

# Data Placement Using a Classifier for Hybrid SSDs

Subjects: Computer Science, Hardware & Architecture

Contributor: Heeseong Cho, Taeseok Kim

Modern Solid-State Drives (SSDs) are increasingly adopting Quad-Level Cell (QLC) flash memory, a technology that allows for the storage of four bits of data in a single cell, as their primary storage medium to significantly enhance storage capacity.

Keywords: data placement ; hybrid SSDs ; storage performance ; machine learning

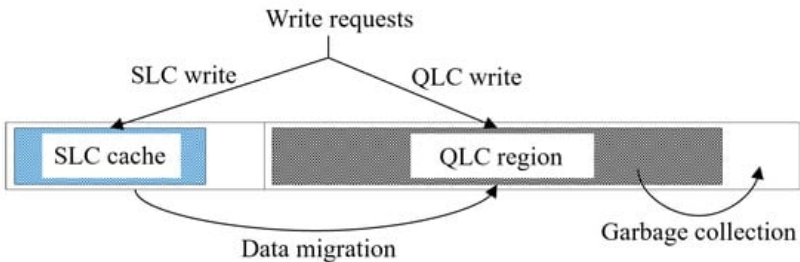
## 1. Introduction

Modern Solid-State Drives (SSDs) are increasingly adopting Quad-Level Cell (QLC) flash memory, a technology that allows for the storage of four bits of data in a single cell, as their primary storage medium to significantly enhance storage capacity [1][2][3][4]. Although this advancement offers a substantial price advantage over SSDs utilizing Single-Level Cell (SLC), Multi-Level Cell (MLC), or Triple-Level Cell (TLC) technologies, QLC SSDs often face limitations in terms of I/O performance and lifespan, a comparison of which is detailed in **Table 1** [5]. To mitigate these drawbacks, recent innovations in SSD technology have led to the exploration of hybrid SSD architectures. These architectures strategically program certain blocks of QLC memory to operate in SLC mode, which involves storing just one bit per cell, therefore harnessing the high-speed advantages of SLC memory [6][7][8]. In these hybrid designs, the SLC region is commonly utilized as a cache layer for the remaining QLC blocks, enhancing the overall efficiency and performance of the storage device [9][10][11].

**Table 1.** Performance of SLC, MLC, TLC, and QLC flash memories.

	SLC	MLC	TLC	QLC
Read time	20–25 us	55–110 us	75–170 us	120–200 us
Write time	50–100 us	0.4–1.5 ms	0.8–2 ms	2–3 ms
Erase time	2–5 ms	5–10 ms	10–15 ms	15–20 ms
Max. P/E	100,000	15,000	3000–5000	800–1500

In the context of hybrid SSD architectures, a notable challenge arises due to the inverse relationship between cache size and available data storage space. As the allocated SLC cache size increases, the available space for data storage correspondingly decreases, therefore emphasizing the need for efficient and strategic management of the SLC cache [12][13]. This management becomes particularly critical when the SLC cache reaches its capacity, necessitating the initiation of processes such as garbage collection or the migration of certain data from the SLC cache to the slower QLC region (**Figure 1**). These processes typically involve employing sophisticated algorithms, and they often result in additional write and erase operations.



**Figure 1.** Operations in the SLC/QLC hybrid SSD architecture.

A specific challenge within this management process is the handling of cold data, which is data that is not frequently accessed and, therefore, is unlikely to be reused in the short term. When such data occupies space in the SLC cache, it eventually needs to be transferred to the QLC region, an action that, in many cases, proves to be unnecessary if the data are not accessed again <sup>[14][15]</sup>. To address this inefficiency, a more effective strategy involves the accurate identification of cold data at the outset, followed by its direct storage in the QLC region, bypassing the SLC cache altogether. This approach was explored by Im et al. <sup>[7]</sup>, who classified data as hot or cold based on the size of the I/O requests, with the assumption that smaller-sized I/O requests usually correspond to hot data. Building on these insights, Yoo et al. <sup>[16]</sup> furthered this field of study by implementing reinforcement learning techniques to determine various parameters in hybrid SSDs, including the optimal size for the SLC cache and the threshold for distinguishing between hot and cold data.

The task of deciding whether to allocate incoming write requests to the SLC cache or to route them directly to the QLC region in hybrid SSDs is both essential and complex. This complexity arises from the dynamic nature of workload characteristics and the fluctuating status of the SSD itself. Despite the intricate nature of this decision-making process, the fundamental goal remains clear: to optimally choose between these two storage options for each write request. Given this challenge, machine learning emerges as a particularly promising approach due to its ability to process and learn from complex and variable data.

