Improving 3D NAND Flash Memory Lifetime by Tolerating Early Retention Loss and Process Variation

Yixin Luo[†] Saugata Ghose[†] Yu Cai[‡] Erich F. Haratsch[◇] Onur Mutlu[⋆†]

[†]Carnegie Mellon University [‡]SK Hynix, Inc. [◇]Seagate Technology [⋆]ETH Zürich

ABSTRACT

Compared to planar (i.e., two-dimensional) NAND flash memory, 3D NAND flash memory uses a new flash cell design, and vertically stacks dozens of silicon layers in a single chip. This allows 3D NAND flash memory to increase storage density using a much less aggressive manufacturing process technology than planar NAND flash memory. The circuit-level and structural changes in 3D NAND flash memory significantly alter how different error sources affect the reliability of the memory.

In this paper, through experimental characterization of real, state-of-the-art 3D NAND flash memory chips, we find that 3D NAND flash memory exhibits *three* new error sources that were not previously observed in planar NAND flash memory: (1) *layer-to-layer process variation*, a new phenomenon specific to the 3D nature of the device, where the average error rate of each 3D-stacked layer in a chip is significantly different; (2) *early retention loss*, a new phenomenon where the number of errors due to charge leakage increases *quickly within several hours* after programming; and (3) *retention interference*, a new phenomenon where the rate at which charge leaks from a flash cell is dependent on the data value stored in the neighboring cell.

Based on our experimental results, we develop new analytical models of layer-to-layer process variation and retention loss in 3D NAND flash memory. Motivated by our new findings and models, we develop four new techniques to mitigate process variation and early retention loss in 3D NAND flash memory. Our first technique, Layer Variation Aware Reading (LaVAR), reduces the effect of layer-to-layer process variation by fine-tuning the read reference voltage separately for each layer. Our second technique, Layer-Interleaved Redundant Array of Independent Disks (LI-RAID), uses information about layer-to-layer process variation to intelligently group pages under the RAID error recovery technique in a manner that reduces the likelihood that the recovery of a group fails significantly earlier than the recovery of other groups. Our third technique, Retention Model Aware Reading (ReMAR), reduces retention errors in 3D NAND flash memory by tracking the retention time of the data using our new retention model and adapting the read reference voltage to data age. Our fourth technique, Retention Interference Aware Neighbor-Cell Assisted Correction (ReNAC), adapts the read reference voltage to the amount of retention interference a page has experienced, in order to re-read the data after a read operation fails. These four techniques are complementary, and can be combined together to significantly improve flash memory reliability. Compared to a state-of-the-art baseline, our techniques, when combined, improve flash memory lifetime by 1.85×. Alternatively, if a NAND flash vendor wants to keep the lifetime of the 3D NAND flash memory device constant, our techniques reduce the storage overhead required to hold error correction information by 78.9%.

CCS CONCEPTS

• **Computer systems organization → Reliability**; **Secondary storage organization**; • **Hardware → Memory test and repair**; **Memory and dense storage**; *Non-volatile memory*;

KEYWORDS

3D NAND flash memory; error correction; fault tolerance; reliability; solid-state drives; storage systems

ACM Reference Format:
Y. Luo et al. 2018. Improving 3D NAND Flash Memory Lifetime by Tolerating Early Retention Loss and Process Variation. In *Proc. ACM Meas. Anal. Comput. Syst.*

1 INTRODUCTION

Solid-state drives (SSDs), which consist of NAND flash memory chips, are a popular data storage medium in modern computer systems. Traditionally, NAND flash memory has employed a *planar* (i.e., two-dimensional) architecture, where the entire chip resides on a single layer of silicon. In planar NAND flash memory, a flash cell is made using a *floating-gate transistor*, where data is represented by the amount of charge stored in the transistor's floating gate. The amount of charge stored in the floating gate determines the *threshold voltage* of the flash cell transistor (i.e., the voltage at which the transistor turns on).

For planar NAND flash memory, to continually increase the SSD capacity and decrease the cost-per-bit of the SSD, flash vendors have been aggressively scaling NAND flash memory to smaller manufacturing process technology nodes. This, however, comes at the cost of lower reliability [9, 13, 69]. Due to a combination of manufacturing process technology limitations and reduced reliability of planar NAND flash memory, it has become increasingly difficult for vendors to continue to scale the density of planar NAND flash memory chips [11, 31, 80].

To overcome this scaling challenge, 3D NAND flash memory has recently been introduced [39, 45, 80]. Although 3D NAND flash memory is already being deployed at large scale in new computer systems, there is a lack of available knowledge on the error characteristics of real 3D NAND flash memory chips, which makes it harder to estimate the reliability characteristics of systems that employ such chips. Previous publicly-available experimental studies on NAND flash memory errors using real flash memory chips

(e.g., [4–9, 11, 13–16, 64, 69, 81]) have mostly been on planar NAND flash memory devices.[1]

We identify that 3D NAND flash memory has *three* fundamental differences from the most recent generation (i.e., 10–15 nm) of planar NAND flash memory, which lead to new error characteristics for 3D NAND flash memory that we observe experimentally: (1) 3D NAND flash memory currently uses a different flash cell architecture than planar NAND flash memory. Instead of using a floating-gate transistor, a cell in 3D NAND flash memory consists of a *charge trap transistor* [86], which stores charge within an insulator. (2) Unlike planar NAND flash memory, 3D NAND flash memory vertically stacks *multiple layers* of silicon together within a single chip. Modern 3D NAND flash memory chips typically contain 24–96 stack layers [1, 39, 45, 50, 80, 90]. Due to the high layer count, 3D NAND flash memory can provide high storage density without needing to scale the process technology as aggressively as was done for planar NAND flash memory. (3) While modern planar NAND flash memory uses a manufacturing process technology node as small as 10–15 nm [58, 90], 3D NAND flash memory currently uses a much larger manufacturing process technology node (e.g., 30–50 nm [86]).

Our goal in this work is to (1) identify and understand the *new* error characteristics of 3D NAND flash memory (i.e., those that did *not* exist previously in planar NAND flash memory), and (2) develop new techniques to mitigate prevailing 3D NAND flash memory errors. We aim to achieve these goals via rigorous experimental characterization of real, state-of-the-art 3D NAND flash memory chips from a major flash vendor. Based on our comprehensive characterization and analysis, we identify *three new error characteristics* that were *not* previously observed in planar NAND flash memory, but are fundamental to the new architecture of 3D NAND flash memory:

(1) 3D NAND flash memory exhibits *layer-to-layer process variation*, a new phenomenon specific to the 3D nature of the device, where the average error rate of each 3D-stacked layer in a chip is significantly different from one another (Section 4.2). We are the *first* to provide detailed experimental characterization results of layer-to-layer process variation in real flash devices in open literature. Our results show that the raw bit error rate in the middle layer can be 6× the raw bit error rate in the top layer.

(2) 3D NAND flash memory experiences *early retention loss*, a new phenomenon where the number of errors due to charge leakage increases *quickly within several hours* after programming, but then increases at a much slower rate (Section 4.3). We are the *first* to perform an extended-duration observation of early retention loss. While a prior study [23] examines the impact of early retention loss over only the first 5 minutes after data is written, we examine the impact of early retention loss over the course of 24 days. Our results show that the retention error rate in a 3D NAND flash memory block quickly increases by an order of magnitude within ∼3 hours after programming.

(3) 3D NAND flash memory experiences *retention interference*, a new phenomenon where the rate at which charge leaks from a flash cell is dependent on the amount of charge stored in neighboring flash cells (Section 4.4). Our results show that charge leaks at a lower rate (i.e., the retention loss speed is slower) when the vertically-adjacent cell is in a state that holds more charge (i.e., a higher-voltage state).

Our experimental observations indicate that we must revisit the error models and the error mitigation mechanisms devised for planar NAND flash memory, as they are no longer accurate for 3D NAND flash memory behavior. To this end, we develop *new analytical models* of (1) the layer-to-layer process variation in 3D NAND flash memory (Section 5.1), and (2) retention loss in 3D NAND flash memory (Section 5.2). Our models estimate the raw bit error rate (RBER), threshold voltage distribution, and the *optimal read reference voltage* (i.e., the voltage at which the RBER is minimized when applied during a read operation) for each flash page. Both models are useful for developing techniques to mitigate raw bit errors in 3D NAND flash memory.

We propose *four new techniques* to mitigate the unique layer-to-layer process variation and early retention loss errors observed in 3D NAND flash memory. Each technique makes use of our new analytical models of layer-to-layer process variation and retention loss in 3D NAND flash memory. Our first technique, Layer Variation Aware Reading (LaVAR), reduces process variation by fine-tuning the read reference voltage independently for each layer. Our second technique, Layer-Interleaved Redundant Array of Independent Disks (LI-RAID), improves reliability by changing how pages are grouped under the RAID error recovery technique. LI-RAID uses information about layer-to-layer process variation to reduce the likelihood that the RAID recovery of a group could fail significantly earlier during the flash lifetime than the recovery of other groups. Our third technique, Retention Model Aware Reading (ReMAR), reduces retention errors in 3D NAND flash memory by tracking the retention time of the data using our retention model and adapting the read reference voltage to data age. Our fourth technique, Retention Interference Aware Neighbor-Cell Assisted Correction (ReNAC), adapts the read reference voltage to the amount of retention interference and re-reads the data after a read operation fails, in order to correct the cells affected by retention interference. These four techniques are complementary, and can be combined together to significantly improve flash memory reliability. Compared to a state-of-the-art baseline, our techniques, when combined, improve flash memory lifetime by 1.85×. Alternatively, if a NAND flash vendor wants to keep the lifetime of the 3D NAND flash memory device constant, our techniques reduce the storage overhead required to hold error correction information by 78.9%.

This paper makes the following **key contributions**:

- It presents the first *comprehensive experimental characterization* of real, state-of-the-art 3D NAND flash memory chips, and provides an *in-depth analysis* of layer-to-layer process variation, early retention loss, and retention interference, which are *three new error characteristics* inherent to 3D NAND flash memory.
- It develops *new analytical models* for (1) layer-to-layer process variation and (2) early retention loss, which can be used to estimate the raw bit error rate, mean and standard deviation of the

[1] With the exception of our very recent prior work [65], which examined two specific important aspects of 3D NAND flash memory reliability: temperature and self-recovery effects.

threshold voltage distribution of each state, and the optimal read reference voltages.
- It develops *four new mechanisms*, LaVAR, LI-RAID, ReMAR, and ReNAC, to mitigate the three new error characteristics we have identified in 3D NAND flash memory. It evaluates these techniques, and shows that, when applied together, they improve 3D NAND flash memory lifetime by 1.85×, or reduce the storage overhead for error correction by 78.9% if we keep the lifetime constant, compared to a state-of-the-art baseline.

2 BACKGROUND

In this section, we first provide necessary background on the basics of NAND flash memory (Section 2.1). Next, we briefly discuss the different known sources of errors within planar NAND flash memory (Section 2.2). For an extended background on NAND flash memory, we refer the reader to our prior works [9–11].

2.1 NAND Flash Memory Basics

In NAND flash memory, each flash cell consists of a transistor that can store charge. A flash cell represents a certain data value based on the *threshold voltage* (V_{th}) of its transistor, which is determined by the amount of charge stored in it. In *multi-level cell* (MLC) flash memory, each cell stores two bits of data. A threshold voltage window (i.e., *state*) is assigned for each possible two-bit value. Figure 1a shows the four possible states (i.e., ER, P1, P2, P3) in MLC NAND flash memory, along with their corresponding bit values. As a result of manufacturing process variation, the threshold voltage of cells programmed to the same state follow a Gaussian-like distribution across the voltage window of the state [9, 14, 64, 81], depicted as a probability density curve in Figure 1a.

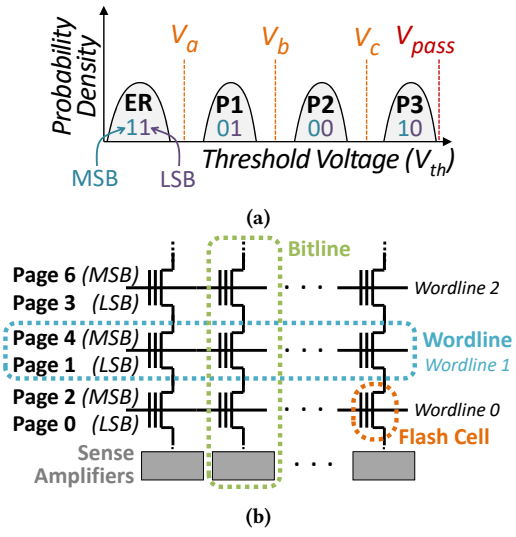


Figure 1: (a) Threshold voltage distribution and read reference voltages for MLC NAND flash memory; (b) Internal organization of a flash block.

A NAND flash memory chip contains thousands of *flash blocks*, which are two-dimensional arrays of flash cells. Figure 1b shows the internal organization of a flash block. Each block contains dozens of rows (i.e., *wordlines*) of flash cells, where each row typically contains 64K to 128K cells. All of the cells on the same wordline are read and programmed together as a group. MLC NAND flash memory partitions the two bits of each flash cell in a wordline across two *pages*, which are the unit of data programmed at a time (typically 8 kB). The *least significant bits* (LSBs) of all cells in one wordline form the *LSB page* of that wordline, and the *most significant bits* (MSBs) of these cells form the *MSB page*. The sources and drains of cells across different wordlines in the same block are connected in series to form a *bitline*.

Reads and writes to the flash memory are managed by an SSD *controller*. The controller reads a page from a flash block by applying a *read reference voltage* (V_{ref}) to the wordline that holds the page. A cell switches on only if $V_{th} > V_{ref}$. Figure 1a shows the three read reference voltages (V_a, V_b, and V_c) that are used to distinguish between each state. A *sense amplifier* is attached to each bitline to detect if the cell is switched on. In order to detect the state of a particular cell on the bitline, the controller applies a *pass-through voltage* (V_{pass}) to the wordlines of *all* unread cells in the flash block. This turns on the unread cells, allowing the value of the cell that is being read to propagate through the bitline to the sense amplifier. To guarantee that all unread cells are *on*, V_{pass} is set to the maximum possible threshold voltage [5, 9].

