Computer Science
UNIVERSITY OF TORONTO

NetApp®

Operational Characteristics of SSDs in Enterprise Storage Systems: A Large-Scale Field Study

Stathis Maneas, Kaveh Mahdaviani, Tim Emami, Bianca Schroeder

USENIX FAST ‘22

Research Questions

- Performance and lifespan of SSDs are affected by operational characteristics.

Goal: Understand operational characteristics of SSDs *in the field*.

- What are write rates in production systems? How close do drives get to reaching wear-out?
- What do R/W ratios look like?
- What are write amplification factors (WAF)? What factors affect them?
- How effective are SSDs at wear-leveling?
- Usage *type*: write-back cache vs. persistent storage.
- How full are production systems?

Systems Description

- We focus on NetApp's *enterprise storage systems*:
  - ~2 million SSDs.
  - 4+ years of data.
  - cMLC, eMLC, 3D-TLC, 3D-eTLC drives.
  - 2 interfaces (SAS and NVMe).
  - 3 manufacturers.
  - 20 drive models:
    - 12 different capacities (200GB – 30TB).
  - Varying age, usage, and system configurations.

What are NAND Usage Rates in the field?

- SSD lifespan is affected by usage due to limited PE cycles:
  - 7K, 10K, 30K.
- Future generations of flash are expected to have lower endurance.

- At what rate do today's drives "use up" their PE cycles?

What are NAND Usage Rates in the field?

- SSD lifespan is affected by usage due to limited PE cycles:
  - 7K, 10K, 30K.
- Future generations of flash are expected to have lower endurance.

- At what rate do today's drives "use up" their PE cycles?

$$NAND\ Usage\ Rate\ (\%)\ per\ Year = \frac{\%\ of\ PE\ Cycle\ Limit\ Used\ So\ Far}{PowerOn\ Years}$$

What are NAND Usage Rates in the field?

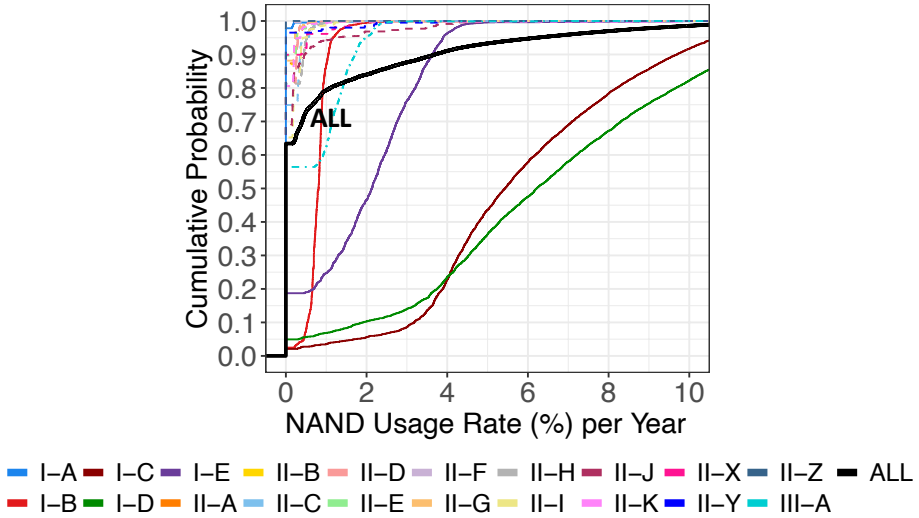
- SSD lifespan is affected by usage due to limited PE cycles:
  - 7K, 10K, 30K.
- Future generations of flash are expected to have lower endurance.

- At what rate do today's drives "use up" their PE cycles?

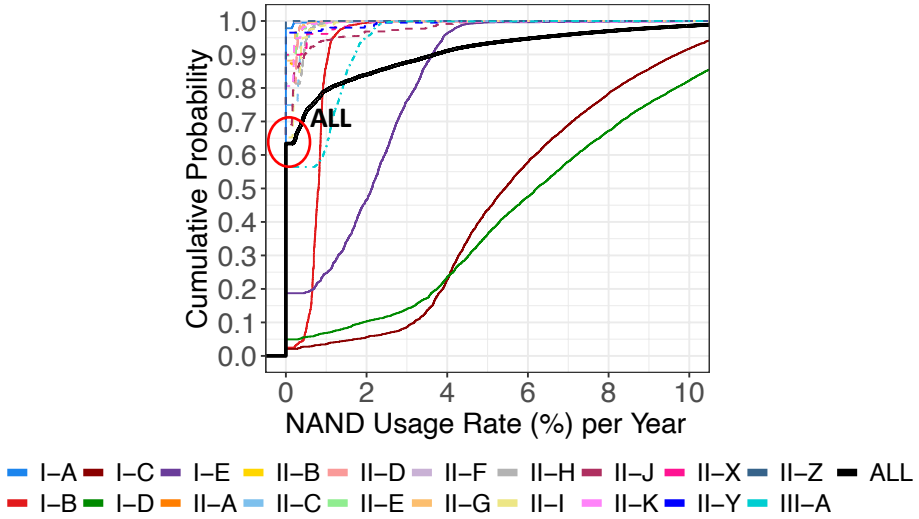
Specified by the manufacturer

$$NAND\ Usage\ Rate\ (\%)\ per\ Year = \frac{\%\ of\ PE\ Cycle\ Limit\ Used\ So\ Far}{PowerOn\ Years}$$

