Computer Science
UNIVERSITY OF TORONTO

NetApp®

A Study of SSD Reliability in Large Scale Enterprise Storage Deployments

Stathis Maneas, Kaveh Mahdaviani, Tim Emami, Bianca Schroeder

USENIX FAST ‘20

Reliability of SSD-based enterprise storage systems

- What we know:
  - Four field studies (distributed data center storage systems).
  - Facebook ‘15, Google ‘16, Microsoft ‘16, Alibaba ‘19.

Reliability of SSD-based enterprise storage systems

- What we know:
  - Four field studies (distributed data center storage systems).
  - Facebook ‘15, Google ‘16, Microsoft ‘16, Alibaba ‘19.
- We focus on *enterprise storage systems*:
  - Different drives, workloads, and reliability mechanisms.
  - High-end drives, reliability is ensured through RAID, etc.

Reliability of SSD-based enterprise storage systems

- What we know:
  - Four field studies (distributed data center storage systems).
  - Facebook ‘15, Google ‘16, Microsoft ‘16, Alibaba ‘19.
- We focus on *enterprise storage systems*:
  - Different drives, workloads, and reliability mechanisms.
  - High-end drives, reliability is ensured through RAID, etc.
- Factors that have not been studied before:
  - 3D-TLC NAND.
  - Large Capacity Drives (e.g., 8TB and 15TB).
  - Firmware Versions.
  - RAID Groups.

Systems Description

- 1.4 million SSDs.
- 2.5 years of data.
- SLC, cMLC, eMLC, 3D-TLC drives.
- 3 manufacturers.
- 18 drive models:
  - 12 different capacities.
- Varying age, usage, and system configurations.

Replacement Types

- Issues can be reported by a drive, the storage layer, the file system, etc.

Category	Type	
SL1	Predictive Failures	
	Threshold Exceeded	
	Recommended Failures	
SL2	Aborted Commands	
	Disk Ownership I/O Errors	
	Command Timeouts	
SL3	Lost Writes	
SL4	SCSI Errors	
	Unresponsive Drive	

Replacement Types

- Issues can be reported by a drive, the storage layer, the file system, etc.

Category	Type	Percentage (%)
SL1	Predictive Failures	**12.78**
	Threshold Exceeded	**12.73**
	Recommended Failures	**8.93**
SL2	Aborted Commands	**13.56**
	Disk Ownership I/O Errors	**3.27**
	Command Timeouts	**1.81**
SL3	Lost Writes	**13.54**
SL4	SCSI Errors	**32.78**
	Unresponsive Drive	**0.60**

Replacement Types

- Issues can be reported by a drive, the storage layer, the file system, etc.

Category	Type	Percentage (%)
SL1	Predictive Failures	**12.78**
	Threshold Exceeded	**12.73**
	Recommended Failures	**8.93**
SL2	Aborted Commands	13.56
	Disk Ownership I/O Errors	3.27
	Command Timeouts	1.81
SL3	Lost Writes	13.54
SL4	SCSI Errors	**32.78**
	Unresponsive Drive	0.60

- **SCSI Errors** dominate!
- One third of drive replacements are merely preventative based on *predictions* (**Category SL1**)!
- SSDs rarely become completely unresponsive!

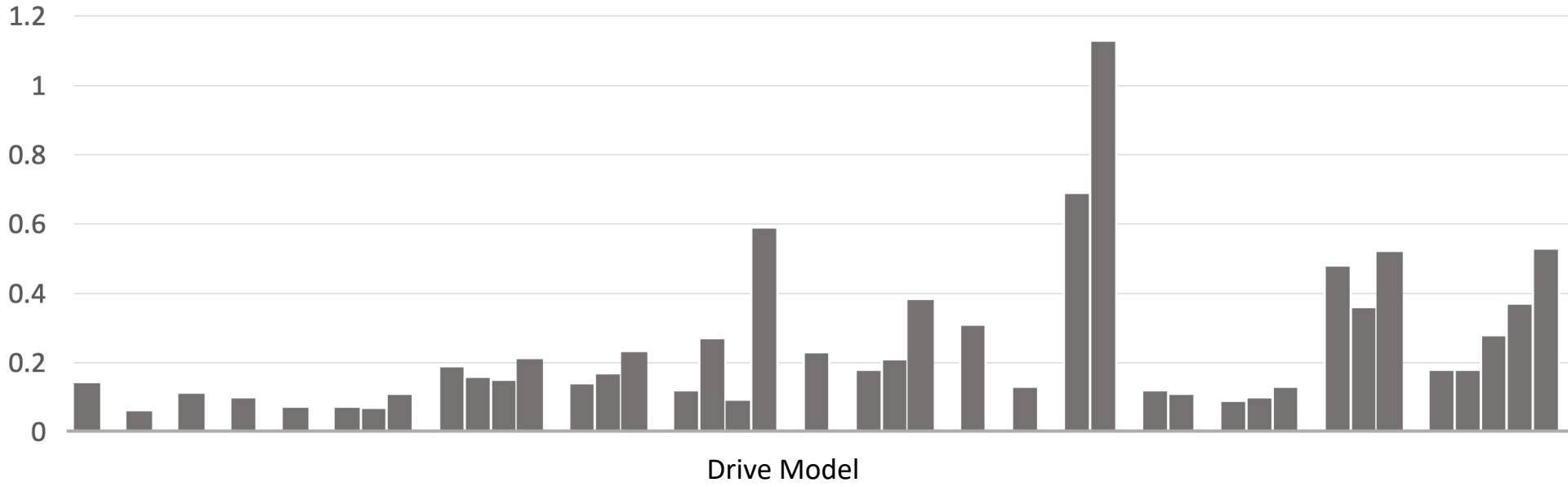
How frequently are SSDs replaced?

- *Annual Replacement Rate (**ARR**)*:

