Don't Let RAID Raid the Lifetime of Your SSD Array

Sangwhan Moon and A. L. Narasimha Reddy

Texas A&M University

sangwhan@tamu.edu reddy@ece.tamu.edu

Electrical & Computer Engineering
Texas A&M University

Solid State Drive (SSD)

High performance
Low power consumption

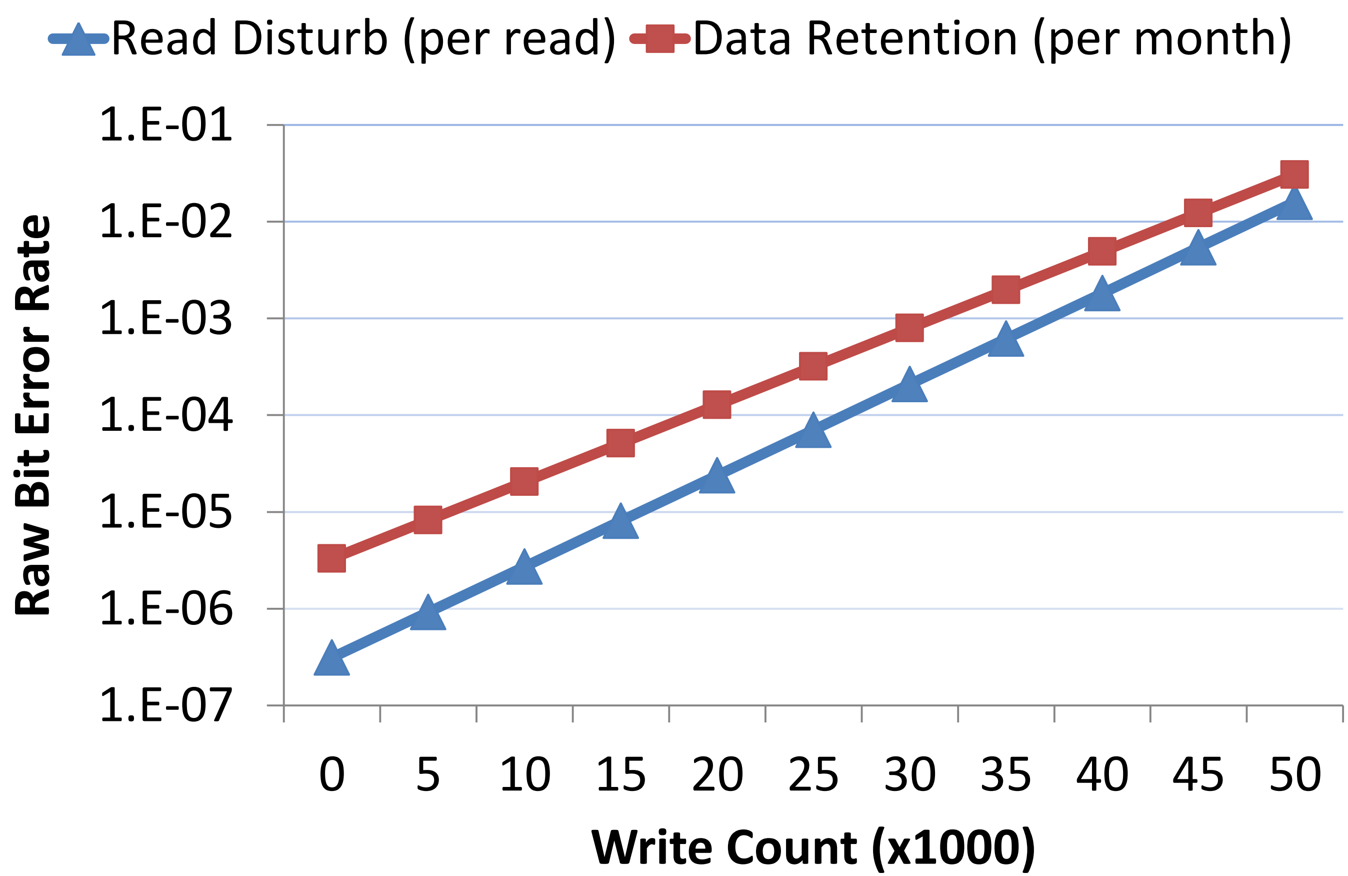
MLC SSD ??

~~Cost-per-bit~~

Write Endurance

Reliability of SSD

- Reliability of MLC Flash Memory



* The measurement data from Hairong Sun et al., "Qualifying Reliability of Solid-State Storage from Multiple Aspects," MSST'11.

