# ZoneLife: How to Utilize Data Lifetime Semantics to Make SSDs Smarter

*Yun-Chih Chen
Advisor: †Tei-Wei Kuo
Estimated graduation date: July, 2023
Department of Computer Science and Information Engineering, National Taiwan University, Taiwan
E-mail: {*f07922039, †ktw}@ntu.edu.tw

This work was published in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, which is available online: 10.1109/TCAD.2022.3224898 and has not been presented at ASP-DAC Ph.D. Forum or DATE Ph.D. Forum. This paper belongs to the track "*ESS3.2 Embedded storage systems organization and management*".

My research focuses on hardware/software co-design for Solid State Drives (SSDs). I'm currently working on the design of peripheral circuits within NAND flash chips that can perform bulk bitwise processing in-memory. The novel circuit will enable index structures in databases to take advantage of SSD's massive parallelism and beyond-DRAM capacity. Here is a list of publications related to my dissertation:

### Authored Related Publications

[1] Hasan Alhasan, Yun-Chih Chen, and Chien-Chung Ho. RVO: Unleashing ssd's parallelism by harnessing the unused power. In *2021 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*, pages 1–6, 2021.

[2] Hasan Alhasan, Yun-Chih Chen, Chien-Chung Ho, and Tei-Wei Kuo. RUSM: Harnessing unused resources in 3d nand ssd to enhance reading performance. In *2022 IEEE 11th Non-Volatile Memory Systems and Applications Symposium (NVMSA)*, pages 63–68, 2022.

[3] Yun-Chih Chen, Chun-Feng Wu, Yuan-Hao Chang, and Tei-Wei Kuo. Reptail: Cutting storage tail latency with inherent redundancy. In *2021 58th ACM/IEEE Design Automation Conference (DAC)*, pages 595–600, 2021.

## I. Potential Impact

Our proposed framework, ZoneLife, has the potential to significantly accelerate I/O operations in workloads that generate a substantial amount of transient data, such as large-scale data analytics, high-performance computing, and database transaction processing. Additionally, ZoneLife can extend the lifespan of their storage backends. ZoneLife is designed to be easily-adoptable, extensible, and allow for modular integration of future storage memory.

## II. Introduction

Modern applications increasingly use Solid State Drives (SSDs) as short-term buffers to absorb bursts of high-volume writes or as low-latency scratchpad memory to store data volumes that are larger than DRAM can accommodate. In both use cases, a significant amount of transient data is written to the SSD, which doesn't need long-term storage.

Since SSD reliability has dropped as a result of increased density, more error protection with more overhead is required. This is crucial for assuring data accuracy when data need to be stored for a long time, because electrons in the memory cell can leak over time and destroy data. If, on the other hand, data is known to have a short retention period, strong error correction would not be required; instead, weaker, less expensive error correction would be sufficient to ensure data accuracy. In other words, if the required storage period is known when the data is written, SSDs can select the best Error-Correction Code (ECC) from a set of codes with varying strengths to store data with varying lifetimes [4]. This has the potential to significantly reduce the processing costs for storing and retrieving short-lived data.

**Limitations of prior works:** The block interface between the host operating system and the disk, originating in the HDD-era, poses a challenge for SSDs to distinguish between short-term and long-term data. Prior research suggests monitoring the frequency of data overwrite or deletion to differentiate between the two [5]. However, this approach can yield imprecise measurements in log-structured storage systems such as RocksDB and F2FS because data deletion is often postponed in favor of batch garbage collection.

Recent research has proposed enriching the block interface by associating each write operation with a stream to indicate that pages with the same stream will be invalidated at a similar time, thus sharing a similar data lifetime [6]. Although this approach effectively improves the efficiency of garbage collection, it is limited in its ability to support a multi-tenant environment due to the limited number of streams.

**Our proposal:** To address the problem of imprecise data lifetime measurements in prior works, we introduce *ZoneLife*, a solution that creates a clear interface for applications to establish a data lifetime contract with the SSD. In many cases, the data lifetime can be predetermined, and even if it is not, the data typically have a bounded and deterministic lifespan, as we will demonstrate later.

ZoneLife offers a flexible interface that allows applications to assist the SSD in avoiding the overhead of over-protecting data while maintaining data safety. Moreover, ZoneLife is easy to adopt by existing software with localized modifications since it builds on the well-established interface. Experimental results show that ZoneLife improves endurance by at least 11% compared to lifetime-oblivious SSDs, while also boosting read

and write throughput by 4% to 15%.