$$ARR = \frac{\#Failed\ Devices}{\#Device\ years}$$

Drive Replacements

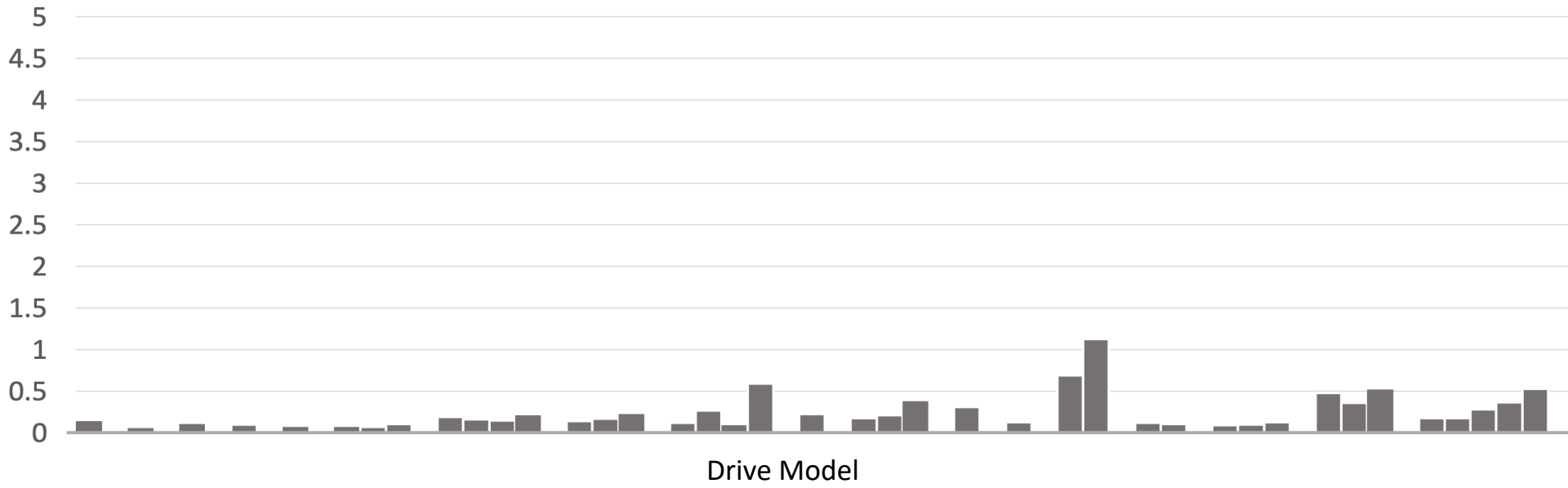
- *Annual Replacement Rate (**ARR**)*:



- The average ARR across the entire population is 0.22%, but rates vary widely (0.07 - 1.2%)!

Drive Replacements

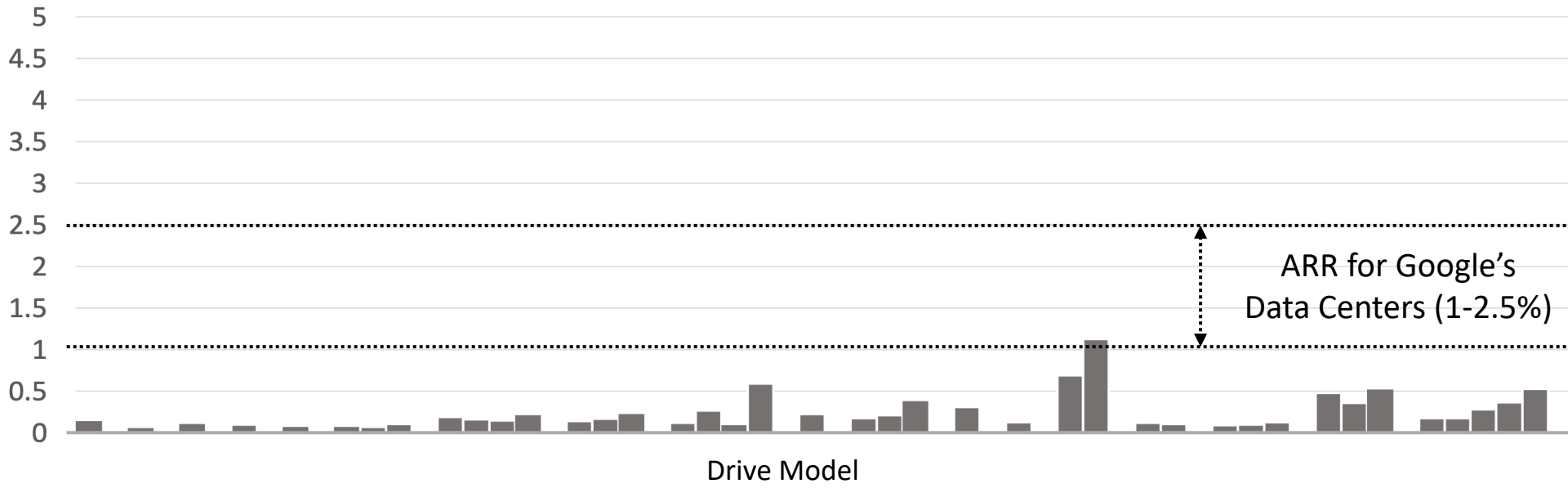
- *Annual Replacement Rate (**ARR**)*:



- The average ARR across the entire population is 0.22%, but rates vary widely (0.07 - 1.2%)!

Drive Replacements

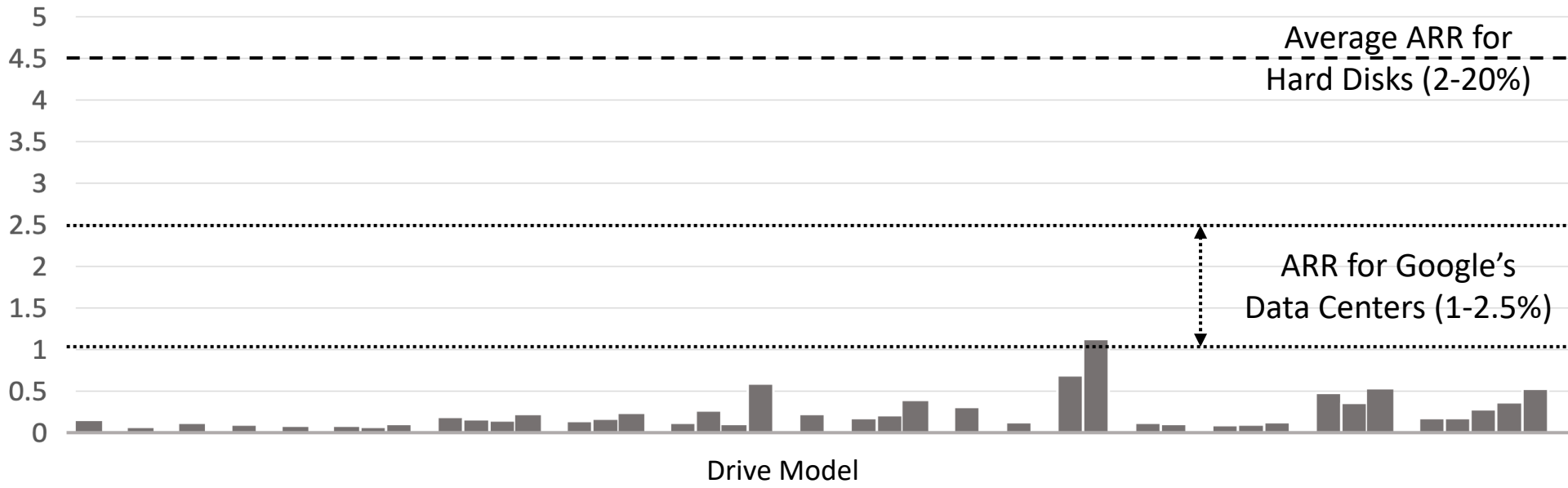
- *Annual Replacement Rate (**ARR**)*:



- The average ARR across the entire population is 0.22%, but rates vary widely (0.07 - 1.2%)!

Drive Replacements

- *Annual Replacement Rate (**ARR**)*:



- The average ARR across the entire population is 0.22%, but rates vary widely (0.07 - 1.2%)!