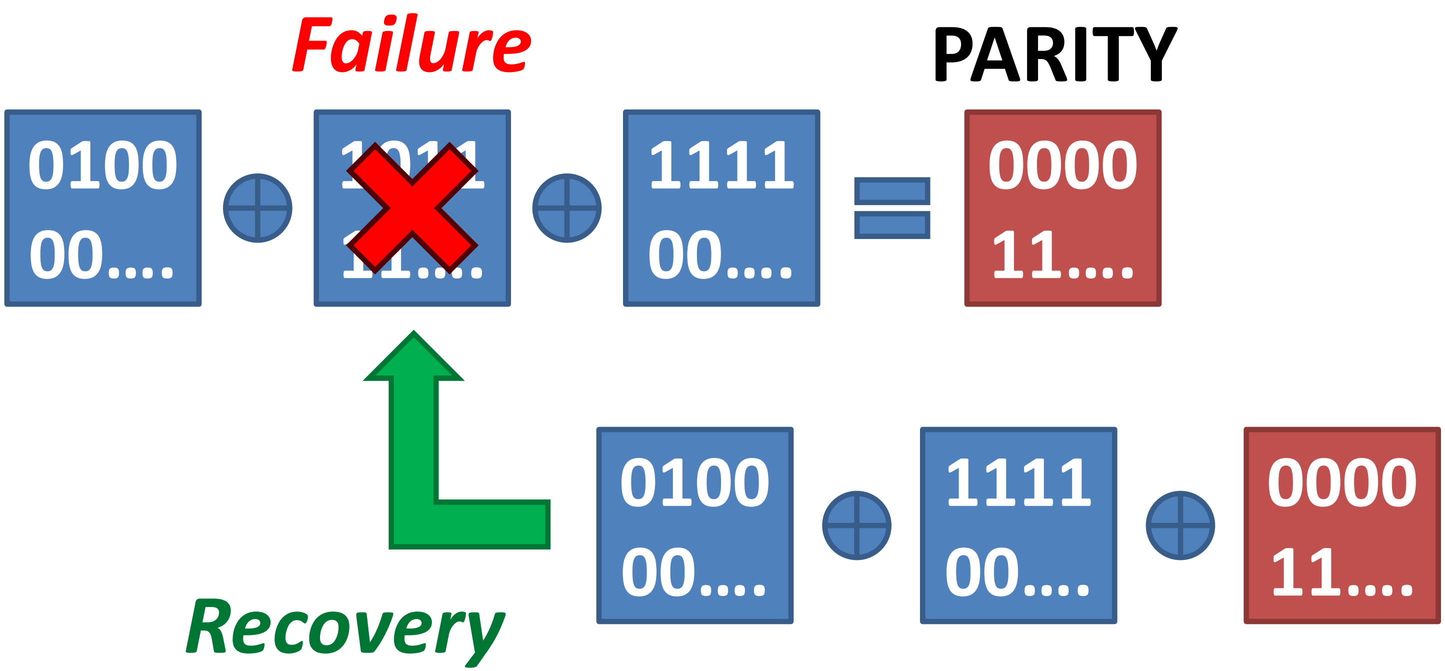
Reliability of SSD

- Device Level Protection Scheme
  - Error Correcting Code (ECC)
  - Flash Translation Layer
    - Log-like Writing and Garbage Collection
    - Wear Leveling
- System Level Protection Scheme
  - **Parity Protection (RAID5)**

These protection schemes require additional writes internally which in turn reduce the lifetime of SSD

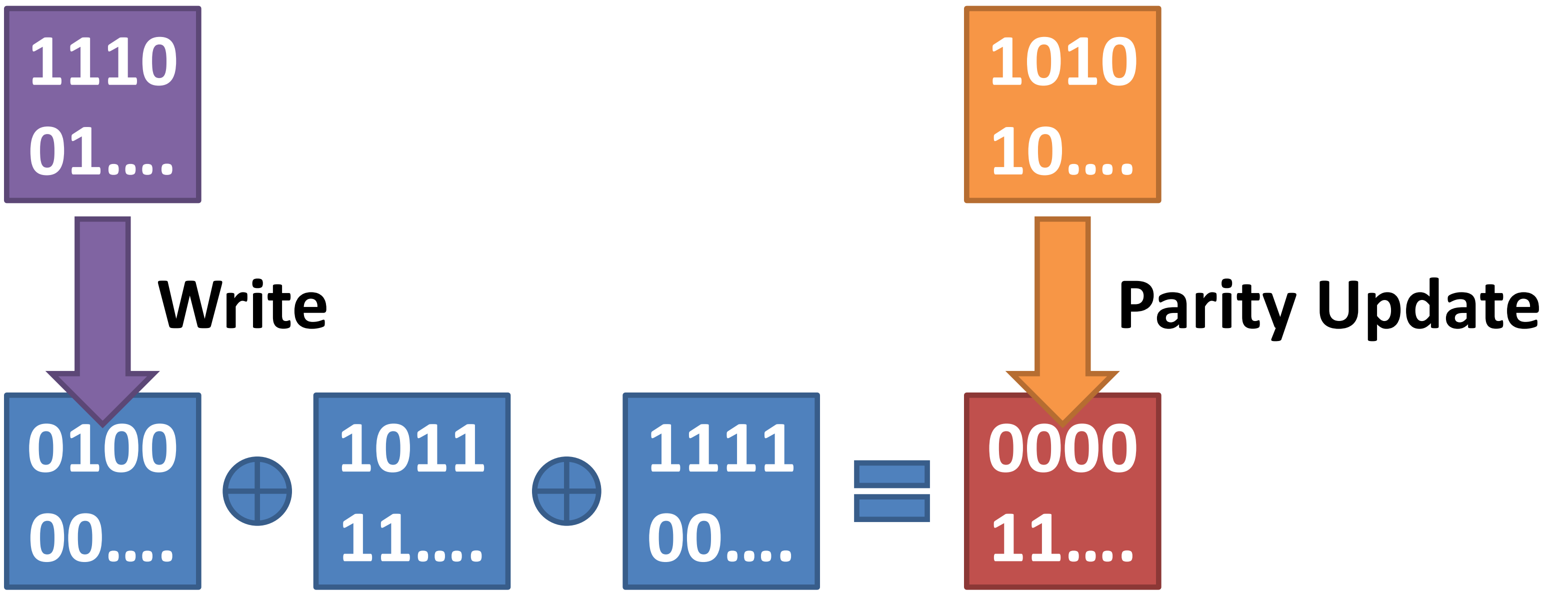
Parity Protection (RAID5)

- Protect a device array from a device failure
  - Protect each page group from a page error
- PARITY = XOR of ALL data



PARITY PROTECTION (RAID5)