What are NAND Usage Rates in the field?

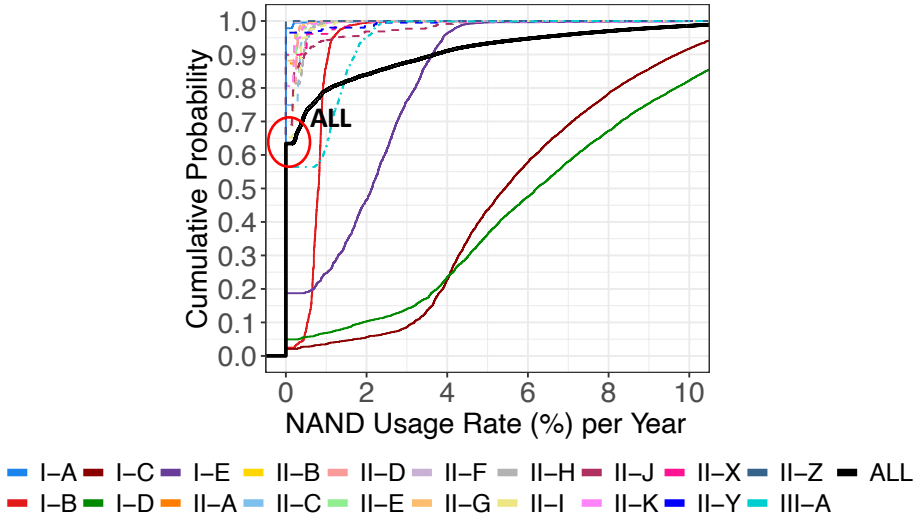


What are NAND Usage Rates in the field?



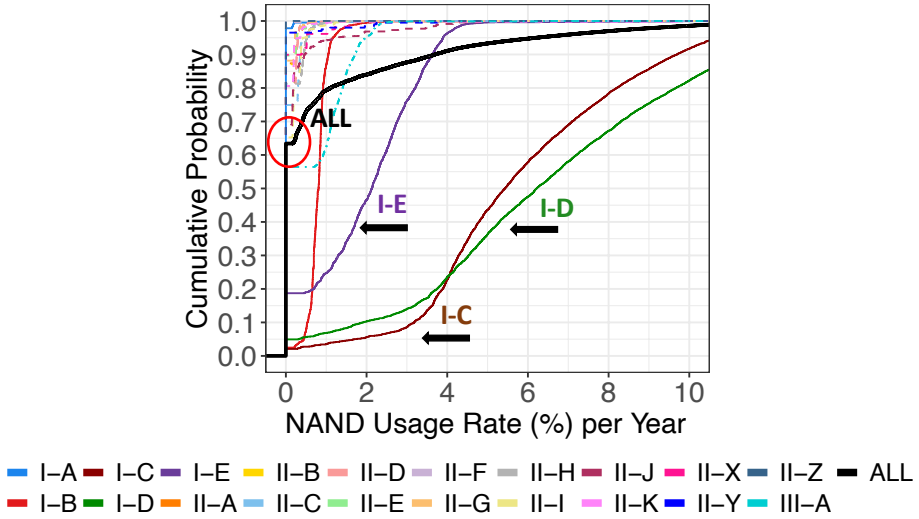
- The majority of SSDs consume PE cycles at a very slow rate!

What are NAND Usage Rates in the field?



- The majority of SSDs consume PE cycles at a very slow rate!
- Huge differences between drive models (not due to higher application writes)!

What are NAND Usage Rates in the field?



