Life-after-Death: Exploring Thermal Annealing Conditions to Enhance 3D NAND SSD Endurance

Matchima Buddhanoy
Colorado State University
Fort Collins, CO, USA
matchima.buddhanoy@colostate.edu

Sudeep Pasricha
Colorado State University
Fort Collins, CO, USA
sudeep@colostate.edu

Biswajit Ray
Colorado State University
Fort Collins, CO, USA
biswajit.ray@colostate.edu

Abstract

In this paper, we evaluate thermal annealing effects on the endurance of commercial off-the-shelf (COTS) 3D NAND flash memory beyond its end-of-life. We systematically evaluate the effects of anneal duration, anneal temperature, and state of the memory cells during annealing on the endurance enhancement. Interestingly, we find that endurance enhancement critically depends on the state of flash memory cells during annealing, with programmed cells showing significantly larger improvements than erased cells. Our experimental evaluation indicates that the post-cycle data retention property of an annealed chip significantly improves after thermal annealing, resulting in ~30% endurance recovery. Our results have significant implications for the future wear-leveling algorithms of SSD-based storage systems.

CCS Concepts: • **Hardware** → **External storage**.

Keywords: Flash memories, 3D NAND, annealing

ACM Reference Format:
Matchima Buddhanoy, Sudeep Pasricha, and Biswajit Ray. 2024. Life-after-Death: Exploring Thermal Annealing Conditions to Enhance 3D NAND SSD Endurance. In *16th ACM Workshop on Hot Topics in Storage and File Systems (HOTSTORAGE '24), July 8–9, 2024, Santa Clara, CA, USA.* ACM, Santa Clara, CA, USA, 7 pages.
https://doi.org/10.1145/3655038.3665949

1 Introduction

3D NAND flash memories are indispensable in the design of storage systems used in a wide range of computing domains, from low-end IoT platforms, smartphones, solid-state drives (SSDs) in workstations and servers, to storage systems in cloud computing. The bit density continues to increase with every new generation of 3D NAND flash chips, driven by an increase in the number of vertical layers (3D scaling) and logical scaling that allows for storing 2, 3, or 4 bits of data per memory cell. As a result of these trends, the endurance (or useful lifetime) of memory blocks is getting lower, which creates additional challenges for the efficient management of storage media and overall system design [8, 24].

It is well-known that thermal annealing can repair worn-out flash memory cells and thereby extend the endurance limit. For example, researchers at Macronix demonstrated flash endurance exceeding 100 million program and erase (PE) cycles by using a built-in thermal heater [12]. They utilized a novel flash cell architecture with word line heaters that create locally high temperatures to self-heal the stress-induced damage in the memory cell. Similarly, Wu et al. [21, 22] proposed an on-board heater die to boost the temperature of an SSD to achieve wear-out recovery.

While numerous studies have delved into the system-level implementation of self-healing SSDs [3–6], there is a notable gap in the literature regarding the evaluation of thermal annealing conditions for the effective recovery of endurance, especially in 3D NAND technology. For example, prior experimental characterizations of annealing effects on flash memory have predominantly focused on 2D NAND technology [12]. However, with the prevalent adoption of 3D NAND in modern SSDs, it becomes imperative to experimentally characterize the effects of thermal annealing on the endurance of 3D NAND. This exploration may unveil new optimization variables that can be leveraged in system-level algorithms.

In this paper, we study the thermal annealing effects on the endurance of COTS 3D NAND flash by exploring endurance recovery effects. To the best of our knowledge, this is the first work that explores annealing effects on the endurance of COTS 64-layer floating-gate (FG) 3D NAND flash memory. A key finding of our study highlights the profound impact of the programmed threshold voltage (V_{th}) of memory cells during thermal annealing on endurance recovery. Specifically, memory cells programmed at the highest V_{th} level exhibit a much steeper recovery rate (more than twice) compared to cells kept in the erased state. Furthermore, we identify a proper anneal duration for a given temperature, beyond which endurance recovery becomes minimal. To validate our findings, we conducted experimental evaluations on multiple COTS 3D FG NAND flash chips from different configurations including multiple-level-cell (MLC) and triple-level-cell (TLC).