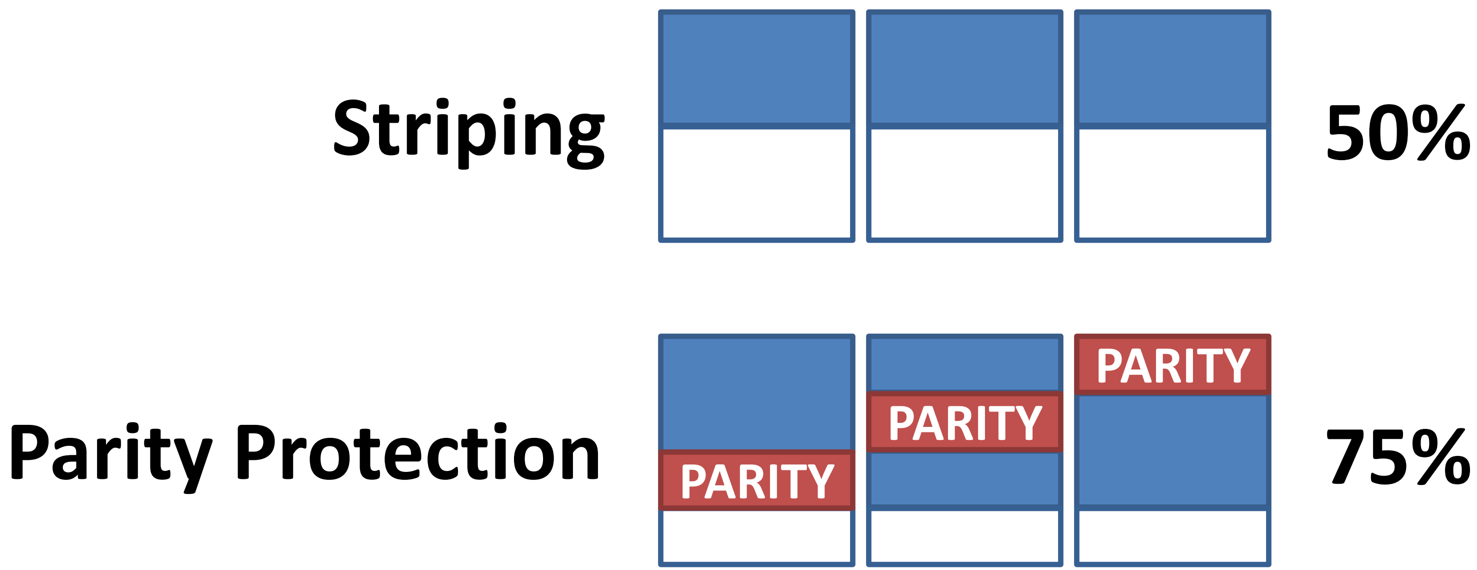
1. Parity update results in additional writes
   – Write amplification: [*N/(N-1), 2]



* N is the number of SSDs in an SSD array

Parity Protection (RAID5)

2. Parity consumes more space
   – Higher space utilization reduces the lifetime



Problem Statement

- Parity protection is supposed to improve the lifetime of SSD array
  1) Parity update amplifies the number of writes by up to 2
  2) Space overhead for parity protection initiates frequent garbage collection

Is parity protection beneficial or not in terms of reliability?

Problem Statement

For the same number of SSDs given, to store the same amount of data, which is better in lifetime, striping (RAID0) or parity protection (RAID5)?

Contributions

Texas A&M University

Markov models

We estimated the lifetime of SSD arrays.

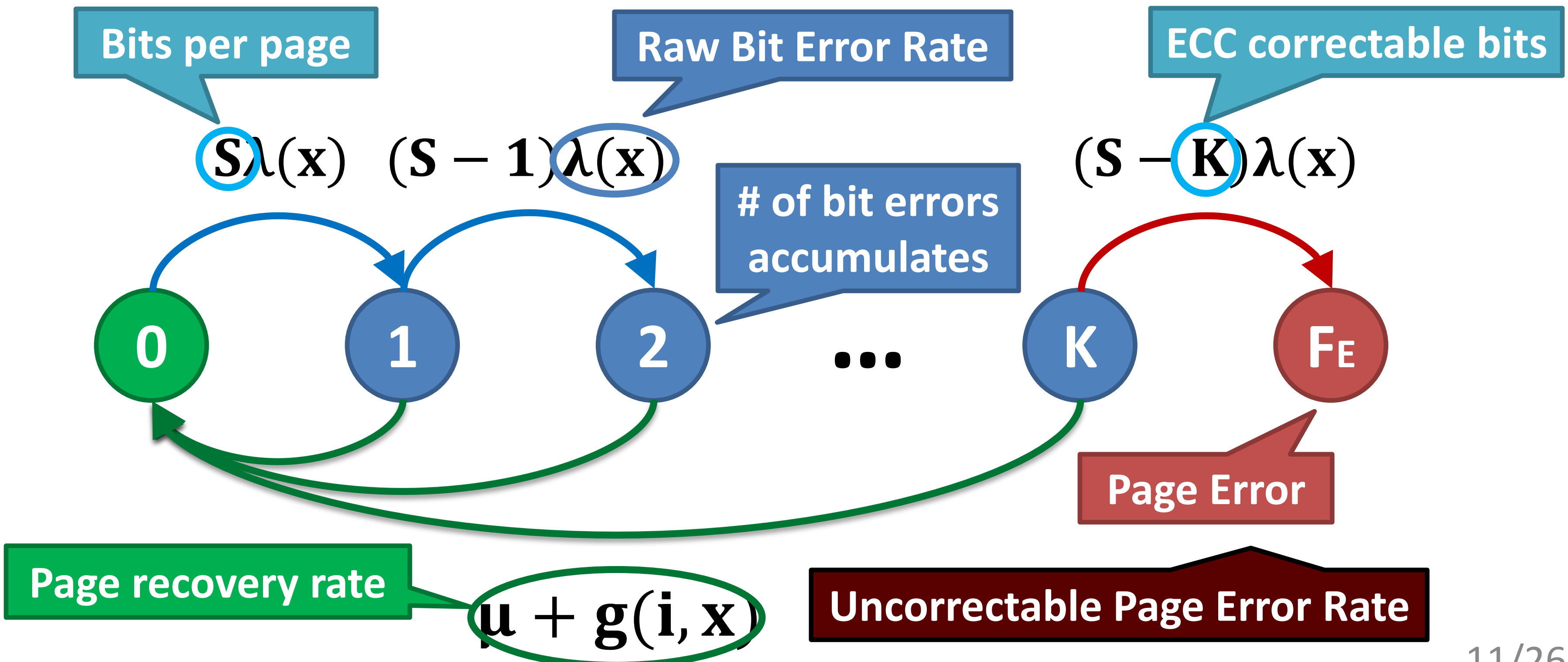
Parity protection vs. Striping

Systems with different parameters are explored.