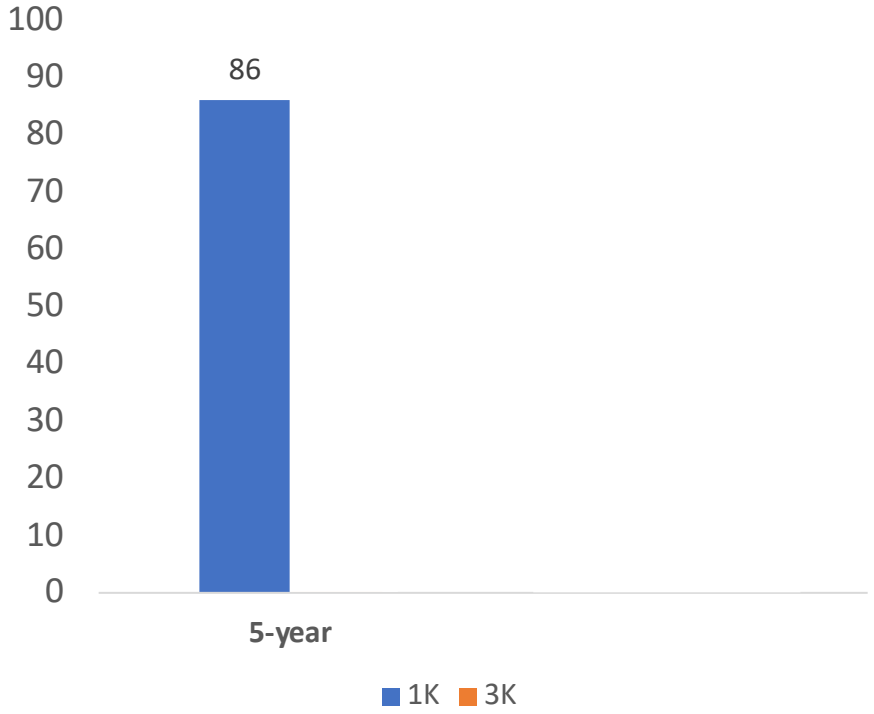
- The majority of SSDs consume PE cycles at a very slow rate!
- Huge differences between drive models (not due to higher application writes)!

How many systems could move toward QLC?

- Assumptions:
  - PE cycle limit is set to 1K or 3K.
  - Warranty periods of 5 and 7 years.

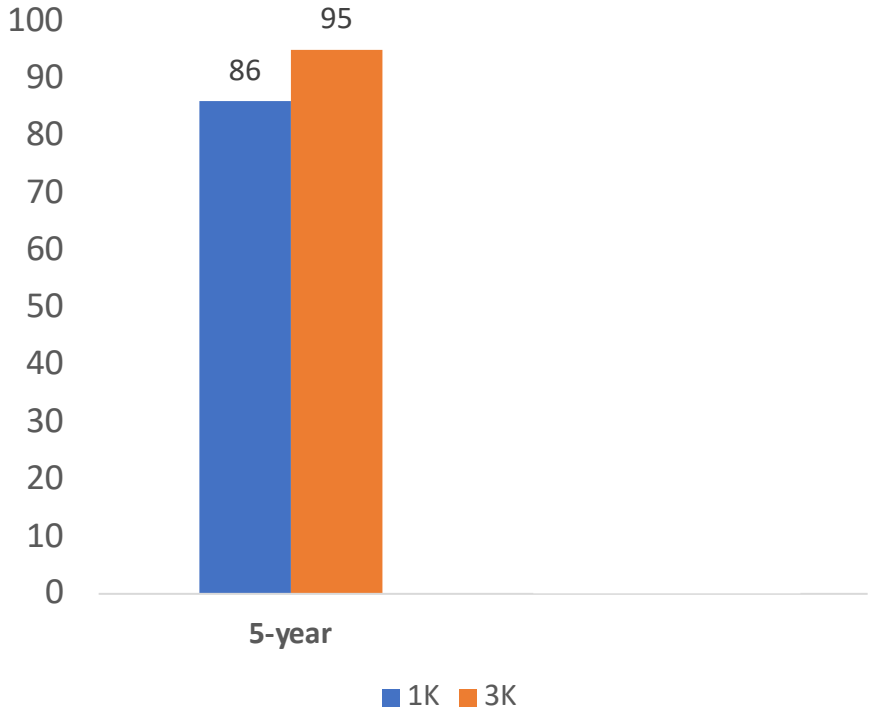
How many systems could move toward QLC?

- Assumptions:
  - PE cycle limit is set to 1K or 3K.
  - Warranty periods of 5 and 7 years.
- **[5 years]**