## **2. Data Placement Using a Classifier for SLC/QLC Hybrid SSDs**

In the evolving landscape of hybrid SSD architecture, two primary configurations have emerged as dominant: hard partitioning and soft partitioning. Hard partitioning involves the use of physically distinct memory chips such as SLC, MLC, TLC, or QLC within a single hybrid SSD, as detailed by Tripathy et al. <sup>[12]</sup>. This configuration enables the effective utilization of different types of memory technologies within the same device, each serving specific storage roles based on their performance characteristics.

In the realm of hard-partitioned hybrid SSDs, Kwon et al. <sup>[8]</sup> implemented an innovative segmentation strategy within the SLC cache. They created separate areas for file system journal data and regular data, aiming to enhance sequential journal writes and, therefore, reduce the overhead associated with garbage collection. During periods of low activity, data are strategically moved from the SLC cache to TLC flash memory. In scenarios where the SLC cache reaches full capacity, normal data bypasses the cache and is written directly to the TLC flash, optimizing space utilization. Furthering this approach, Jung et al. <sup>[17]</sup> proposed a method that employs multiple SLC chips as a cache layer for MLC flash memory. In this system, a single SLC chip is designated for the storage of both random and sequential data. This chip is utilized until it runs out of free blocks. Upon reaching its maximum capacity, with no available space for additional logs or data blocks, another SLC chip is activated for continued data storage. Yao et al. <sup>[9]</sup> discussed a different system, where SLC and MLC pages are allocated using two distinct mapping caches, one for hot data and the other for normal data. This system prioritizes the use of SLC pages for hot data that register hits in its dedicated cache while defaulting to MLC pages for other data types.

Chang et al. <sup>[18]</sup> proposed a unique management technique for SLC-MLC hybrid SSDs. In their approach, hot data are preferentially stored in SLC flash, whereas cold data are directly written to MLC flash. This method employs a K-means algorithm to periodically determine and adjust the threshold for estimating the hotness of data, ensuring optimal data placement. Similarly, Hachiya et al. <sup>[19]</sup> introduced a cache management technique tailored for MLC-TLC hybrid SSDs. In their system, hot data are retained in MLC flash, and cold data, once identified, are transferred to TLC flash. This methodology reflects an ongoing trend in hybrid SSD technology, where smart data management and strategic cache utilization play crucial roles in enhancing overall storage efficiency and performance.

Soft partitioning presents a contrasting approach within hybrid SSD architecture, characterized by its flexibility in allowing various types of flash memory to emulate SLC-type flash characteristics. This versatility offers a dynamic solution to the rigid physical constraints of hard partitioning. Expanding on this concept, Im et al. <sup>[7]</sup> put forward a cache management strategy for SLC-MLC hybrid SSDs. Their approach involves classifying data as either hot or cold based on the size of the incoming requests. Hot data, typically smaller in size and more frequently accessed, is allocated to the SLC cache for fast retrieval. In contrast, cold data, which is less likely to be accessed frequently, is directly written to the MLC flash, therefore optimizing the usage of the SLC cache. Crucially, this technique incorporates a dynamic adjustment of the hotness threshold, which is calibrated based on the observed frequency of data migrations to the MLC flash, ensuring efficient cache utilization and data flow.

Yoo et al. <sup>[16]</sup> introduced another management policy for SLC-QLC hybrid SSDs, utilizing the principles of reinforcement learning. This policy is specifically tailored to distinguish between hot and cold data, directing hot data to the SLC cache

while routing cold data straight to the QLC flash. The innovative aspect of this policy lies in its dynamic nature, allowing for continuous modification of both the size of the SLC cache and the threshold that differentiates hot from cold data, adapting in real time to changing data access patterns. Furthering the advancements in soft partitioning, Zhang et al. [20] developed a method tailored for SLC-TLC hybrid SSDs. Their approach focuses on spreading SLC pages across multiple planes to significantly enhance parallelism, therefore reducing latency associated with data migration. This strategy includes a sophisticated policy for allocating these planes, which takes into account not only the availability of SLC pages but also the capabilities for parallel processing. This method represents a strategic balance between enhancing data access speed and optimizing the use of available storage resources, a key consideration in the evolving landscape of hybrid SSD technology.

Most existing research in the realm of hybrid SSDs has predominantly concentrated on the identification and management of hot data for the effective utilization of the SLC cache. In addition to this, several studies have ventured into the application of machine-learning techniques, aiming to ascertain or fine-tune the optimal size of the SLC cache and to establish precise thresholds for evaluating the hotness level of incoming data. These thresholds are crucial in determining which data benefits most from the faster SLC cache and which can be relegated to slower storage.