Our preliminary results

Parity protection ***conditionally*** provides benefit in lifetime over striping.

- Page Error Rate Model at x write count
  - Bit errors accumulate until access time
  - ECC detects/corrects the bit errors

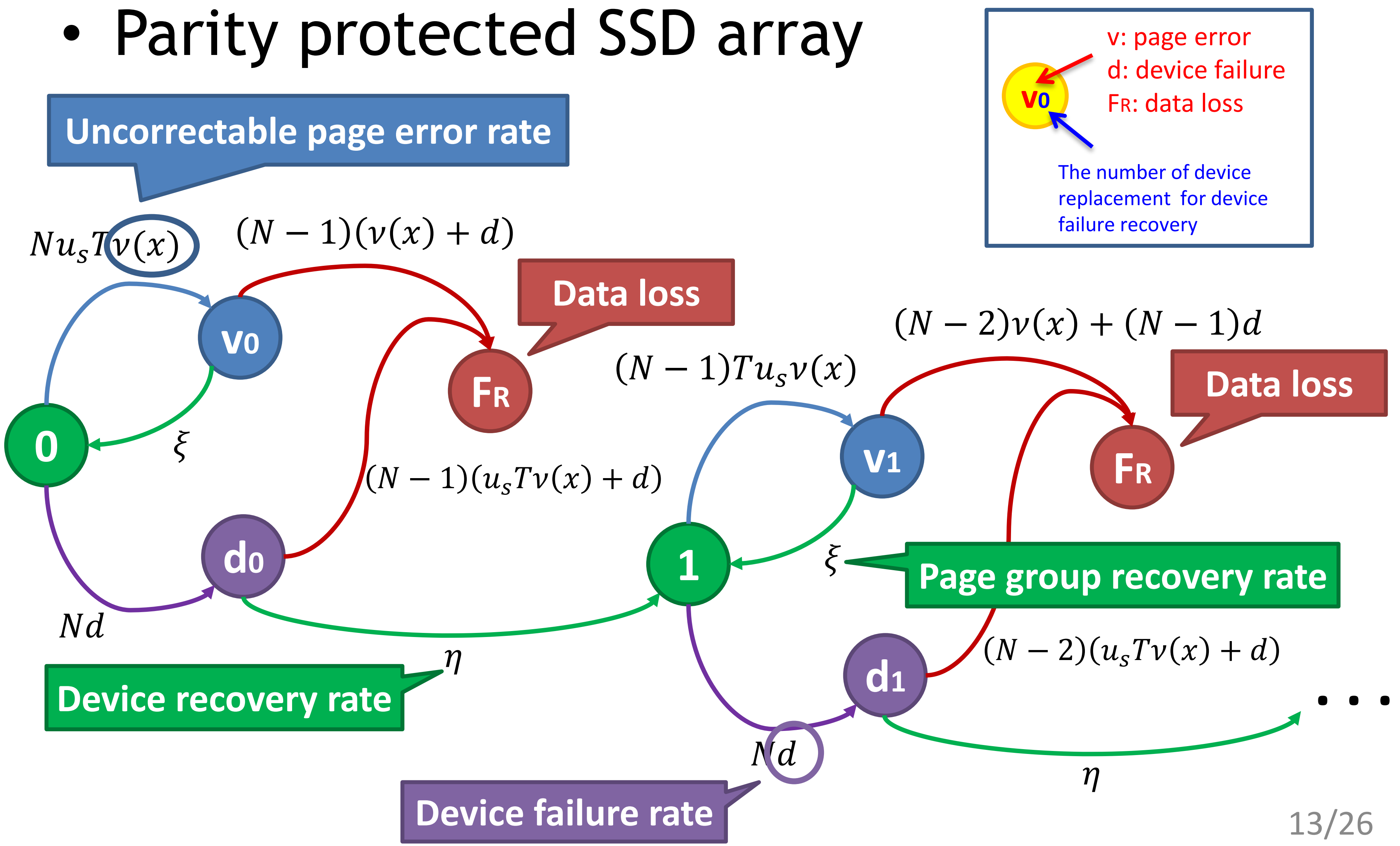


Lifetime Model

- Source of failures
  - Page error
  - Device failure
- Any failure results in data loss in striping
- Parity protection loses data when
  - Two page errors *in the same page group*
  - Two device failures
  - Page error + device failure
  - Device failure + page error

Lifetime Model

- Parity protected SSD array



Lifetime Model

- Mean Time to Data Loss (MTTDL)
  - The expected time to *encounter* ***the first data loss*** in an SSD array

$$\mathrm{MTTDL} = \sum_{j=1}^{\infty}\left(j\,p(j)\prod_{i=1}^{j-1}(1 - p(i))\right)t_w$$