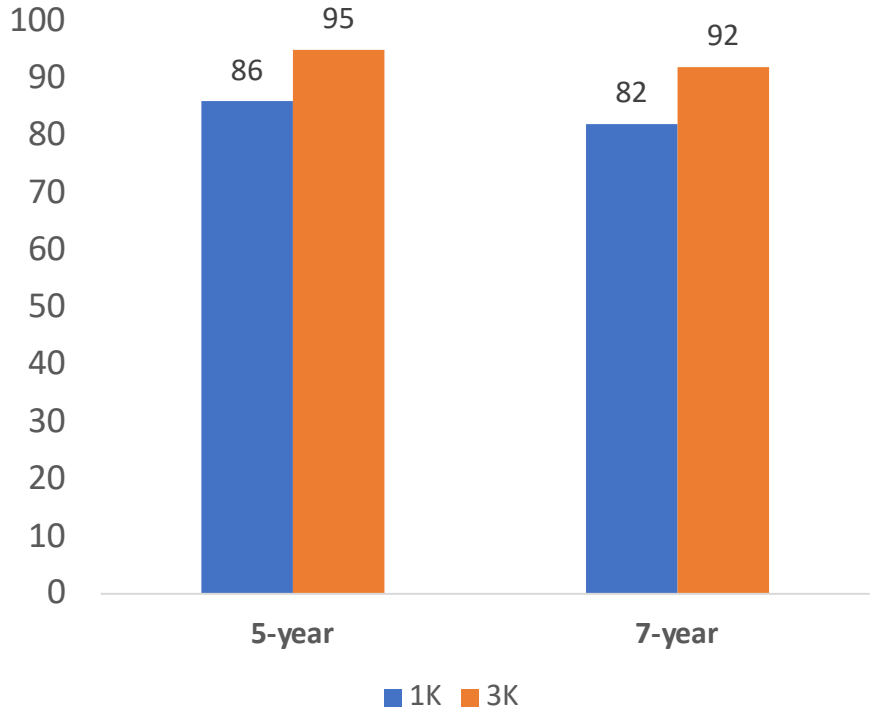
How many systems could move toward QLC?

- Assumptions:
  - PE cycle limit is set to 1K or 3K.
  - Warranty periods of 5 and 7 years.
- **[5 years]** —> 86-95% of the SSD population would not wear out prematurely.



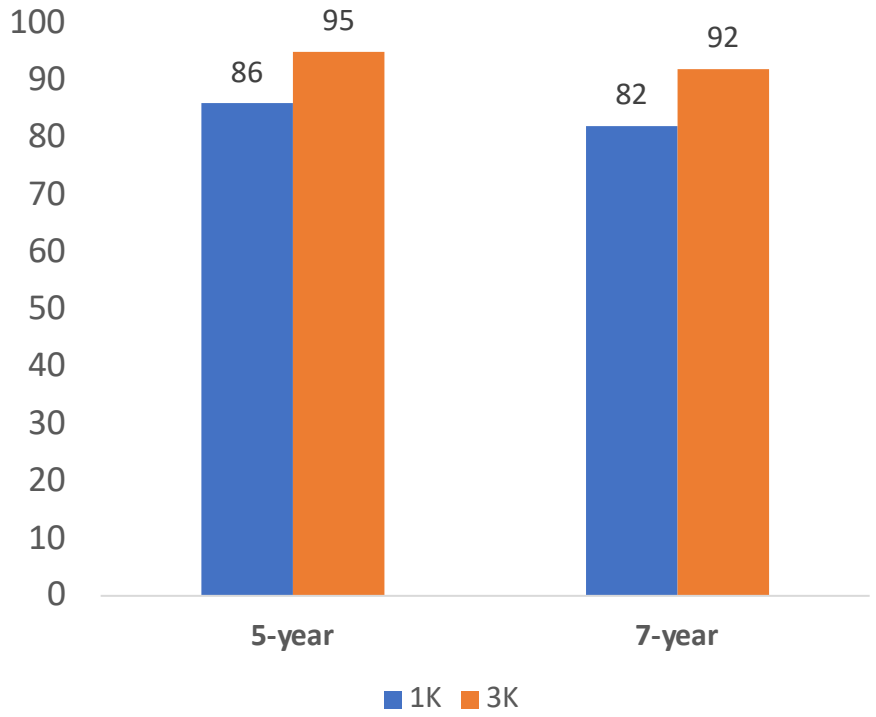
How many systems could move toward QLC?

- Assumptions:
  - PE cycle limit is set to 1K or 3K.
  - Warranty periods of 5 and 7 years.
- **[5 years]** —> 86-95% of the SSD population would not wear out prematurely.
- **[7 years]** —> 82-92% of the SSD population would not wear out prematurely.



How many systems could move toward QLC?

- Assumptions:
  - PE cycle limit is set to 1K or 3K.
  - Warranty periods of 5 and 7 years.
- **[5 years]** —> 86-95% of the SSD population would not wear out prematurely.
- **[7 years]** —> 82-92% of the SSD population would not wear out prematurely.



- Endurance-wise: the systems in our data set can transition to QLC!