Before new data can be written (i.e., *programmed*) to a flash page, the controller must first *erase* the *entire block* (i.e., 512 to 1024 pages) that the page belongs to, due to wiring constraints. After erase, all of the cells in the erased block are reset to the ER state. To program a flash cell, the controller sends the data to be programmed to the flash chip, which repeatedly pulses a high programming voltage on a cell to increase a cell's threshold voltage until the cell reaches its target state. This iterative programming approach is called *incremental step pulse programming* (ISPP) [3, 69, 89, 91]. Each pair of erase and program operations is referred to as a *program/erase* (P/E) *cycle*.

2.2 Errors in NAND Flash Memory

As vendors work to increase the density of NAND flash memory, they use aggressive manufacturing process technology scaling to reduce the size of a flash cell. As a result, each cell has a smaller capacity to store charge, and the cells move closer to each other. These changes reduce the reliability of the NAND flash memory, thereby increasing the probability of flash memory errors in newer generations of planar (i.e., two-dimensional) NAND flash memory. Errors occur when the cell threshold voltage (V_{th}) unintentionally changes or is read incorrectly, which can alter the cell state observed by the controller. Errors can be induced by a range of sources [4–9, 11, 13–16, 65, 69], which we divide into four categories: process variation errors, retention errors, write-induced errors, and read-induced errors. We briefly describe each error source below, and refer the reader to the prior work cited below for detailed explanations of each error source. A comprehensive treatment of different types of NAND flash memory errors and mitigation mechanisms for them can be found in our recent survey papers [9, 11].

Process variation errors occur as a result of the fabrication process. Within a single chip, different flash cells have different attributes, due to the lithography limitations of modern manufacturing process technologies [13, 84]. As a result, there is inherent variation among the cells, and some cells have a higher error rate than other cells.

Retention errors [6–8] are a type of error that increase and accumulate over time after a flash cell is programmed. A retention error occurs because charge leaks out of the transistor over time. As charge leaks from a cell, the cell's threshold voltage (V_{th}) decreases. In planar NAND flash memory, retention errors are the dominant source of all flash memory errors [6–8, 13], if aggressive refresh techniques [7, 8, 63] are not employed.

Write-induced errors occur during program or erase operations. *P/E cycling errors* (or program/erase variation errors) [14, 64, 81] are errors that occur immediately after erasing and programming a flash page. These errors occur because of the inaccuracy of each program and erase operation. This inaccuracy causes some cells to be programmed into a state other than its desired target state. As more P/E cycles take place over the lifetime of a flash cell, the repeated stress causes more electrons to become trapped within the transistor, which is known as *wearout*. Wearout increases the inaccuracy during program and erase operations, thereby increasing the number of P/E cycling errors. *Cell-to-cell program interference errors* [15, 16] are another type of write-induced error that increases the threshold voltage of a cell and thereby increases the RBER, when an *adjacent* cell in *another* wordline is being programmed. Since parasitic capacitance coupling exists between cells within close proximity of each other, when a high programming voltage is applied on one cell, the capacitance coupling adds charge to the transistors of the adjacent cells, increasing the program interference errors.

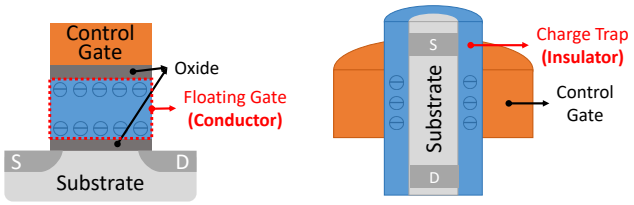
Read-induced errors occur during read operations. *Read errors* [24, 29, 42] are a type of read-induced error where two reads to a flash cell may return different data values. A read error occurs when the read reference voltage is close to the cell's threshold voltage. Such an error occurs when random fluctuations on the bitline cause the sense amplifier to detect the wrong data. *Read disturb errors* [5, 81] are another type of read-induced error where reading a page in a flash block may change the values stored in (i.e., increase the RBER) of *other* pages in the same block. This type of error occurs due to the application of the *pass-through voltage* (V_{pass}) to unread cells. When one cell on a bitline is being read, applying V_{pass} to the unread cells can induce a weak programming effect on the *unread* cells, slowly transferring electrons into the unread cells' transistors and increasing the threshold voltage of the unread cells.

To mitigate these errors, SSDs use error-correcting codes (ECC) on the data. ECC has a fixed *error correction capability*: it can correct only a limited number of errors, beyond which the data is no longer correctable. When a flash page is uncorrectable, we say that the SSD has reached the end of its *lifetime*.

3 ARCHITECTURAL DIFFERENCES BETWEEN 3D NAND AND PLANAR NAND

3D NAND flash memory (or 3D NAND) has three *fundamental* differences from the most recent generation (i.e., 10–15 nm) of planar NAND flash memory: (1) the flash cell design, (2) the organization of flash cells within a chip, and (3) the manufacturing process technology node.

Flash Cell Design. In both planar and 3D NAND flash memory, each flash cell consists of a transistor that can store charge, where the amount of charge determines the threshold voltage of the cell (i.e., the voltage at which the cell turns on). The vast majority of planar NAND flash memory uses a *floating-gate transistor* (FG) for each cell. Figure 2a illustrates the design of a floating-gate cell. A *control gate* sits at the top of the transistor. Read, program, and erase operations all apply a voltage onto the control gate to turn on the cell or to add charge to the transistor. A *floating gate* sits in the middle of the transistor. The floating gate is a conductor that stores the transistor's charge, and is sandwiched by *oxide* layers. The oxide layers minimize the amount of charge that leaks out of the floating gate. At the bottom of the cell is the *substrate*, which has two terminals on either end, marked *source* (S) and *drain* (D). When the voltage applied on the control gate is higher than the voltage of the charge stored in the floating gate, an electrical *channel* forms between the source and drain, connecting them together. The floating gate voltage can be increased or decreased by applying a large positive or negative voltage, respectively, to the control gate, which induces Fowler-Nordheim tunneling [27] of electrons through the oxide.



(a) Floating-Gate Cell (b) 3D Charge Trap Cell

Figure 2: The design of (a) a floating-gate cell, and (b) a 3D charge trap cell.

Instead of floating-gate transistors, most existing 3D NAND flash memory designs use a *charge trap transistor* (CT) for each cell. Figure 2b illustrates the design of a charge trap cell. The substrate, and therefore the channel between source and drain, sits vertically in the center of the cell. A *charge trap layer* wraps around the substrate. The charge trap layer takes the place of the floating gate, storing the transistor's charge. However, unlike the floating gate, the charge trap layer is an insulator. The control gate still exists in a charge trap cell, but it now wraps around the charge trap layer.

Flash Chip Organization. Figure 3 illustrates the physical organization of flash cells in 3D NAND flash memory. The charge trap transistor design allows the bitline (BL in Figure 3) of a block to stand *vertically* (i.e., along the z-axis) in the chip. In other words, the bitline now connects together one charge trap cell from *each layer* of the chip, as the cells are stacked on top of each other. Note that all of the cells along the z-axis share the same charge trap insulator, akin to how transistors are connected together on a bitline in planar NAND flash memory. The control gates of cells in the same layer, along the y-axis, are connected together to form a wordline.

In this figure, we show a simple example where the cells in the same y–z plane form a flash block. In reality, to form larger flash blocks, multiple stacks of flash cells are connected together to form longer bitlines, thus increasing the number of wordlines within a block. Multiple such flash blocks are aligned along the x-axis to form a flash chip.

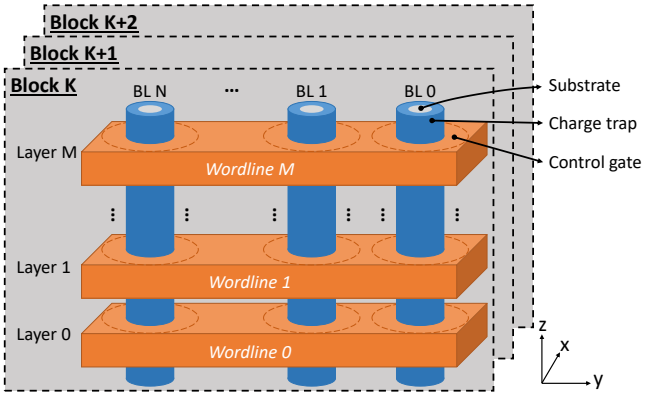


Figure 3: 3D NAND flash memory organization.

Manufacturing Process Technology. Compared with the most recent generation of planar NAND flash memory (i.e., 10–15 nm), 3D NAND flash memory uses a much larger manufacturing process technology node (e.g., 30–50 nm) [86]. Because 3D NAND flash memory has a large number of layers (typically 24–96 [1, 39, 45, 50, 80, 90]), it can reach the same storage density of the most recent planar NAND flash memory generation while using much larger flash cells.

4 CHARACTERIZATION OF 3D NAND FLASH MEMORY ERRORS

Our goal is to identify and understand new error characteristics in 3D NAND flash memory, through rigorous experimental characterization of real, state-of-the-art 3D NAND flash memory chips. We use the observations and analyses obtained from such characterization to (1) compare how the reliability of a 3D NAND flash memory chip differs from that of a planar NAND flash memory chip, (2) develop a model of how each new error source affects the error rate of 3D NAND flash memory, (3) understand if and how these reliability characteristics will change with future generations of 3D NAND flash memory, and (4) develop mechanisms that can mitigate new error sources in 3D NAND flash memory.

For our characterization, we use the methodology discussed in Section 4.1. First, we perform a detailed characterization and analysis of three error characteristics that are drastically different in 3D NAND flash memory than in planar NAND flash memory: layer-to-layer process variation (Section 4.2), early retention loss (Section 4.3), and retention interference (Section 4.4). In addition to identifying *new* error sources in 3D NAND flash memory, we use our methodology to corroborate and quantify 3D NAND error characteristics that are a result of error sources that were *previously* identified in planar NAND flash memory, including retention loss [6–9, 11, 23, 80], P/E cycling [9, 11, 14, 64, 80, 81], program interference [4, 9, 11, 15, 16, 80], read disturb [5, 9, 11, 81], and process variation [13, 84]. We summarize our findings for these error types in Section 4.5, and provide detailed results on our characterization of these previously-identified error sources in Appendix A.

4.1 Methodology

We experimentally characterize several real, state-of-the-art 3D MLC NAND flash memory chips from a single vendor.[2,3] We use a NAND flash characterization platform similar to prior work [4–9, 11–16, 64, 65, 81], which allows us to issue *read-retry* commands directly to the flash chip. The read-retry command [9, 14] allows us to fine-tune the read reference voltage used for each read operation. The smallest amount by which we can change the read reference voltage is called a *voltage step*. We conduct all experiments at room temperature (20 °C).

We use two metrics to evaluate 3D NAND flash memory reliability. First, we show the *raw bit error rate* (RBER), which is the rate at which errors occur in the data *before error correction*. We show the RBER for when we read data using the *optimal read reference voltage* (V_{opt}), which is the read reference voltage that generates the fewest errors in the data.[4]

Second, we show how the various error sources change the *threshold voltage distribution*. These changes (i.e., shifting and widening) in threshold voltage distribution directly lead to raw bit errors in the flash memory. To obtain the distribution, we first use the read-retry command to sweep over all possible voltage values, to identify the threshold voltage of each cell.[5] Then, we use this data to calculate the probability density of each state at every possible threshold voltage value. As part of our analysis, we fit the threshold voltage distribution of each state to a Gaussian distribution. We use the *mean* of the Gaussian model to represent how the distribution shifts as a result of errors, and we use the *standard deviation* of the model to represent how the distribution widens. Throughout this paper, we present normalized voltage values, as the actual voltage values are proprietary to NAND flash memory vendors. A normalized voltage of 1 represents a single fixed voltage step.

We show two examples in Figure 4 to visualize how well this simple Gaussian model captures the change in the measured threshold voltage distribution. Figure 4 shows the measured and modeled distributions under two conditions: (1) after 0 P/E cycles, 0-day retention time [6], and 0 read disturbs (i.e., the data contains few errors); and (2) after 10K P/E cycles, 3-day retention time [6], and 900K read disturbs (i.e., the data contains a high number of errors). Dotted points plot the measured threshold voltage distributions from the real 3D NAND memory chips. Note that we are unable to show the ER state distribution when the P/E cycle count is low

[2]The trends we observe from the characterization are expected be similar for 3D charge trap flash memory manufactured by different vendors, as their 3D flash memory organizations are similar in design.
[3]We normalize the actual number of stacked layers of the chips and leave out the exact process technology to protect the anonymity of the flash vendor and to avoid revealing proprietary information.
[4]We show RBER at the optimal read reference voltage to accurately represent the reliability of NAND flash memory, as SSD controllers tune the read reference voltage to a near-optimal point to extend the NAND flash lifetime [6, 9, 64, 76].
[5]We refer to prior work for more detail on the methodology to obtain the threshold voltage distribution [14, 64, 81].

(i.e., the black dots), because the erase operation cleanly resets the threshold voltage to a negative value that is lower than the observable voltage range under a low P/E cycle count. We use a solid line to show a fitted Gaussian distribution for each state. The Kullback-Leibler divergence error values [64, 81] of the fitted Gaussian distributions are 0.034 and 0.23.[6] We observe, from this figure, that after the chip is used, the threshold voltage distribution shifts due to P/E cycling, retention loss, and read disturb, reducing the error margins between neighboring states, and leading to more raw bit errors in the data. Thus, depicting and understanding how threshold voltage distributions are affected by various factors helps us understand how raw bit errors occur and thus devise mechanisms to mitigate various errors more effectively.

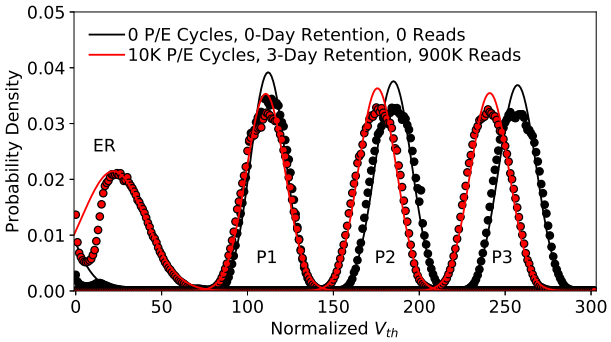


Figure 4: 3D NAND threshold voltage distribution before (black) and after (red) the data is subject to a high number of errors (due to P/E cycling, retention loss, and read disturb).