The probability of data loss at *j* write count

Time to write an SSD array

Simulation Environment

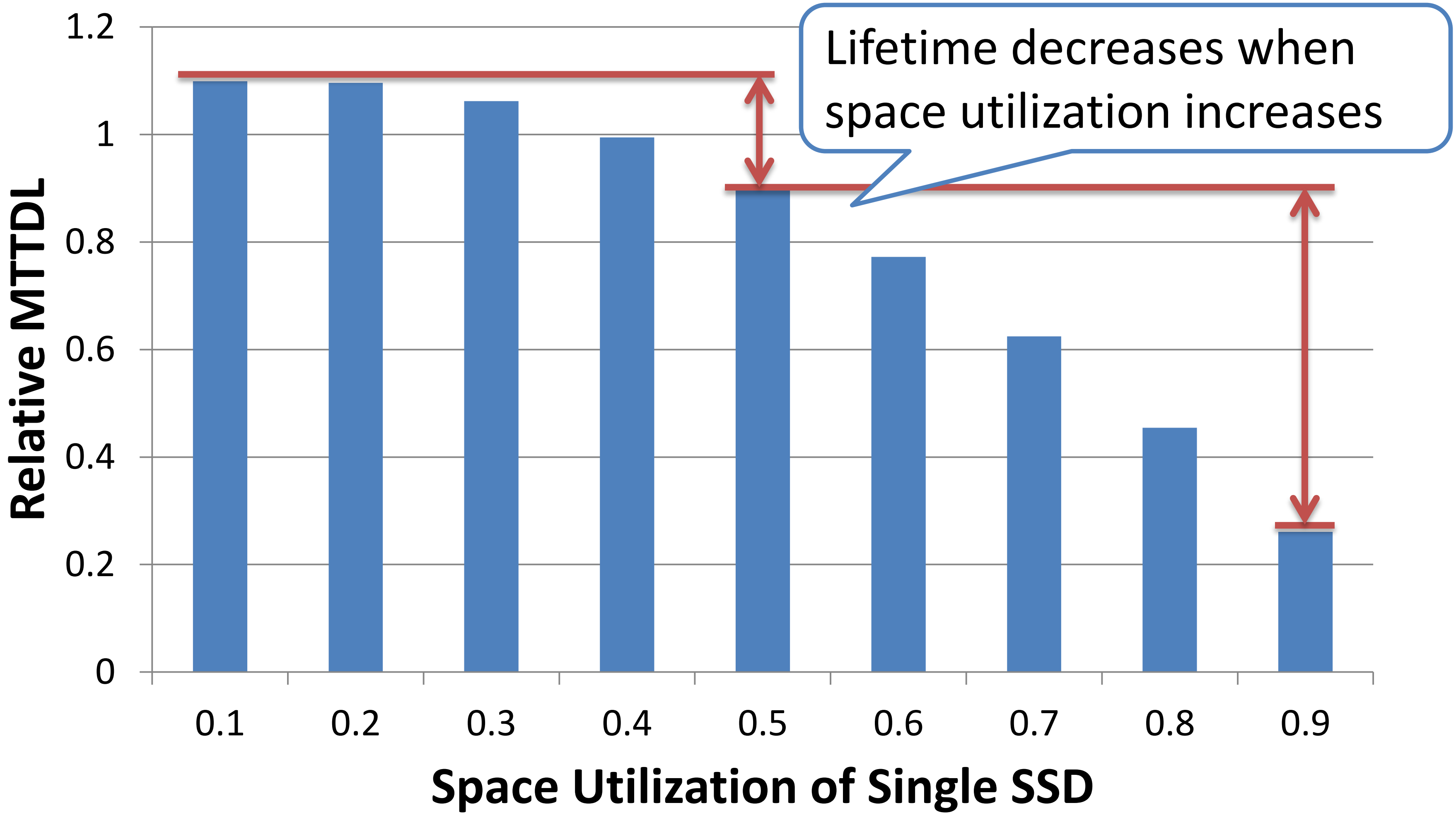
- SSD Parameters
  - 3x nm MLC flash memory
  - Capacity = 80GB
  - Page size = 4KB
    - ECC: 61-bit errors correctable BCH code
  - Annual device failure rate = 3%
  - TRIM command is exploited

Simulation Environment

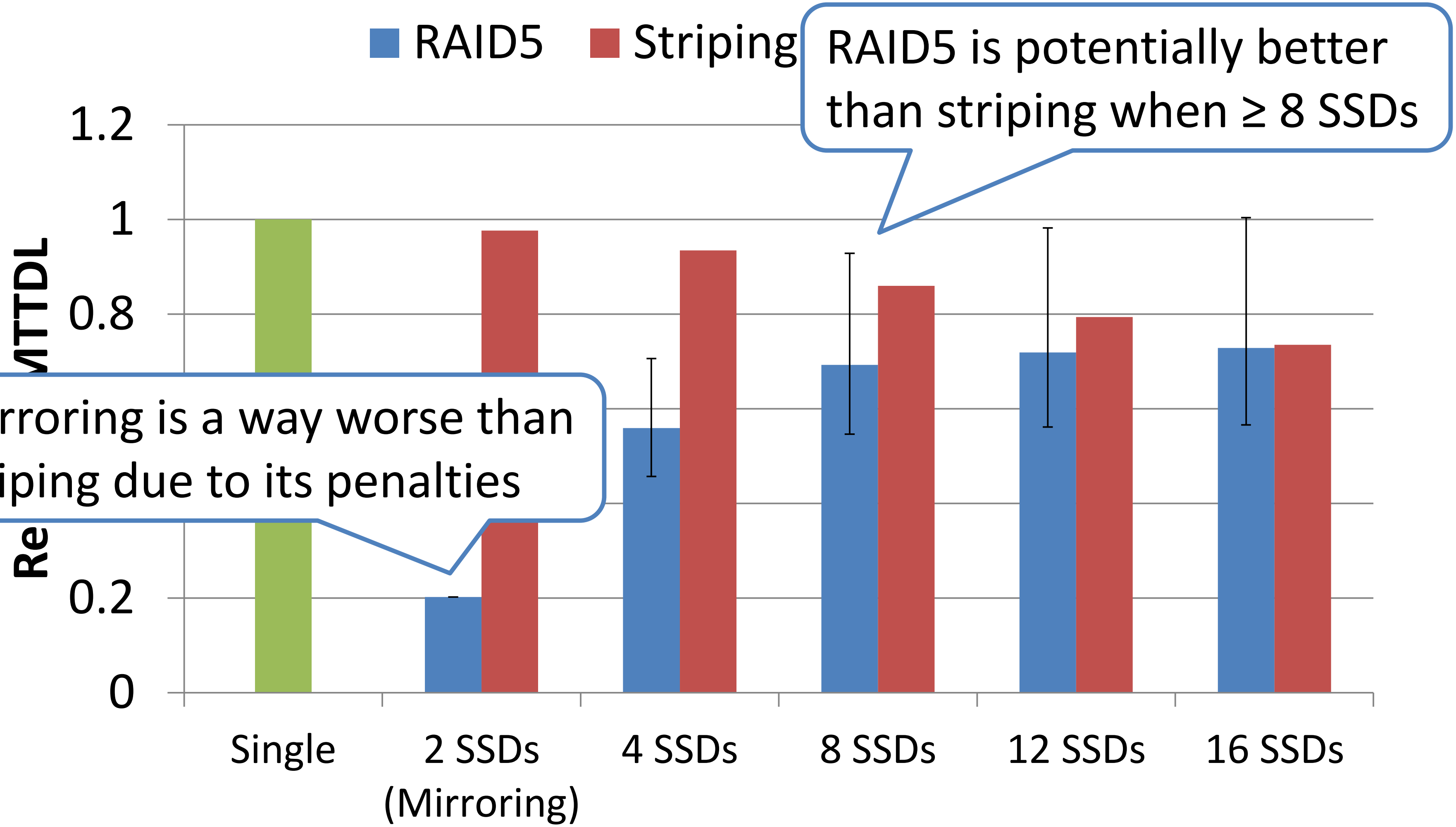
- Simulation Parameters
  - The amount of data = 30GB/SSD
  - Workload
    - Read + Write = 125 MB/s/SSD
    - Read : Write = 3:1
  - 8 SSDs in an SSD array