Write Amplification Factors (WAF)

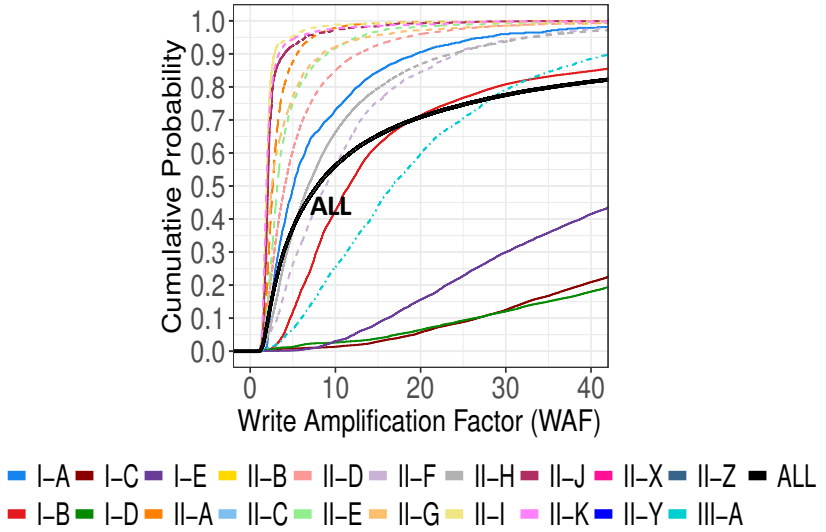
SSDs internally perform housekeeping tasks (garbage collection, wear levelling, etc.) resulting in Write Amplification:

—> impacting endurance & performance.

- What we know:
  - Existing field studies report modest WAF of 1.3 - 1.5.
    - However, very limited scope: single application type or flash technology.
  - Trace-driven simulation studies report WAF up to 7.

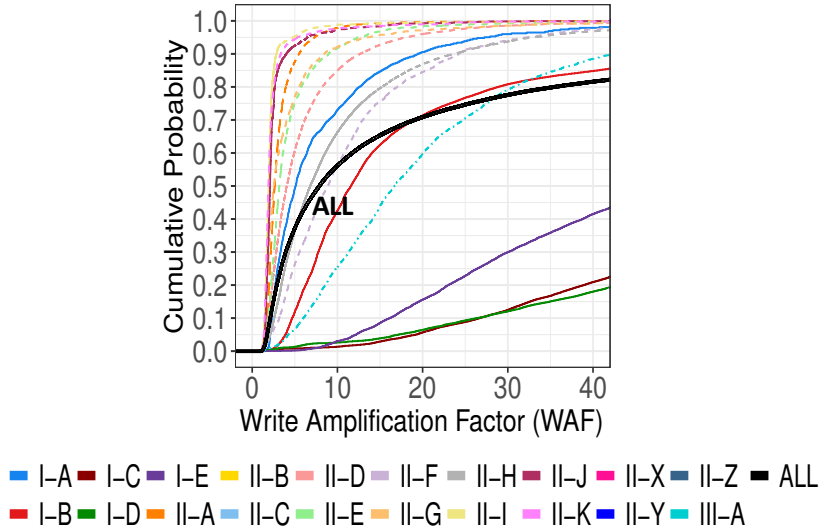
- How effective are SSDs in production systems in controlling WAF under real-world workloads?

Write Amplification Factors (WAF) in production?



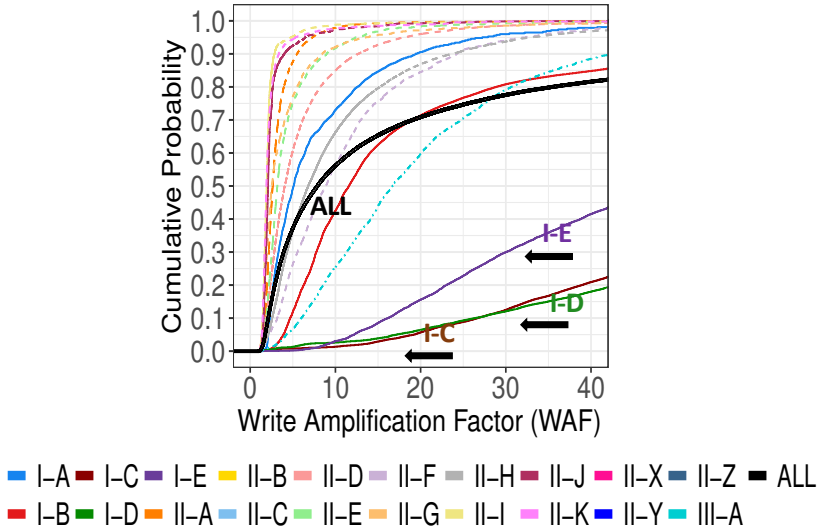
Write Amplification Factors (WAF) in production?

- Huge variation across drive models and manufacturers!