## III. Our contributions

- **Novel integration:** ZoneLife combines two storage components originally designed for different purposes, resulting in a perfect match when considering data lifetime. ZoneLife leverages the multi-rate ECC technique, designed to address the complex error characteristics of SSDs, and the Zoned Namespace (ZNS) command set, designed to improve garbage collection efficiency. ZoneLife inherits their optimization and overcomes the problems arising from limited lifetime hints.
- **Low-overhead implementation:** ZoneLife abstracts away the underlying heterogeneous memory I/O units and integrates with existing SSD firmware components with low memory overheads. By leveraging the sequential write constraints of Zoned Namespace (ZNS), ZoneLife prevents write amplification caused by mismatched write units between the host OS and the memory. Our ECC encoder/decoder implements a systematic multi-ECC construction that balances three crucial design considerations: the theoretical feasibility of the error correction code, the cost of hardware implementation, and the ability to extend to multiple types of flash memory.
- **Data safety as the top priority:** ZoneLife's data allocation strategy ensures sufficient safety margins to guarantee data safety. It derives the optimal ECC by considering both the flash memory reliability and the required data lifetime. While most applications tend to specify a conservative data lifetime to prevent data loss, we show that resource savings are still possible even with a pessimistic data lifetime. We also provide a fallback strategy to deal with the rare case that data must be retrieved after the specified expiration time.

## IV. Key insights

- *Short-lived data is prevalent and can be accurately determined*: Our comprehensive characterization reveals the dominance of short-lived data in several crucial real-world workloads and the significant resource savings that can be achieved by optimizing for them. We also observe that many applications write or invalidate data at deterministic intervals, resulting in stable and concentrated data lifetimes bounded within minutes.
- *Optimizing for short-lived data storage can extend SSD's lifespan:* Our experiments reveal that ZoneLife can double the number of Tensorflow checkpoints that can be stored on an SSD, compared to lifetime-oblivious SSDs, before the SSD fails. This is because an SSD can only endure a finite number of writes before it becomes unusable. By storing short-lived data with weaker ECC, less physical space is required to store the ECC's parity, which can then be utilized for user-writable space.
- *Optimizing for short-lived data storage can enhance application throughput:* Using a weaker ECC for short-lived data enables faster encoding and decoding compared to a stronger ECC. Additionally, since writing to flash memory is significantly more expensive than reading and can cause significant resource competition, reducing the amount of writes can significantly improve overall performance.
- *Aged memory can be resurrected:* Conventional SSDs consider extremely-aged memory pages as unusable, but we demonstrate that it is possible to resurrect them for safe storage of short-lived data. We also found a non-linear relationship between data retention time and the required strength of ECC to maintain data integrity. Strong ECCs offer diminishing returns and are necessary only when the data lifetime requirement is sufficiently large. However, for data with short retention requirements, even extremely aged memory pages can use weak ECCs. This reduces the amount of parity overhead and decelerates the aging of these memory pages. Compared to lifetime-oblivious SSDs without multi-ECC, a significant number of old memory pages become usable.

## V. Key Results

We compare ZoneLife with three existing techniques: *Single-ECC*, *Multi-ECC*, and *SR-FTL*. Single-ECC represents the baseline SSD that employs only one ECC throughout its lifetime. *Multi-ECC* [7] represents the state-of-the-art technique in which the SSD uses a weaker ECC in the beginning of life and stronger ECCs as the SSD ages. SR-FTL [5] is the state-of-the-art data-lifetime-aware technique that utilizes old flash blocks to store short-lived data.

**Performance Improvement:** In the evaluated workloads, ZoneLife exhibits a write throughput improvement of 5% to 15% compared to Single-ECC, and a 1% to 4% improvement compared to Multi-ECC. As an SSD approaches the end of its lifespan, Multi-ECC utilizes stronger ECCs while ZoneLife continues to use weaker ECCs for short-lived data. This means that the advantage of ZoneLife becomes more evident in an aged SSD.

**Lifetime Improvement:** ZoneLife reduces the amount of written space, as measured by *Program/Erase cycles*. In seven out of the eight tested workloads, ZoneLife reduced P/E cycle consumption by at least 11% compared to Single-ECC. Furthermore, ZoneLife's P/E cycle consumption is up to 10% better than Multi-ECC. Before the SSD reaches its end of life, ZoneLife can write 21% to 71% more data than Single-ECC, up to 52% more data than Multi-ECC, and 16% to 51% more data than SR-FTL. Moreover, experiments indicate that ZoneLife exhibits a lower aging rate than Multi-ECC near the end of the SSD's life.

## References


[4] Y. Sun, M. Karkooti, and J. R. Cavallaro, "Vlsi decoder architecture for high throughput, variable block-size and multi-rate ldpc codes," in *IEEE International Symposium on Circuits and Systems*, 2007.

[5] P. Huang, G. Wu, X. He, and W. Xiao, "An aggressive worn-out flash block management scheme to alleviate ssd performance degradation," in *EuroSys*, New York, NY, USA, 2014.

[6] J.-U. Kang, J. Hyun, H. Maeng, and S. Cho, "The multi-streamed solid-state drive," in *HotStorage*, 2014.

[7] Y. Cai, "Error correction code (ECC) selection using probability density functions of error correction capability in storage controllers with multiple error correction codes," U.S. Patent, 2016.