- Relative MTTDL
  - The ratio of the lifetime of the target SSD array to that of **single SSD with default parameters**

Analysis of Single SSD

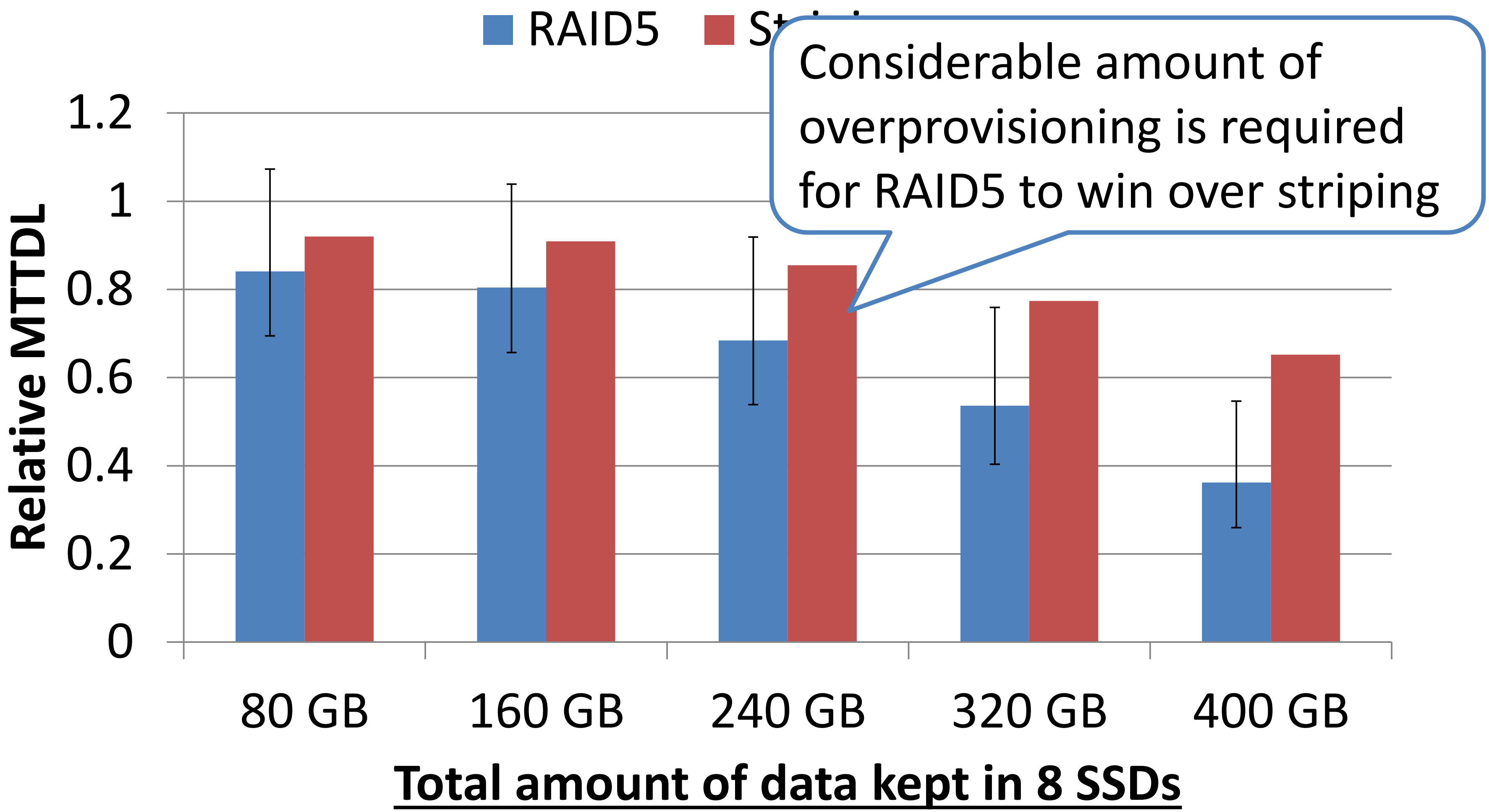


Evaluation: Different Number of Devices



* The amount of data = 30GB/SSD

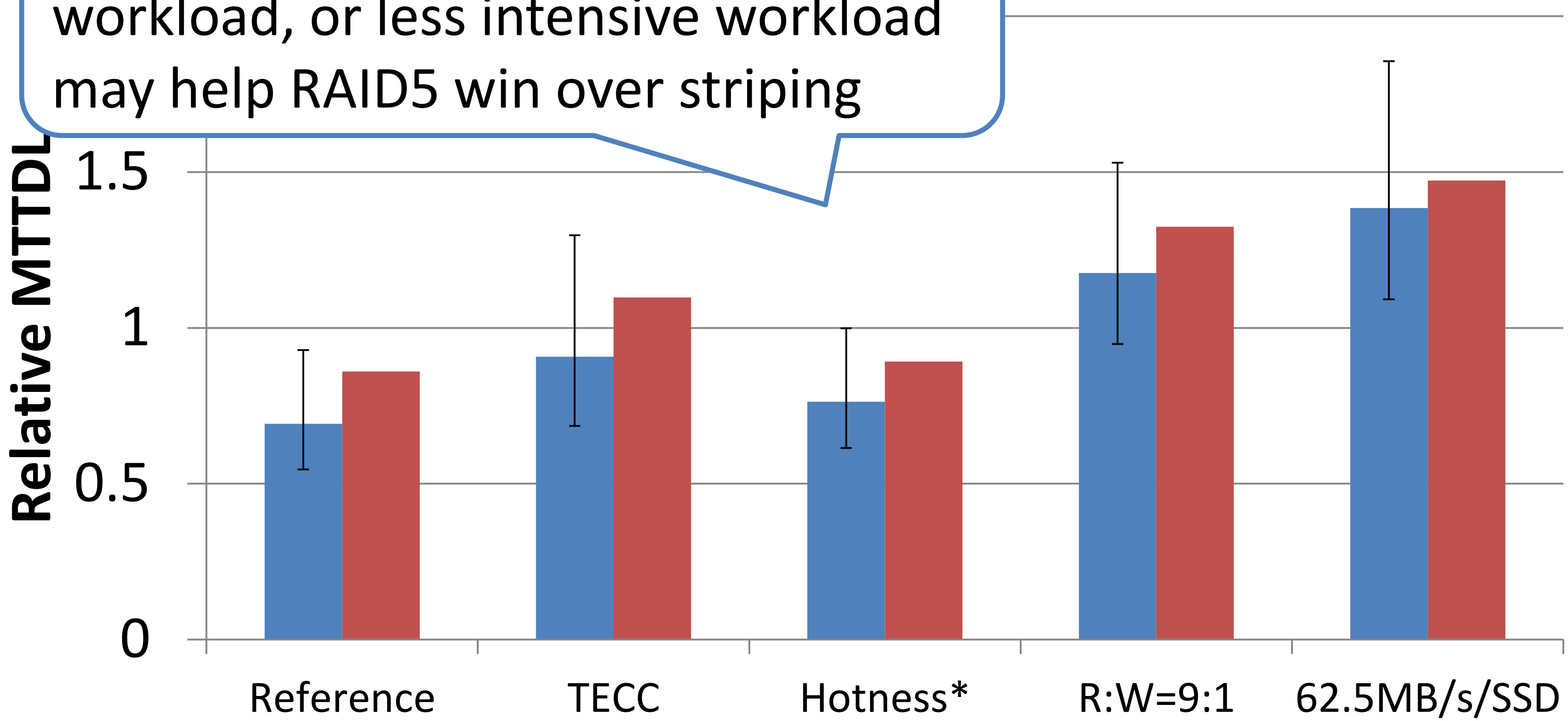
Evaluation: Different Amount of Data