In the following sections, we directly show the mean and the standard deviation of the *fitted* threshold voltage distributions instead of the distribution itself, to simplify the presentation of our results.

Limitations. In our experiments, we randomly sampled 27 flash blocks throughout our characterizations. Note that each sampled flash block consists of tens of millions of flash cells. Thus, we believe that our observations are representative of the general behavior that takes place in the model of 3D NAND chips that we tested. While adding more data samples (i.e., flash blocks to test) can add to the statistical strength of our results, we do not believe that this would change the *general qualitative findings* that we make and the *models* that we develop in this work. This is because the new error characteristics we observe are caused by the underlying architecture of 3D NAND flash memory (see Section 3).

Note that we do not characterize *chip-to-chip* process variation, as an accurate study of such variation requires a large-scale study of a large number (e.g., hundreds) of 3D NAND flash memory chips, which we do not have access to. Hence, we leave such a large-scale study for future work.

4.2 Layer-to-Layer Process Variation

Process variation refers to the variation in the attributes of flash cells when they are fabricated (see Section 2.2). Due to process variation, some flash cells can have a higher RBER than others, making these cells the limiting factor of overall flash memory reliability. In 3D NAND flash memory, process variation can occur along all three axes of the memory (see Figure 3). Among the three axes, we expect the variation along the z-axis (i.e., layer-to-layer variation) to be the most significant, due to the new challenge of stacking multiple flash cells across layers. Prior work has shown that current circuit etching technologies are unable to produce identical 3D NAND cells when punching through multiple stacked layers, leading to significant variation in the error characteristics of flash cells that reside in different layers [38, 92].

To characterize layer-to-layer process variation errors within a flash block, we first wear out the block by programming random data to each page in the block until the block endures 10K P/E cycles. Then, we compare the collective characteristics of the flash cells in one layer with those in another layer. We repeat this experiment for flash blocks on multiple chips to verify all of our findings.

Observations. Figure 5 shows the RBER variation along the z-axis (i.e., across layers) for a flash block that has endured 10K P/E cycles. The chips we use for characterization have between 30 and 40 layers. We normalize the number of layers from 0 (the top-most layer) to 100 (the bottom-most layer) by multiplying the actual layer number with a constant, to maintain the anonymity of the chip vendors. Figure 5a breaks down the errors according to the originally-programmed state and the current state of each cell; Figure 5b breaks down the errors into MSB and LSB page errors. In Figure 5b, the solid curve and the dotted curve show the results for two blocks that were randomly selected from two different flash chips. We make five observations from Figure 5. First, ER $\leftrightarrow$ P1 and P1 $\leftrightarrow$ P2 errors vary significantly across layers, while P2 $\leftrightarrow$ P3 errors remain similar across layers. The variation in ER $\leftrightarrow$ P1 errors is mainly caused by the large variation in mean threshold voltage of the ER state across layers; the variation in P1 $\leftrightarrow$ P2 is caused by the variation in the threshold voltage distribution width of the P1 state across layers (Section A.4). Second, both the MSB and LSB error rates vary significantly across layers. We call this phenomenon *layer-to-layer process variation*. For example, MSB page on normalized layer 55 in the middle (i.e., *Max MSB*) has an RBER 21× that of normalized layer 0. Third, MSB error rates are much higher than LSB error rates in a majority of the layers, on average by 2.4×. We call this phenomenon *MSB–LSB RBER variation*. MSB error rates are usually higher than LSB error rates because reading an MSB page requires two read reference voltages (V_a and V_c), whereas reading an LSB page requires only one (V_b). Fourth, the top half of the layers have lower error rates than the bottom half. This is likely caused by the variation in the flash cell size across layers. Fifth, the RBER variation we observe is consistent across two randomly-selected blocks from two different chips. This indicates that layer-to-layer process variation and MSB–LSB RBER variation are consistent characteristics of 3D NAND flash memory.

Figure 6 shows how the optimal read reference voltages vary across layers. Three subfigures show the optimal read reference voltages for V_a, V_b, and V_c. We make two observations from Figure 6. First, the optimal voltages for V_a and V_b vary significantly across layers, but the optimal voltage for V_c does not change by much. This is because process variation mainly affects the threshold voltage distributions of the ER and P1 states, whereas the threshold voltage

[6] A KL-divergence error of x means that the model loses x natural units of information (i.e., nats) due to modeling error.

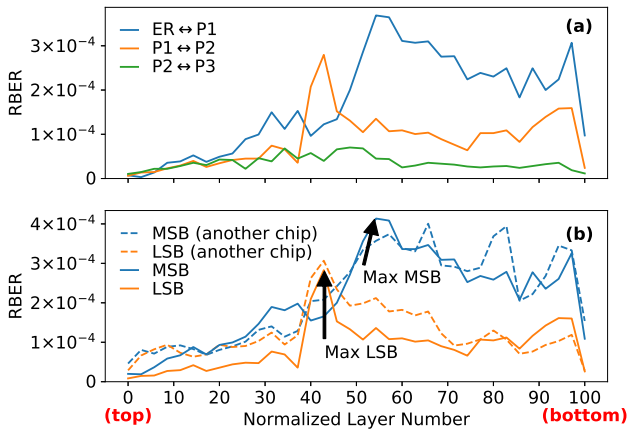


Figure 5: Variation of RBER across layers.

distributions of the P2 and P3 states, which are more accurately controlled by ISPP (see Section 2), are similar across layers. We discuss this further in Appendix A.4. Second, the optimal read reference voltages for V_a and V_b are lower for cells in the top half of the layers than for cells in the bottom half. This is because process variation significantly affects the threshold voltage of the ER and P1 states (see Appendix A.4).

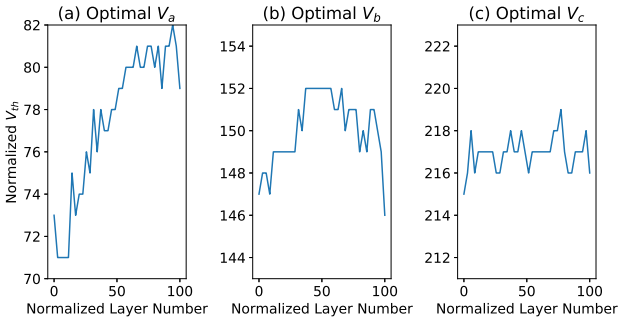


Figure 6: Variation of optimal read reference voltage across layers.

Insights. We show that the phenomena of layer-to-layer process variation and MSB-LSB RBER variation, which are unique to 3D NAND flash memory, are significant. We refer to Appendix A.4 for a comparison between layer-to-layer process variation and bitline-to-bitline process variation. In the future, as 3D NAND flash devices scale along the z-axis, more layers will be stacked vertically along each bitline. This will likely further exacerbate the effect of layer-to-layer process variation, making it even more important to study and mitigate its negative effects.

4.3 Early Retention Loss

Retention errors are flash memory errors that accumulate after data has been programmed to the flash cells [6–8] (see Section 2.2). Because 3D NAND flash memory typically uses a different cell design (i.e., the charge trap cell described in Section 3) than planar NAND flash memory (which uses floating-gate cells), it has drastically different retention error characteristics. The charge trap flash cells used in 3D NAND flash memory suffer from *early retention loss*, i.e., fast charge loss within a few seconds. This phenomenon has been observed by prior works using circuit-level characterization [21, 23]. However, due to limitations of the circuit-level characterization methodology used by these prior works, openly-available characterizations of early retention loss in 3D charge trap NAND flash devices document retention loss behavior for up to only 5 minutes after the data is written (i.e., for a maximum *retention time* of 5 minutes). This limited window is insufficient for understanding early retention loss under real workloads, which typically have much longer *retention time* requirements [63], i.e., the length of time that has elapsed since programming until the data is accessed again.

Our goal is to experimentally characterize early retention loss in 3D NAND flash memory for a large range of retention times (e.g., from several minutes to several weeks). First, we randomly select 11 flash blocks within each chip and write pseudo-random data to each page within the block to wear the blocks out. We wear out each block to a different P/E cycle count, so that we have error data for every 1K P/E cycles between 0 and 10K P/E cycles.[7] Then, we program pseudo-random data to each flash block, and wait for up to 24 days under room temperature. To characterize retention loss, we measure the RBER and the threshold voltage distribution at nine different retention times, ranging from 7 minutes to 24 days. To minimize the impact of other errors, and to allow us to include very low retention times, we characterize only the first 72 flash pages within each block. We believe that the observations we make on these flash cells are representative of the entire chip, and we can generalize the observations to a majority of 3D NAND flash memory cells. We analyze the threshold voltage distribution in Appendix A.2.

Observations. Figure 7 shows the comparison between the retention error rate of 3D NAND and planar NAND flash memory at 10,000 P/E cycles using both a logarithmic time scale on the x-axis (Figure 7a) and a linear time scale on the x-axis (Figure 7b) for different retention times after programming. To make this comparison, we perform the same experiment as above for planar NAND flash memory chips. Due to limitations of the available data, we extend our data to the same retention time range using a linear model that was proposed by prior work [65, 69]: $\log(RBER) = A \cdot \log(t) + B$, where t is the retention time, and A and B are parameters of the linear model. The dotted portions of the lines represent the RBER that is predicted by the linear model.

We make two observations from this figure. First, in Figure 7a, we observe that the retention error rate changes much more slowly for planar NAND flash memory than for 3D NAND flash memory. Although the 3D NAND flash memory chip has lower RBER than the planar NAND flash memory chip shortly after programming, the RBER becomes higher on the 3D NAND flash memory chip after 7×10^3 seconds (~2 hours) of retention time. This means that 3D NAND flash memory is more susceptible to the retention loss phenomenon than planar NAND flash memory. Second, in Figure 7b, we observe that the RBER of 3D NAND flash memory quickly

[7] For all experiments throughout the paper, we consistently assume a 0.5-second *dwell time*, which is the length of time between consecutive program/erase operations [65].

increases by an order of magnitude in 10^4 seconds (~3 hours), and by another order of magnitude in 10^6 seconds (~11 days). However, we do *not* observe a large difference in retention loss between low and high retention times for *planar* NAND flash memory (also shown by prior works [6, 69]). This shows that the retention loss is *steep* when retention time is *low*, but the retention loss flattens out when the retention time is high. This is a result of the early retention loss phenomenon in 3D NAND flash memory.

Early retention loss can be caused by two possible reasons. First, the tunnel oxide layer is thinner in 3D NAND flash memory than in planar NAND flash memory [86, 97]. Since a 3D charge trap cell uses an insulator to store charge, which is immune to the short circuiting caused by stress-induced leakage current (SILC) [26, 73], the tunnel oxide layer in 3D NAND flash memory is designed to be thinner to improve programming speed [80]. This causes charge to leak very fast soon after programming. Second, cells connected on the same bitline share the same charge trap layer. As a result, charge that is programmed to a flash cell quickly leaks to adjacent cells that are on the same bitline due to *electron diffusion* through the shared charge trap layer [23], which we discuss further in Section 4.4.

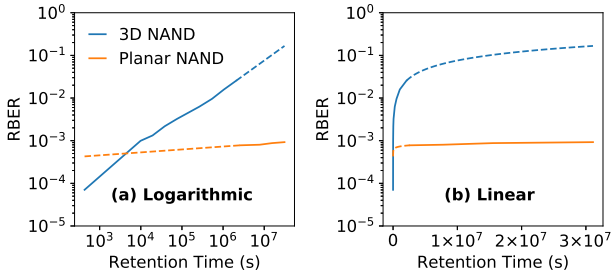


Figure 7: Retention error rate comparison between 3D NAND and planar NAND flash memory at 10K P/E cycles. Dotted portions of lines represent the RBER predicted by the linear model proposed by prior work [65, 69]. We show the retention time on the x-axis using both (a) a *logarithmic* time scale and (b) a *linear* time scale.

Figure 8 plots how the optimal read reference voltage changes with retention time. The three subfigures show the optimal voltages for V_a, V_b, and V_c. We make three observations from this figure. First, the relation between the optimal read reference voltages of V_b or V_c and the retention time can be modeled as [65, 69]: $V = A \cdot \log(t) + B$, similar to the logarithm of RBER (which we discuss above). Second, the optimal read reference voltages for V_b and V_c decrease significantly as retention time increases, whereas V_a remains relatively constant. Third, due to the early retention loss phenomenon, the optimal read reference voltages for V_b and V_c change rapidly when the retention time is low (e.g., V_c changes by 5 voltage steps within the first 3 hours), but they change slowly when the retention time is high (e.g., V_c changes by another 5 voltage steps after 11 days).

Insights. We compare the errors caused by retention loss in 3D NAND flash memory to that in planar NAND flash memory, using our results in Figure 7 and the results reported in prior

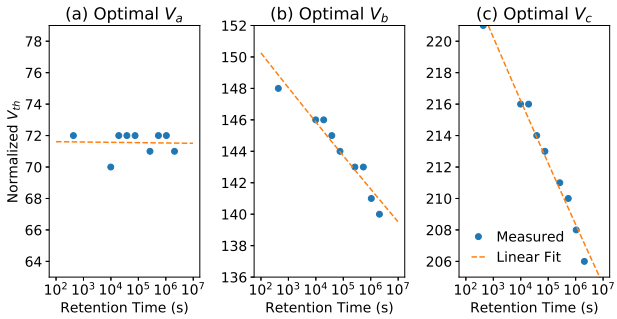


Figure 8: Optimal read reference voltages for different retention times. Note that the x-axis uses a logarithmic time scale.

work [6, 7, 69]. We find two major differences in 3D NAND flash memory, which we summarize below. More results and insights are in Appendix A.2. First, 3D NAND flash memory is more susceptible to retention errors than planar NAND flash memory, and its error rate increases much faster when the retention time is low than when the retention time is high. This is a result of the early retention loss phenomenon in 3D NAND flash memory, which is due to the use of a different flash cell design and thus is likely to remain in future generations of 3D NAND flash memory. Second, the optimal read reference voltages for V_b and V_c in 3D NAND flash memory change significantly with retention time. However, in planar NAND flash memory, the optimal voltage for V_b does *not* change by much [6], indicating that retention loss is a more pressing phenomenon in 3D NAND flash memory. This makes adjusting the optimal read reference voltages even more important for 3D NAND flash memory than for planar NAND flash memory. We conclude that it is necessary to develop novel mechanisms to mitigate the early retention loss phenomenon in 3D NAND flash memory.