Drive Replacements

- *Annual Replacement Rate (**ARR**)*:

$$ARR = \frac{\#Failed\ Devices}{\#Device\ years}$$

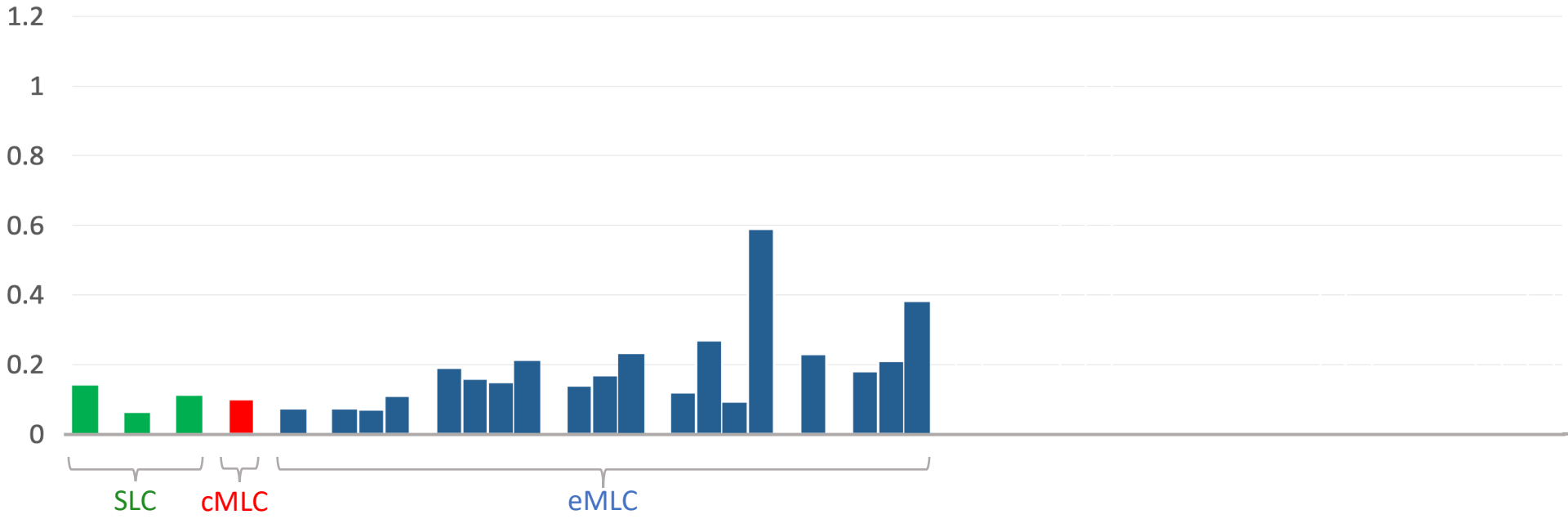
- **Which factors impact flash reliability?**
  - Flash Type (SLC, cMLC, eMLC, 3D-TLC).
  - Lithography.
  - Usage and Age.
  - Firmware Version.
  - Other factors (see the paper).

Flash Type

- **Common expectation:** Lower failure rates for SLC ($$$) versus cMLC/eMLC and 3D-TLC.

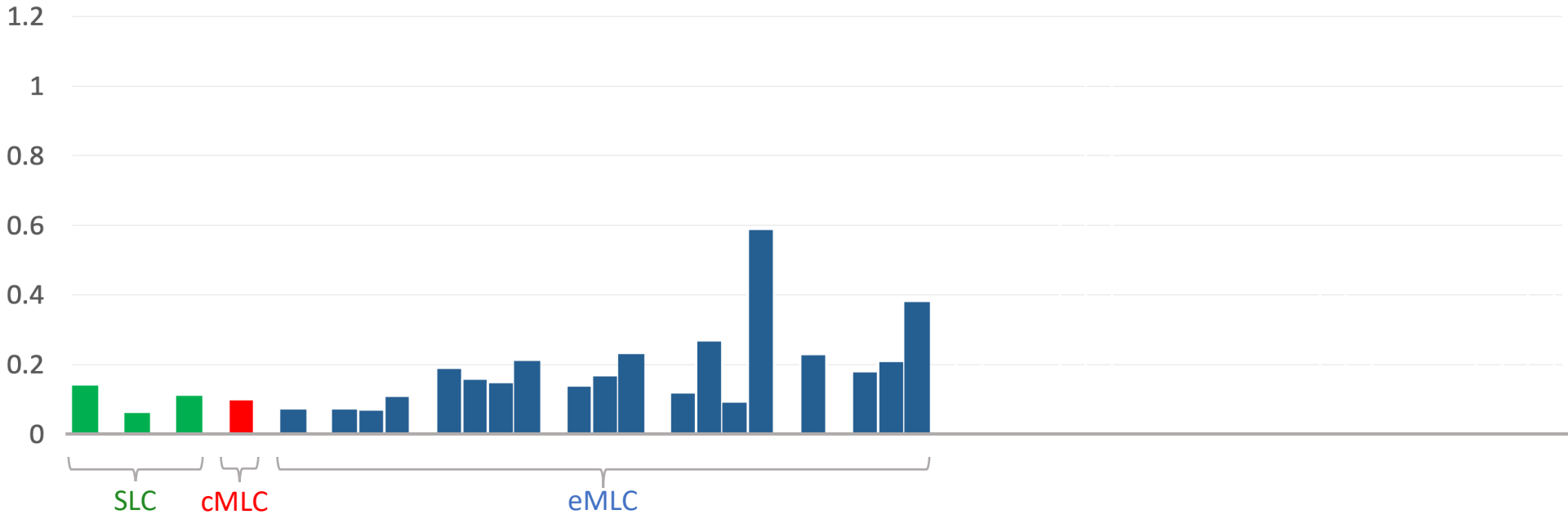
Flash Type

- **Common expectation:** Lower failure rates for SLC ($$$) versus cMLC/eMLC and 3D-TLC.



