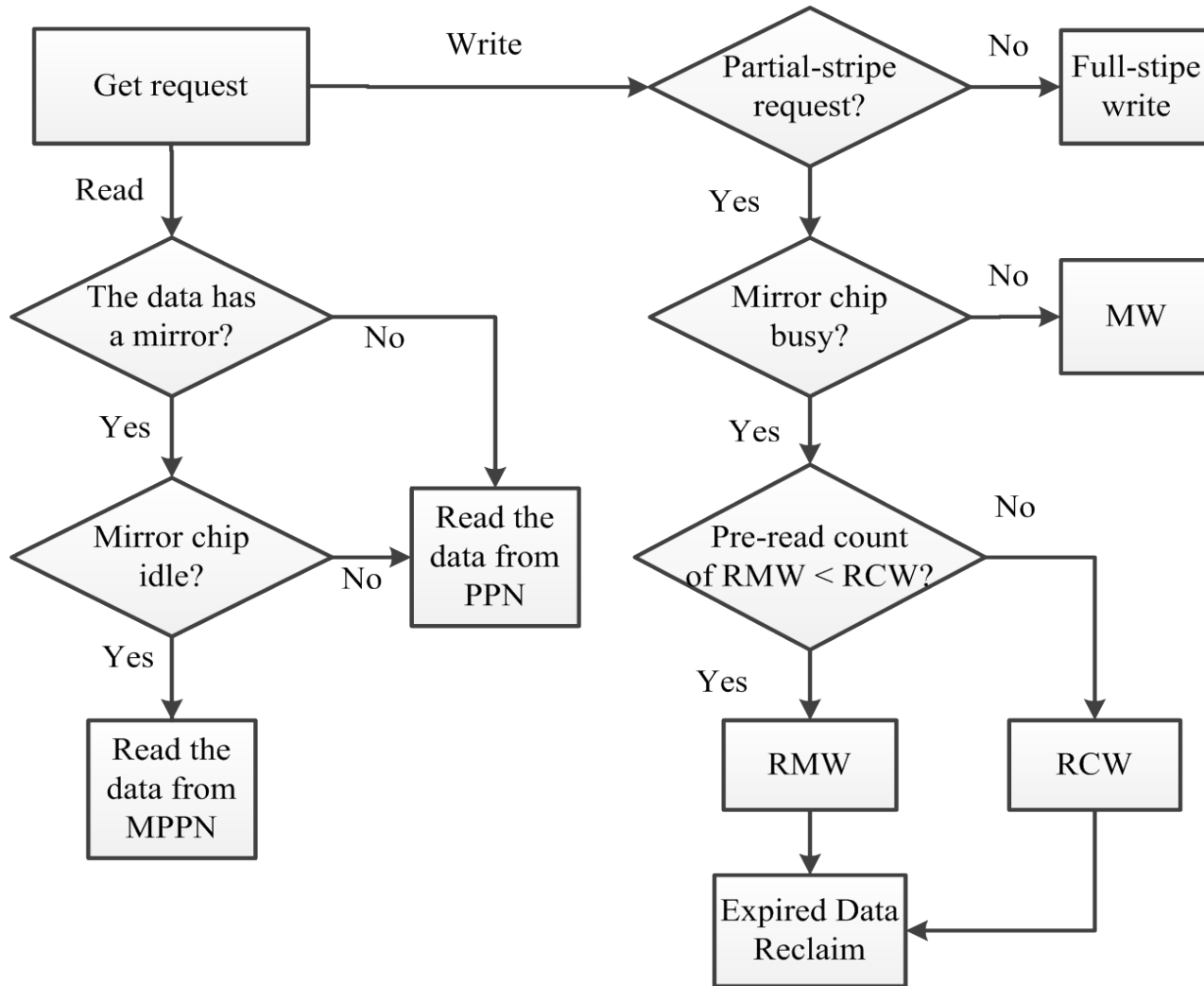


Revised Mapping Table

- Mirroring Address (MA) is appended to each entry. Its value tells the existence of mirroring data for current entry.