EVALUATION: DIFFERENT PARAMETERS

TEXAS A&M UNIVERSITY

TECC, hotness, read intensive workload, or less intensive workload may help RAID5 win over striping



Lifetime of 8 SSDs with different parameters

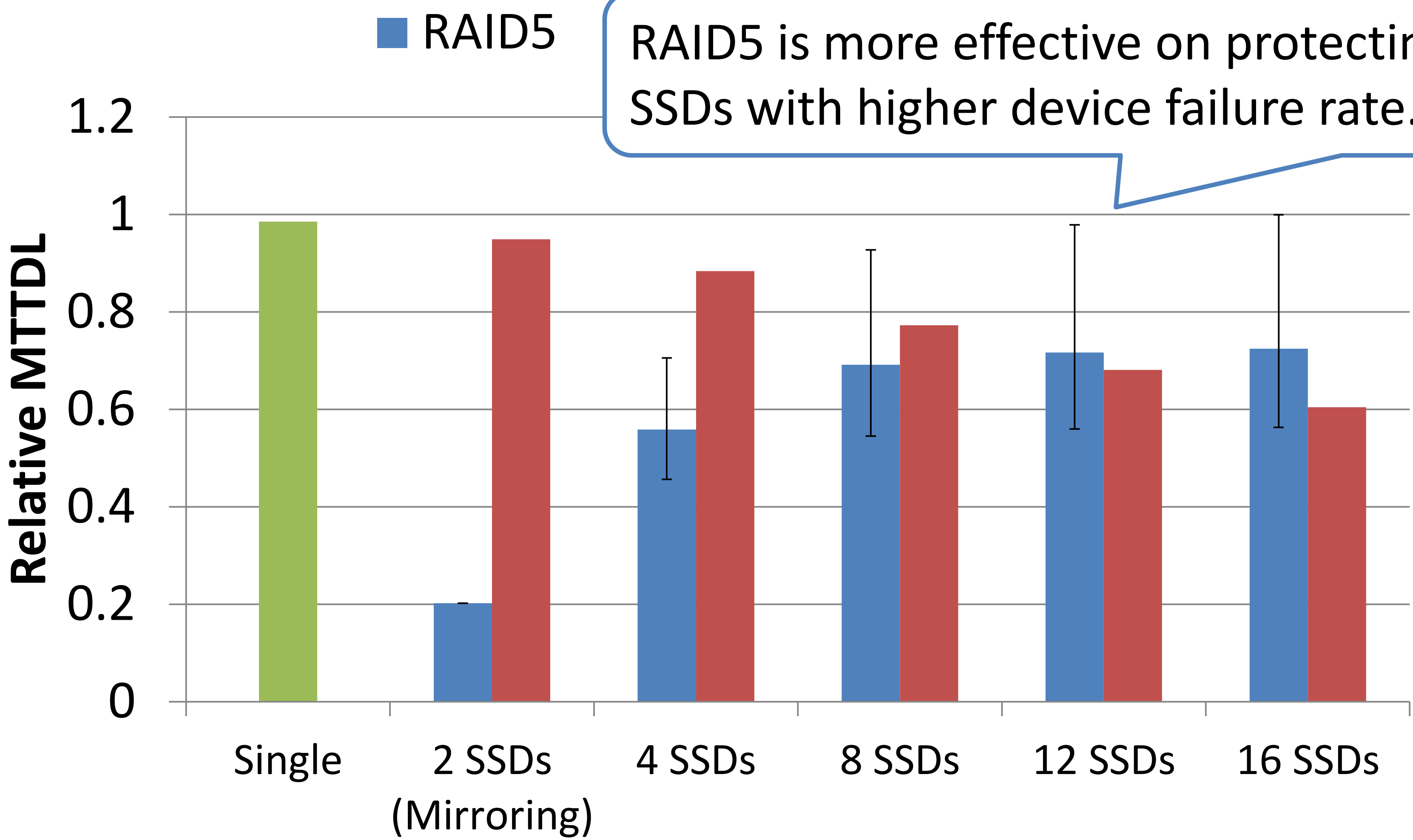
[TECC] S. Moon and A. Reddy, “Write amplification due to ECC on flash memory or leave those bit errors alone,” in MSST’12

