How the Common Retention Acceleration Method of 3D NAND Flash Memory Goes Wrong?

Qiao Li
City University of Hong Kong

Min Ye
City University of Hong Kong

Tei-Wei Kuo
City University of Hong Kong
National Taiwan University

Chun Jason Xue
City University of Hong Kong

ABSTRACT

The reliability of solid-state drives (SSDs) has become increasingly important as SSDs are now widely applied in data centers. Retention error is a major source of impact on the reliability of SSDs. Even though the common practice in understanding the retention errors of an SSD is done by high-temperature baking processes, their characterization accuracy is not yet rigidly reviewed. In this paper, we first present how the common retention acceleration method goes wrong. Through a one-year study of 3D flash error behaviors, we found that the retention errors through baking with high temperatures have very different characteristics from the real long-retention errors. These differences come from the inherent structure and the materials of 3D NAND flash. Several findings regarding the retention errors characterized through baking are presented, followed by the analysis of the error behaviors. Finally, the retention errors of one year on 3D flash memory are provided with real data.

ACM Reference Format:
Qiao Li, Min Ye, Tei-Wei Kuo, and Chun Jason Xue. 2021. How the Common Retention Acceleration Method of 3D NAND Flash Memory Goes Wrong?. In *13th ACM Workshop on Hot Topics in Storage and File Systems (HotStorage '21).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3465332.3470877

1 INTRODUCTION

High-density solid-state drives (SSDs) are now widely used in data centers because of their performance and cost advantages. At the same time, the reliability of SSDs and their data endurance over the years are always of great concern [6, 21, 23, 24]. The raw bit error rate (RBER) of SSDs' 3D NAND flash-memory chips is dominated by retention errors [4, 5, 32], which are caused by the charge leakage of flash cells over retention time. The understanding of the RBER and retention loss has huge impacts on not only the lifetime, but also on the access performance of a flash-memory chip.

Many previous studies have conducted experiments to characterize the performance and reliability features of planar and 3D NAND flash and make meticulous analyses [23, 26, 32, 33]. It is a common practice to understand the flash retention errors through high-temperature baking processes [2, 3, 7, 32, 33]. According to Arrhenius' model [2, 3, 19, 28], the retention loss process can be accelerated through baking since high temperatures increase the mobility of electrons in flash cells, which accelerates the charge leakage speed. Baking flash chips with a high temperature shortens the experimental time of retention error characterizations and is widely accepted and applied in both academia and industry. Meanwhile, previous studies present that there might exist differences between the real retention errors and those obtained through baking [5, 28]. To obtain the accurate estimations, efforts have been made to calibrate the value of activation energy E_a, which is a key parameter in Arrhenius' model [13, 22]. Through a one-year study, this work shows that this retention characterization method is not suitable for 3D NAND flash memories.

In this work, we show that the retention errors based on baking processes have different characteristics from the real retention errors of 3D NAND flash-memory chips over a long retention period. The main findings are as follows:

- The real long-retention RBER at a low temperature is underestimated through baking with high temperatures (Finding 1);
- The difference between the real and estimated (through baking) retention errors on floating-gate flash is smaller than that of charge-trap flash, and the difference changes with the number of P/E cycles (Finding 2 and 3);

HotStorage '21, July 27–28, 2021, Virtual, USA

ACM ISBN 978-1-4503-8550-3/21/07…$15.00
https://doi.org/10.1145/3465332.3470877

- Baking flash chips affects the RBER variations that exist on 3D NAND flash memory, including the variations among pages and among wordlines of a flash block (Finding 4 and 5).

These new characteristics come from the inherent structure design and materials of 3D NAND flash. First, most planar flash chips are based on floating gate (FG) cells. While today's 3D NAND flash memories mainly adopt charge trap (CT) technology for flash cells [11, 12, 14, 17, 29, 30]. Secondly, the new block structure design in connecting flash cells along a channel on 3D flash also provides new charge loss paths, which amplifies retention errors. Thirdly, there are variations of RBER among layers and among pages on 3D NAND flash memories, and baking with high temperatures could change such variations. As a result, current retention error characterization based on baking processes fails to present the real retention errors over a long retention time on 3D flash. To address the issue, we provide several insights to explore the long retention errors, along with the real retention error data of one year on 3D NAND flash memory.