4.4 Retention Interference

Retention interference is the phenomenon that the speed of retention loss for a cell depends on the threshold voltage of a *vertically-adjacent neighbor cell* whose charge trap layer is directly connected to the victim cell along the bitline. Retention interference is unique to 3D NAND flash memory, as cells along the *same* bitline in 3D NAND flash memory share the same charge trap layer. If two neighboring cells have different threshold voltages over time, charge can leak away from the cell with a higher threshold voltage to the cell with a lower threshold voltage [23]. Figure 9 shows an example of this phenomenon, where charge leaks from the top cell (which is in a higher-voltage state) to the bottom cell (which is in a lower-voltage state) through the shared charge trap layer. This charge leakage reduces the threshold voltage of the top cell while increasing the threshold voltage of the bottom cell.

We use the same data used for retention loss in Section 4.3 to observe the effects of retention interference. To eliminate any noise due to program interference, we use only the neighboring cells that are programmed *before* the victim cells to establish the retention

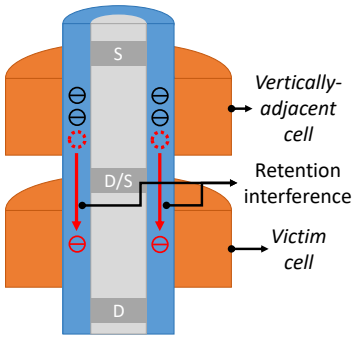


Figure 9: Retention interference phenomenon: a vertically-adjacent cell leaks charge into a victim cell.

interference correlation, as these cells do *not* induce program interference on the victim cells. We also ignore victim cells that are in the ER state, as they are significantly affected by program interference even though they are programmed after their neighbors [4]. Once program interference is eliminated, the cells should experience a similar threshold voltage shift due to retention loss *except for* the effects of retention interference. To find the retention interference, we first group all of the victim cells based on their threshold voltage states and the states of their neighboring cells. Then, we compare the amount by which the threshold voltages shift over a 24-day retention time, for each group, to observe how the cells are affected by the retention interference caused by neighboring cells.

Observations. Figure 10 shows the average threshold voltage shift over a 24-day retention time, broken down by the state of the victim cell (V) and the state of the neighboring cell (N). Each bar represents a different (V, N) pair. Different shades represent the different states of the neighboring cell, as labeled in the legend. Every 4 bars are grouped by the state of the victim cell, as labeled on the y-axis. The length of each bar represents the amount of threshold voltage shift over the 24-day retention time. From Figure 10, we observe that the threshold voltage shift over retention time is lower when the neighboring cell is in a higher-voltage state (e.g., the P3 state).

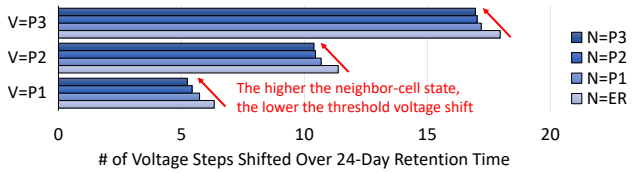


Figure 10: Retention interference phenomenon observed at 10K P/E cycles.

Insights. We are the first to quantify the retention interference phenomenon in 3D NAND flash memory. Our observation from Figure 10 shows that the amount of retention loss for a flash cell is correlated with its neighboring cell's state. We expect retention interference to become stronger as we shrink the manufacturing process technology node in future 3D NAND flash memory devices. This is because the distance between neighboring cells will decrease, and fewer electrons will be stored within each flash cell, increasing the susceptibility of a cell to interference from neighboring cells.

4.5 Other Error Characteristics

In addition to the three *new* error sources we find in 3D NAND flash memory, we also characterize the behavior of other *known* error sources in 3D NAND flash memory and compare them to their behavior in planar NAND flash memory. We present a high-level summary of our findings for these errors here, and provide detailed results and analyses for them in Appendix A:

- Unlike in planar NAND flash memory, we do *not* find any evidence of *program errors* [4, 64, 81] in 3D NAND flash memory (Section A.1.1).
- P/E cycling error in 3D NAND flash memory follows a linear trend, which is similar to that in planar NAND flash memory using an older manufacturing process technology node (e.g., 20–24 nm) [14]. However, in sub-20 nm planar NAND flash memory, P/E cycling error exhibits a *power law* trend [64, 81] (Appendix A.1.2).
- 3D NAND flash memory experiences 40% *less program interference* than 20–24 nm planar NAND flash memory (Appendix A.1.3).
- 3D NAND flash memory experiences 96.7% *weaker read disturb* than 20–24 nm planar NAND flash memory. The impact of read disturb is low enough in 3D NAND flash memory that it does *not* require significant error mitigation (Appendix A.3.2).

Note that these differences are mainly due to the larger manufacturing process technology nodes currently used in 3D NAND flash memory, and thus are not the focus of this paper. In comparison, the new error characteristics that we focus on (layer-to-layer process variation, early retention loss, and retention interference) are caused by the architectural and circuit-level changes introduced in 3D NAND flash memory.

4.6 Summary

We summarize the key differences between 3D NAND and planar NAND flash memory, in terms of error characteristics and the expected trends for future 3D NAND flash memory devices, in Table 1. The first column of this table lists each attribute that we study. The second column shows the key difference in the observation that we find in 3D NAND flash memory versus planar NAND flash memory, for each attribute that we study. The third column shows the fundamental cause of each difference. The last column describes the expected trend of this difference in future 3D NAND flash memory devices. We provide the necessary characterizations and models that help us quantitatively understand these differences in Appendix A.1.2, A.1.3, A.2, A.3.1, A.3.2, and A.4.

5 3D NAND FLASH MEMORY ERROR MODELS

In the previous sections, we have established a basic understanding of the similarities and differences between 3D NAND and planar NAND flash memory in terms of error characteristics and reliability. In this section, we quantify these differences by developing analytical models of the process variation (Section 5.1) and retention loss (Section 5.2) phenomena in 3D NAND flash memory. These models are useful for at least two major purposes. First, the insights

Attribute	Observation in 3D NAND	Cause of Difference in 3D vs. Planar	Future Trend
Process Variation (Section 4.2, Appendix A.4, A.5)	Layer-to-layer process variation is significant	Vertical stacking of flash cells	Process variation will increase as we stack more cells vertically
Retention Loss (Sections 4.3, 4.4, Appendix A.2)	Early retention loss	Charge trap cell	Early retention loss will continue if charge trap cell is used
	Retention interference	Vertical stacking of flash cells	Retention interference will increase when smaller process technology node is used
P/E Cycling (Appendix A.1.2)	Distribution parameters change with P/E cycle count following a linear trend instead of a power-law trend	Larger manufacturing process technology node	P/E cycle trend will go back to power-law trend when smaller process technology node is used
Program Interference (Appendix A.1.3)	Wordline-to-wordline interference along the z-axis	Vertical stacking of flash cells	Will continue to exist in 3D NAND
	40% lower program interference	Larger manufacturing process technology node	Program interference will increase when smaller process technology node is used
V_{th} *Distribution* (Section 4.1)	ER and P1 states have no programming errors	Use of one-shot programming instead of two-step programming	Programming errors may start occurring if two-step programming is used
Read Disturb (Appendix A.3.2)	96.7% smaller read disturb effect	Larger manufacturing process technology node	Read disturb effect will increase when smaller process technology node is used

Table 1: Summary of error characteristics of 3D NAND and planar NAND flash memory.

obtained from using these models can motivate and enable us to develop new error mitigation mechanisms for 3D NAND flash memory. Second, the retention model and the model parameters are also useful for comparing the reliability of newer or older generations of planar NAND flash memory with our tested 3D NAND flash memory chips. We focus on developing these models using our existing characterization data from real 3D NAND flash memory chips (some of which was presented in Section 4). In Section 6, we discuss (1) how to efficiently *learn* the models for each chip *online* within the SSD controller by performing the characterization and model fitting online, and (2) how to use the online models to develop mechanisms that improve the lifetime of 3D NAND flash memory.

5.1 RBER Variation Model

Since the layer-to-layer variation in 3D NAND flash memory causes variation in RBER within a flash block, it is no longer sufficient to use a single RBER value to represent the reliability of *all* pages in that block. Instead, we model the variation in per-page RBER within a flash block as a gamma distribution (i.e., $gamma(x, a, s) = \frac{x^{a-1}e^{-\frac{x}{s}}}{\Gamma(a)s^a}$). In this model, x is the RBER; a is the shape parameter, which controls how the RBER distribution is skewed; and s is the scale parameter, which controls the width of the RBER distribution.

Figure 11 shows the probability density for per-page RBER within a block that has endured 10K P/E cycles. The bars show the measured per-page RBERs categorized into 50 bins, and the blue and orange curves are the fitted gamma distributions whose parameters are shown on the legend. The blue bars and curve represent the measured and fitted RBER distributions when the pages are read using the *variation-agnostic* V_{opt}. To find the variation-agnostic V_{opt}, we use techniques designed for planar NAND flash memory to learn a single optimal read reference voltage (V_{opt}) for each flash block, such that the chosen voltage minimizes the overall RBER *across the entire block* [64, 76]. The orange bars and curve represent the measured and fitted RBER distributions when the pages are read using the *variation-aware* V_{opt}, on a per-page basis. To find the variation-aware V_{opt}, we use techniques that are described in Section 6.1 to efficiently learn an optimal read reference voltage *for each page in the block*, such that we minimize the *per-page* RBER.

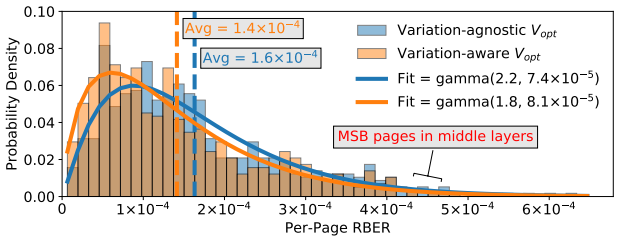


Figure 11: RBER distribution across pages within a flash block.

We make three observations from the figure. First, the gamma distribution fits well with the measured probability density function of RBER variation across layers: the Kullback-Leibler divergence error value [53] between the measured and fitted distributions is only 0.09. Second, the average RBER reduces from 1.6×10^{-4} to 1.4×10^{-4} when we use the variation-aware V_{opt}. Third, some flash pages have a much higher RBER than the average RBER (e.g., $> 4 \times 10^{-4}$) even when we use the variation-aware V_{opt}. This large gap between the worst-case RBER and the average RBER is caused by both layer-to-layer process variation and MSB–LSB RBER variation (see Figure 5 in Section 4.2). The pages that have the highest RBER are MSB pages that reside in the middle layers. This observation indicates that there is potential to significantly improve reliability by minimizing the RBER variation across flash pages (for which we describe a mechanism in Section 6.2).

5.2 Retention Loss Model

We construct a model to describe the early retention loss phenomenon and its impact on RBER ($\log(RBER)$) and threshold voltage (V) in 3D NAND flash memory, as a function of retention time (t) and the P/E cycle count (PEC): $\log(RBER) = A \cdot \log(t) + B$; $V = A \cdot \log(t) + B$. For both equations, $A = \alpha \cdot PEC + \beta$ and $B = \gamma \cdot PEC + \delta$, where α, β, γ, and δ are constants that change depending on which variable we are solving for. We use ordinary least squares method implemented in Statsmodel [88] to fit the model to our real characterization data described in Section 4.3. Recall that this data is collected from 72 flash pages belonging to 11 randomly-selected flash blocks. Following the experimental observations in Section 4.3 and in prior work [65, 69], we break down our model into two parts. The first part (A) models the retention loss at a certain P/E cycle count as a logarithmic function of retention time. The second part (B) models how the P/E cycle count changes the parameters of retention loss.

Table 2 shows all of the parameters we use to model the RBER and the threshold voltage as a function of the retention time (t) and the P/E cycle count (PEC). In this table, the first column shows the modeled variable for each row. The second to fifth columns show the parameters (i.e., α, β, γ, and δ) fitted to our model. Note that the model for the optimal V_a does not have α and β parameters because V_a is insensitive to retention time. The last column shows the adjusted coefficient of determination (*adjusted* R^2) of our model. We find that our model achieves high adjusted R^2 values for all variables except for σ_{ER} and V_a, meaning that our model explains >89% of the variation in the characterized data. The adjusted R^2 values are relatively small for σ_{ER} and V_a because these two variables do not change much with the retention time or the P/E cycle count. We conclude that our model is accurate and easy to compute (as it can be computed using simple linear regression). Thus, our model is suitable to use online in the SSD controller (for which we will describe a mechanism in Section 6.3).