* 80% of workload accumulates on 20 % of space

Evaluation: Annual Device Failure Rate = 5%

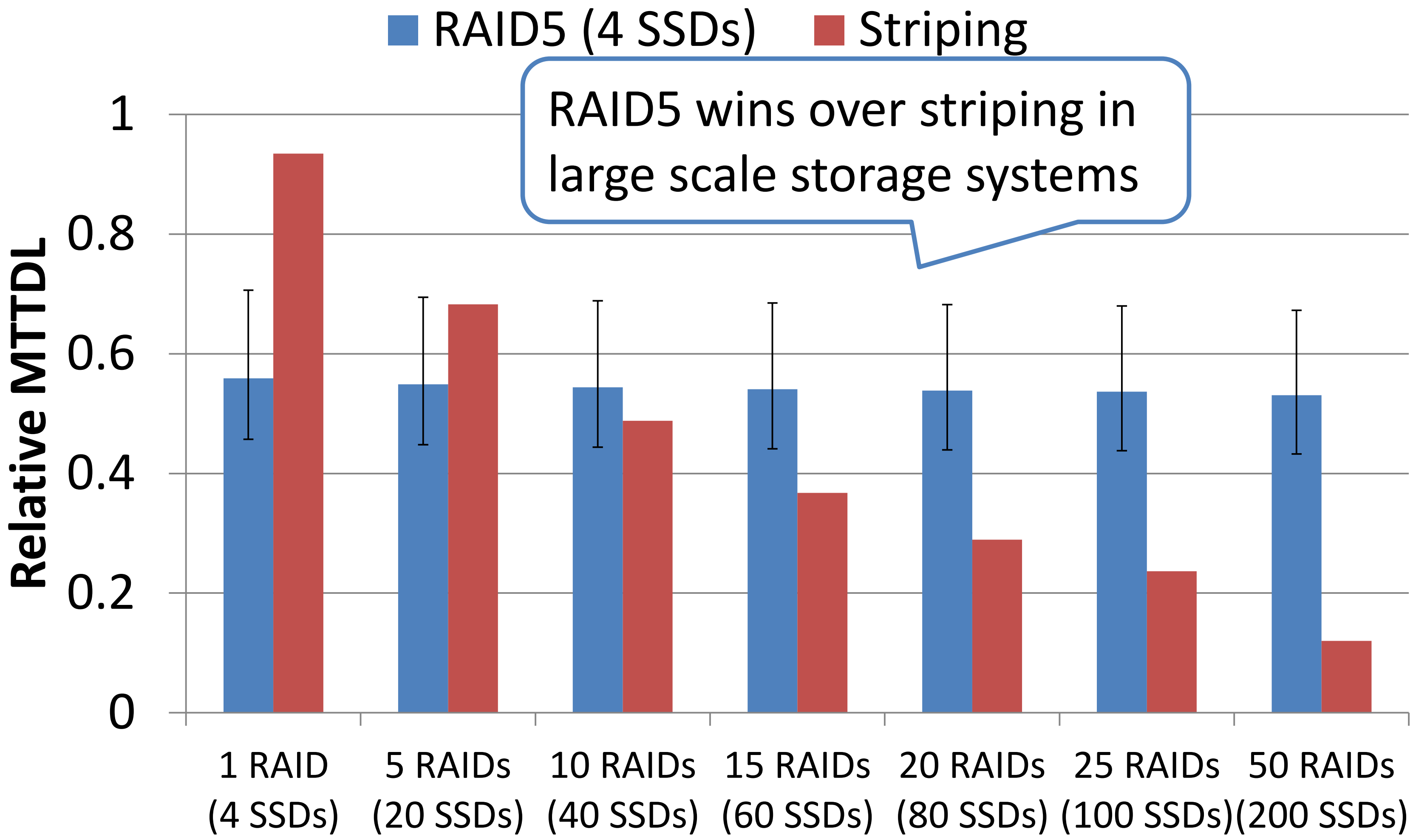
Texas A&M University

RAID5 is more effective on protecting SSDs with higher device failure rate.



Annual Device Failure Rate = 5%

Evaluation: Large Scale Storage System



SUMMARY OF EVALUATIONS

- Parity protection is potentially worse than striping with small number of SSDs
- Parity protection wins against striping when
  1) considerably lower space utilization is guaranteed.
  2) TECC, hotness, read intensive workload, or less intensive workload is provided.
  3) SSDs have higher device failure rate.
- Parity protection wins against striping in large scale storage systems

Future Work

- Other lifetime evaluations
  - Different write sizes
  - Other storage systems (e.g. RAID6)
- Monetary cost of ownership
- Validation of our analytic models
- Advanced techniques to reduce write amplification from parity protection

Conclusion

- Markov models to estimate the lifetime of an SSD array with protection schemes
- Lifetime comparison of striping (RAID0) and parity protection (RAID5) with different parameters
- Parity protection is conditionally superior to striping.

Thank You for Listening!

Questions?