This work is licensed under a Creative Commons Attribution International 4.0 License.
HOTSTORAGE '24, July 8–9, 2024, Santa Clara, CA, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0630-1/24/07
https://doi.org/10.1145/3655038.3665949

2 Background

Figure 1(a) illustrates the physical organization of a 3D NAND flash memory block. The green layers represent the word lines (WLs) of the memory array. The red bars represent the BLs, while the purple pillars are the poly-silicon channels. Figure 1(b) illustrates the circuit schematic of a physical sub-block of a NAND flash block. Each sub-block contains several flash pages, while each flash page consists of thousands of flash cells. All cells in the same row share the same WL while all cells in the same column share the same bitlines (BLs). Traditional flash memory cells store 1 bit of data and are known as SLC (Single-Level Cell). Advances in controlling the amount of charge on the floating gate and in sensing the charge during a read operation have allowed for logical scaling, further increasing bit density. Thus, modern flash memories store 2 bits (MLC), 3 bits (TLC), or 4 bits (QLC – Quad-Level Cell).

Flash-based storage systems consist of a flash controller and one or more flash chips. The flash controller interfaces and manages the flash media on one side and provides an interface to the host on the other side as shown in Figure 1(c). It executes an intermediate firmware layer called Flash Translation Layer (FTL). FTL encompasses multiple modules responsible for: (a) mapping of logical block addresses into physical addresses in the flash media; (b) wear leveling to ensure that all blocks wear out evenly; (c) garbage collection that tracks invalid pages within flash blocks and selects blocks to be erased when the number of free blocks falls below a certain threshold; and (d) error correction, allowing a certain number of errors in the flash media to be corrected.

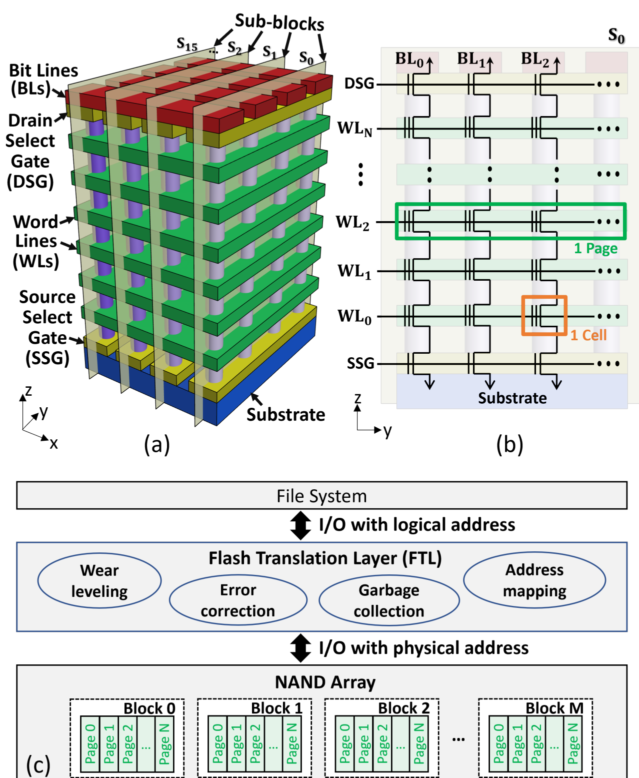

Figure 1. (a) Schematic of a 3D NAND block. (b) Circuit diagram of a sub-block. (c) System view of flash storage.

The endurance of NAND flash storage is determined by the total number of PE cycles [11, 13]. When a flash cell completes its designated number of PE cycles, it becomes worn-out, losing the ability to retain data effectively [14]. The degradation of the intrinsic oxide layer is identified as a primary cause of this limited endurance [11]. With an increasing number of PE cycles, a higher concentration of defects accumulates in the oxide layer [23]. This increased defect density allows stored electrons to leak out, compromising post-cycled data retention in memory cells. Consequently, over time, data corruption becomes a risk as bit errors may surpass the correction threshold of the error collection code (ECC) engine. Moreover, worn-out memory cells pose challenges during the erasure process, leading to prolonged erase times and increased rates of erase failures [10, 18].

3 Related Work and Motivation

It is well-known that thermal annealing can repair worn-out flash memory cells and thereby extend the endurance limit. For example, researchers at Macronix demonstrated flash endurance exceeding 100 million PE cycles by using a built-in thermal heater [12]. They utilized a novel flash cell architecture with word line heaters that create locally high temperature to self-heal the stress-induced damage in the memory cell. Similarly, Wu et al. [21, 22] proposed an on-board heater die to boost the temperature of the self-healing SSD to achieve wear-out recovery. Since the heating process requires a significant amount of time and consumes a substantial amount of power and energy, several innovative schemes have been proposed to reduce the heating overhead of self-healing SSDs. For example, Chang et al. [4] introduced a new wear-leveling design that evenly distributes heating cycles on flash memory blocks and minimizes the live-data copying overhead during heating. Chen et al. [5] proposed a novel wear-leveling technique called DHeating, which disperses the heating operation over different time periods, thereby avoiding the concentrated heating problem of self-healing SSDs. Similarly, Chang et al. [3] proposed a virtual wear-leveling technique to address the concentrated heating overhead or the heal-storm problem. They introduced the concept of virtual erase count, replacing the classical (or real) erase count of a block in the wear-leveling algorithm to balance the wear of a flash block and disperse heating operations. Cui et al. [6] proposed a SmartHeating technique that eliminates unnecessary heating operations by exploiting dwell time variation and write hotness variation.