6 3D NAND ERROR MITIGATION TECHNIQUES

Motivated by our new findings in Section 4, we aim to design new techniques that mitigate the three unique error effects (i.e., layer-to-layer process variation, early retention loss, and retention interference) in 3D NAND flash memory. We propose four error mitigation mechanisms. To mitigate layer-to-layer process variation, we propose LaVAR and LI-RAID. LaVAR learns our new RBER variation model (see Section 5.1) online in the SSD controller, and uses

Variable		Model Parameters for: $Variable = (\alpha \cdot PEC + \beta) \cdot \log(t) + \gamma \cdot PEC + \delta$				Adjusted R^2
		α	β	γ	δ	
MSB RBER	$\log(RBER_{MSB})$	5.49×10^{-6}	0.16	1.33×10^{-4}	-13.11	97.17%
LSB RBER	$\log(RBER_{LSB})$	7.92×10^{-6}	0.25	3.28×10^{-5}	-12.72	90.05%
ER Mean	μ_{ER}	1.01×10^{-4}	0.74	1.52×10^{-3}	-27.27	96.86%
P1 Mean	μ_{P1}	-1.94×10^{-5}	-0.40	3.51×10^{-4}	114.47	95.88%
P2 Mean	μ_{P2}	-4.71×10^{-5}	-0.70	3.23×10^{-4}	189.58	98.50%
P3 Mean	μ_{P3}	-7.37×10^{-5}	-1.20	5.75×10^{-4}	264.85	98.29%
ER Stdev	σ_{ER}	1.20×10^{-5}	-0.10	1.63×10^{-6}	17.01	56.33%
P1 Stdev	σ_{P1}	-1.34×10^{-6}	9.83×10^{-3}	7.55×10^{-5}	10.20	93.20%
P2 Stdev	σ_{P2}	-2.12×10^{-6}	9.85×10^{-3}	6.69×10^{-5}	10.65	89.02%
P3 Stdev	σ_{P3}	2.87×10^{-6}	1.40×10^{-2}	3.30×10^{-5}	10.83	93.00%
Optimal V_a	V_a	—	—	1.20×10^{-3}	60.52	71.20%
Optimal V_b	V_b	-3.72×10^{-5}	-0.57	4.20×10^{-4}	150.56	94.27%
Optimal V_c	V_c	-6.51×10^{-5}	-1.06	4.81×10^{-4}	227.24	97.72%

Table 2: Retention loss model for 3D NAND flash memory and its model parameters. *PEC* is P/E cycle lifetime, *t* is retention time.

this model to predict and apply an optimal read reference voltage that is fine-tuned to each layer (Section 6.1). LI-RAID is a new RAID scheme that reduces the RBER variation induced by layer-to-layer process variation in 3D NAND flash memory (Section 6.2). To mitigate retention loss in 3D NAND flash memory, we propose ReMAR, a new technique that tracks the retention time information within the SSD controller and uses our new retention loss model (see Section 5.2) to predict and apply the optimal read reference voltage that is fine-tuned to the retention time of the data (Section 6.3). To mitigate retention interference, we propose ReNAC, which is adapted from neighbor-cell assisted correction (NAC) [16], an existing technique originally designed to reduce program interference in planar NAND flash memory, to also account for retention interference in 3D NAND flash memory (Section 6.4).

6.1 LaVAR: Layer Variation Aware Reading

In planar NAND flash memory, existing techniques assume that the RBER is the *same* across all pages within a flash memory block, and, thus, a single V_{opt} value can be used for all pages in the block [6, 76]. This approach is called *variation-agnostic* V_{opt}. However, as our results in Section 4.2 show, this assumption no longer holds in 3D NAND flash memory due to layer-to-layer process variation, as each page in a block resides in a different layer. We aim to improve flash memory lifetime by mitigating layer-to-layer process variation and reducing the RBER. The key idea is to identify how much the read reference voltage must be offset by for *each* layer in a flash chip, to account for the layer-to-layer process variation, instead of using a single read reference voltage for the *entire* block irrespective of layers. When the SSD controller performs a read request, it accounts for (1) per-block variation in RBER, by predicting a variation-agnostic V_{opt} based on the P/E cycle count of the flash block; and (2) layer-to-layer variation, by adding the layer-specific offset to the variation-agnostic V_{opt} for the target block. This generates a *variation-aware* V_{opt} that the controller uses as the read reference voltage.

Mechanism. We devise a new mechanism called *Layer Variation Aware Reading* (LaVAR), which (1) learns the voltage offsets for each layer and records them in per-chip tables in the SSD controller, and (2) uses the variation-aware V_{opt} during a read operation by reading the appropriate voltage offset for the request from the per-chip table that corresponds to the layer of the request. LaVAR constructs a model of the optimal read reference voltage (V_{opt}) variation across different layers. Since there are only a limited number of layers, this model can be represented as a table (i.e., it is a non-parametric model) of the offset between the V_{opt} for each layer (*variation-aware* V_{opt}) and the overall V_{opt} for the entire flash block (*variation-agnostic* V_{opt}). Any previously-proposed model for V_{opt} [6, 64, 76] can be used to calculate the variation-agnostic V_{opt}. Since the layer-to-layer process variation is similar across blocks and is consistent across P/E cycle counts, the V_{opt} variation model can be learned *offline* for each chip through an extensive characterization of a single flash block. To do this, the SSD controller randomly picks a flash block and records the difference between the variation-aware V_{opt} and the variation-agnostic V_{opt}. LaVAR uses the existing read-retry functionality in modern NAND flash memory chips (see Section 4.1) to find the variation-aware V_{opt} online. The controller then computes and stores the average V_{opt} offset for each layer in a lookup table stored for each chip. Note that V_c variation does not need to be modeled, since V_c is unaffected by layer-to-layer process variation (see Figure 6 in Section 4.2).

When performing a read operation, the SSD controller simply looks up the V_{opt} offset that corresponds to the layer and the chip that contains the data being read, and adds the offset to the per-block V_{opt} predicted by existing techniques [6, 64, 76]. By using variation-aware V_{opt}, LaVAR enables the use of a more accurate

V_{opt} for 3D NAND flash memory than existing techniques, and thus reduces the RBER (see Figure 11 in Section 5.1).

Overhead. LaVAR can be implemented fully in the SSD controller firmware, and, thus, does not require any modification to the hardware. Assuming that the 3D NAND flash memory chip has N layers and that it takes 1 Byte to store each V_{opt} offset for each layer, the memory overhead of storing the lookup table for V_a and V_b in the SSD controller is $2N$ Bytes. The latency overhead of each read operation is negligible as LaVAR requires only a table lookup and an addition to obtain variation-aware V_{opt}, which take less than 100 ns. Since the lookup table is shared across *all* blocks in a chip, it needs to be learned only once, and it can be constructed gradually in the background. Thus, the performance overhead of LaVAR is negligible.

Evaluation. Figure 12 compares the RBER obtained by using LaVAR (*variation-aware* V_{opt}) [6, 64, 76] to that obtained by using an existing read reference voltage tuning technique (*variation-agnostic* V_{opt}) designed for planar NAND flash memory. We evaluate the average RBER obtained by each mechanism by simulating read operations using our characterization data in Section 4.2. Averaged across all P/E cycle counts, LaVAR reduces the RBER by 43.3%. The benefit comes from tuning the read reference voltage towards the variation-aware V_{opt} by an offset learned by our model. The RBER reduction becomes smaller as the P/E cycle count increases, because the overall RBER increases exponentially as the NAND flash memory wears out, decreasing the fraction of process variation errors. While the flash lifetime improvements produced by LaVAR might seem small (as we show in Section 6.5), (1) they are achieved with negligible overhead, and (2) the RBER reduction enabled by LaVAR throughout the flash memory lifetime reduces the average flash read latency [6]. As the number of layers within a 3D NAND flash memory chip grows (e.g., vendors are already bringing chips with 96 layers to the market [1]), we expect that layer-to-layer process variation will increase, which in turn will increase the magnitude of the lifetime benefits provided by LaVAR.

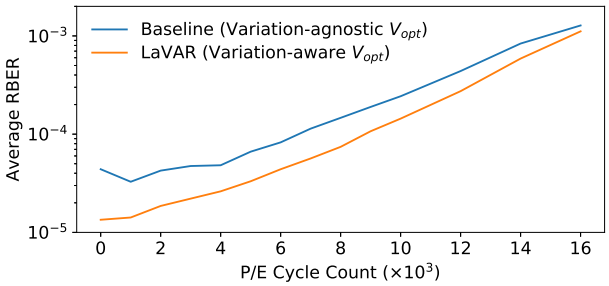


Figure 12: RBER reduction using LaVAR.

6.2 LI-RAID: Layer-Interleaved RAID

As we observe in Section 5.1, even after applying the variation-aware V_{opt}, the per-page RBER is distributed over a wide range according to a fitted gamma distribution due to layer-to-layer process variation and MSB–LSB RBER variation (see Figure 5 in Section 4.2). In enterprise SSDs, in addition to ECC, the Redundant Array of Independent Disks (RAID) [2, 83] error recovery technique is used across *multiple* flash chips to tolerate chip-to-chip process variation in error rates. RAID in modern SSDs typically combines one flash page from each flash chip into a logical unit called a *RAID group*, and uses one of the pages to store the parity information for the entire group. However, state-of-the-art RAID schemes do *not* consider layer-to-layer process variation and MSB–LSB RBER variation. These schemes group MSB or LSB pages in the *same* layer together in a RAID group. As a result, the reliability of the SSD is limited by the RBER of the weakest (i.e., the least reliable) RAID group that contains the MSB or LSB pages from the least reliable layer across all chips. We devise a new RAID scheme called *Layer-Interleaved RAID* (LI-RAID), which eliminates these low-reliability RAID groups by equalizing the RBER among different RAID groups. LI-RAID makes use of two key ideas: (1) group flash pages in less reliable layers with pages in more reliable layers, and (2) group MSB pages with LSB pages.

Mechanism. Instead of grouping pages in the same layer together in the same RAID group, we select pages from different chips and different layers and group them together, such that the low-reliability pages (either due to layer-to-layer process variation or MSB–LSB RBER variation) are *distributed* to different RAID groups. Thus, the new groups formed by LI-RAID have a more evenly-distributed RBER than the groups formed using traditional layer-unaware RAID schemes. We assume, without loss of generality, that there are m chips in the SSD, and each RAID group contains m pages, one from each chip. We also assume that each block contains n wordlines, and that the layer numbers of each wordline are in ascending order (e.g., the wordline in layer i has a lower wordline number than its neighboring wordline in layer $i+1$). Thus, LI-RAID groups together the MSB page of wordline 0, the LSB page of wordline $\frac{n}{m}$, the MSB page of wordline $2 \cdot \frac{n}{m}$, the LSB page ..., the MSB page of wordline $(m-2) \cdot \frac{n}{m}$, the LSB page of wordline $(m-1) \cdot \frac{n}{m}$. Figure 13 shows an example LI-RAID layout on an SSD with 4 chips and with 4 wordlines within each flash block. Flash pages in the same RAID group are highlighted in the same color. In this way, LI-RAID distributes the less reliable pages within each chip across different RAID groups, thereby avoiding the formation of significantly less reliable RAID groups that bottleneck SSD reliability.

Note that, since the order of RAID group number is different in each flash chip, the LI-RAID layout may potentially violate the program sequence recommended by flash vendors, where wordlines within each flash block must be programmed *in order* to minimize harmful program interference [9, 15, 16, 77]. For example, in Chip 2 in Figure 13, Wordline 3 (Groups 2 and 3) is programmed *after* Wordline 2 (Groups 0 and 1). In Chip 2, we leave Wordline 1 *blank* (marked as"Blank" in Figure 13). Otherwise, Wordline 1 would cause program interference to the data in Wordline 2, which already experiences program interference when Wordline 3 is programmed, significantly increasing the error rate of Wordline 1 [15, 16] (see Appendix A.1.3). By laying out the data in the proposed manner, LI-RAID provides the same reliability guarantee as the recommended program sequence, by guaranteeing that any data stored in a flash page experiences program interference from *at most* one neighboring wordline.

Wordline #	Layer #	Page	Chip 0	Chip 1	Chip 2	Chip 3
0	0	MSB	Group 0	Blank	Group 4	Group 3
0	0	LSB	Group 1	Blank	Group 5	Group 2
1	1	MSB	Group 2	Group 1	Blank	Group 5
1	1	LSB	Group 3	Group 0	Blank	Group 4
2	2	MSB	Group 4	Group 3	Group 0	Blank
2	2	LSB	Group 5	Group 2	Group 1	Blank
3	3	MSB	Blank	Group 5	Group 2	Group 1
3	3	LSB	Blank	Group 4	Group 3	Group 0

Figure 13: LI-RAID layout example for an SSD with 4 chips and with 4 wordlines in each flash block.

Overhead. The grouping of flash pages by LI-RAID is implemented entirely in the SSD controller firmware. This requires the firmware to be aware of the physical-page-to-layer mapping. The flash pages left blank in LI-RAID incur a small additional storage overhead compared to a conventional RAID scheme. Only *one* wordline (i.e., two pages in MLC NAND flash memory) within a flash block is left blank, to mitigate the impact of program interference on Groups 0 and 1. Without this blank wordline, the data in Groups 0 and 1 would be the only data to experience program interference twice: once when Groups 2 and 3 are programmed, and once when the last two groups are programmed. In modern NAND flash memory, each flash block typically contains at least 256 flash pages. Thus, the additional storage overhead for the blank pages is less than 0.8%. LI-RAID does not incur additional computational overhead because it computes parity in the same way as a conventional RAID scheme, and only *reorganizes* the RAID groups differently. Because we do *not* change the data layout across flash blocks, the flash translation layer (FTL) and the garbage collection (GC) algorithms remain the same as in a conventional RAID scheme.

Evaluation. Figure 14 plots the *worst-case RBER* (i.e., the highest per-page RBER within a flash block) when we use different error mitigation techniques at 10,000 P/E cycles. Recall that the per-page RBER within a flash block follows a gamma distribution (see Figure 11 in Section 5.1). Thus, several least-reliable flash pages within a block may become unusable (i.e., their RBER exceeds the ECC correction capability) before the *overall* RBER of the flash chip exceeds the ECC correction capability. We use the worst-case RBER to represent the reliability of these least-reliable flash pages. In this figure, the baseline uses the per-block variation-agnostic optimal read reference voltage (i.e., variation-agnostic V_{opt}), achieving a worst-case RBER of $4.8 \cdot 10^{-4}$. When we use the variation-aware V_{opt} proposed in Section 6.1, the worst-case RBER is reduced by 9.6% over the baseline, to $4.3 \cdot 10^{-4}$. LI-RAID reduces the worst-case RBER by 66.9% over the baseline, to only $1.6 \cdot 10^{-4}$. Thus, by grouping flash pages on less reliable layers with pages on more reliable layers, and by grouping MSB pages with LSB pages, LI-RAID reduces the probability of unusable pages within a block, thereby reducing the number of retired flash blocks due to ECC failures.

Note that LaVAR and LI-RAID do *not* rely on whether the RBER variation is consistent across all chips. LaVAR learns a different lookup table for each chip. So, even if there is some chip-to-chip process variation that is present, our models are effective at dynamically capturing the behavior of *any* NAND flash memory chips. Conventional RAID tolerates only chip-to-chip process variation. LI-RAID improves flash reliability over conventional RAID by eliminating the strong correlation between RBER and layer number, which we show in Figure 5. We conclude that both LaVAR and LI-RAID are effective at reducing the impact of layer-to-layer variation on the RBER.