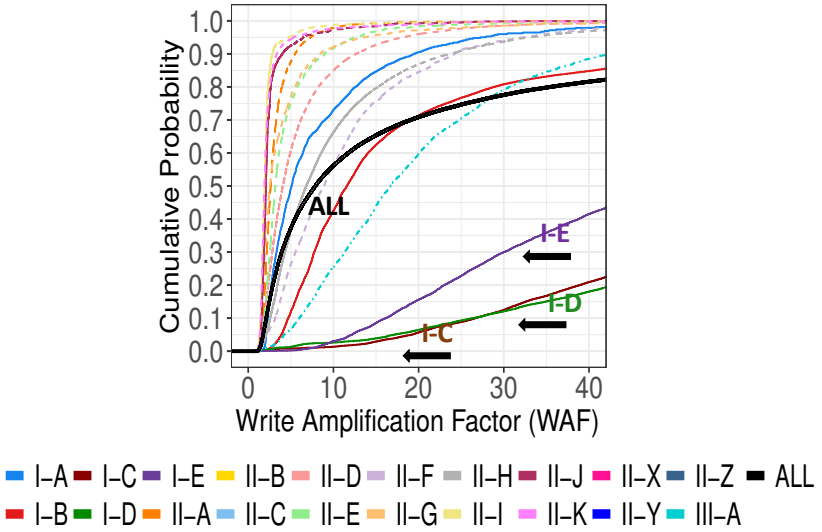
Write Amplification Factors (WAF) in production?

- Huge variation across drive models and manufacturers!
  - Aggressive rewriting of blocks can be the main contributor of WAF!



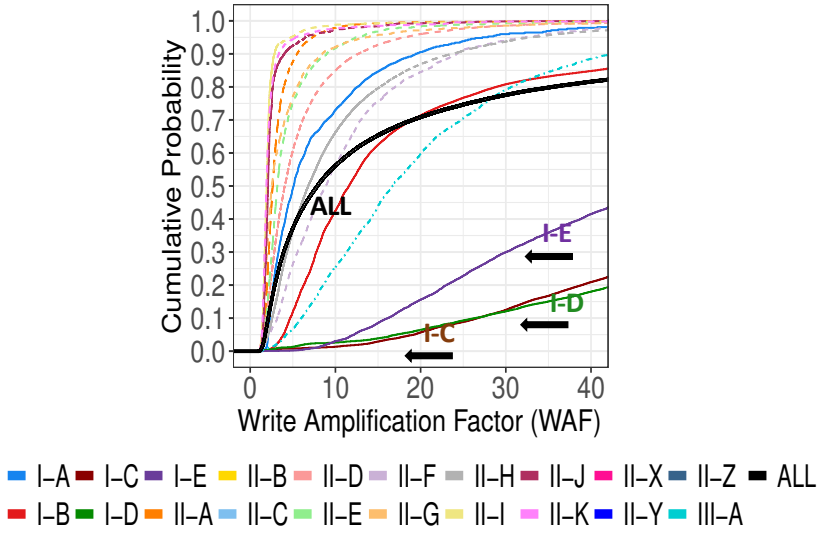
Write Amplification Factors (WAF) in production?

- Huge variation across drive models and manufacturers!
  - Aggressive rewriting of blocks can be the main contributor of WAF!
- Higher WAF than prior field studies (which saw 1.3 & 1.5):
  - Limited scope, not representative of all real-world applications and systems.



Write Amplification Factors (WAF) in production?

- Huge variation across drive models and manufacturers!
  - Aggressive rewriting of blocks can be the main contributor of WAF!
- Higher WAF than prior field studies (which saw 1.3 & 1.5):
  - Limited scope, not representative of all real-world applications and systems.
- Higher WAF than prior trace-driven simulations:
  - Challenging to simulate all complexities of modern FTLs.
  - Importance of getting SSD-based traces into the public domain!



Which factors impact WAF?

- Several factors can affect an SSD's write amplification.

Factor	Impact on WAF
Flash Translation Layer (FTL)	✓
Workload Characteristics	✓

Which factors impact WAF?

- Several factors can affect an SSD's write amplification.

Factor	Impact on WAF
Flash Translation Layer (FTL)	✓
Workload Characteristics	✓
Fullness	X
Over-provisioning (OP)	X

- Fullness and OP have little impact on WAF in practice, unlike commonly assumed!

Which factors impact WAF?

- Several factors can affect an SSD's write amplification.

Factor	Impact on WAF
Flash Translation Layer (FTL)	✓
Workload Characteristics	✓
Fullness	X
Over-provisioning (OP)	X
Multi-stream writes (MSW)	Unclear

- Fullness and OP have little impact on WAF in practice, unlike commonly assumed!
- See the paper for all our results!

Why focus on Read/Write (R/W) Ratios?

