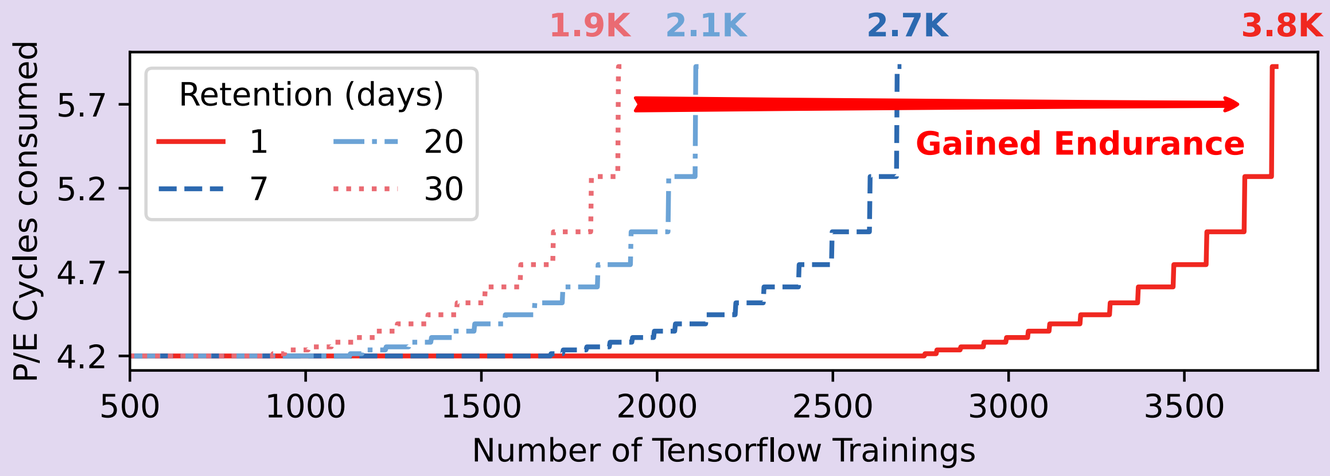
Short-lived data?
Safely Cut Data Protection Tax From Your SSD

Yun-Chih Chen & Tei-Wei Kuo, National Taiwan University

if data retain for **< 1 day**
- ✓ Use simple ECC
- ✓ Reduce storage overhead
- ✓ Expand SSD lifespan



2X more Tensorflow trainings before *wearing out your SSD*

Short-lived data are everywhere

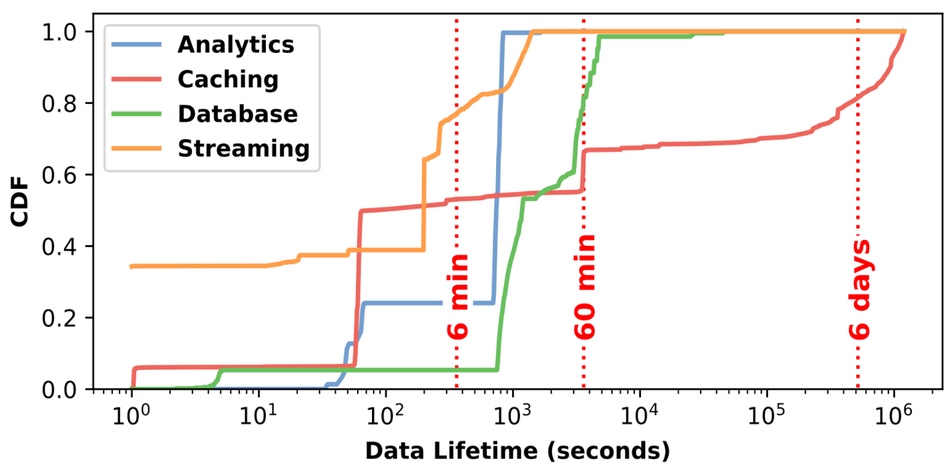
3 Months: required retention for enterprise SSDs

- Microsoft Object Cache & Video Analytics:

  > 50% expires by *6 minutes*

- PageRank (Apache Spark), Facebook Key-Value Database:

  > 80% expires by *60 minutes*

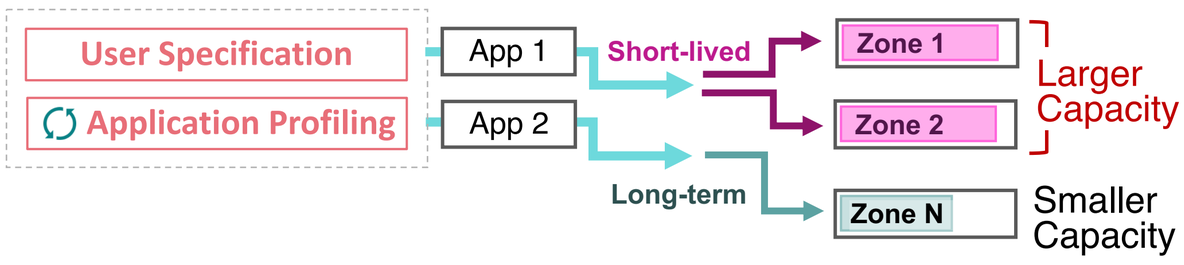


Data Lifetime Cumulative Distribution (CDF)