2 FLASH RETENTION ACCELERATION

NAND flash memory stores charge in flash cells and represents data based on the threshold voltage (V_{th}) of cells. Figure 1 shows an example of the V_{th} distribution of triple-level cell (TLC) flash memory. The V_{th}s of flash cells are separated into eight states by seven read voltages (V_1 to V_7). Each TLC cell stores three bits, most significant bit (MSB), center significant bit (CSB), and least significant bit (LSB). The LSB (CSB/MSB) of all cells in one wordline form the LSB (CSB/MSB) page of that wordline. Several types of error sources will impact

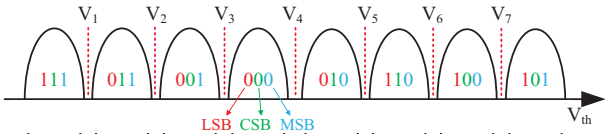

Figure 1: V_{th} distributions of TLC flash cells.

the reliability of flash memory, such as program/erase (P/E) cycling, retention time, program interference, and so on. Retention errors are one of the most important flash errors, which are caused by charge leakage of flash cells over retention time. To guarantee the reliability of stored data, the flash controller provides error correction codes (ECC) that are strong enough for the raw bit error rate (RBER) after a long retention age, for example, one year. However, it is time-consuming to observe the retention errors over one year, or even longer. Therefore, it is essential to model and study the long-retention-age errors of NAND flash memories.

To study the retention errors on flash memory, it is a common practice to bake flash-memory chips where Arrhenius' Law is widely adopted [1, 5, 13, 22, 22]. Electrons become more active at high temperatures, and the charge leakage over retention time will be accelerated by high temperatures. Based on the model, the equivalence between the baking time at a high temperature and the corresponding retention time at room temperature can be modeled:

$$AF = \frac{t_1}{t_2} = exp(\frac{E_a}{k} * (\frac{1}{T_1} - \frac{1}{T_2})) \quad (1)$$

AF is the acceleration factor between t_1 and t_2, where t_1 is the retention time at room temperature T_1, and t_2 is the retention time at high temperature T_2. k is the Boltzmann constant, which is 8.62×10^{-5}eV/K. E_a is the activation energy, which is a manufacturing-process dependent parameter. Currently, E_a is set as 1.1 eV for planar flash memory obtained through estimation [31]. Although the acceleration method of retention errors based on this theory is widely applied in both academia and industry, it is generally accepted that this method cannot accurately represent the retention characteristics of planar flash memory at room temperature [10, 15, 20]. Currently, retention error characterization methods still apply the same model and parameters that are obtained on planar flash to high-density 3D NAND flash memories [33–35]. In this work, we will show that this model works even worse on 3D NAND flash memory, via studying long-retention errors of 3D flash chips with temperature considerations based on real data.

3 HOW 3D FLASH RETENTION ACCELERATION GOES WRONG?

In this section, we conduct evaluations on real 3D NAND flash memory chips to study the retention characteristics.

3.1 Evaluation Setup

Table 1: The baking time by setting $E_a = 1.1eV$.

Temperature (°C)	60	80	100	120
Time (hour)	97.65	11.16	1.61	0.28

The experiments are conducted on YEESTOR 9083 flash memory testing platform [36]. It is a flash controller that enables users to implement specified flash memory solutions. Several 64-layer and 96-layer 3D flash-memory chips from different vendors are evaluated, including two CT TLC flash, one FG TLC flash and one FG QLC (Quad-level cell) flash. The one-year retention experiments over 3D NAND flash-memory chips here were conducted over multiple temperatures. For one-year retention at room temperature, flash chips are placed in the office with air condition, and the temperature is set at 25°C. For high-temperature experiments, we calculated the baking time according to Arrhenius' Law, where E_a was set as 1.1eV. Table 1 shows the baking hours of different high temperatures so as to achieve the expected equivalent one-year retention time at 25°C.