Revised mapping table			Mirroring table	
LPN	PPN	MA	EPPN	MPPN
D ₃ '	42	●	40	160
D ₄	240	Null	80	161
D ₂	640	Null		
D ₆ '	82	●		
D ₇	1046	Null		
P ₂	322	Null		

Workflow of CR5M



Experimental Setup

- The Characteristics of Traces

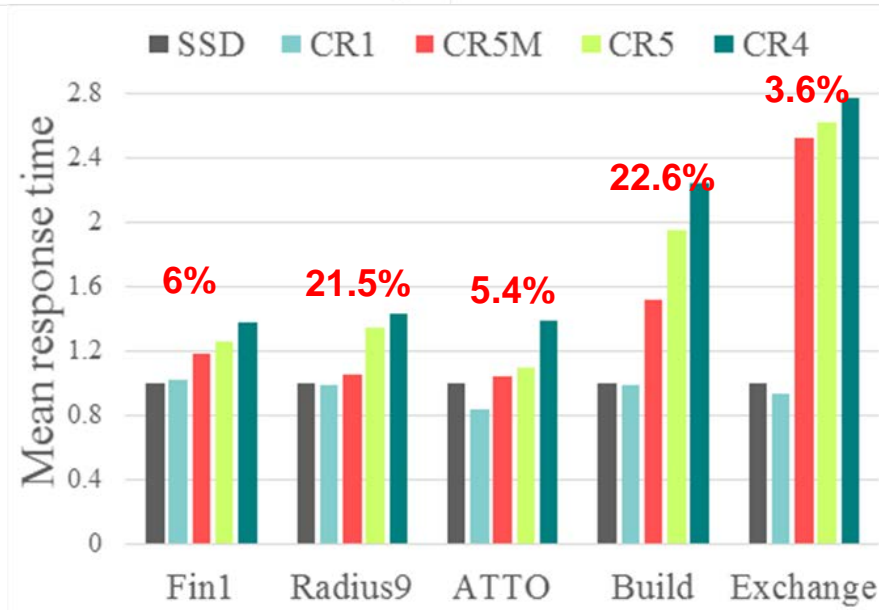
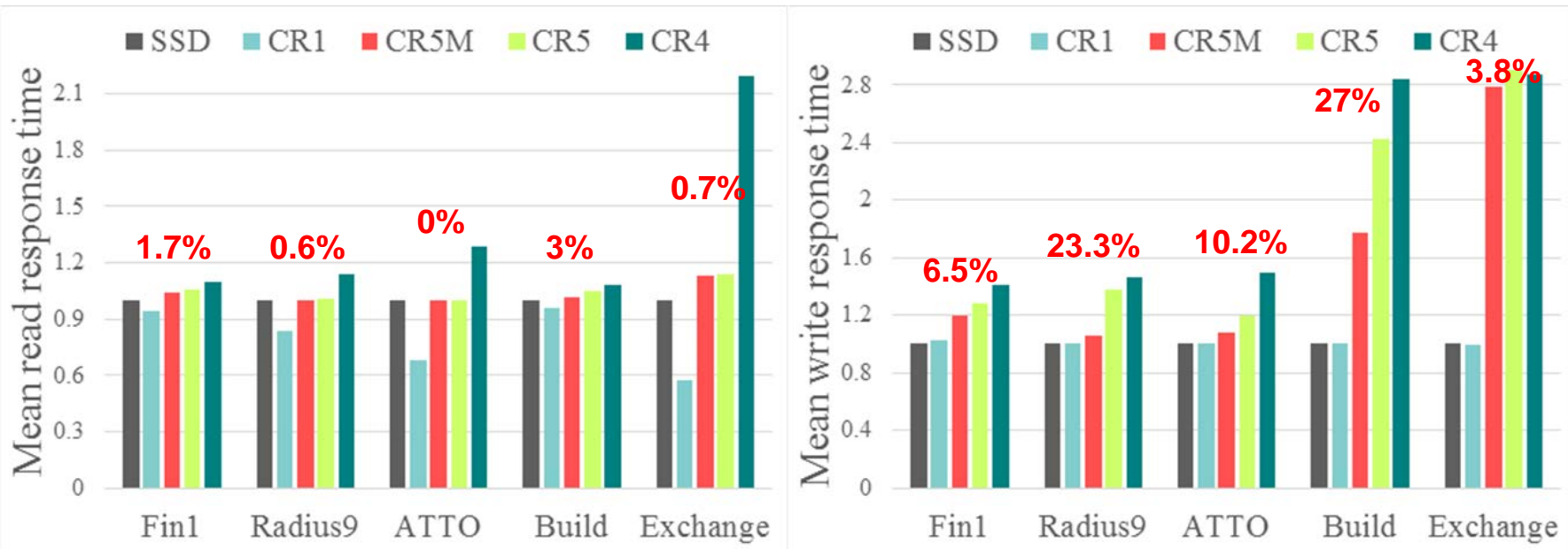
Trace Name	Write Ratio (%)	Ave.Size (KB)	Access Rate (req/sec.)	Duration (mins.)
Financial1	77.88	3.46	129	515
Radius9	88.46	6.8	57	35.2
ATTO	47.45	23.1	792.4	2.5
Build	45.71	6.5	372	15
Exchange	46.43	12.5	166	15