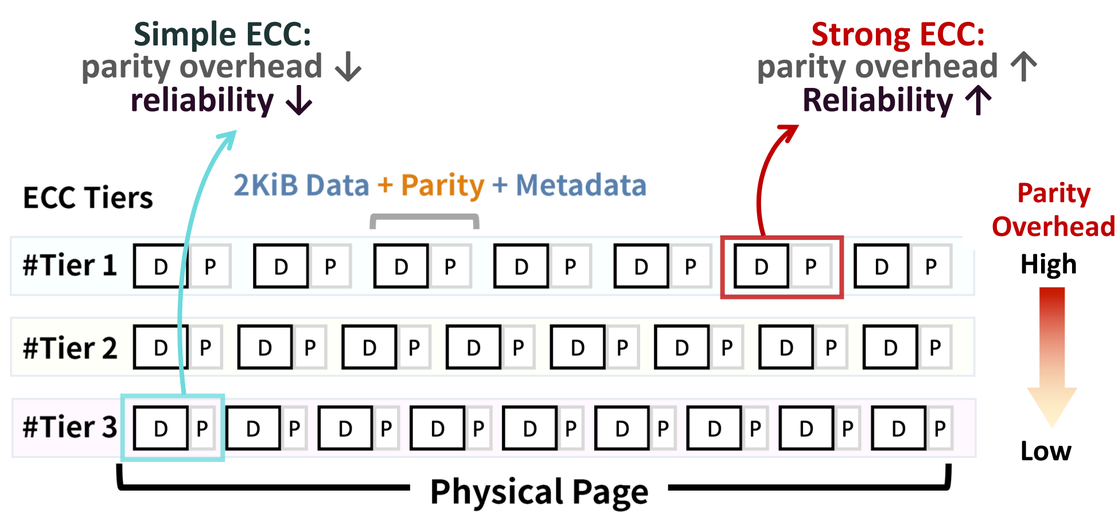
Larger capacity for short-lived data

❶ Get data lifetime
(w/ safety margin)

❸ Cut ECC overhead to store more user data



❷ Lifetime-aware Interface
(Extend NVMe Zone-Namespace)



3-tier Logical Page Layout w/ Adaptive Error Correction

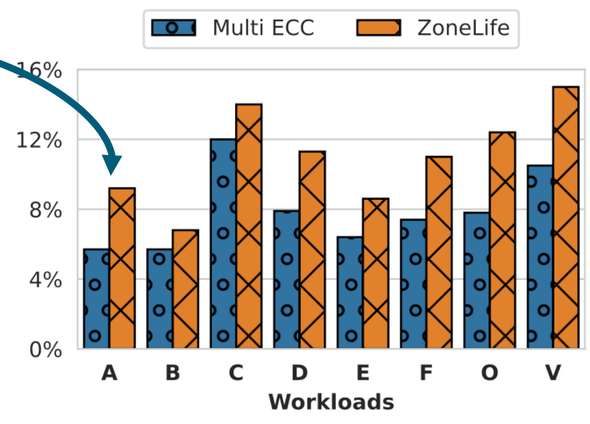
Data Safety ✓ SSD Life 70% ↑ Write Performance 15% ↑

Evaluated Benchmark
- DB & E-mail workloads on F2FS
- ZNS-based RocksDB

Comparisons
- Multi-ECC (Seagate Patent, 2016): weaker ECC in beginning of life, stronger ECC as SSD ages.
- SR-FTL (EuroSys, 2014): uses old blocks to store short-lived data.

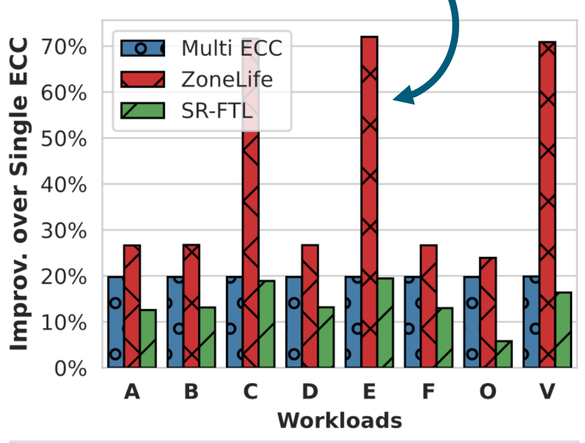
Limitations
- Inapplicable: Data lifetimes fluctuate indefinitely & short-lived data inseparable from long-lived one.

Less write overhead



Write Perf. Improvement

Lower aging rate near end of the SSD's life



Lifespan Improvement

Aged memory can be utilized

Acknowledgments
This work was published in IEEE TCAD in Nov. 2022

Scan the QR code to download the full paper & poster

f07922039@ntu.edu.tw
https://github.com/yunchih