4 Overview of annealing method

The goal of this paper is to explore the annealing effect by monitoring the raw bit error rate (RBER). The experimental evaluation is performed on multiple COTS NAND flash memory chips. Table 1 summarizes the sample details, which include 3D TLC and MLC FG memories from a major NAND vendor. Each memory chip contains thousands of flash blocks while each block consists of thousands of logical flash pages.

Table 1. Sample details.

Sample	3D FG TLC	3D FG MLC
Capacity	256GB	256GB
Number of layer	64	32
Block count	1,008	2,192
Total pages per block	2,304	1,024
Total byte per page	16,384	16,384
Endurance (PE cycles)	1,000	3,000

Figure 2(a) shows a custom-designed test board used in our experimental evaluation. The board consists of an FT2232H mini module from Future Technology Devices International (FTDI) and a socket to hold a raw flash chip. We connect the FT2232H mini module to a workstation through a Universal Serial Bus (USB). This hardware setup allows us to perform basic memory operations such as page read, page write, and block erase. Our software emulating FTL functions determines the RBER for page read operations. To evaluate the RBER, we first erase and write a full block with random data. Then, we read all pages in a block and calculate the RBER by comparing the data read from the chip with the original data written to the flash block.

The experimental flow is shown in Figure 2(b). Each flash chip first gets PE cycles repeatedly at room temperature

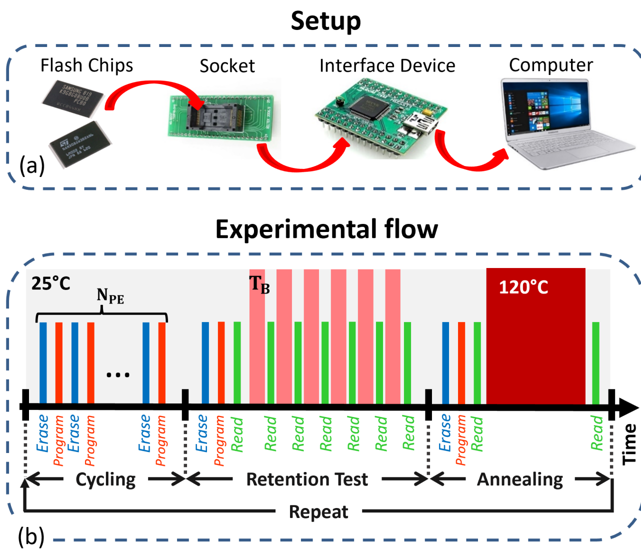

Figure 2. (a) Our setup. (b) Experimental flow.

(T_{ROOM}) of 25°C to reach its endurance. Note that due to the internal randomizer within the memory chip, 3D FG TLC chips are cycled with the highest state pattern while 3D FG MLC chip is cycled with random data pattern. After each chip reaches its endurance, we erase the full flash block and then program a new set of random data patterns into the chips at T_{ROOM}. Immediately after programming, we read the data and calculate RBER. Next, we perform an accelerated retention test by baking the chip at high temperature. We choose the baking temperature (T_B) to be 120°C while the baking time interval is 5 minutes, repeated over 6 episodes (for a total bake time of half an hour). After every 5 minutes of baking, we cool down all the chips to T_{ROOM} and then perform a read operation. We estimate the equivalent retention time at T_{ROOM} by using the Arrhenius model [1, 13]. Under the assumption that the activation energy (E_a) for 3D NAND is 1 eV [1, 13], we find that 30 minutes of data retention at 120°C is equivalent to ~8 months of data retention at T_{ROOM}. After the retention test, we perform thermal annealing of the entire memory chip by baking it at high temperature. Next, we repeat all steps again to quantify the endurance recovery.

5 Experimental results

For brevity, we present detailed experimental results for 3D TLC NAND in Sections 5.1-5.4, and summarize results for 3D MLC NAND in Section 5.5.

5.1 Annealing conditions evaluation

Efficient implementation of a self-healing SSD solution necessitates the optimization of thermal annealing conditions. The annealing process involves several variables, including anneal duration, anneal temperature, and the state of the memory cells during annealing. Experimental characterization of these variables specific to a given memory technology is crucial for ensuring the efficient implementation of the solution.