Flash Type

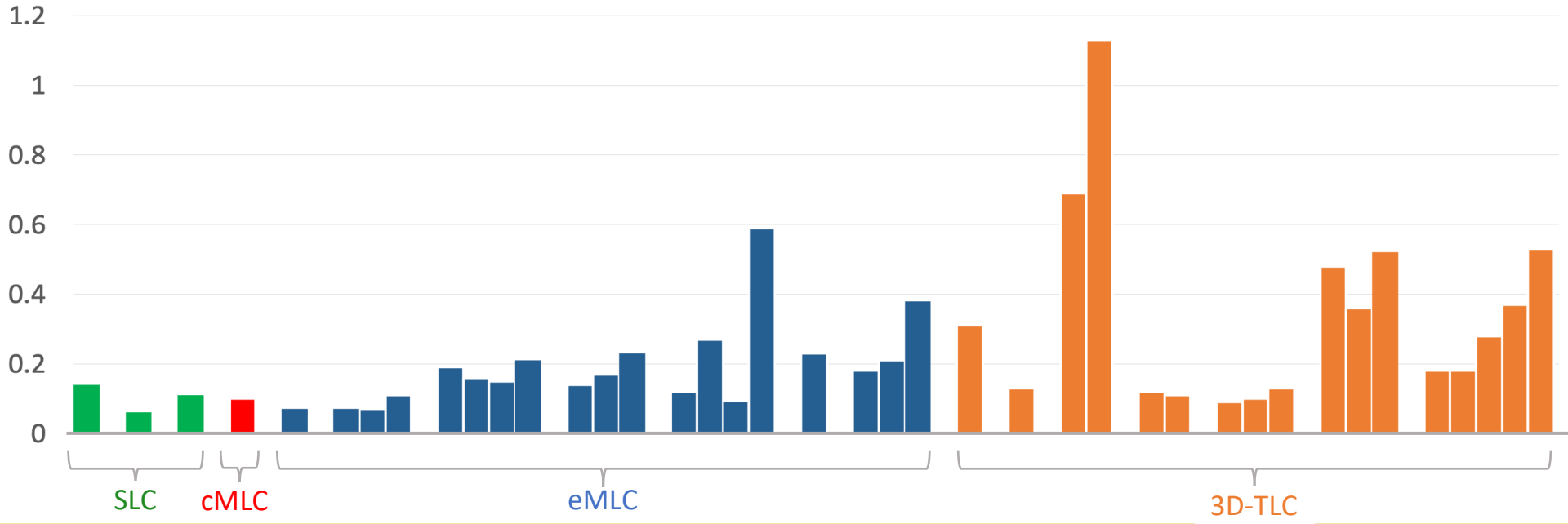
- **Common expectation:** Lower failure rates for SLC ($$$) versus cMLC/eMLC and 3D-TLC.



- *SLC* drives not necessarily better than *MLC* drives.
- *eMLC* drives not necessarily better than *cMLC* drives.

Flash Type

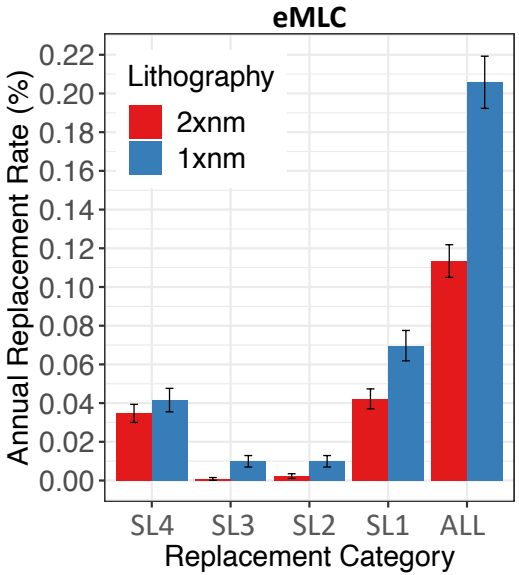
- **Common expectation:** Lower failure rates for SLC ($$$) versus cMLC/eMLC and 3D-TLC.

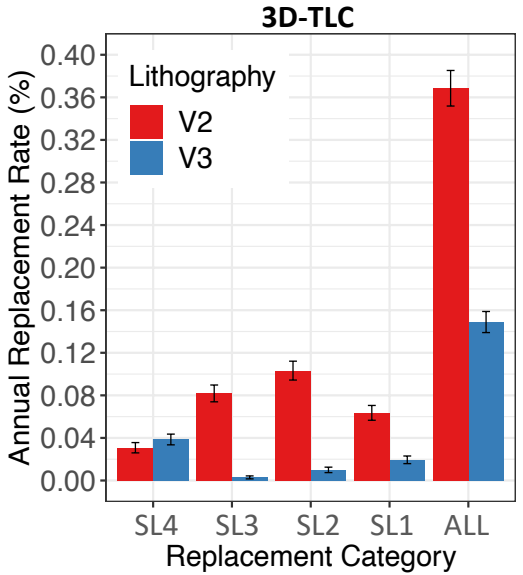


- *SLC* drives not necessarily better than *MLC* drives.
- *eMLC* drives not necessarily better than *cMLC* drives.
- *3D-TLC* drives have the highest replacement rates.

Lithography

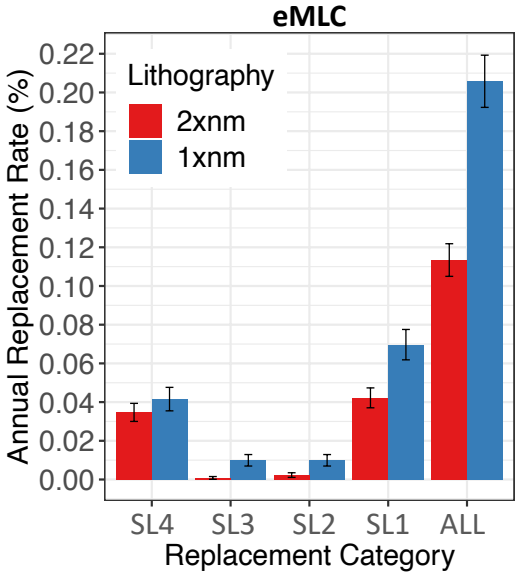
- Compare models with the same flash type.
- **Common expectation:** Higher failures rates for higher densities.

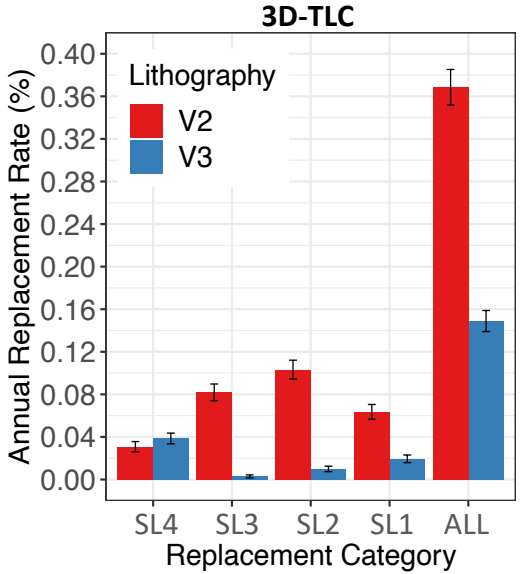




Lithography

- Compare models with the same flash type.
- **Common expectation:** Higher failures rates for higher densities.