- The Varied Experiment Parameters

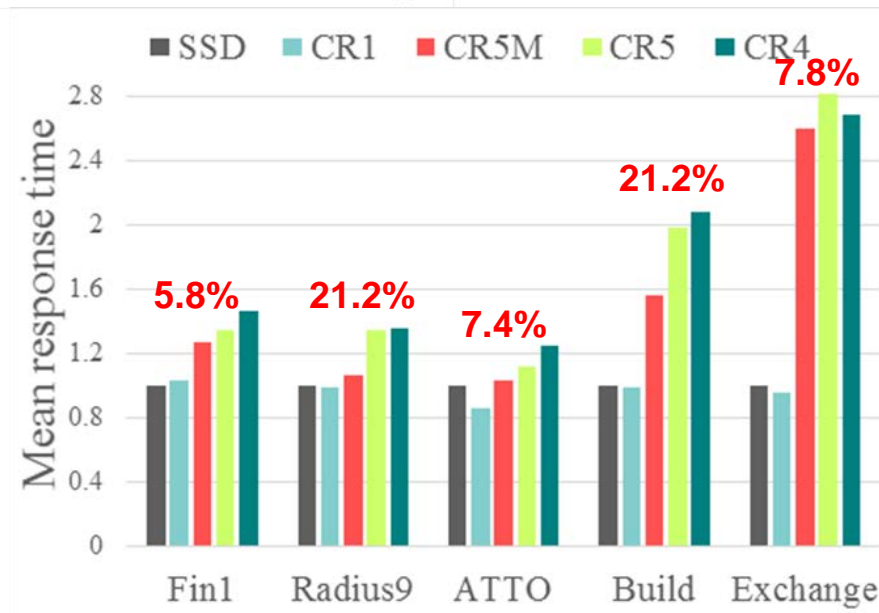
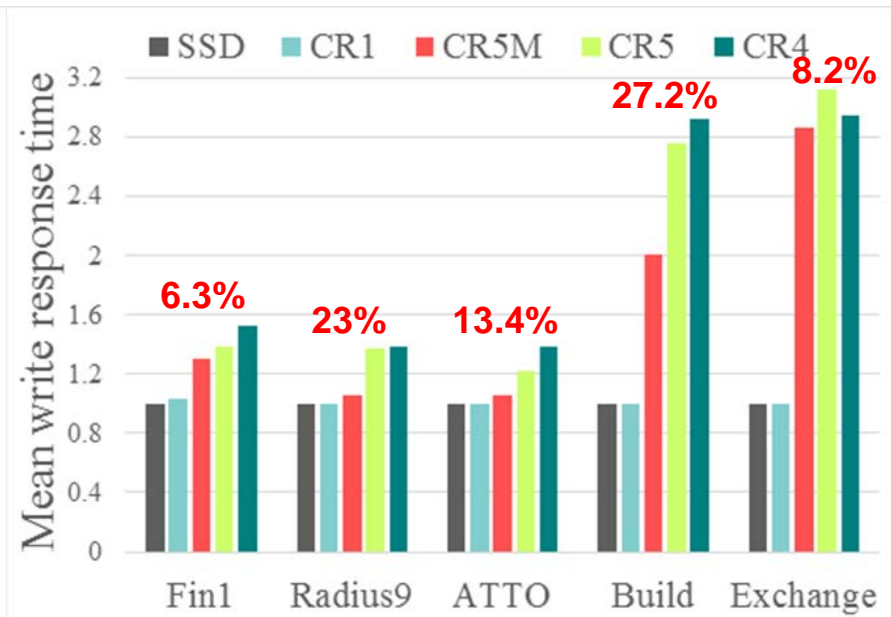
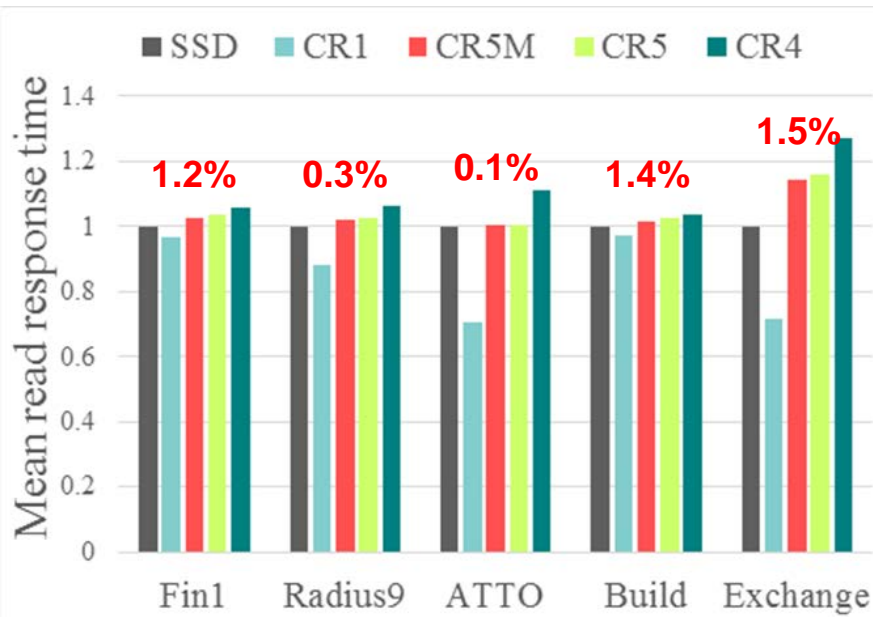
Conf.	Pure SSD	CR1	CR4 & CR5	CR5M
SSD1	4cl-6cp	8cl-6cp	4cl-6cp	4cl-7cp
SSD2	6cl-4cp	12cl-4cp	6cl-4cp	6cl-5cp
SSD3	8cl-3cp	16cl-3cp	8cl-3cp	8cl-4cp