Figure 3 presents a summary of our characterization results regarding the effects of anneal duration on the post-cycled (N_{PE} = 1k) retention performance. Retention performance is quantified by the RBER value, which is measured through the retention test, as detailed in section 4. The data point corresponding to zero anneal duration represents the RBER obtained from an unannealed memory block with a 1k PE cycle count. The remaining data points correspond to annealed memory blocks, with the anneal duration indicated on the x-axis of the plot. The graph demonstrates that a longer anneal duration is more effective in enhancing the RBER of the block. However, it is noteworthy that there is a diminishing return observed after an anneal duration of 30 minutes. Note that the anneal temperature is kept fixed at 120°C during this analysis. If a higher anneal temperature is used, the anneal duration needed for the same retention

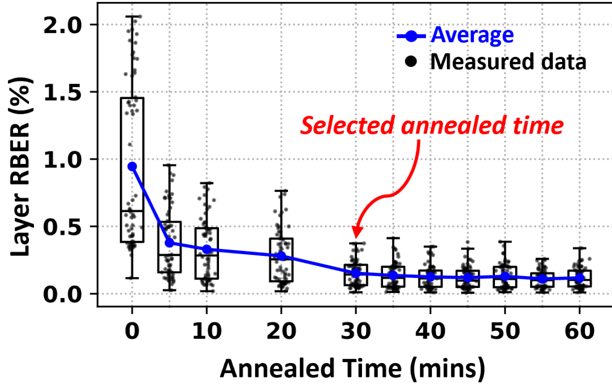

Figure 3. Layer RBER of worn-out (1k PE) cells after different annealed time.

improvement might get lowered. Therefore, in the subsequent analysis of the paper, we fix our annealing condition at 120°C for 30 minutes.

Next, we explore the effects of the programmed V_{th} level of memory cells during annealing on the subsequent retention characteristics, and the results are shown in Figure 4. It is important to note that TLC memory encompasses eight different V_{th} levels: L_0, L_1, ..., and L_7, where L_0 is the erased memory state and L_7 is the highest programmed V_{th} level. The x-axis of the figure illustrates the V_{th} levels, while the y-axis represents post-cycling retention RBER.

Data is collected from several worn-out memory blocks cycled up to 1k PE cycles. After PE cycling, we program all memory cells of a given block to one of the eight memory states and perform annealing. Following annealing, we erase and re-write memory blocks with random data and conduct the retention test, collecting RBER from each memory block. Our findings indicate that memory cells in the erased state during annealing exhibit the highest RBER during the retention test. Moreover, the higher the programmed V_{th} level of a memory block during annealing, the lower its retention RBER. However, we observe diminishing improvement in RBER beyond the L_2 state. Similar trend on program state dependent retention improvement was previously reported after irradiation effects on 3D NAND flash memory [1, 2].

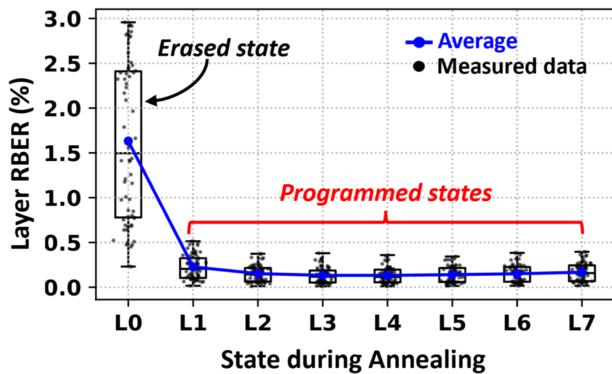

Figure 4. Layer RBER of worn-out (1k PE) cells after 30 minutes annealed time under different cell states during annealing.

5.2 Effects on 3D vertical layers

In Figure 5, a comparison is presented for the page RBER before and after annealing across all vertical layers of a 3D NAND chip. Since TLC memory has three different types of logical pages - LSB (Least Significant Bit), CSB (Central Significant Bit), and MSB (Most Significant Bit), we separately plot the RBER for each of these three different pages in Figure 5(a), (b), and (c), respectively. The green data points depict the page RBER from a fresh memory block following the retention test, while the blue data points represent the page RBER from a worn-out memory block subjected to 1k PE cycles. Lastly, the red points correspond to another worn-out (1k PE cycles) memory block of the same chip after annealing. Clearly, annealing has a substantial impact on improving the RBER of the worn-out memory block. Post-annealing, the RBER of the LSB pages becomes closely comparable to that of the fresh memory block (green), while for the CSB and MSB pages, it is slightly higher than that of the fresh block.