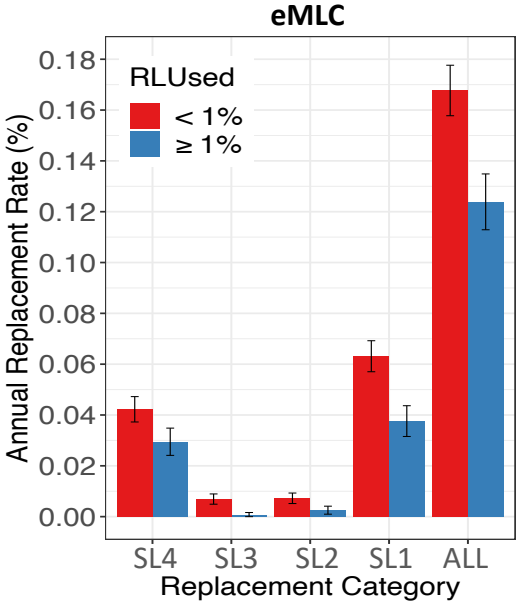


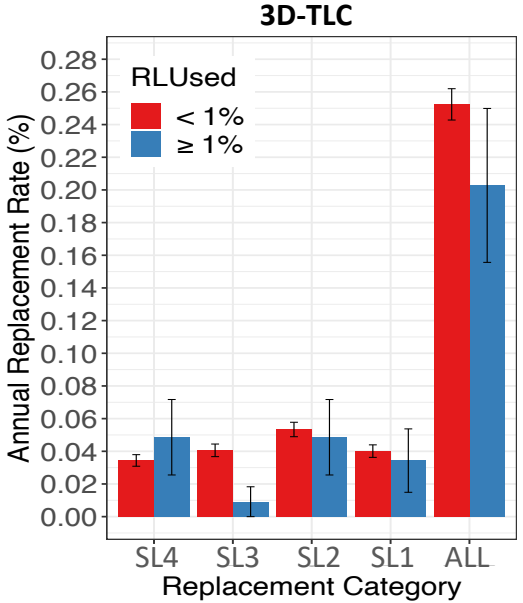


- **eMLC:** models with higher densities (1xnm) have higher replacement rates.
- **3D-TLC:** models with lower densities (V2) have higher replacement rates (the trend is reversed)!

Usage

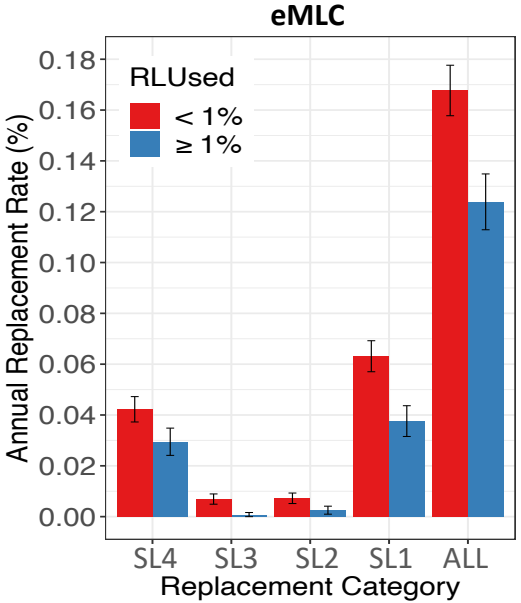
- Usage affects the reliability of SSDs, due to wear-out of their cells.
- Percentage of P/E cycles limit used so far.

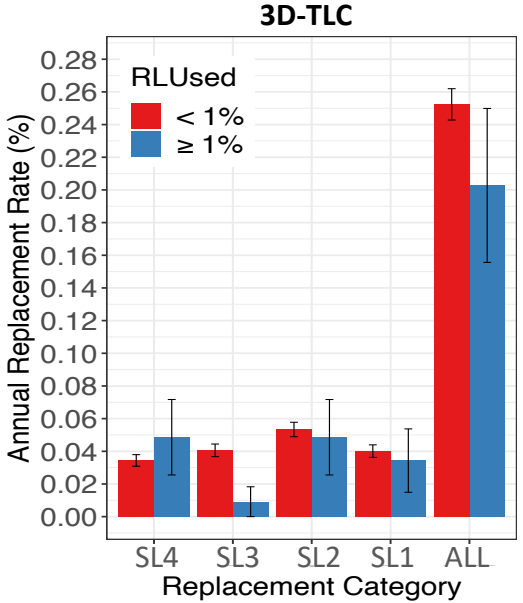




Usage

- Usage affects the reliability of SSDs, due to wear-out of their cells.
- Percentage of P/E cycles limit used so far.





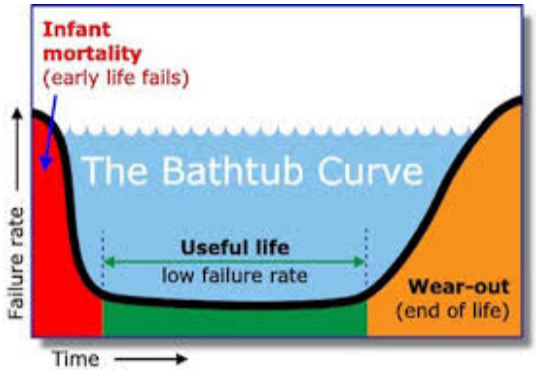
- **eMLC:** The effect of infant mortality is evident!
- **3D-TLC:** The differences are not pronounced, other effects at play (capacity, age).

Age

- Usage affects the reliability of SSDs, due to wear-out of their cells.
- Drive’s age (time deployed in production), as an indicator of wear-out.

Age

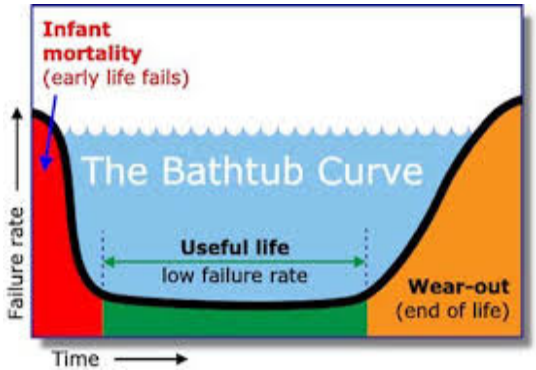
- Usage affects the reliability of SSDs, due to wear-out of their cells.
- Drive's age (time deployed in production), as an indicator of wear-out.



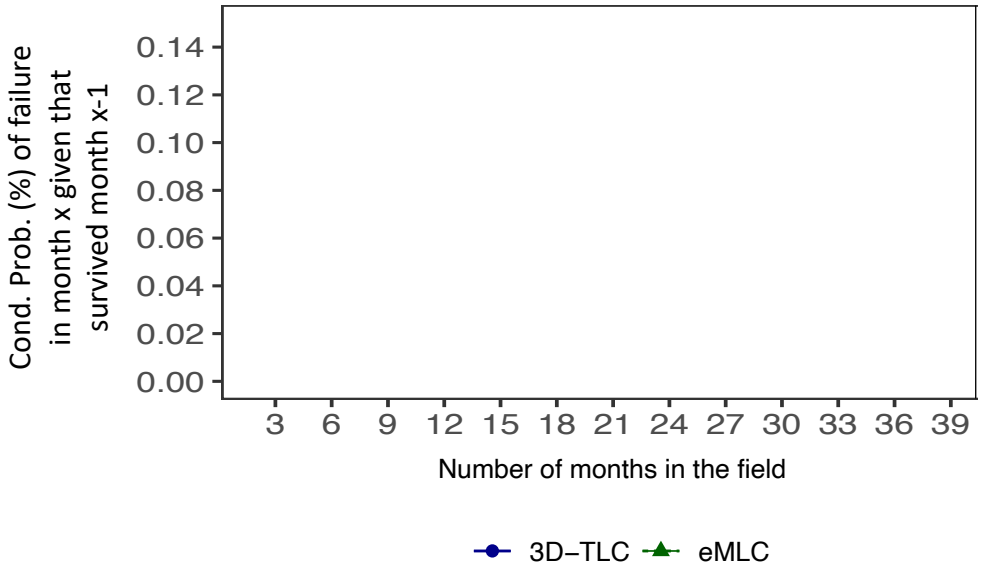
Source: https://www.nrel.gov/

Age

- Usage affects the reliability of SSDs, due to wear-out of their cells.
- Drive's age (time deployed in production), as an indicator of wear-out.