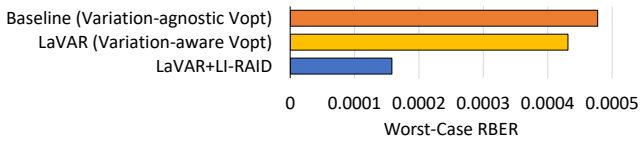


Figure 14: Effect of LaVAR and LI-RAID on worst-case RBER at 10,000 P/E cycles.

6.3 ReMAR: Retention Model Aware Reading

As we show in Section 4.3, due to early retention loss, retention errors increase much faster after programming a page in 3D NAND flash memory than they do in planar NAND flash memory. Thus, mitigating retention errors has become more important in 3D NAND than in planar NAND flash memory, as the errors have a greater impact on SSD reliability. However, as we show in our model in Section 5.2, the RBER impact of early retention loss is proportional to the logarithm of retention time. This means that a large majority of the retention errors and threshold voltage shifts happen *shortly* after programming. As a result, traditional retention error mitigation techniques developed for planar NAND flash memory, which are optimized for much larger retention times, may become less effective on 3D NAND flash memory. For example, Flash Correct-and-Refresh (FCR) [7, 8], a mechanism that remaps all data periodically, allows planar NAND to tolerate 50× more P/E cycles with a 3-day refresh period. However, according to our evaluations, the P/E cycle lifetime improvement of FCR reduces to *only* 2.7× for 3D NAND flash memory due to the early retention loss phenomenon. This motivates us to explore new ways to mitigate retention errors in 3D NAND flash memory.

Mechanism. We propose a new mechanism called *Retention Model Aware Reading* (ReMAR), whose key idea is to accurately track the retention time of the data and apply the optimal read

reference voltage predicted by our model in Section 5.2. First, ReMAR constructs the same linear models proposed in Section 5.2 *online* to accurately predict the optimal V_a, V_b, and V_c. Similar to the distribution parameter model used in Section 5.2, we model the optimal V_b and V_c as: $V = (\alpha \cdot PEC + \beta) \cdot \log(t) + \gamma \cdot PEC + \delta$. We model the optimal V_a as: $V_a = \gamma \cdot PEC + \delta$, since V_a is *not* affected by retention time (as we show empirically in Section 4.3). To construct this model *online*, the controller randomly selects a flash block and records the optimal read reference voltage of the block (which the controller learns by sweeping the read reference voltages, as done in prior work [6]), along with the block's P/E cycle count (PEC) and retention time (t). Over time, these data samples would cover a range of P/E cycle counts and retention times.[8] Note that as the P/E cycle count of the SSD increases, the accuracy of the model increases, because more data samples are collected. Once this online model is constructed, it is used in the controller to predict the optimal read reference voltage to be used for each read operation. To do this, the SSD controller stores the P/E cycle count and the program time of each block as metadata. During each read operation, the controller computes the retention time for each read by subtracting the program time from the read time. Using the recorded P/E cycle count and the computed retention time of the data, ReMAR applies the online model to predict V_a, V_b, and V_c. By accurately predicting and applying the optimal read reference voltages, ReMAR increases the accuracy of read operations and thereby decreases the raw bit error rate.

Overhead. Like LaVAR, ReMAR is implemented fully in the SSD controller firmware, and does *not* require any modifications to the hardware. Assuming that the flash block size is 5 MB, and that ReMAR stores the program time in the UNIX Epoch time format [67], which takes up 4 B, the memory and storage overhead of ReMAR is 800KB for a 1TB SSD. The performance overhead of each read operation is small, as ReMAR needs only a few dozen CPU cycles (on the order of 100 ns in total) in the SSD controller to compute V_{opt}, which is negligible compared to flash read latency (on the order of 10 µs). The performance overhead of learning the model can be hidden by (1) performing learning in the background and (2) deprioritizing the requests issued for characterization purposes.

The controller uses the UNIX Epoch time format [67] for program and read times, such that the recorded time is valid after reboot. To do this, the controller needs a real-time clock to keep track of the current time. Without a power source on the SSD, the controller needs a special command to synchronize the current time with the host when it boots up. The program time of each block is stored in the memory of the controller, along with other metadata that already exists such as the logical address map and the P/E cycle count of each block.

Evaluation. Figure 15 compares the RBER achieved by ReMAR to that of the state-of-the-art read reference voltage tuning technique [64] designed for planar NAND flash memory (*Baseline*). The results are based on the characterization data in Section 4.3. We assume that the average retention time of the data is 24 days. The *Baseline* technique is unaware of the retention time. Thus, *Baseline* uses a *retention-agnostic* V_{opt} based on only the P/E cycle count of the flash page. ReMAR uses a *retention-aware* V_{opt} based on both the P/E cycle count and the retention time of the flash page. On average across all P/E cycle counts, ReMAR reduces the RBER by 51.9%. As the P/E cycle count increases, the benefit of ReMAR (i.e., the RBER improvement of ReMAR over *Baseline*) also increases. We conclude that, by accurately tracking retention time, and by using our retention loss model, ReMAR accurately adapts the read reference voltage to the threshold voltage shifts that occur due to retention loss, and hence it effectively reduces the RBER.

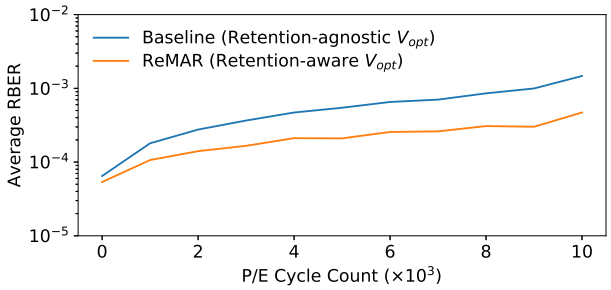


Figure 15: RBER reduction using ReMAR.

6.4 ReNAC: Retention Interference Aware Neighbor-Cell Assisted Correction

As we observe in Section 4.4, due to retention interference, the amount of threshold voltage shift of a victim cell during a certain amount of retention time is affected by the value stored in a vertically-adjacent neighbor cell. This phenomenon presents a similar data dependency as that induced by program interference, where the amount of the threshold voltage shift of a victim cell during programming operation also depends on the value stored in the directly-neighboring cells [15, 16]. To mitigate program interference errors, prior work proposes neighbor-cell assisted correction (NAC) [16]. The goal of NAC is to reduce the raw bit error rate by reading each cell at the read reference voltage optimized for the amount of program interference induced by its directly-neighboring cells. To achieve this goal, after error correction fails on a flash page, NAC reads the data stored in the *neighboring* wordline and re-reads the failed page using a set of read reference voltage values that are adjusted based on the data values stored in the directly-neighboring cells [16]. However, this mechanism does *not* account for *retention interference* induced by the neighboring cells, which is new in 3D NAND flash memory. We adapt NAC for 3D NAND flash memory to account for the new retention interference phenomenon, and call this adapted mechanism *Retention Interference Aware Neighbor-Cell Assisted Correction* (ReNAC).

Mechanism. The key idea of ReNAC is to use the data stored in a vertically-adjacent neighbor cell to predict the amount of retention interference on a victim cell. Using similar techniques from Section 5.2, ReNAC first develops an online model of retention interference as a function of the retention time and the neighbor cell's state. The SSD controller obtains the retention time of each block using a mechanism similar to ReMAR, and computes and applies the neighbor-cell-dependent read offset at that retention time from

[8] The SSD controller can also perform additional characterization if a certain data range is missing.

the model. For ReNAC, we are currently unable to show any meaningful improvements in flash lifetime for the current generation of 3D NAND flash memory, because retention interference shifts the threshold voltage by only less than two voltage steps (Figure 10), which is much smaller than the voltage changes due to process variation (Figure 6) and early retention loss (Figure 8). However, we expect that retention interference will increase in future 3D NAND flash memory devices due to decreasing cell sizes and decreasing distances between neighboring cells (Table 1), which, in turn, will likely increase the benefit of using ReNAC. We also expect ReNAC to have a relatively larger benefit in 3D NAND flash memory chips that use *triple-level cell* (TLC) or *quadruple-level cell* (QLC) technologies. A TLC or QLC NAND flash memory chip stores more bits in a cell than an MLC NAND flash memory chip, by splitting up the same voltage range into a greater number of states (eight for TLC and sixteen for QLC). Doing so reduces the voltage margin between neighboring threshold voltage distributions. Therefore, shifting the read reference voltage by two voltage steps may affect more cells in TLC and QLC 3D NAND flash memory than in MLC 3D NAND flash memory, and, thus, ReNAC can reduce a greater number of raw bit errors in future TLC or QLC NAND flash memory. We leave a quantitative evaluation of ReNAC on future 3D NAND flash memory chips to future work.

6.5 Putting It All Together: Effect on System Reliability and Performance

The mechanisms we propose in this section can be combined together to achieve significant reductions in average and worst-case RBER. For a consumer-class 3D NAND flash memory device, these reductions improve *flash memory lifetime*, i.e., the device can tolerate more P/E cycles before failing. For an enterprise-class device which is expected to be replaced after a fixed amount of time, these reductions improve the sustainable workload write intensity or reduce the ECC storage overhead. We evaluate these potential effects of our mechanisms on storage system reliability and performance.

Flash Lifetime (or Performance) Improvement. In Figure 16, we compare and contrast the reliability (i.e., the RBER) of five example SSDs: (1) *Baseline*, an SSD that uses a fixed, default read reference voltage and employs a conventional RAID scheme; (2) *State-of-the-art*, an SSD that uses the optimal read reference voltage predicted by existing mechanisms designed for planar NAND flash memory [6, 64, 76, 81] and employs a conventional RAID scheme; (3) *LaVAR*, an SSD that uses the optimal read reference voltage for each layer predicted by LaVAR in addition to *State-of-the-art*; (4) *LaVAR+LI-RAID*, an SSD that uses the LI-RAID scheme in addition to *LaVAR*; and (5) *This Work* (LaVAR + LI-RAID + ReMAR), an SSD that uses the optimal read reference voltage predicted by LaVAR and ReMAR, and also employs the LI-RAID scheme. In this figure, we plot the *worst-case RBER* (i.e., the highest per-page RBER within a flash block) instead of the average RBER, because the worst-case RBER *limits* the flash memory lifetime. Because RBER increases with P/E cycle count, if the worst-case RAID group has a high enough worst-case RBER, NAND flash memory can no longer guarantee reliable operation.

Assuming that the ECC deployed on the SSD can correct errors up to an RBER of $3 \cdot 10^{-3}$ [6, 9] (i.e., the *ECC limit*, shown as a purple dashed line in Figure 16), we can calculate the lifetime of each SSD we evaluate.[9] In our evaluations, the flash memory lifetime ends when the worst-case RBER exceeds the ECC limit. We find that *State-of-the-art*, *LaVAR*, *LaVAR+LI-RAID*, and *This Work* improve flash memory lifetime by 23.8%, 25.3%, 57.2%, and 85.0%, respectively, over the *Baseline*. When the SSD is used in a server, which has a fixed device lifetime, the server has to throttle the write frequency to a certain *drive writes per day* (DWPD) to ensure that the SSD can operate reliably during the fixed lifetime. In this case, our combined mechanisms (*This Work*) increase the maximum write frequency (i.e., the maximum DWPD) of the SSDs in a server by 85.0%. Thus, our mechanisms either improve lifetime or improve performance under a fixed lifetime.

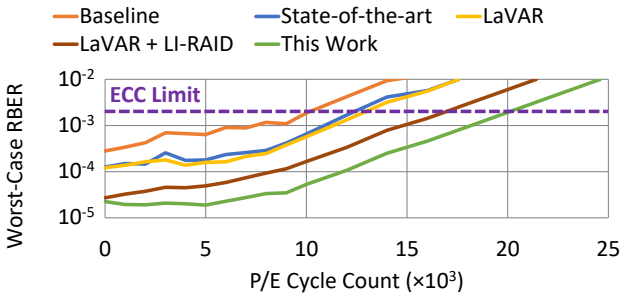


Figure 16: Effect of LaVAR, LI-RAID, and ReMAR on worst-case RBER experienced by any flash block.

ECC Storage Overhead Reduction. In modern SSDs, the storage overhead for error correction increases in each generation to better tolerate the degraded flash reliability due to aggressive scaling. For example, to tolerate an RBER of up to $3 \cdot 10^{-3}$ for the *Baseline* SSD at the end of its lifetime, a modern BCH code [36] requires 12.8% storage overhead for the redundant ECC bits [25] (i.e., *ECC redundancy*). By deploying all of our proposed error mitigation techniques in an enterprise-class SSD, the RBER at the end of the fixed flash memory lifetime is significantly lower compared to *Baseline*. Thus, we can redesign the ECC deployed in the SSD to tolerate only up to the reduced RBER, which requires fewer ECC bits and, thus, lower ECC redundancy than the ECC required for the *Baseline*. Assuming all five of the evaluated SSDs achieve the same lifetime, and the same reliability (i.e., uncorrectable error rate) at the end of their lifetime, *State-of-the-art*, *LaVAR*, *LaVAR+LI-RAID*, and *This Work* reduce ECC redundancy by 42.2%, 45.3%, 68.8%, and 78.9%, respectively, over *Baseline*. We leave the evaluation of the performance improvements due to a weaker ECC requirement [22, 59] for future work.

We conclude that by combining LaVAR, LI-RAID, and ReMAR, we can (1) achieve significant improvements in the lifetime of 3D NAND flash memory, (2) enable higher write intensity in workloads

[9] Note that we are *unable* to *directly* measure the flash lifetime improvements on real devices, because manufacturers do *not* provide us with the ability to modify the SSD firmware directly, which prevents us from evaluating our techniques on the real devices themselves. Unfortunately, we also do *not* have the resources to measure the lifetime of a large number of real flash chips by emulating the behavior of our mechanisms, as this would require many additional months to years of effort. Instead, we follow the precedent of prior work to evaluate the flash memory lifetime based on real RBER characterization data we obtain from the testing of real flash memory devices.

within a given lifetime requirement, or (3) keep the lifetime constant but greatly reduce the storage cost of reliability in 3D NAND flash memory.