Figure 5 also reveals an interesting trend of page RBER with layer number [16, 17]. The RBER profile exhibits two distinct halves, with the first 32 layers (lower layers) displaying significantly lower RBER compared to the remaining 32 layers (upper layers). This pattern may be attributed to the two-tier process employed by 3D NAND vendors for fabricating larger numbers of vertical layers [7]. It is evident that the upper layers manifest more RBER compared to the lower layers. Lastly, we observe that MSB pages incur higher RBER among all the three logical page types. Note that the RBER of the MSB page is influenced by the transition of the highest V_{th} state of the TLC memory which typically experiences the highest V_{th} loss during the retention test. Hence, we observe higher RBER from the MSB pages in comparison to the corresponding LSB and CSB page types.

Since the worn-out memory block has been subjected to PE cycling until reaching the manufacturer-specified end-of-life limit, the RBER values for this block (depicted in blue) are considerably higher than what a typical ECC engine in SSDs can manage. Such elevated RBER levels may lead to uncorrectable read failure events in the SSD. It is essential to note that the RBER is measured using the default read settings of the chip. If a Read-Retry operation is employed [9], the RBER values might fall below the ECC threshold, which is ~0.1%, allowing correction of the page RBER. However, for the specific focus of this paper, which is to demonstrate the benefits of thermal annealing on endurance enhancement, we exclusively utilize the default read operation throughout the study.

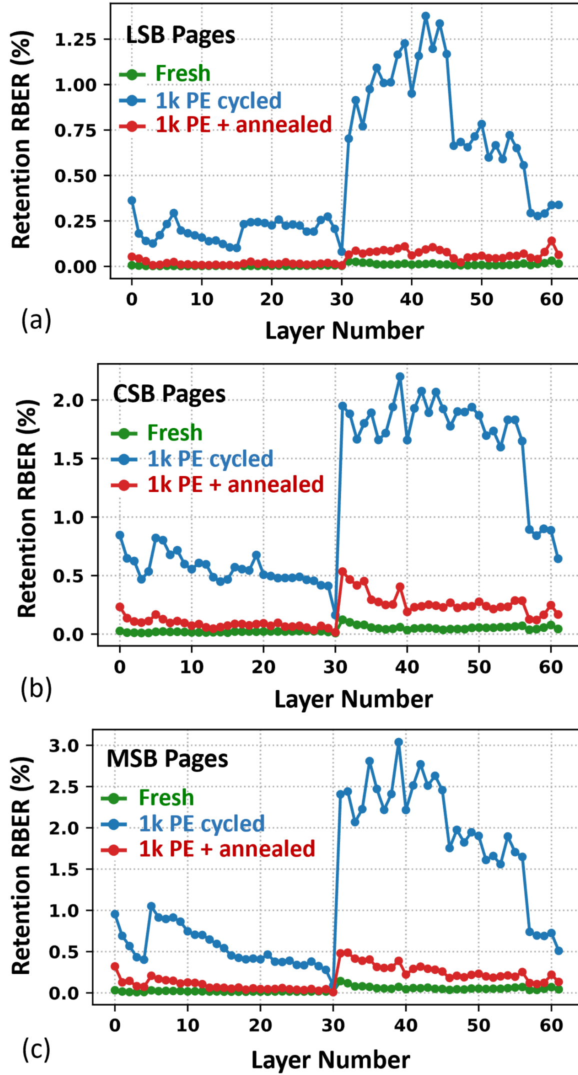

Figure 5. Comparison of RBER for logical memory pages from a fresh, worn-out (1k PE) and post-annealed (1k PE) memory block from the same chip. The plots show the average retention RBER values as a function of layer number of 3D stack for three different types of logical pages: (a) LSB pages, (b) CSB pages, and (c) MSB pages.

5.3 Effects on endurance improvement

In this section, we investigate the endurance improvement of flash memory chips after applying the thermal annealing method by keeping memory in a programmed state at 120°C for half an hour. The endurance improvement is quantified by comparing the retention RBER of the post-annealed memory chip with the baseline RBER vs. PE cycles curve obtained from unannealed memory.

The baseline retention RBER is determined through the following process. Before any retention tests, 10 consecutive fresh blocks in an unannealed chip are chosen which were prepared for different PE cycling conditions starting from 100 to 1,000 with 100 PE cycle increments. Note that we maximize the stress in the memory cells by programming all the memory cells to the highest memory state (L_7) during PE cycling. Additionally, we select a fresh block for 1 PE cycle condition. We perform erase, program, and read operations on all the selected memory blocks with random data pattern. Then, we perform the high-temperature data retention test and evaluate retention RBER. The baseline retention RBER is shown as the blue group in Figure 6.