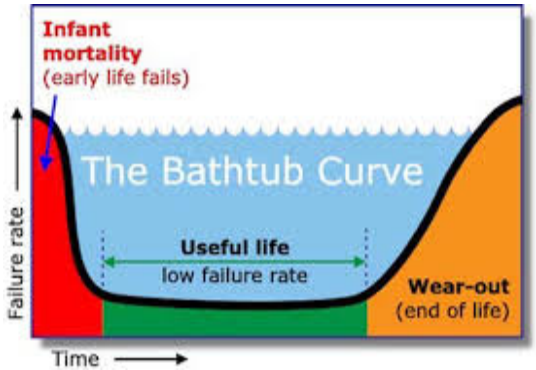


Source: https://www.nrel.gov/

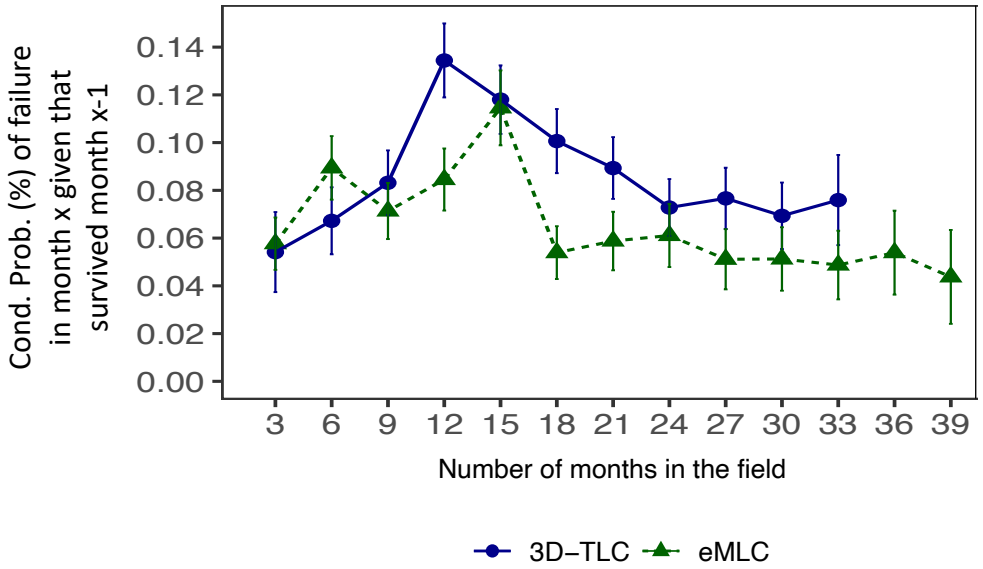


Age

- Usage affects the reliability of SSDs, due to wear-out of their cells.
- Drive's age (time deployed in production), as an indicator of wear-out.

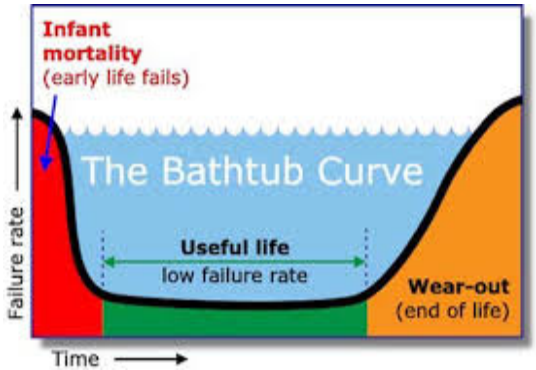


Source: https://www.nrel.gov/

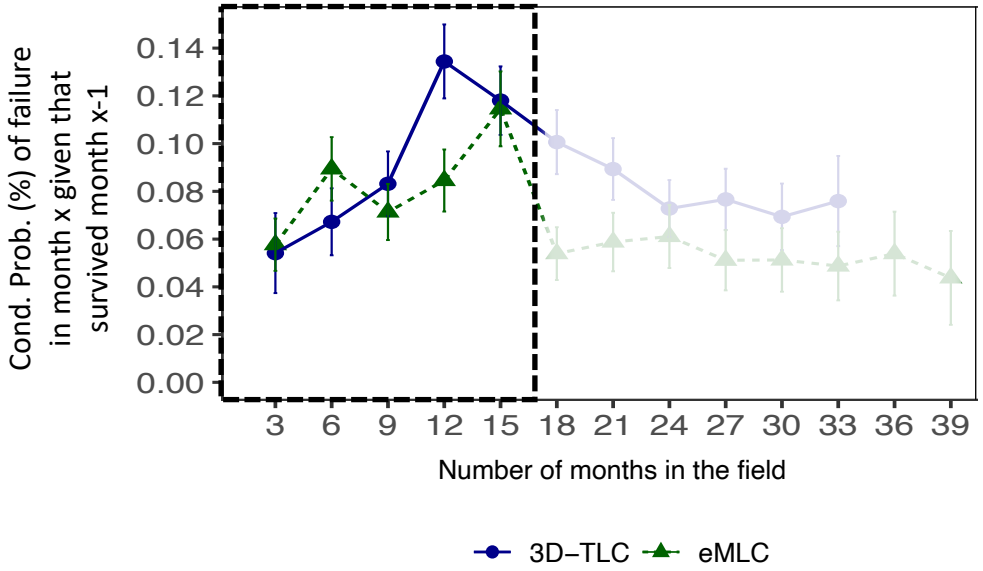


Age

- Usage affects the reliability of SSDs, due to wear-out of their cells.
- Drive's age (time deployed in production), as an indicator of wear-out.