cl: the channel number in an SSD
cp: the chip number on each channel

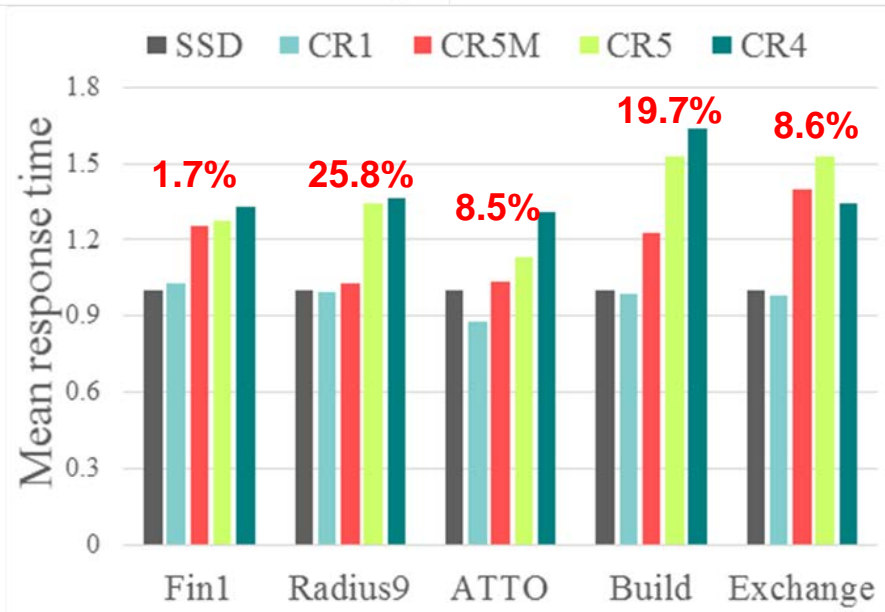
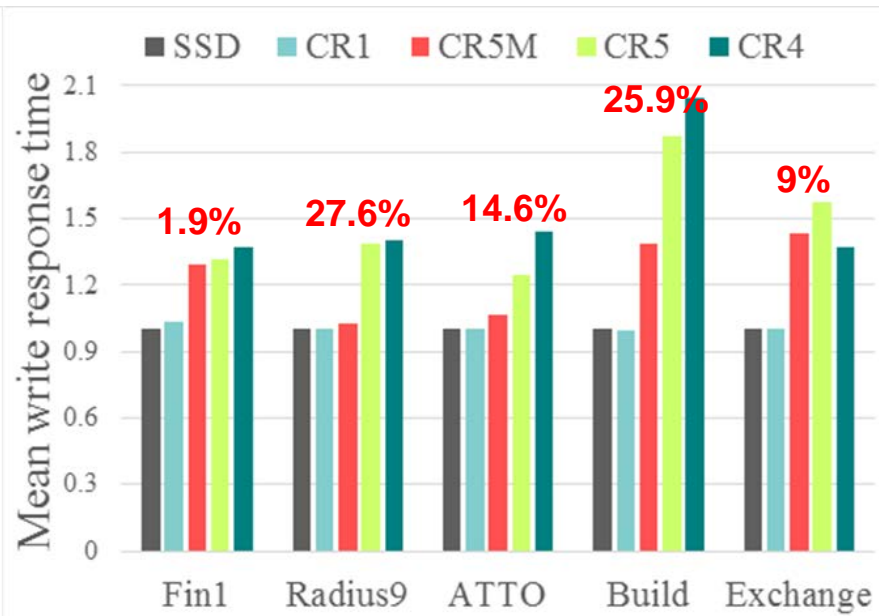
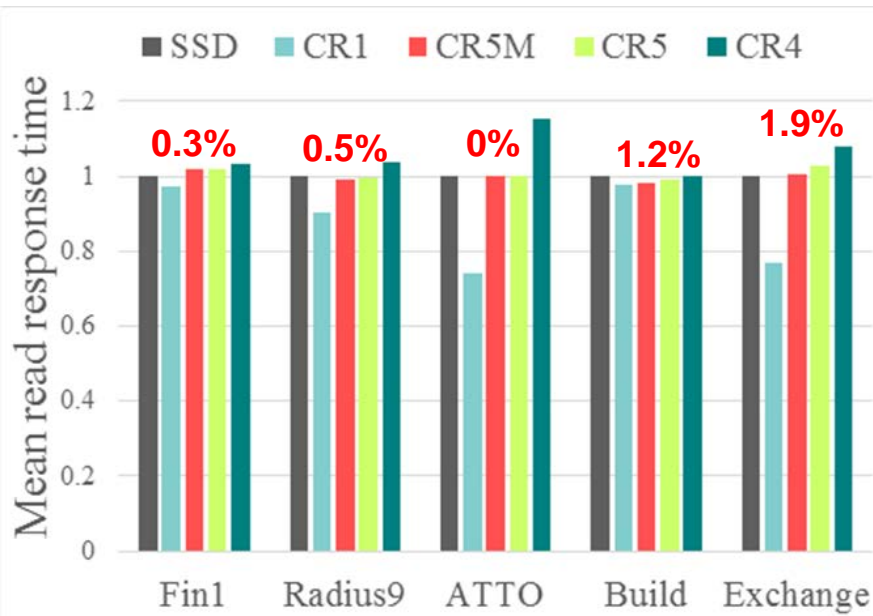
Performance Evaluation on SSD1