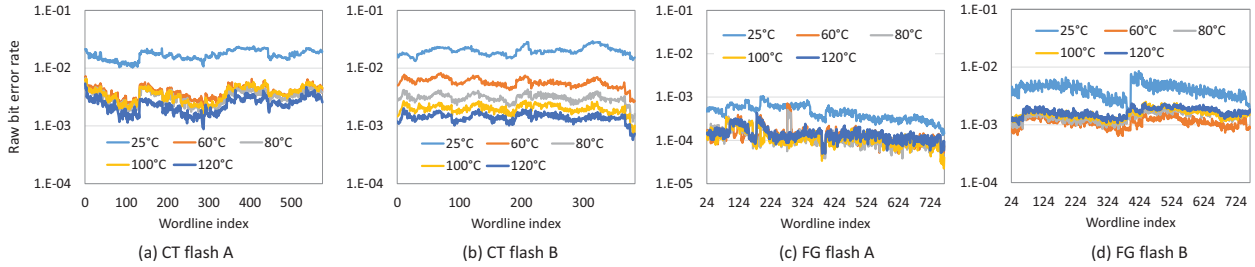

Figure 2: The default RBER of one-year retention time at 25°C and that through baking with high temperatures.

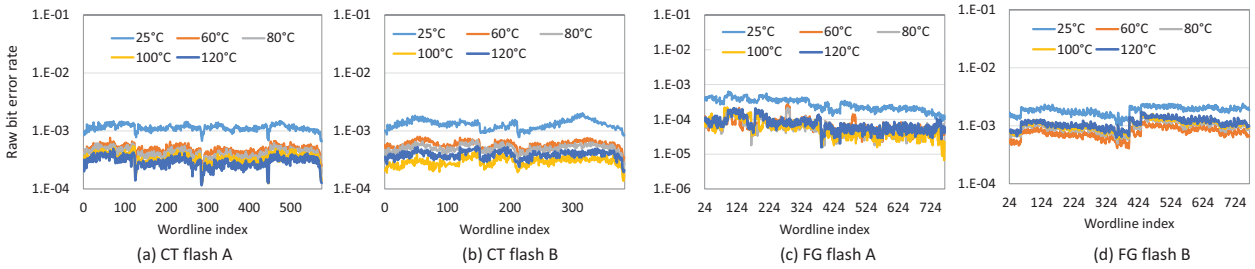

Figure 3: The optimal RBER of one-year retention time at 25°C and that through baking with high temperatures.

3.2 New Findings

Two types of RBER before ECC decoding are collected: The RBER introduced when data are read with default read voltages (default RBER), and the lowest RBER when data are read through retrying with calibrated read reference voltages (optimal RBER). We present five new findings that are contrary to the common beliefs.

Finding 1: The errors through baking with high temperatures significantly underestimate the real long-retention errors. Figures 2 and 3 show the default and optimal RBER of one flash block on different flash chips. The results are the average of multiple evaluations. On all the evaluated flash chips, the real RBERs after one-year retention time are much higher than the RBER through baking, more than 20 times in some cases. Existing researches have reported that the retention loss on 3D NAND flash could be more severe than the planar flash. On 3D NAND flash, cells are stacked in vertical layers to increase the storage density, which creates additional leakage paths for the charge, from each cell's active area towards other cells on the same string. As a result, the retention loss on 3D NAND flash becomes more significant. However, current retention characterizations still adopt the same model and parameters of planar flash [33, 35]. This leads to huge deviations in the characterized retention errors through baking with high temperatures, as shown in Figures 2 and 3. We can note that the estimation deviation of higher temperatures tends to be larger. One exception is 120°C, which could have higher RBERs than 100°C in some cases. This is because the baking time at 120°C is very short, less than 17 minutes, which could easily introduce significant estimation deviations.

On the other hand, the optimal RBERs are much lower than the default RBERs, and thus the difference between the real and estimated (through baking) optimal RBERs is also smaller. This is because the calibrated read reference voltages could greatly counteract the threshold voltage shifting due to charge loss over retention time. This observation confirms that the charge leakage on 3D NAND flash is significant, especially for CT NAND flash chips. As a result, we mainly present the optimal RBER in the following discussions.