---

## References

1. Liang, S.; Qiao, Z.; Tang, S.; Hochstetler, J.; Fu, S.; Shi, W.; Chen, H.B. An empirical study of quad-level cell (QLC) nand flash ssds for big data applications. In Proceedings of the 2019 IEEE International Conference on Big Data, Los Angeles, CA, USA, 9–12 December 2019; pp. 3676–3685.
2. Chen, Q.; Wang, S.; Zhou, Y.; Wu, F.; Li, S.; Wang, Z.; Xie, C. PACA: A page type aware read cache scheme in QLC flash-based SSDs. In Proceedings of the 2022 IEEE 40th International Conference on Computer Design, Olympic Valley, CA, USA, 23–26 October 2022; pp. 59–66.
3. Luo, L.; Li, S.; Lv, Y.; Shi, L. Performance and reliability optimization for high-density flash-based hybrid SSDs. *J. Syst. Archit.* 2023, 136, 102830.
4. Li, S.; Luo, L.; Lv, Y.; Shi, L. Latency aware page migration for read performance optimization on hybrid SSDs. In Proceedings of the 2022 IEEE 11th Non-Volatile Memory Systems and Applications Symposium (NVMSA), Taipei, Taiwan, 23–25 August 2022; pp. 33–38.
5. Stoica, R.; Pletka, R.; Ioannou, N.; Papandreou, N.; Tomic, S.; Pozidis, H. Understanding the design trade-offs of hybrid flash controllers. In Proceedings of the 2019 IEEE 27th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), Rennes, France, 21–25 October 2019; pp. 152–164.
6. Liu, D.; Yao, L.; Long, L.; Shao, Z.; Guan, Y. A workload-aware flash translation layer enhancing performance and lifespan of TLC/SLC dual-mode flash memory in embedded systems. *Microprocess. Microsyst.* 2017, 52, 343–354.
7. Im, S.; Shin, D. ComboFTL: Improving performance and lifespan of MLC flash memory using SLC flash buffer. *J. Syst. Archit.* 2010, 56, 641–653.
8. Kwon, K.; Kang, D.H.; Eom, Y.I. An advanced SLC-buffering for TLC NAND flash-based storage. *IEEE Trans. Consum. Electron.* 2017, 63, 459–466.
9. Yao, Y.; Fan, J.; Zhou, J.; Kong, X.; Gu, N. Hdftl: An on-demand flash translation layer algorithm for hybrid solid state drives. *IEEE Trans. Consum. Electron.* 2021, 67, 50–57.
10. Wei, Q.; Li, Y.; Jia, Z.; Zhao, M.; Shen, Z.; Li, B. Reinforcement learning-assisted management for convertible SSDs. In Proceedings of the 2023 60th ACM/IEEE Design Automation Conference (DAC), San Francisco, CA, USA, 9–13 July 2023; pp. 1–6.
11. Li, J.; Li, M.; Cai, Z.; Trahay, F.; Wahib, M.; Gerofi, B.; Liao, J. Intra-page cache update in SLC-mode with partial programming in high density SSDs. In Proceedings of the 50th International Conference on Parallel Processing, Lemont, IL, USA, 9–12 August 2021; pp. 1–10.
12. Tripathy, S.; Satpathy, M. SSD internal cache management policies: A survey. *J. Syst. Archit.* 2022, 122, 102334.
13. Xiao, C.; Qiu, S.; Xu, D. PASM: Parallelism aware space management strategy for hybrid SSD towards in-storage DNN training acceleration. *J. Syst. Archit.* 2022, 128, 102565.
14. Lee, J.; Kim, J.S. An empirical study of hot/cold data separation policies in solid state drives (SSDs). In Proceedings of the 6th International Systems and Storage Conference, Haifa, Israel, 30 June–2 July 2013; pp. 1–6.

15. Wang, Y.L.; Kim, K.T.; Lee, B.; Youn, H.Y. A novel buffer management scheme based on particle swarm optimization for SSD. *J. Supercomput.* 2018, 74, 141–159.
16. Yoo, S.; Shin, D. Reinforcement learning-based SLC cache technique for enhancing SSD write performance. In *Proceedings of the 12th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 20)*, Virtual Event, 13–14 July 2020.
17. Jung, S.; Song, Y.H. Hierarchical use of heterogeneous flash memories for high performance and durability. *IEEE Trans. Consum. Electron.* 2009, 55, 1383–1391.
18. Chang, L.P. A hybrid approach to NAND-flash-based solid-state disks. *IEEE Trans. Comput.* 2010, 59, 1337–1349.
19. Hachiya, S.; Johguchi, K.; Miyaji, K.; Takeuchi, K. Hybrid triple-level-cell/multi-level-cell NAND flash storage array with chip exchangeable method. *Jpn. J. Appl. Phys.* 2014, 53, 04EE04.
20. Zhang, W.; Cao, Q.; Jiang, H.; Yao, J.; Dong, Y.; Yang, P. SPA-SSD: Exploit heterogeneity and parallelism of 3D SLC-TLC hybrid SSD to improve write performance. In *Proceedings of the 2019 IEEE 37th International Conference on Computer Design (ICCD)*, Abu Dhabi, United Arab Emirates, 17–20 November 2019; pp. 613–621.

---

Retrieved from <https://encyclopedia.pub/entry/history/show/125359>