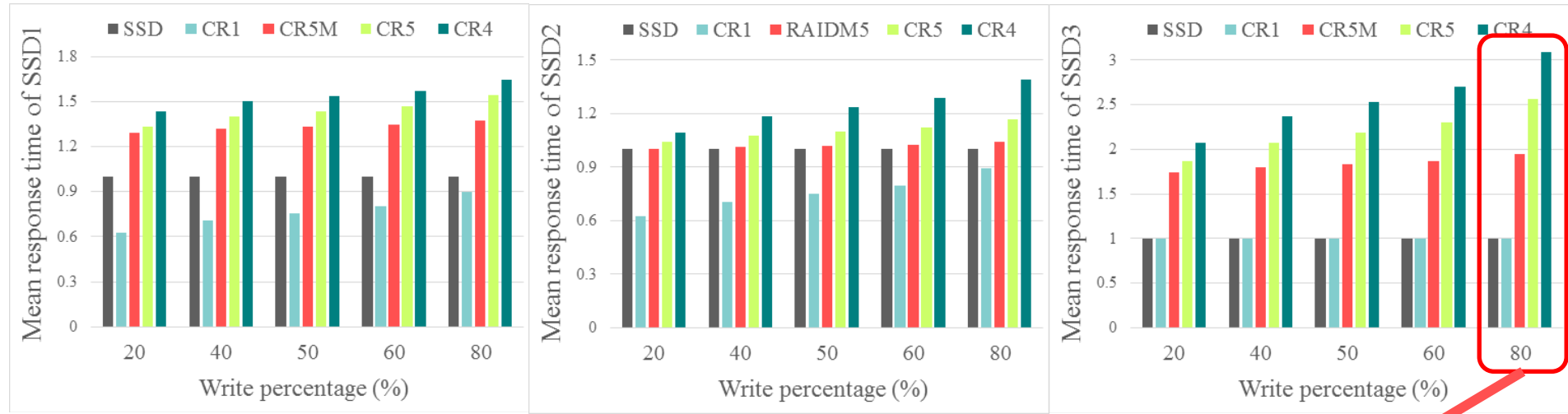
Performance Evaluation on SSD2



Performance Evaluation on SSD3

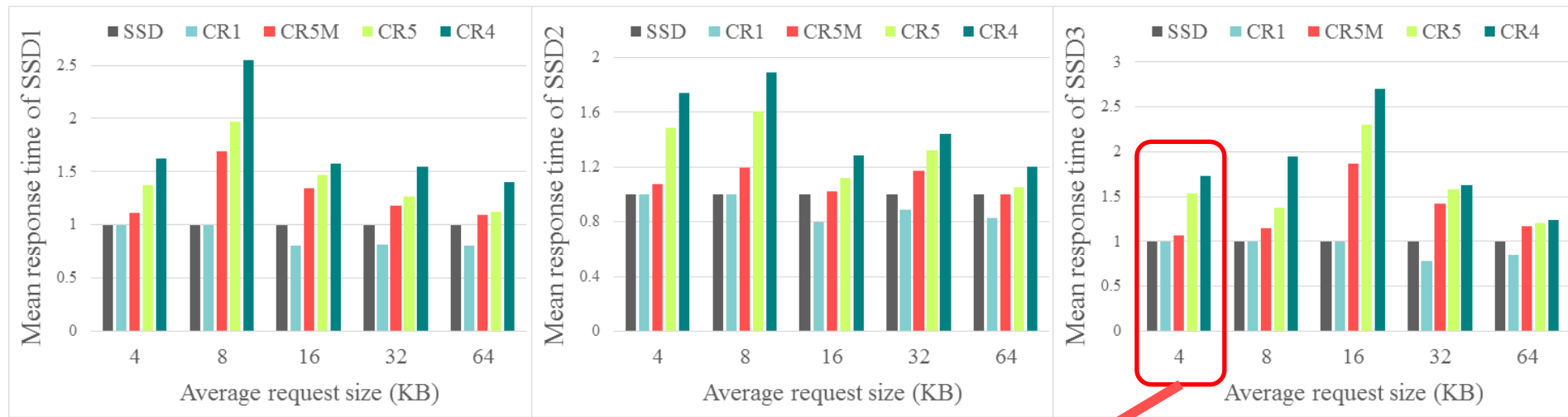