Finding 2: The difference between the real and estimated retention errors on 3D FG flash is smaller than that of 3D CT flash. As shown in Figures 2 and 3, the retention errors of CT flash chips and FG flash chips change in different ranges. The real retention errors on 3D CT flash chips could be more than 20 times the estimated ones through baking. While the difference between the real and estimated on FG flash is much smaller. The reasons are two-fold. First, in the planar flash era, most flash chips are based on floating gate cells. While today's 3D NAND flash memories mainly adopt CT technology as the fundamental flash cells. The acceleration parameters that used to be adopted on planar FG flash chips introduce less deviations on 3D FG flash chips than 3D CT flash chips. Second, 3D CT structure is more prone to retention errors. The CT nitride layer in a string of the 3D CT flash is continuously connected from top to bottom control gates along the channel side, and it acts as a charge spreading path [16, 18]. As a result, 3D CT flash has a fast charge leakage speed, and the leakage accumulates to a significant amount over a long retention time, for example, one year in this work. Though high temperatures accelerate the leakage speed, the leaked charge is not comparable

because of the short baking time. In 3D FG cells, on the contrary, the floating gate is completely isolated by the tunnel oxide and the Inter Poly Dielectric. Such block structure is reliable to contain charges without any problem related to leakage paths. This is also why 3D FG flash, no matter TLC or QLC, has a lower RBER than 3D CT flash, as shown in Figures 2 and 3.

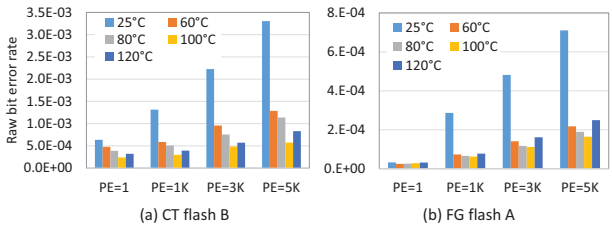

Figure 4: The average RBER of a flash block at different P/E cycles.

Finding 3: The difference between the real and estimated retention errors changes with the number of P/E cycles. Figure 4 shows the average RBER at different numbers of P/E cycles on CT flash B and FG flash A, and the data on other flash chips have a similar trend. As shown in the figure, with the increase of P/E cycles, the difference between the real and estimated retention errors becomes larger. During its lifetime, a flash memory chip endures a large number of program/erase cycles. Every P/E cycle involves high electrical fields applied to the tunnel oxide of 3D NAND flash cells. The program and erase pulses have constant amplitude and duration, which leads to oxide degradation and flash cell damage [27]. As a result of consecutive electron tunneling, traps are generated into the oxide [25]. When filled by electrons, charged traps can increase the potential barrier, thus changing the physical characteristics of flash cells. After enduring different numbers of P/E cycles, flash chips could present different retention characteristics. Therefore, the retention-acceleration impact from high temperatures changes with the increase of P/E cycles.

Finding 4: The temperature affects the RBER variations among read voltages, as well as pages. According to existing work, we know that higher voltages suffer more from retention errors [4, 21]. This is because higher voltage states have more charge, and thus, the leakage speed is higher, which leads to more charge leakage. However, as shown in Figure 5, our evaluation shows that the real difference between low voltage states and high voltage states could be larger than the estimated ones through baking. The difference between V_1 and V_7 at 25°C after a one-year retention time can exceed one order of magnitude. While through baking, the RBER differences among seven read voltages are reduced. The reason might be that high temperatures increase the mobility of electrons in flash cells, and all the voltage states have an extremely high charge leakage speed at high temperatures. After a baking period, which is very short, the charge leakages of different states present smaller variations than that of long retention time and low temperature. Therefore, the RBERs introduced by seven read voltages become more uniform if baking flash chips with high temperatures.

The variations among seven read voltages also affect the variations among pages, as reading each page requires one or several specific read voltages. Figures 5 and 6 show that at 25°C, the MSB page of TLC flash chips always has the highest RBER among three pages, because V_7 is required to read an MSB page and V_7 introduces a much higher RBER than other read voltages after one-year retention time. However, after baking flash chips with high temperatures, the CSB page has the highest RBER because the variations among read voltages are changed due to high temperatures. For the QLC flash chip, there are 15 read voltages and 4 different pages, the variations among them are also changed. Similarly, at 25°C, the read voltages to separate higher voltage states introduce higher RBER and the differences among the four pages are very large. While at 80°C, the differences become much smaller because of the same reason discussed above.