Next, we evaluate the data retention property of the post-annealed worn-out memory blocks. We choose 3 post-annealed worn-out (N_{PE} = 1k) blocks from the annealed chip. The first block remains unchanged while we add extra 250 and 500 PE cycles to the remaining 2 blocks. Following the cycling process, we erase, program, and read all the selected memory blocks with random data pattern. Then, we conduct the same high-temperature data retention test procedures as performed on the unannealed chip to calculate the retention RBER. The retention RBER of the annealed chip after adding different PE cycles is shown as the red group in Figure 6.

The experimental results in Figure 6 reveal that the data retention property of the annealed chip significantly improves after thermal annealing. At 1,000 PE cycles, the average block retention RBER of the post-annealed worn-out block remarkably drops from ~1% to ~0.4%, which is equivalent to the block retention RBER of the unannealed chip after ~500 PE cycles. This behavior is illustrated with the solid vertical arrow in Figure 6. Even though the RBER value of the post-annealed worn-out block is comparable to the block RBER of the baseline at ~500 PE cycles, we find that the post-annealed worn-out memory block can store only ~300 PE cycles extra, resulting in ~1,300 PE cycles total before reaching the end-of-life again. We use the solid horizontal and the dashed vertical arrows to show the overall endurance improvement trend in Figure 6. Thus, we conclude that the post-annealed

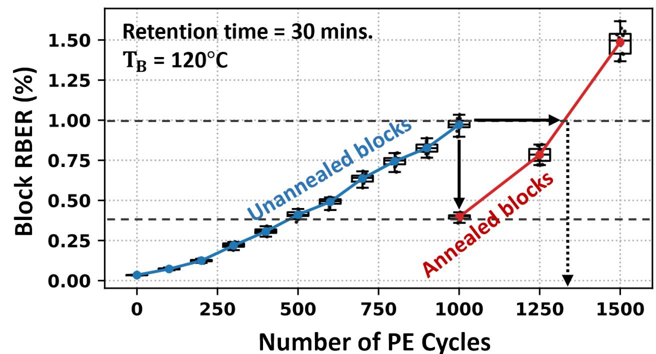

Figure 6. Block RBER vs. number of PE cycles from unannealed and annealed blocks.

worn-out memory block can experience ~30% endurance enhancement after applying the thermal annealing method.

5.4 Effects of annealing on erase time

Erase time of a memory block monotonically increases with PE cycling conditions and hence it is referred to as a good indicator of flash memory aging [10, 19]. In this section, we evaluate the improvement of erase time in flash memory blocks after applying thermal annealing, and the results are shown in Table 2. The detail of erase time measurement is discussed in our previous paper [19]. Our experimental results show that the erase time of the post-annealed worn-out 3D TLC memory block slightly drops after thermal annealing. Therefore, we conclude that the thermal annealing method helps improve the erase time of the flash block by reducing the trap density within the oxide layer.

5.5 Evaluation of MLC NAND chips

We further implemented the thermal annealing method in 3D MLC NAND chips. The experimental results are summarized in Table 2. Note that flash cells are worn-out when N_{PE} = 1k for 3D FG TLC chip and N_{PE} = 3k for 3D FG MLC chip. Based on the evaluation results of the MLC chip, we find that the average RBER value of the post-annealed worn-out cells condition significantly drops after the annealing method; however, it is still higher than the fresh condition. In addition, we find that the erase time of the post-annealed worn-out memory block slightly decreases after the high-temperature annealing, which indicates the improvement in erase time. Since the same behaviors can be observed in the 3D TLC type of memory, we conclude that the thermal annealing method is applicable to both 3D FG TLC and MLC NAND flash memories.

Table 2. Summary of erase time and average RBER.

	Cell conditions	3D FG TLC	3D FG MLC
Erase time	Fresh	7.94 ms	7.19 ms
	Worn-out	15.93 ms	14.09 ms
	Post-annealed	11.88 ms	11.76 ms
RBER	Fresh	0.0265%	0.0005%
	Worn-out	0.9722%	0.0131%
	Post-annealed	0.3600%	0.0038%

6 Plausible explanation for baking improvement

It is generally known that a high number of PE cycles creates trap states or defects in the tunnel oxide layer [11]. Several studies have investigated the physical origin of these trap states. According to the commonly accepted model, holes are initially injected into the tunnel oxide layer during erase operations. These injected holes are trapped at oxygen vacancy sites by breaking strained Si-Si bonds, which are predominantly found near the oxide-Si interface of the tunnel oxide. During subsequent program operations, electrons tunnel from the channel and neutralize the trapped positive charge. However, this process does not permanently or "truly" anneal the oxide defect through Si-Si bond reformation, as the electron can reversibly tunnel out of the trap site under reverse voltage. True annealing of the defects relies on the appropriate voltage across the oxide, temperature, and annealing duration. If the defect site is not truly annealed, it acts as a neutral electron trap site, which can subsequently capture another electron, leading to a negative trap charge in the tunnel oxide.