The Impact Of Write Percentage



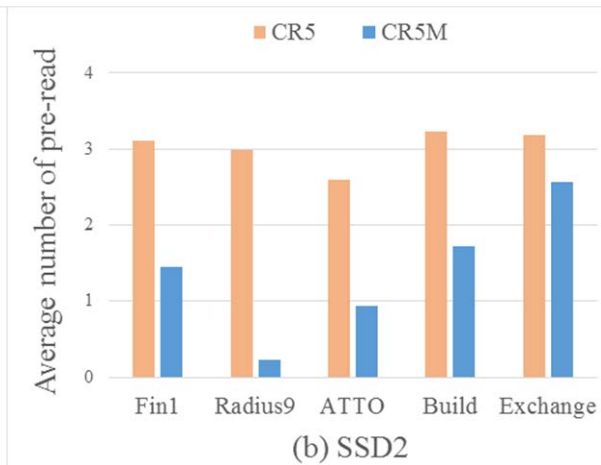
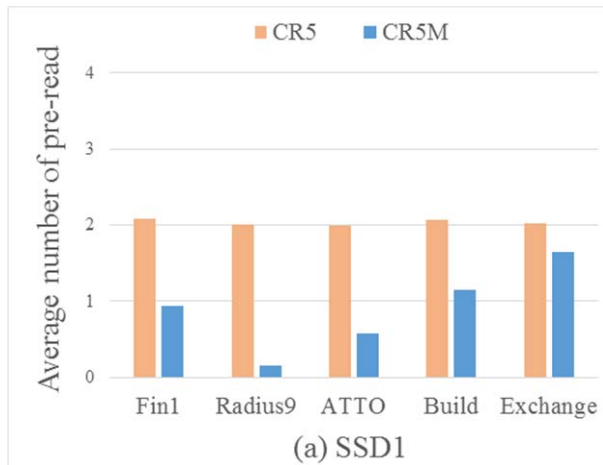
CR5M outperforms CR5 by up to 24.1%.