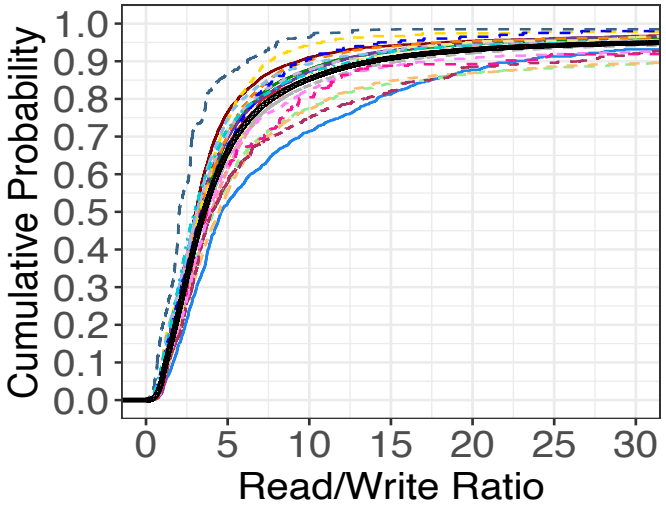
- Mix of reads and writes impact read performance, as reads compete internally with writes.

- Parameterize simulations and experimental testbeds.

What R/W ratios look like in production?

Observations

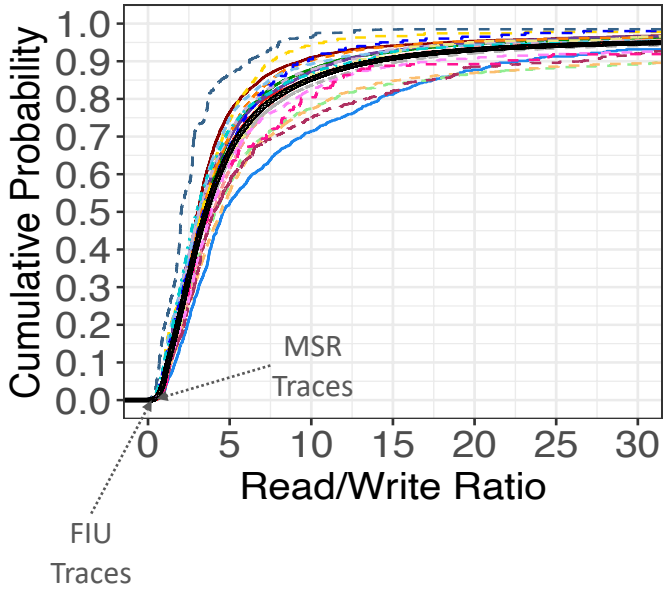
- Most SSDs experience more reads than writes (median ratio is $4:1$)!



What R/W ratios look like in production?

Observations

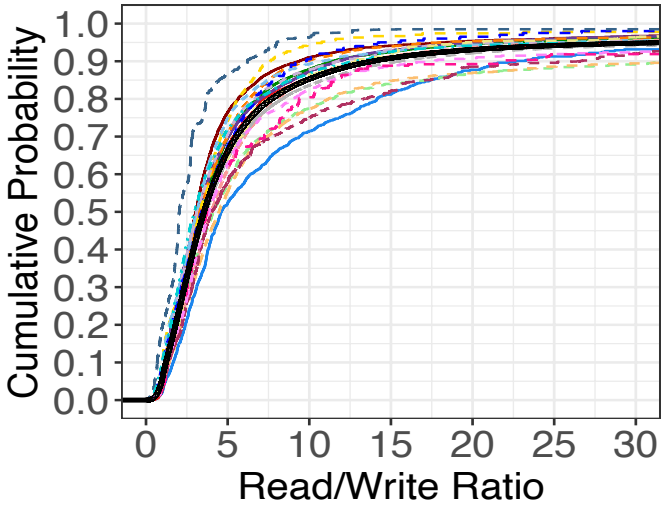
- Most SSDs experience more reads than writes (median ratio is 4 : 1)!
- Stark contrast to HDD-based storage systems, which generally have more writes than reads!



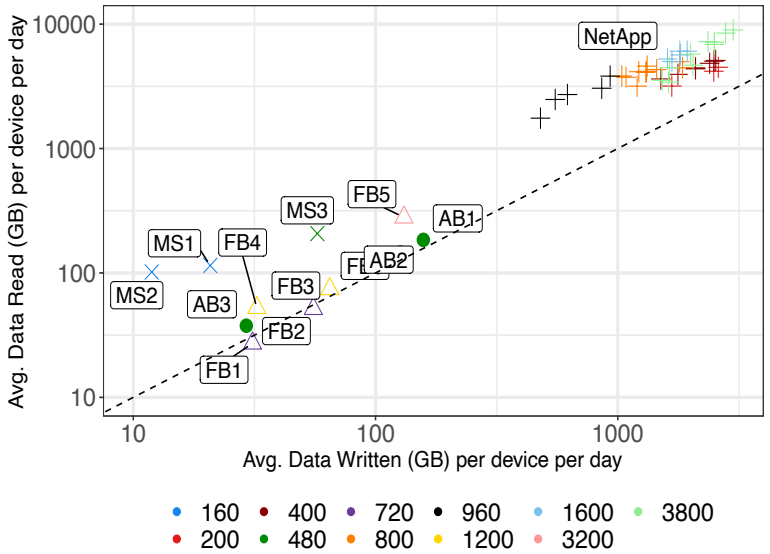
What R/W ratios look like in production?