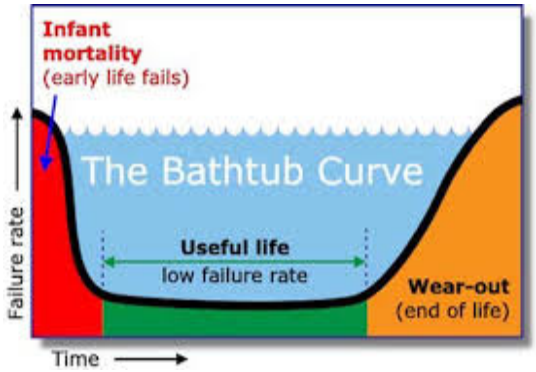
Source: https://www.nrel.gov/



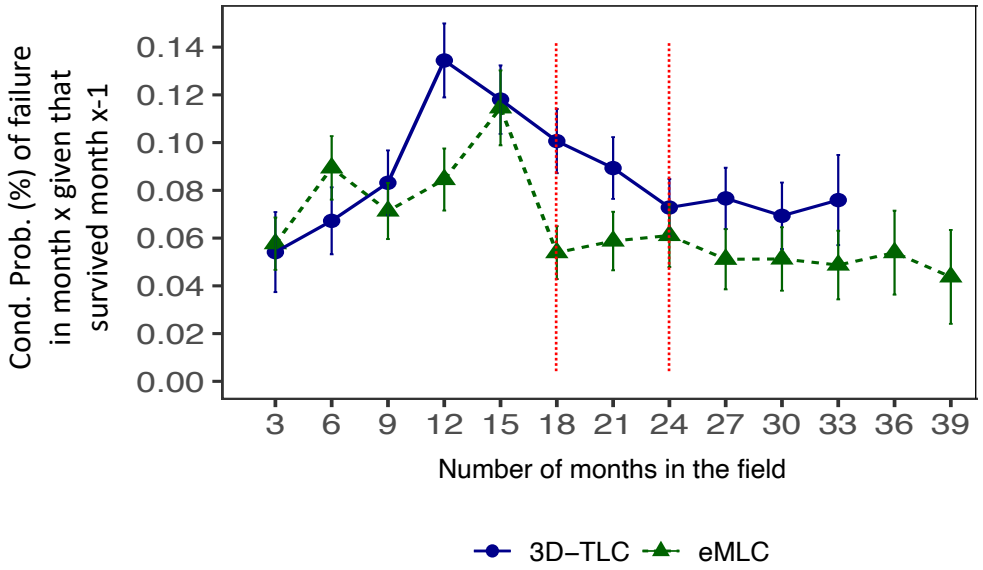
- Infant mortality is significant (**12–15** months)!

Age

- Usage affects the reliability of SSDs, due to wear-out of their cells.
- Drive's age (time deployed in production), as an indicator of wear-out.



Source: https://www.nrel.gov/



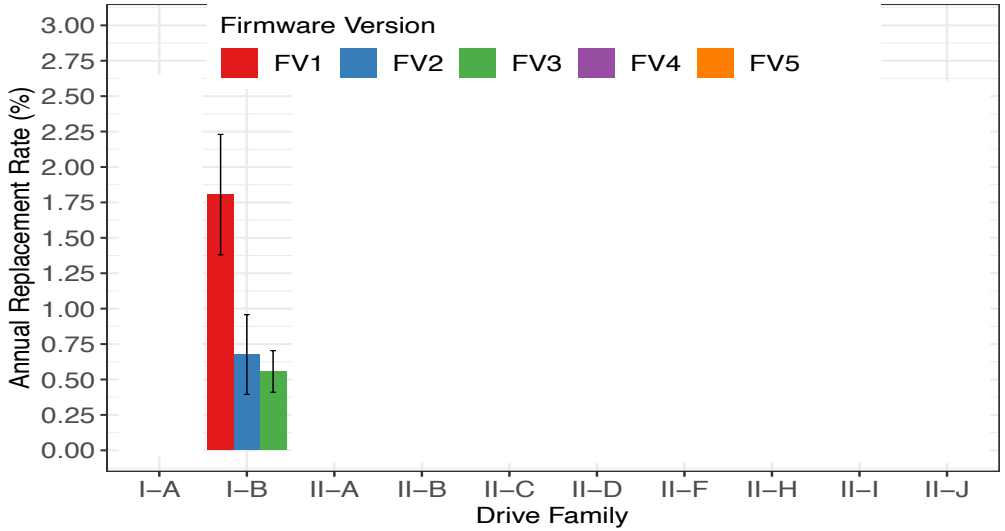
- Infant mortality is significant (**12–15** months)!
- It takes a long time to stabilize (1.5–2 years)!

Firmware Version

- Compare individual firmware versions within the same model:
  - Most SSDs (70%) have the same firmware version in our observation window.
- Consider SSDs which have seen little usage (< 1%).

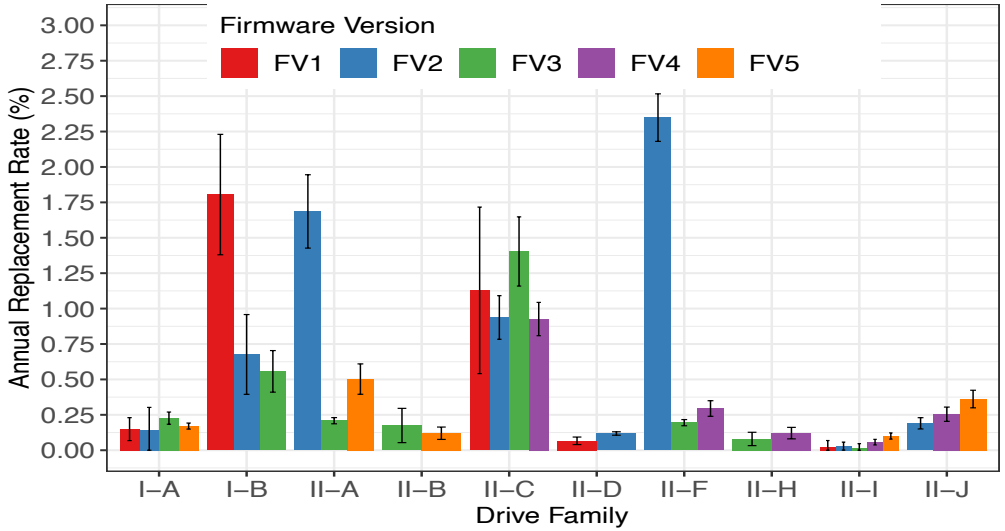
Firmware Version

- Compare individual firmware versions within the same model:
  - Most SSDs (70%) have the same firmware version in our observation window.
- Consider SSDs which have seen little usage (< 1%).



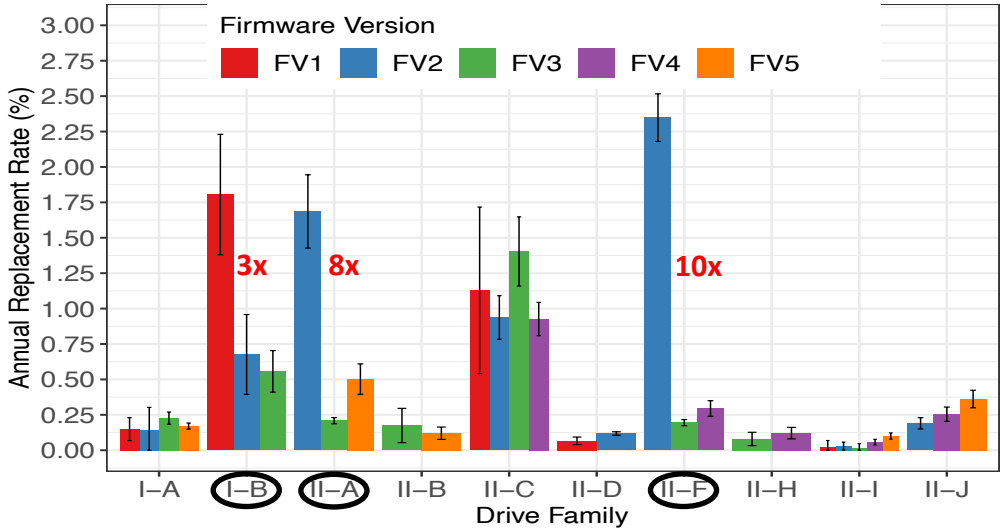
Firmware Version

- Compare individual firmware versions within the same model:
  - Most SSDs (70%) have the same firmware version in our observation window.
- Consider SSDs which have seen little usage (< 1%).