The Impact Of Average Request Size



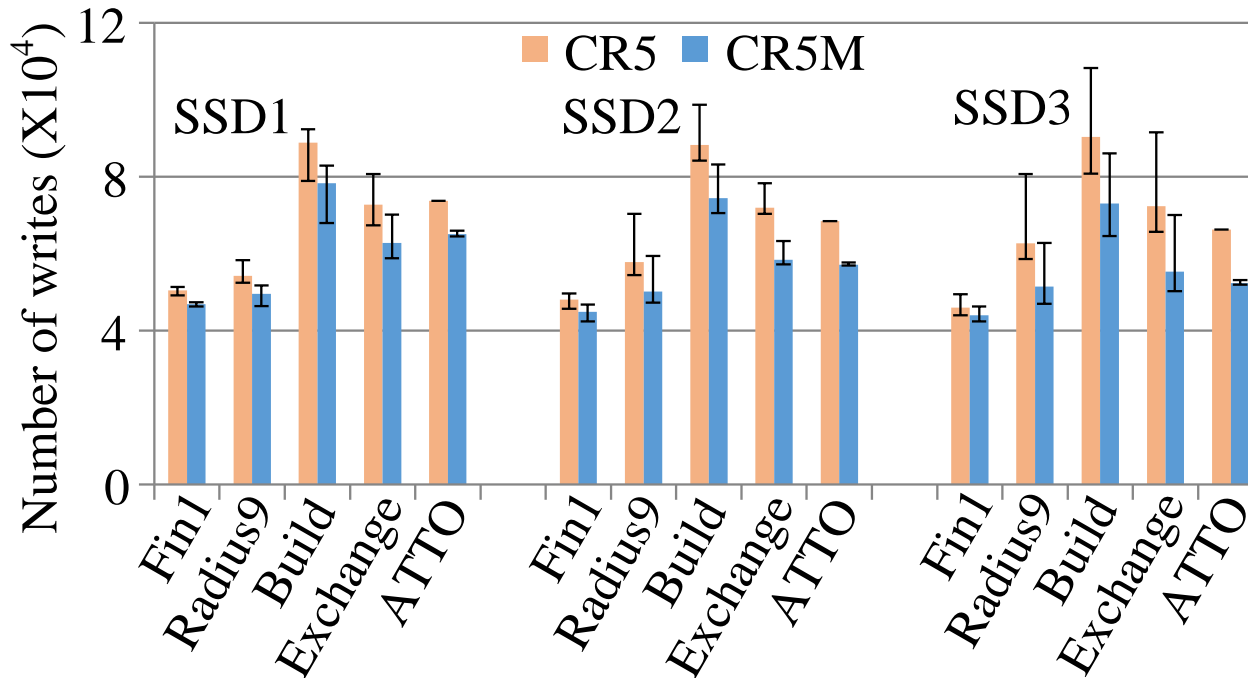
CR5M outperforms CR5 by up to 31.7%.

Parity Pre-Read Overhead



- On average CR5M reduces the number of pre-reads by 56%.


Wear-Leveling Evaluation



- CR5M can reduce the number of writes per channel by 14% compared with CR5.



Conclusions

- ECC scheme has its own capacity limitation, above which it can no longer work.
 - We implement several common RAID structures in the channel level of a single SSD to understand their impact on an SSD's performance.
 - We propose a new data redundancy architecture for a single SSD called CR5M
 - We largely extend the validated SSD simulator SSDSim
 - Experimental results demonstrate that CR5M outperforms CR5 by up to 25.8%.
-
- 



Future Work

- We will implement and study the channel-RAID architecture on a hardware evaluation board.





Acknowledgments

- This work is sponsored in part by the U.S. National Science Foundation under grant CNS-(CAREER)-0845105 and Key Technologies R&D Program of Anhui Province (China)-11010202190.





Thank you!