Observations

- Most SSDs experience more reads than writes (median ratio is $4:1$)!
- Stark contrast to HDD-based storage systems, which generally have more writes than reads!
- How do the ratios in enterprise storage compare against data center drives?



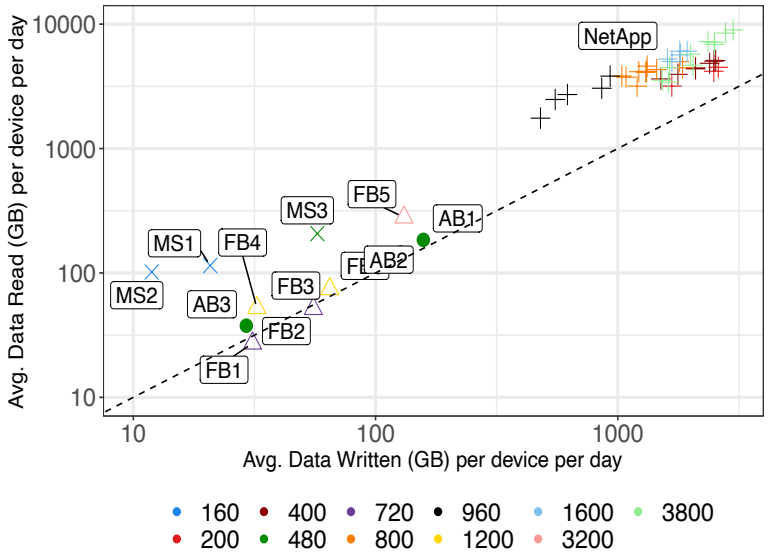
Enterprise storage systems versus data centers



Enterprise storage systems versus data centers

Observations

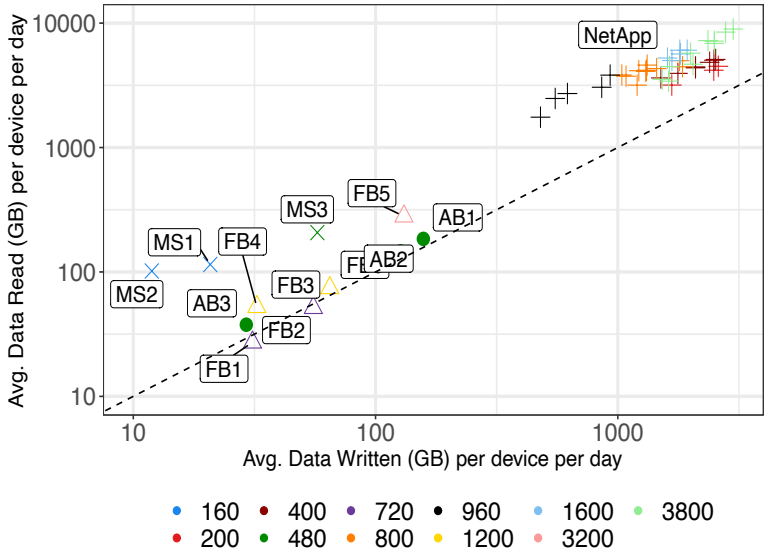
- Workload characteristics vary significantly!
- The workloads in our enterprise storage systems are significantly more intensive:
  - Both read and write rates are higher!



Enterprise storage systems versus data centers

Observations

- Workload characteristics vary significantly!
- The workloads in our enterprise storage systems are significantly more intensive:
  - Both read and write rates are higher!
- Comparison of R/W ratios:
  - **[NetApp]** —> read-dominant.
  - **[FB, AB]** —> ~1:1.
  - **[MS]** —> read-dominant, higher than those in our systems.



Conclusions

- Most SSDs consume PE cycles at a very slow rate (i.e., < 1% / year).
- WAF varies significantly across drive models and manufacturers, and is much larger than previously reported:
  - Larger than 10 at 95% tail.
- Impact of several factors on WAF:
  - FTL, write rates, etc.
- The research community would benefit greatly from representative **SSD-based** traces:
  - Read-dominant workloads (in contrast to HDD-based traces).
  - Trace-driven WAF simulation studies.
- Many more in the paper!