7 RELATED WORK

To our knowledge, this paper is the first in open literature to (1) show the differences between the error characteristics of 3D NAND flash memory and that of planar NAND flash memory through extensive characterization using real 3D NAND flash memory chips, (2) develop models of layer-to-layer process variation and early retention loss for 3D NAND flash memory, and (3) propose and show the benefits of four new mechanisms based on the new error characteristics of 3D NAND flash memory. Due to the importance of NAND flash memory reliability in storage systems, there is a large body of related work. We treat this related work in five different categories.

3D NAND Flash Memory Error Characterization. Two recent works compare the retention loss phenomenon between 3D NAND and planar NAND flash memory [65, 70] through real device characterization, and report findings similar to our work regarding the early retention loss phenomenon. Two other recent works use a methodology similar to ours to characterize 3D NAND devices based on *different* 3D NAND flash memory cell technologies (i.e., 3D floating-gate cell and 3D vertical gate cell) [38, 94, 95], which are *less common* than the 3D charge trap NAND flash memory cell technology that we test in this paper. Other recent works [23, 31, 78, 80, 92] report several differences of 3D NAND flash memory from planar NAND flash memory. These differences include (1) smaller program variation at high P/E cycle counts [80], (2) smaller program interference [80], (3) layer-to-layer process variation [92], (4) early retention loss [23, 31, 78], and (5) retention interference [23]. While prior works have reported on the existence of these errors, *none* of them provide a comprehensive characterization of *all* of the different errors using the *same* chips. Only one of these prior works [23] provides a detailed analysis based on *circuit-level* measurements and characterizations, and does so only for early retention loss and retention interference. Other works provide only a high-level *summary* of real device characterization [80] or do *not* provide any real device characterization results at all [31, 78, 92]. Our work performs an extensive detailed analysis of *all* known sources of error in 3D NAND flash memory chips, which allows us to understand the relative impact of each error source on the same chip. We report the first set of extensive results on three error characteristics that are new in 3D NAND flash memory: layer-to-layer process variation, early retention loss, and retention interference.

Planar NAND Flash Memory Error Characterization. A large body of prior work studies all types of error sources on planar NAND flash memory, including P/E cycling errors [9, 14, 64, 81], programming errors [4, 64, 81], cell-to-cell program interference errors [15, 16], retention errors [6, 7, 9, 28], and read disturb errors [5, 9]. These works characterize how the raw bit error rate and threshold voltage change due to various types of error sources. A detailed survey of such prior works on planar NAND flash memory can be found in our recent survey articles [9, 11]. Our paper experimentally studies all of these error mechanisms in the new 3D NAND flash memory context, and compares 3D NAND flash memory error characteristics with results in these prior works to show the differences between 3D NAND and planar NAND flash memory. Prior work demonstrates the early retention loss phenomenon in planar NAND flash memory based on charge trap transistors [21], which is similar to, but not as severe as, the early retention loss phenomenon in 3D NAND flash memory. We investigate retention interference and process variation related errors, in addition to these other error types discovered before in planar NAND flash memory.

Planar NAND Error Modeling and Mitigation. Based on characterization results, prior work proposes models for planar NAND flash memory threshold voltage distribution, and models for estimating the effect of P/E cycling on the threshold voltage distribution [14, 64, 81]. Our work uses a simpler threshold voltage distribution model, since more complex models are designed to handle programming errors in planar NAND flash memory that do not exist in the 3D NAND flash memory chips that we test. We develop a unified model of retention loss and wearout for the RBER, threshold voltage distribution, and V_{opt} in 3D NAND flash memory. There is a large body of prior work that proposes mechanisms to mitigate planar NAND flash memory errors [4–9, 11, 15, 16, 32, 33, 37, 40, 41, 60, 63, 64, 74, 75, 93, 98]. In Section 6, we have already compared our mechanisms to several of these techniques that are state-of-the-art, and have shown that prior techniques developed for planar NAND flash memory are less effective in 3D NAND flash memory than our techniques due to the new error characteristics of 3D NAND flash memory.

3D NAND Flash Memory Error Mitigation. Prior work proposes circuit-level and system-level techniques to tolerate layer-to-layer process variation in 3D NAND flash memory. Two recent works propose to use different read reference voltages for different layers [38, 96], which is similar to the LaVAR technique that we propose in Section 6.1. Unlike our work, these prior works do not (1) design a detailed mechanism like LaVAR to learn and use the V_{opt} in a lookup table, or (2) evaluate their techniques using real characterization data. Wang et al. propose to apply different read reference voltages for less-reliable pages storing critical metadata [92]. As we have shown in Section 6.1, while these prior techniques improve average RBER, they do *not* significantly reduce worst-case RBER, which limits the flash memory lifetime. In this work, we propose a series of mitigation techniques that not only significantly reduce the average and worst-case RBER but also tolerate other new error characteristics we find in 3D NAND flash memory, such as early retention loss and retention interference.

Large-Scale SSD Error Characterization. Prior work performs large-scale studies of errors found in flash memories deployed in data centers [68, 72, 87]. Since the operating system is unaware of the raw bit errors in the NAND flash memory devices, these studies can only use drive-level statistics provided by the SSD controller, such as overall RBER and uncorrectable error rate, average P/E cycle count, and a coarse estimation of retention time and read disturb counts. In contrast, in our studies, we have complete access to the physical location, P/E cycle count, retention time, and read disturb count of each read/write operation, and thus can provide deeper insights and controlled experimental results compared to large-scale studies, which have to be correlational in nature.

DRAM Error Characterization. Like a flash memory cell, a DRAM cell stores charge to represent a piece of data. Hence, DRAM has many error characteristics that are similar to NAND flash memory. For example, charge leaks from a DRAM cell over time, at a speed much faster than that for NAND flash memory (i.e., on the order of *milliseconds* to *seconds* in DRAM [61, 62]), leading to *data retention errors*. This phenomenon in DRAM is analogous to the retention loss phenomenon in NAND flash memory (see Section 4.3 and Appendix A.2), and its effect has been studied through extensive experimental characterization of DRAM chips [34, 35, 44, 46–49, 51, 56, 61, 82, 85]. Similar to the retention interference phenomenon found in 3D NAND flash memory (see Section 4.4), DRAM exhibits data-dependent retention behavior, or *data pattern dependence* (DPD) [61], where the retention time of a DRAM cell is dependent on the values written to *nearby* DRAM cells [46–49, 61, 82]. Conceptually similar to the read disturb errors found in NAND flash memory (see Appendix A.3.2), commodity DRAM chips that are sold and used in the field today exhibit read disturb errors [52], also called *RowHammer*-induced errors [71]. These errors are affected by *process variation*, which we comprehensively examine in 3D NAND flash memory (see Section 4.2 and Appendix A.4). *Process variation* in DRAM is shown to also affect access latency, retention time, and power consumption [17–20, 30, 34, 35, 43, 44, 46–49, 51, 54–56, 61, 62, 66, 82, 85].

8 CONCLUSION

We develop a new understanding of three new error characteristics in 3D NAND flash memory through rigorous experimental characterization of real, state-of-the-art 3D NAND flash memory chips: layer-to-layer process variation, early retention loss, and retention interference. We analyze and show that these new error characteristics are fundamentally caused by changes introduced in the 3D NAND flash memory architecture compared to the planar NAND flash memory architecture. To handle these three new error characteristics in 3D NAND flash memory, we develop new analytical models for layer-to-layer process variation and early retention loss in 3D NAND flash memory. Our models can accurately predict/estimate the optimal read reference voltage and the raw bit error rate based on the retention time and the layer number of each flash memory page. We propose four new error mitigation techniques that utilize our new models to improve the reliability of 3D NAND flash memory. Our evaluations show that our newly-proposed techniques successfully mitigate the new error patterns that we discover in 3D NAND flash memory. We hope that the rigorous and comprehensive error characterization and analyses performed in this work motivate future rigorous studies on 3D NAND flash memory reliability, and that they inspire new error mitigation mechanisms that cater to the new error characteristics found in 3D NAND flash memory.

ACKNOWLEDGMENTS

We thank our shepherd, Benny Van Houdt, the anonymous reviewers, and SAFARI members for their feedback. This work is partially supported by grants from Huawei and Seagate, and gifts from Huawei, Intel, Microsoft, and Samsung.

REFERENCES

[1] AnandTech, "Western Digital Announce BiCS4 3D NAND: 96 Layers, TLC & QLC, Up to 1 Tb per Chip," https://www.anandtech.com/show/11585/western-digital-announce-bics4-96-layer-nand, 2017.
[2] M. Balakrishnan, A. Kadav, V. Prabhakaran, and D. Malkhi, "Differential RAID: Rethinking RAID for SSD Reliability," *TOS*, 2010.
[3] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, "Introduction to Flash Memory," *Proc. IEEE*, 2003.
[4] Y. Cai, S. Ghose, Y. Luo, K. Mai, O. Mutlu, and E. F. Haratsch, "Vulnerabilities in MLC NAND Flash Memory Programming: Experimental Analysis, Exploits, and Mitigation Techniques," in *HPCA*, 2017.
[5] Y. Cai, Y. Luo, S. Ghose, E. F. Haratsch, K. Mai, and O. Mutlu, "Read Disturb Errors in MLC NAND Flash Memory: Characterization and Mitigation," in *DSN*, 2015.
[6] Y. Cai, Y. Luo, E. F. Haratsch, K. Mai, and O. Mutlu, "Data Retention in MLC NAND Flash Memory: Characterization, Optimization, and Recovery," in *HPCA*, 2015.
[7] Y. Cai, G. Yalcin, O. Mutlu, E. F. Haratsch, A. Cristal, O. Unsal, and K. Mai, "Flash Correct and Refresh: Retention Aware Management for Increased Lifetime," in *ICCD*, 2012.
[8] Y. Cai, G. Yalcin, O. Mutlu, E. F. Haratsch, A. Cristal, O. Unsal, and K. Mai, "Error Analysis and Retention-Aware Error Management for NAND Flash Memory," *Intel Technology J.*, 2013.
[9] Y. Cai, S. Ghose, E. F. Haratsch, Y. Luo, and O. Mutlu, "Error Characterization, Mitigation, and Recovery in Flash-Memory-Based Solid-State Drives," *Proc. IEEE*, 2017.
[10] Y. Cai, S. Ghose, E. F. Haratsch, Y. Luo, and O. Mutlu, "Errors in Flash-Memory-Based Solid-State Drives: Analysis, Mitigation, and Recovery," arxiv:1711.11427 [cs.AR], 2017.
[11] Y. Cai, S. Ghose, E. F. Haratsch, Y. Luo, and O. Mutlu, "Reliability Issues in Flash-Memory-Based Solid-State Drives: Experimental Analysis, Mitigation, Recovery," in *Inside Solid State Drives (SSDs)*, 2nd ed. Springer Nature, 2018.
[12] Y. Cai, E. F. Haratsch, M. McCartney, and K. Mai, "FPGA-Based Solid-State Drive Prototyping Platform," in *FCCM*, 2011.
[13] Y. Cai, E. F. Haratsch, O. Mutlu, and K. Mai, "Error Patterns in MLC NAND Flash Memory: Measurement, Characterization, and Analysis," in *DATE*, 2012.
[14] Y. Cai, E. F. Haratsch, O. Mutlu, and K. Mai, "Threshold Voltage Distribution in MLC NAND Flash Memory: Characterization, Analysis, and Modeling," in *DATE*, 2013.
[15] Y. Cai, O. Mutlu, E. F. Haratsch, and K. Mai, "Program Interference in MLC NAND Flash Memory: Characterization, Modeling, and Mitigation," in *ICCD*, 2013.
[16] Y. Cai, G. Yalcin, O. Mutlu, E. F. Haratsch, O. Unsal, A. Cristal, and K. Mai, "Neighbor-Cell Assisted Error Correction for MLC NAND Flash Memories," in *SIGMETRICS*, 2014.
[17] K. Chandrasekar, S. Goossens, C. Weis, M. Koedam, B. Akesson, N. Wehn, and K. Goossens, "Exploiting Expendable Process-Margins in DRAMs for Run-Time Performance Optimization," in *DATE*, 2014.
[18] K. K. Chang, "Understanding and Improving the Latency of DRAM-Based Memory Systems," Ph.D. dissertation, Carnegie Mellon Univ., 2017.
[19] K. K. Chang, A. Kashyap, H. Hassan, S. Ghose, K. Hsieh, D. Lee, T. Li, G. Pekhimenko, S. Khan, and O. Mutlu, "Understanding Latency Variation in Modern DRAM Chips: Experimental Characterization, Analysis, and Optimization," in *SIGMETRICS*, 2016.
[20] K. K. Chang, A. G. Yaglikci, A. Agrawal, N. Chatterjee, S. Ghose, A. Kashyap, H. Hassan, D. Lee, M. O'Connor, and O. Mutlu, "Understanding Reduced-Voltage Operation in Modern DRAM Devices: Experimental Characterization, Analysis, and Mechanisms," in *SIGMETRICS*, 2017.
[21] C.-P. Chen, H.-T. Lue, C.-C. Hsieh, K.-P. Chang, K.-Y. Hsieh, and C.-Y. Lu, "Study of Fast Initial Charge Loss and Its Impact on the Programmed States Vt Distribution of Charge-Trapping NAND Flash," in *IEDM*, 2010.
[22] C.-L. Chen, "High-Speed Decoding of BCH Codes (Corresp.)," *TIT*, 1981.
[23] B. Choi, S. H. Jang, J. Yoon, J. Lee, M. Jeon, Y. Lee, J. Han, J. Lee, D. M. Kim, D. H. Kim *et al.*, "Comprehensive Evaluation of Early Retention (Fast Charge Loss Within a Few Seconds) Characteristics in Tube-Type 3-D NAND Flash Memory," in *VLSIT*, 2016.
[24] C. M. Compagnoni, M. Ghidotti, A. L. Lacaita, A. S. Spinelli, and A. Visconti, "Random Telegraph Noise Effect on the Programmed Threshold-Voltage Distribution of Flash Memories," *IEEE EDL*, 2009.
[25] E. Deal, "Trends in NAND Flash Memory Error Correction," *Cyclic Design*, 2009.
[26] R. Degraeve, F. Schuler, B. Kaczer, M. Lorenzini, D. Wellekens, P. Hendrickx, M. van Duuren, G. J. M. Dormans, J. van Houdt, L. Haspeslagh, G. Groeseneken, and G. Tempel, "Analytical Percolation Model for Predicting Anomalous Charge Loss in Flash Memories," *TED*, 2004.
[27] R. H. Fowler and L. Nordheim, "Electron Emission in Intense Electric Fields," in *Proc. Royal Society of London A*, 1928.
[28] A. Fukami, S. Ghose, Y. Luo, Y. Cai, and O. Mutlu, "Improving the Reliability of Chip-Off Forensic Analysis of NAND Flash Memory Devices," *Digital Investigation*, 2017.