The above discussion provides a plausible explanation for the trend observed in Figure 4. When a flash memory cell is in the programmed state during thermal annealing, the electrons in the FG layer create a reverse voltage across the oxide layer, which prevents the formation of negative trap sites in the tunnel oxide layer. This condition is conducive to the reformation of broken Si-Si bonds, resulting in the repair of defects in the oxide layer [15, 20]. Consequently, this leads to an improvement in endurance after thermal annealing. In contrast, a flash memory cell in the erased state has an opposite voltage across the tunnel oxide layer, inducing a high density of electrons in the channel. During thermal annealing, the channel electrons tunnel into the oxide layer, creating negative trap charges [11]. The presence of negative trapped charges prevents bond reformation, and therefore, the oxide defects are not repaired, resulting in little or no improvement in endurance.

7 Conclusion

This paper explores the thermal annealing effects on the endurance of 3D NAND flash memories when reaching the end of their lifetime. The experimental evaluation is carried out on a COTS 3D FG TLC and MLC NAND flash memory chips. Our experimental results on 3D TLC chips reveal that thermal annealing of the chip in the programmed states offers the highest endurance recovery. Moreover, we find that retention RBER of the post-annealed worn-out (N_{PE} = 1k) block remarkably drops to a value equivalent to ~500 PE cycles. However, the post-annealed worn-out block can take only ~300 PE cycles extra, which gives ~1,300 PE cycles total before reaching its second lifetime. This leads to ~30% endurance improvement. In addition, we find that the erase time slightly drops after thermal annealing. Finally, we conclude that the annealing method is applicable to both TLC and MLC 3D NAND flash memory chips.

Acknowledgments

This work was supported in part by the National Science Foundation under Grant #2403540 and #2346853.

References

[1] Matchima Buddhanoy, Preeti Kumari, Umeshwarnath Surendranathan, Maryla Wasiolek, Khalid Hattar, and Biswajit Ray. 2022. Total Ionizing Dose Effects on Long-Term Data Retention Characteristics of Commercial 3-D NAND Memories. *IEEE Transactions on Nuclear Science* 69, 3 (2022), 390–396.

[2] Matchima Buddhanoy and Biswajit Ray. 2023. Electrostatic Shielding of NAND Flash Memory from Ionizing Radiation. In *2023 IEEE International Reliability Physics Symposium (IRPS)*. 1–5.

[3] Li-Pin Chang, Sheng-Min Huang, and Kun-Lin Chou. 2017. Relieving self-healing SSDs of heal storms. In *Proceedings of the 10th ACM International Systems and Storage Conference (SYSTOR)*. Association for Computing Machinery, New York, NY, USA.

[4] Yu-Ming Chang, Yuan-Hao Chang, Jian-Jia Chen, Tei-Wei Kuo, Hsiang-Pang Li, and Hang-Ting Lue. 2014. On Trading Wear-leveling with Heal-leveling. In *Proceedings of the 51st Design Automation Conference (DAC)*. Association for Computing Machinery, New York, NY, USA, 1–6.

[5] Renhai Chen, Yi Wang, Duo Liu, Zili Shao, and Song Jiang. 2017. Heating Dispersal for Self-Healing NAND Flash Memory. *IEEE Trans. Comput.* 66, 2 (2017), 361–367.

[6] Jinhua Cui, Junwei Liu, Jianhang Huang, and Laurence T. Yang. 2021. SmartHeating: On the Performance and Lifetime Improvement of Self-Healing SSDs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 40, 1 (2021), 52–65.

[7] Akira Goda and Krishna Parat. 2012. Scaling directions for 2D and 3D NAND cells. In *2012 International Electron Devices Meeting (IEDM)*.

[8] Tzu-Hsuan Hsu, Hang-Ting Lue, Pei-Ying Du, Wei-Chen Chen, Teng-Hao Yeh, Roger Lo, Hung-Sheng Chang, Keh-Chung Wang, and Chih-Yuan Lu. 2019. Study of Self-Healing 3D NAND Flash with Micro Heater to Improve the Performances and Lifetime for Fast NAND in NVDIMM Applications. In *2019 IEEE 11th International Memory Workshop (IMW)*. 1–4.