Finding 5: The temperature affects the RBER variations among wordlines. Figure 6 shows the optimal RBER of three pages on CT flash B. The RBER of different wordlines inside the same flash block presents significant variations. The variation comes from the physical structure of 3D flash blocks, as current circuit etching technologies are unable to produce identical 3D NAND cells when punching through multiple stacked layers. Such variation has been reported by previous studies and considered as constant. The variation among wordlines is consistent across different P/E cycles and different retention time [23]. However, in this work, we found that such variation is not consistent, especially CSB and MSB pages, when the temperature is changed. For example, at 25°C, the MSB page of WL 313 has the highest RBER, which is 986 times larger than the LSB page RBER of WL 383, the lowest RBER. While at high temperature, the difference between the highest and the lowest is 12×, and the WL that has the largest RBER is the CSB page of WL 140. Therefore, the worst wordlines or worst pages that have the highest RBER at room temperature may not be found through baking with high temperatures.

Based on the above findings, we can conclude that the error behaviors over a long retention time at low temperatures are very different from those through baking with high temperatures. As a result, we cannot achieve the equivalent error behaviors even if we calibrate the model parameter E_a. Current retention acceleration methods should be revised for 3D NAND flash memories.

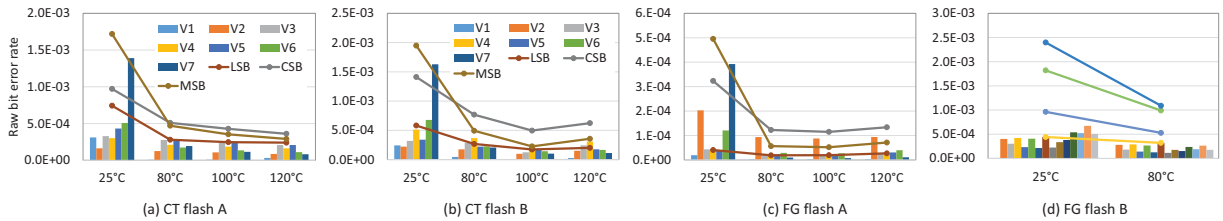

Figure 5: The RBER of read voltages and of pages. The left three figures are TLC flash, and the right figure is QLC flash with 15 read voltages (bars in the figure) and 4 pages (lines in the figure).

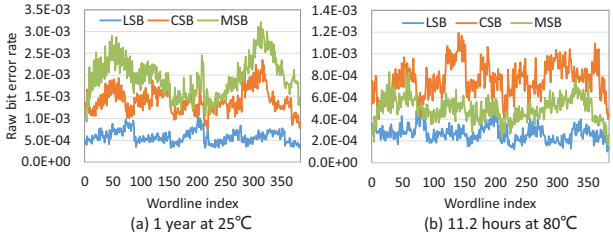

Figure 6: The page RBER of 1-year retention time at 25°C and through baking at 80°C.

4 DISCUSSIONS

Long-Retention Errors: We first present the real retention errors of 3D flash-memory chips over one year at room temperature (25°C in our case), where current long-retention errors are all estimated through baking with high temperatures. Figure 7 shows how the RBER introduced by different

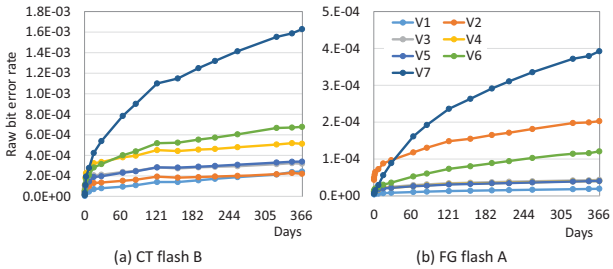

Figure 7: The increasing of retention errors over retention time.

read voltages changes with the retention time. The increasing rate of RBER decreases with the retention time. The electric charge near the flash cell's edge, after programming, is easy to leak, and it results in a fast leakage speed in the beginning. This is called the early retention loss of 3D NAND flash memory [8, 9, 23]. After the leakage of the charges near the cell's edge, the leakage speed becomes slow because the charges near the flash cells' centers are hard to leak out.

Characterization of retention errors: The above findings reveal that the retention errors of 3D NAND flash are much more severe than our expectation, especially on CT flash. The behaviors of accelerated retention loss are very different from that over long retention time, regarding read voltages, pages, wordlines, P/E cycles, and so on. Because of that, we cannot achieve the equivalent error behaviors through calibrating the parameter E_a in Arrhenius' model. In the coming work, we shall propose new methods to accurately characterize retention errors with considerations of these characteristics.