Firmware Version

- Compare individual firmware versions within the same model:
  - Most SSDs (70%) have the same firmware version in our observation window.
- Consider SSDs which have seen little usage (< 1%).



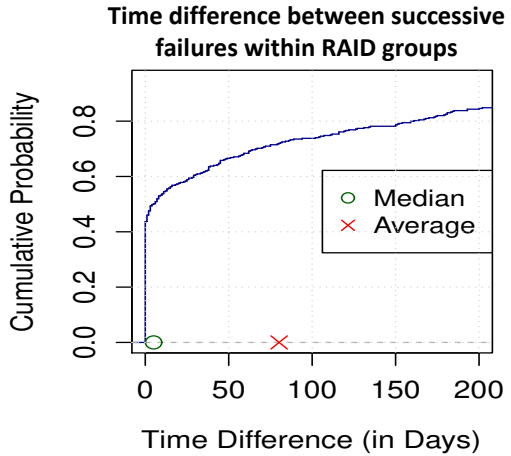
- A drive's firmware version has a tremendous impact on reliability (by a factor of 3-10X)!
- Firmware updates must be made as easy as possible for customers!

Failure correlations within a RAID group

- How frequently do double failures occur?
  - 2% of RAID groups see > 1 failure in our observation window.

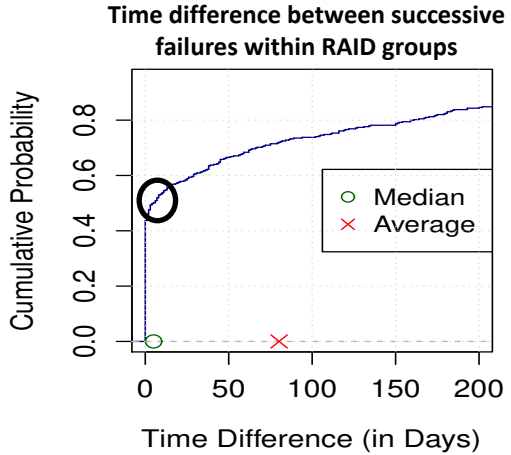
Failure correlations within a RAID group

- How frequently do double failures occur?
  - 2% of RAID groups see > 1 failure in our observation window.
- How quickly after the first does the second failure happen?



Failure correlations within a RAID group

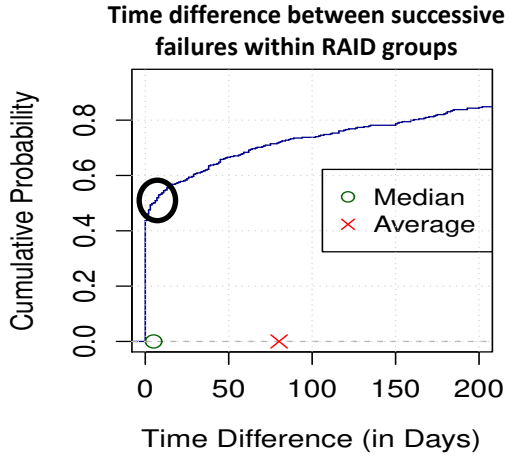
- How frequently do double failures occur?
  - 2% of RAID groups see > 1 failure in our observation window.
- How quickly after the first does the second failure happen?



- 46% of successive failures occur on the same day!

Failure correlations within a RAID group

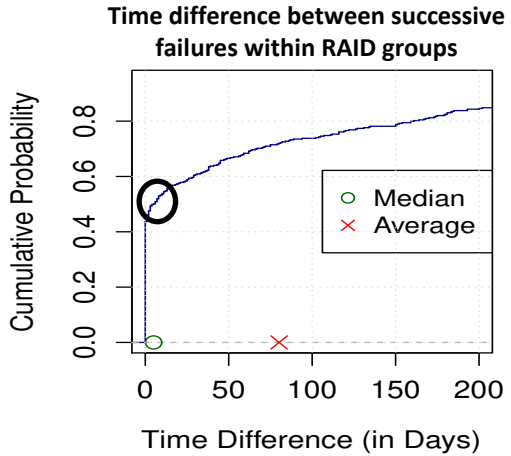
- How frequently do double failures occur?
  - 2% of RAID groups see > 1 failure in our observation window.
- How quickly after the first does the second failure happen?



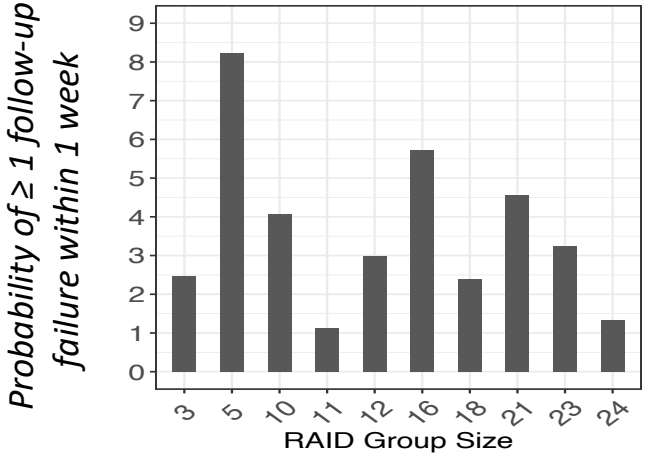
- 46% of successive failures occur on the same day!
- Probability of 2nd failure within a week: 2.54%!

Failure correlations within a RAID group

- How frequently do double failures occur?
  - 2% of RAID groups see > 1 failure in our observation window.
- How quickly after the first does the second failure happen?



- How are they related to RAID group size?



- 46% of successive failures occur on the same day!
- Probability of 2nd failure within a week: 2.54%!
- The chance of a follow-up failure does not show a direct relationship with RAID group size!

Conclusion – Final Remarks

- Many aspects different from expectations:
  - A long period of infant mortality!
  - Higher densities not always experience higher replacement rates.
  - SLC not generally more reliable than MLC.
- Firmware versions can have a significant impact on replacements:
  - Make firmware updates as easy and painless as possible!
- Temporally correlated failures within the same RAID group:
  - No evidence that follow-up failures are correlated with RAID group size.
  - Single-parity RAID configurations, data loss analysis, etc.
- Several other metrics and factors that were not presented:
  - Capacity, Bad Blocks, Spare Blocks consumed, etc.
  - Statistical tests.