[9] Preeti Kumari, Umeshwarnath Surendranathan, Maryla Wasiolek, Khalid Hattar, Narayana P. Bhat, and Biswajit Ray. 2021. Radiation-Induced Error Mitigation by Read-Retry Technique for MLC 3-D NAND Flash Memory. *IEEE Transactions on Nuclear Science* 68, 5 (2021), 1032–1039.

[10] Preeti Kumari, B. M. S. Bahar Talukder, Sadman Sakib, Biswajit Ray, and Md Tauhidur Rahman. 2018. Independent detection of recycled flash memory: Challenges and solutions. In *2018 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*. 89–95.

[11] Jae-Duk Lee, Jeong-Hyuk Choi, Donggun Park, and Kinam Kim. 2003. Degradation of tunnel oxide by FN current stress and its effects on data retention characteristics of 90 nm NAND flash memory cells. In *2003 IEEE International Reliability Physics Symposium Proceedings, 2003. 41st Annual.* 497–501.

[12] Hang-Ting Lue, Pei-Ying Du, Chih-Ping Chen, Wei-Chen Chen, Chih-Chang Hsieh, Yi-Hsuan Hsiao, Yen-Hao Shih, and Chih-Yuan Lu. 2012. Radically extending the cycling endurance of Flash memory (to > 100M Cycles) by using built-in thermal annealing to self-heal the stress-induced damage. In *2012 International Electron Devices Meeting (IEDM)*.

[13] Yixin Luo, Saugata Ghose, Yu Cai, Erich F. Haratsch, and Onur Mutlu. 2018. HeatWatch: Improving 3D NAND Flash Memory Device Reliability by Exploiting Self-Recovery and Temperature Awareness. In *2018 IEEE International Symposium on High Performance Computer Architecture (HPCA)*. 504–517.

[14] M. Ceylan Morgul, Xinfei Guo, and Mircea Stan. 2022. Towards Everlasting Flash: Preventing Permanent Flash Cell Damage using Circadian Rhythms. In *2022 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. 376–379.

[15] Timothy R Oldham. 2004. Switching oxide traps. *International journal of high speed electronics and systems* 14, 02 (2004), 581–603.

[16] Md Raquibuzzaman, Md Mehedi Hasan, Aleksandar Milenkovic, and Biswajit Ray. 2022. Layer-to-Layer Endurance Variation of 3D NAND Flash Memory. In *2022 IEEE International Reliability Physics Symposium (IRPS)*. 1–5.

[17] Md Raquibuzzaman, Aleksandar Milenkovic, and Biswajit Ray. 2023. Intrablock Wear Leveling to Counter Layer-to-Layer Endurance Variation of 3-D NAND Flash Memory. *IEEE Transactions on Electron Devices* 70, 1 (2023), 70–75.

[18] Sadman Sakib, Preeti Kumari, BMS Bahar Talukder, Md Tauhidur Rahman, and Biswajit Ray. 2018. Non-invasive detection method for recycled flash memory using timing characteristics. *Cryptography* 2, 3 (2018), 17.

[19] S M Nazmuz Sakib, Preeti Kumari, Bashir M. Sabquat Bahar Talukder, Md. Tauhidur Rahman, and Biswajit Ray. 2018. Non-Invasive Detection Method for Recycled Flash Memory Using Timing Characteristics †. *Cryptogr.* 2 (2018), 17.

[20] M. Walters and A. Reisman. 1991. Radiation-Induced Neutral Electron Trap Generation in Electrically Biased Insulated Gate Field Effect Transistor Gate Insulators. *Journal of the Electrochemical Society* 138, 9 (1991), 2756.

[21] Qi Wu, Guiqiang Dong, and Tong Zhang. 2011. Exploiting Heat-Accelerated Flash Memory Wear-Out Recovery to Enable Self-Healing SSDs. In *3rd Hot Topics in Storage and File Systems (HotStorage'11)*.

[22] Qi Wu, Guiqiang Dong, and Tong Zhang. 2011. A First Study on Self-Healing Solid-State Drives. In *2011 IEEE 3rd International Memory Workshop (IMW)*. 1–4.

[23] Zhiliang Xia, Dae Sin Kim, Narae Jeong, Young-Gu Kim, Jae-Ho Kim, Keun-Ho Lee, Young-Kwan Park, Chilhee Chung, Hwan Lee, and Jungin Han. 2012. Comprehensive modeling of NAND flash memory reliability: Endurance and data retention. In *2012 IEEE International Reliability Physics Symposium (IRPS)*.

[24] Yang Xiaoshan, Zhu Ligu, and Zhang Qicong. 2017. Research on endurance evaluation for NAND flash-based solid state drive. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. 523–526.