Influence on the reliability: The mechanisms designed to guarantee data correctness heavily depend on the estimation of flash errors. For example, the frequencies of data refresh are determined when the RBER exceeds the error correction capability of ECC. If the estimation is not correct, the data stored in NAND flash might suffer from the risk of data corruption. Instead of merely relying on the estimation from baking through high temperature, which could be incorrect, it is suggested to update the estimation online during the usage. This can be achieved through the triggering of data checking or scrubbing processes to guarantee the reliability.

Optimizations of the reliability: Currently, there are many approaches proposed to optimize the reliability of flash memory, based on the error behaviors observed through baking. These methods could have sub-optimal performance. For example, the lifetime is determined by the worst-case RBER. Among the three pages of 3D TLC NAND, CSB page is treated as the page that has the highest RBER [32, 33]. However, based on our study, the MSB page has the highest RBER in some normal usage at room temperature. Therefore, reliability optimization approaches to optimize the reliability of 3D NAND should be carefully redesigned.

5 CONCLUSION

Retention errors, caused by charge leakage over time, are the dominant source of flash errors on 3D NAND flash memory. Understanding and characterizing retention errors, especially long retention, are critical for the reliability and endurance enhancement of flash memory. This paper conducted a one-year experiment over the latest 3D NAND flash-memory chips to study the long-retention errors. Our new findings show that the common retention characterization method through baking flash chips with high temperatures has severe problems. The real error behaviors over a long retention time are very different from the accelerated retention errors with high-temperature baking. Finally, we provide several insights to explore the long retention errors, along with real long-retention errors on 3D flash memory.

ACKNOWLEDGMENTS

The work was partial supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (CityU 11219319).

REFERENCES

[1] Svante Arrhenius. 1889. Über die Dissociationswärme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte. *Zeitschrift für physikalische Chemie* 4, 1 (1889), 96–116.

[2] David A Baglee. 1984. Characteristics & Reliability of 100Å Oxides. In *22nd International Reliability Physics Symposium*. IEEE, 152–155.

[3] Franck Bayle and Adamantios Mettas. 2010. Temperature acceleration models in reliability predictions: Justification & improvements. In *2010 Proceedings-Annual Reliability and Maintainability Symposium (RAMS)*. IEEE, 1–6.

[4] Yu Cai, Saugata Ghose, Erich F Haratsch, Yixin Luo, and Onur Mutlu. 2017. Errors in flash-memory-based solid-state drives: Analysis, mitigation, and recovery. *arXiv preprint arXiv:1711.11427* (2017).

[5] Yu Cai, Yixin Luo, Erich F Haratsch, Ken Mai, Saugata Ghose, and Onur Mutlu. 2018. Experimental Characterization, Optimization, and Recovery of Data Retention Errors in MLC NAND Flash Memory. *arXiv preprint arXiv:1805.02819* (2018).

[6] Yu Cai, Yixin Luo, Erich F Haratsch, Ken Mai, and Onur Mutlu. 2015. Data retention in MLC NAND flash memory: Characterization, optimization, and recovery. In *2015 IEEE 21st International Symposium on High Performance Computer Architecture (HPCA)*. IEEE, 551–563.

[7] Yu Cai, Gulay Yalcin, Onur Mutlu, Erich F Haratsch, Adrian Cristal, Osman S Unsal, and Ken Mai. 2012. Flash correct-and-refresh: Retention-aware error management for increased flash memory lifetime. In *2012 IEEE 30th International Conference on Computer Design (ICCD)*. IEEE, 94–101.

[8] Chih-Ping Chen, Hang-Ting Lue, Chih-Chang Hsieh, Kuo-Pin Chang, Kuang-Yeu Hsieh, and Chih-Yuan Lu. 2010. Study of fast initial charge loss and it's impact on the programmed states Vt distribution of charge-trapping NAND flash. In *2010 International Electron Devices Meeting*. IEEE, 5–6.

[9] Bongsik Choi, Sang Hyun Jang, Jinsu Yoon, Juhee Lee, Minsu Jeon, Yongwoo Lee, Jungmin Han, Jieun Lee, Dong Myong Kim, Dae Hwan Kim, et al. 2016. Comprehensive evaluation of early retention (fast charge loss within a few seconds) characteristics in tube-type 3-D NAND flash memory. In *2016 IEEE Symposium on VLSI Technology*. IEEE, 1–2.

[10] Chung-Hao Fu, Hang-Ting Lue, Tzu-Hsuan Hsu, Wei-Chen Chen, Guan-Ru Lee, Chia-Jung Chiu, Keh-Chung Wang, and Chih-Yuan Lu. 2019. A Novel Confined Nitride-Trapping Layer Device for 3D NAND Flash with Robust Retention Performances. In *2019 Symposium on VLSI Technology*. IEEE, T212–T213.

[11] Akira Goda. 2020. 3-D NAND Technology Achievements and Future Scaling Perspectives. *IEEE Transactions on Electron Devices* 67, 4 (2020), 1373–1381.

[12] Hwang Huh, Wanik Cho, Jinhaeng Lee, Yujong Noh, Yongsoon Park, Sunghwa Ok, Jongwoo Kim, Kayoung Cho, Hyunchul Lee, Geonu Kim, et al. 2020. 13.2 A 1Tb 4b/Cell 96-Stacked-WL 3D NAND Flash Memory with 30MB/s Program Throughput Using Peripheral Circuit Under Memory Cell Array Technique. In *2020 IEEE International Solid-State Circuits Conference-(ISSCC)*. IEEE, 220–221.

[13] JEDEC Standard JESD91A. 2003. Method for developing acceleration models for electronic component failure mechanisms. *Arlington, VA, JEDEC Solid State Technology Association* 3, 6 (2003), 7–2.

[14] Dongku Kang, Woopyo Jeong, Chulbum Kim, Doo-Hyun Kim, Yong Sung Cho, Kyung-Tae Kang, Jinho Ryu, Kyung-Min Kang, Sungyeon Lee, Wandong Kim, et al. 2016. 256 Gb 3 b/cell V-NAND flash memory with 48 stacked WL layers. *IEEE Journal of Solid-State Circuits* 52, 1 (2016), 210–217.

[15] Duckseoung Kang, Kyunghwan Lee, Seongjun Seo, Shinhyung Kim, Ji-Seok Lee, Dong-Seok Bae, Dong Hua Li, Yuchul Hwang, and Hyungcheol Shin. 2013. Generation dependence of retention characteristics in extremely scaled NAND flash memory. *IEEE Electron Device Letters* 34, 9 (2013), 1139–1141.

[16] Bryan S. Kim, Jongmoo Choi, and Sang Lyul Min. 2019. Design Tradeoffs for SSD Reliability. In *17th USENIX Conference on File and Storage Technologies (FAST)*. USENIX Association, Boston, MA, 281–294. https://www.usenix.org/conference/fast19/presentation/kim-bryan

[17] Chulbum Kim, Doo-Hyun Kim, Woopyo Jeong, Hyun-Jin Kim, Il Han Park, Hyun-Wook Park, JongHoon Lee, JiYoon Park, Yang-Lo Ahn, Ji Young Lee, et al. 2017. A 512-Gb 3-b/cell 64-stacked WL 3-D-NAND flash memory. *IEEE Journal of Solid-State Circuits* 53, 1 (2017), 124–133.

[18] Hyunsuk Kim, Su-Jin Ahn, Yu Gyun Shin, Kyupil Lee, and Eunseung Jung. 2017. Evolution of NAND flash memory: From 2D to 3D as a storage market leader. In *2017 IEEE International Memory Workshop (IMW)*. IEEE, 1–4.

[19] Kyunghwan Lee, Myounggon Kang, Seongjun Seo, Duckseoung Kang, Shinhyung Kim, Dong Hua Li, and Hyungcheol Shin. 2013. Activation energies (E_{a}) of failure mechanisms in advanced NAND Flash cells for different generations and cycling. *IEEE Transactions on Electron Devices* 60, 3 (2013), 1099–1107.

[20] Kyunghwan Lee, Myounggon Kang, Seongjun Seo, Dong Hua Li, Jungki Kim, and Hyungcheol Shin. 2012. Analysis of failure mechanisms and extraction of activation energies (E_{a}) in 21-nm nand flash cells. *IEEE Electron Device Letters* 34, 1 (2012), 48–50.

[21] Qiao Li, Min Ye, Yufei Cui, Liang Shi, Xiaoqiang Li, Tei-Wei Kuo, and Chun Jason Xue. 2020. Shaving Retries with Sentinels for Fast Read over High-Density 3D Flash. In *2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*. IEEE, 483–495.

[22] 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)*. IEEE, 504–517.

[23] Yixin Luo, Saugata Ghose, Yu Cai, Erich F Haratsch, and Onur Mutlu. 2018. Improving 3D NAND flash memory lifetime by tolerating early retention loss and process variation. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 2, 3 (2018), 37.

[24] Neal Mielke, Todd Marquart, Ning Wu, Jeff Kessenich, Hanmant Belgal, Eric Schares, Falgun Trivedi, Evan Goodness, and Leland R Nevill. 2008. Bit error rate in NAND flash memories. In *2008 IEEE International Reliability Physics Symposium (IRPS)*. IEEE, 9–19.

[25] A Modelli, A Visconti, and R Bez. 2004. Advanced flash memory reliability. In *2004 International Conference on Integrated Circuit Design and Technology (IEEE Cat. No. 04EX866)*. IEEE, 211–218.

[26] Nikolaos Papandreou, Haralampos Pozidis, Thomas Parnell, Nikolas Ioannou, Roman Pletka, Sasa Tomic, Patrick Breen, Gary Tressler, Aaron Fry, and Timothy Fisher. 2019. Characterization and Analysis of Bit Errors in 3D TLC NAND Flash Memory. In *2019 IEEE International Reliability Physics Symposium (IRPS)*. IEEE, 1–6.

[27] Young-Bog Park and Dieter K Schroder. 1998. Degradation of thin tunnel gate oxide under constant Fowler-Nordheim current stress for a flash EEPROM. *IEEE Transactions on Electron Devices* 45, 6 (1998), 1361–1368.

[28] D Resnati, Aurelio Mannara, G Nicosia, GM Paolucci, P Tessariol, AL Lacaita, AS Spinelli, and C Monzio Compagnoni. 2017. Temperature activation of the string current and its variability in 3-D NAND Flash arrays. In *2017 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 4–7.

[29] Noboru Shibata, Kazushige Kanda, T Shimizu, J Nakai, Osamu Nagao, N Kobayashi, M Miakashi, Yasushi Nagadomi, Takeshi Nakano, T Kawabe, et al. 2019. A 1.33 Tb 4-bit/Cell 3D-Flash Memory on a 96-Word-Line-Layer Technology. In *2019 IEEE International Solid-State Circuits Conference-(ISSCC)*. IEEE, 210–212.

[30] Chang Siau, Kwang-Ho Kim, Seungpil Lee, Katsuaki Isobe, Noboru Shibata, Kapil Verma, Takuya Ariki, Jason Li, Jong Yuh, Anirudh Amarnath, et al. 2019. A 512Gb 3-bit/Cell 3D Flash Memory on 128-Wordline-Layer with 132MB/s Write Performance Featuring Circuit-Under-Array Technology. In *2019 IEEE International Solid-State Circuits Conference-(ISSCC)*. IEEE, 218–220.

[31] JEDEC Standard. 2010. Solid-state drive (SSD) requirements and endurance test Method. *JESD218, September* (2010).

[32] Fei Wu, Meng Zhang, Yajuan Du, Weihua Liu, Zuo Lu, Jiguang Wan, Zhihu Tan, and Changsheng Xie. 2019. Using Error Modes Aware LDPC to Improve Decoding Performance of 3-D TLC NAND Flash. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 39, 4 (2019), 909–921.

[33] Fei Wu, Yue Zhu, Qin Xiong, Zhonghai Lu, You Zhou, Weizhen Kong, and Changsheng Xie. 2018. Characterizing 3D charge trap NAND flash: Observations, analyses and applications. In *2018 IEEE 36th International Conference on Computer Design (ICCD)*. IEEE, 381–388.

[34] Qin Xiong, Fei Wu, Zhonghai Lu, Yue Zhu, You Zhou, Yibing Chu, Changsheng Xie, and Ping Huang. 2017. Characterizing 3D floating gate NAND flash. In *2017 ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS)*. ACM, 31–32.

[35] Qin Xiong, Fei Wu, Zhonghai Lu, Yue Zhu, You Zhou, Yibing Chu, Changsheng Xie, and Ping Huang. 2018. Characterizing 3D floating gate NAND flash: Observations, analyses, and implications. *ACM Transactions on Storage (TOS)* 14, 2 (2018), 16.

[36] YEESTOR. 2018. YS9083XT/YS9081XT SSD Platform. http://www.yeestor.com/index-product-info-id-946-cid-33-pid-3-infopid-33.html.