[29] A. Ghetti, C. M. Compagnoni, A. S. Spinelli, and A. Visconti, "Comprehensive Analysis of Random Telegraph Noise Instability and Its Scaling in Deca-Nanometer Flash Memories," *IEEE TED*, 2009.

[30] S. Ghose, A. G. Yağlıkçı, R. Gupta, D. Lee, K. Kudrolli, W. X. Liu, H. Hassan, K. K. Chang, N. Chatterjee, A. Agrawal, M. O'Connor, and O. Mutlu, "What Your DRAM Power Models Are Not Telling You: Lessons from a Detailed Experimental Study," in *SIGMETRICS*, 2018.

[31] A. Grossi, C. Zambelli, and P. Olivo, "Reliability of 3D NAND Flash Memories," in *3D Flash Memories*. Springer, 2016.

[32] K. Ha, J. Jeong, and J. Kim, "A Read-Disturb Management Technique for High-Density NAND Flash Memory," in *APSys*, 2013.

[33] K. Ha, J. Jeong, and J. Kim, "An Integrated Approach for Managing Read Disturbs in High-Density NAND Flash Memory," *TCAD*, 2016.

[34] T. Hamamoto, S. Sugiura, and S. Sawada, "On the Retention Time Distribution of Dynamic Random Access Memory (DRAM)," *IEEE TED*, 1998.

[35] H. Hassan, N. Vijaykumar, S. Khan, S. Ghose, K. Chang, G. Pekhimenko, D. Lee, O. Ergin, and O. Mutlu, "SoftMC: A Flexible and Practical Open-Source Infrastructure for Enabling Experimental DRAM Studies," in *HPCA*, 2017.

[36] A. Hocquenghem, "Codes Correcteurs d'Erreurs," *Chiffres*, 1959.

[37] J. Huang, A. Badam, L. Caulfield, S. Nath, S. Sengupta, B. Sharma, and M. K. Qureshi, "FlashBlox: Achieving Both Performance Isolation and Uniform Lifetime for Virtualized SSDs," in *FAST*, 2017.

[38] C.-H. Hung, M.-F. Chang, Y.-S. Yang, Y.-J. Kuo, T.-N. Lai, S.-J. Shen, J.-Y. Hsu, S.-N. Hung, H.-T. Lue, Y.-H. Shih *et al.*, "Layer-Aware Program-and-Read Schemes for 3D Stackable Vertical-Gate BE-SONOS NAND Flash Against Cross-Layer Process Variations," *JSSC*, 2015.

[39] J. Im, W. Jeong, D. Kim, S. Nam, D. Shim, M. Choi, H. Yoon, D. Kim, Y. Kim, H. W. Park, D. Kwak, S. Park, S. Yoon, W. Hahn, J. Ryu, S. Shim, K. Kang, S. Choi, J. Ihm, Y. Min, I. Kim, D. Lee, J. Cho, O. Kwon, J. Lee, M. Kim, S. Joo, J. Jang, S. Hwang, D. Byeon, H. Yang, K. Park, K. Kyung, and J. Choi, "7.2 A 128Gb 3b/Cell V-NAND Flash Memory with 1Gb/s I/O Rate," in *ISSCC*, 2015.

[40] J. Jeong, S. S. Hahn, S. Lee, and J. Kim, "Lifetime Improvement of NAND Flash-Based Storage Systems Using Dynamic Program and Erase Scaling," in *FAST*, 2014.

[41] X. Jimenez, D. Novo, and P. Ienne, "Wear Unleveling: Improving NAND Flash Lifetime by Balancing Page Endurance," in *FAST*, 2014.

[42] S.-M. Joe, J.-H. Yi, S.-K. Park, H. Shin, B.-G. Park, Y. J. Park, and J.-H. Lee, "Threshold Voltage Fluctuation by Random Telegraph Noise in Floating Gate NAND Flash Memory String," *IEEE TED*, 2011.

[43] M. Jung, D. M. Mathew, É. F. Zulian, C. Weis, and N. Wehn, "A New Bank Sensitive DRAMPower Model for Efficient Design Space Exploration," in *PATMOS*, 2016.

[44] M. Jung, D. M. Mathew, C. C. Rheinländer, C. Weis, and N. Wehn, "A Platform to Analyze DDR3 DRAM's Power and Retention Time," *IEEE Design and Test*, 2017.

[45] D. Kang, W. Jeong, C. Kim, D. Kim, Y. Cho, K. Kang, J. Ryu, K. Kang, S. Lee, W. Kim, H. Lee, J. Yu, N. Choi, D. Jang, J. Ihm, D. Kim, Y. Min, M. Kim, A. Park, J. Son, I. Kim, P. Kwak, B. Jung, D. Lee, H. Kim, H. Yang, D. Byeon, K. Park, K. Kyung, and J. Choi, "7.1 256Gb 3b/Cell V-NAND Flash Memory with 48 Stacked WL Layers," in *ISSCC*, 2016.

[46] S. Khan, D. Lee, Y. Kim, A. Alameldeen, C. Wilkerson, and O. Mutlu, "The Efficacy of Error Mitigation Techniques for DRAM Retention Failures: A Comparative Experimental Study," in *SIGMETRICS*, 2014.

[47] S. Khan, D. Lee, and O. Mutlu, "PARBOR: An Efficient System-Level Technique to Detect Data-Dependent Failures in DRAM," in *DSN*, 2016.

[48] S. Khan, C. Wilkerson, D. Lee, A. R. Alameldeen, and O. Mutlu, "A Case for Memory Content-Based Detection and Mitigation of Data-Dependent Failures in DRAM," *IEEE CAL*, 2016.

[49] S. Khan, C. Wilkerson, Z. Wang, A. R. Alameldeen, D. Lee, and O. Mutlu, "Detecting and Mitigating Data-Dependent DRAM Failures by Exploiting Current Memory Content," in *MICRO*, 2017.

[50] C. Kim, D.-H. Kim, W. Jeong, H.-J. Kim, I. H. Park, H.-W. Park, J. Lee, J. Park, Y.-L. Ahn, J. Y. Lee *et al.*, "A 512-Gb 3-b/Cell 64-Stacked WL 3-D-NAND Flash Memory," *JSSC*, 2018.

[51] J. S. Kim, M. Patel, H. Hassan, and O. Mutlu, "The DRAM Latency PUF: Quickly Evaluating Physical Unclonable Functions by Exploiting the Latency–Reliability Tradeoff in Modern DRAM Devices," in *HPCA*, 2018.

[52] Y. Kim, R. Daly, J. Kim, C. Fallin, J. H. Lee, D. Lee, C. Wilkerson, K. Lai, and O. Mutlu, "Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors," in *ISCA*, 2014.

[53] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, 1951.

[54] D. Lee, "Reducing DRAM Energy at Low Cost by Exploiting Heterogeneity," Ph.D. dissertation, Carnegie Mellon Univ., 2016.

[55] D. Lee, S. Khan, L. Subramanian, S. Ghose, R. Ausavarungnirun, G. Pekhimenko, V. Seshadri, and O. Mutlu, "Design-Induced Latency Variation in Modern DRAM Chips: Characterization, Analysis, and Latency Reduction Mechanisms," in *SIGMETRICS*, 2017.

[56] D. Lee, Y. Kim, G. Pekhimenko, S. Khan, V. Seshadri, K. Chang, and O. Mutlu, "Adaptive-Latency DRAM: Optimizing DRAM Timing for the Common-Case," in *HPCA*, 2015.

[57] J.-D. Lee, S.-H. Hur, and J.-D. Choi, "Effects of Floating-Gate Interference on NAND Flash Memory Cell Operation," *IEEE EDL*, 2002.

[58] S. Lee, J. Lee, I. Park, J. Park, S. Yun, M. Kim, J. Lee, M. Kim, K. Lee, T. Kim, B. Cho, D. Cho, S. Yun, J. Im, H. Yim, K. Kang, S. Jeon, S. Jo, Y. Ahn, S. Joe, S. Kim, D. Woo, J. Park, H. W. Park, Y. Kim, J. Park, Y. Choi, M. Hirano, J. Ihm, B. Jeong, S. Lee, M. Kim, H. Lee, S. Seo, H. Jeon, C. Kim, H. Kim, J. Kim, Y. Yim, H. Kim, D. Byeon, H. Yang, K. Park, K. Kyung, and J. Choi, "7.5 A 128Gb 2b/Cell NAND Flash Memory in 14nm Technology with tPROG=640μs and 800MB/s I/O Rate," in *ISSCC*, 2016.

[59] Y. Lee, H. Yoo, I. Yoo, and I.-C. Park, "6.4 Gb/s Multi-Threaded BCH Encoder and Decoder for Multi-Channel SSD Controllers," in *ISSCC*, 2012.

[60] J. Li, K. Zhao, X. Zhang, J. Ma, M. Zhao, and T. Zhang, "How Much Can Data Compressibility Help to Improve NAND Flash Memory Lifetime?" in *FAST*, 2015.

[61] J. Liu, B. Jaiyen, Y. Kim, C. Wilkerson, and O. Mutlu, "An Experimental Study of Data Retention Behavior in Modern DRAM Devices: Implications for Retention Time Profiling Mechanisms," in *ISCA*, 2013.

[62] J. Liu, B. Jaiyen, R. Veras, and O. Mutlu, "RAIDR: Retention-Aware Intelligent DRAM Refresh," in *ISCA*, 2012.

[63] Y. Luo, Y. Cai, S. Ghose, J. Choi, and O. Mutlu, "WARM: Improving NAND Flash Memory Lifetime with Write-Hotness Aware Retention Management," in *MSST*, 2015.

[64] Y. Luo, S. Ghose, Y. Cai, E. F. Haratsch, and O. Mutlu, "Enabling Accurate and Practical Online Flash Channel Modeling for Modern MLC NAND Flash Memory," *JSAC*, 2016.

[65] Y. Luo, S. Ghose, Y. Cai, E. F. Haratsch, and O. Mutlu, "HeatWatch: Improving 3D NAND Flash Memory Device Reliability by Exploiting Self-Recovery and Temperature Awareness," in *HPCA*, 2018.

[66] D. M. Mathew, M. Schultheis, C. C. Rheinländer, C. Sudarshan, C. Weis, N. Wehn, and M. Jung, "An Analysis on Retention Error Behavior and Power Consumption of Recent DDR4 DRAMs," in *DATE*, 2018.

[67] N. Matthew and R. Stones, *Beginning Linux Programming*. John Wiley & Sons, 2008.

[68] J. Meza, Q. Wu, S. Kumar, and O. Mutlu, "A Large-Scale Study of Flash Memory Failures in the Field," in *SIGMETRICS*, 2015.

[69] N. Mielke, T. Marquart, N.Wu, J.Kessenich, H. Belgal, E. Schares, and F. Triverdi, "Bit Error Rate in NAND Flash Memories," in *IRPS*, 2008.

[70] K. Mizoguchi, T. Takahashi, S. Aritome, and K. Takeuchi, "Data-Retention Characteristics Comparison of 2D and 3D TLC NAND Flash Memories," in *IMW*, 2017.

[71] O. Mutlu, "The RowHammer Problem and Other Issues We May Face as Memory Becomes Denser," in *DATE*, 2017.

[72] I. Narayanan, D. Wang, M. Jeon, B. Sharma, L. Caulfield, A. Sivasubramaniam, B. Cutler, J. Liu, B. Khessib, and K. Vaid, "SSD Failures in Datacenters: What? When? And Why?" in *SYSTOR*, 2016.

[73] K. Naruke, S. Taguchi, and M. Wada, "Stress Induced Leakage Current Limiting to Scale Down EEPROM Tunnel Oxide Thickness," *IEDM Tech. Digest*, 1988.

[74] Y. Pan, G. Dong, Q. Wu, and T. Zhang, "Quasi-Nonvolatile SSD: Trading Flash Memory Nonvolatility to Improve Storage System Performance for Enterprise Applications," in *HPCA*, 2012.

[75] Y. Pan, G. Dong, and T. Zhang, "Exploiting Memory Device Wear-Out Dynamics to Improve NAND Flash Memory System Performance," in *FAST*, 2011.

[76] N. Papandreou, T. Parnell, H. Pozidis, T. Mittelholzer, E. Eleftheriou, C. Camp, T. Griffin, G. Tressler, and A. Walls, "Using Adaptive Read Voltage Thresholds to Enhance the Reliability of MLC NAND Flash Memory Systems," in *GLSVLSI*, 2014.

[77] J. Park, J. Jeong, S. Lee, Y. Song, and J. Kim, "Improving Performance and Lifetime of NAND Storage Systems Using Relaxed Program Sequence," in *DAC*, 2016.

[78] J. K. Park, D.-I. Moon, Y.-K. Choi, S.-H. Lee, K.-H. Lee, S. H. Pyi, and B. J. Cho, "Origin of Transient Vth Shift After Erase and Its Impact on 2D/3D Structure Charge Trap Flash Memory Cell Operations," in *IEDM*, 2012.

[79] K.-T. Park, M. Kang, D. Kim, S.-W. Hwang, B. Y. Choi, Y.-T. Lee, C. Kim, and K. Kim, "A Zeroing Cell-to-Cell Interference Page Architecture with Temporary LSB Storing and Parallel MSB Program Scheme for MLC NAND Flash Memories," *JSSC*, 2008.

[80] K. Park, S. Nam, D. Kim, P. Kwak, D. Lee, Y. Choi, M. Choi, D. Kwak, D. Kim, M. Kim, H. W. Park, S. Shim, K. Kang, S. Park, K. Lee, H. Yoon, K. Ko, D. Shim, Y. Ahn, J. Ryu, D. Kim, K. Yun, J. Kwon, S. Shin, D. Byeon, K. Choi, J. Han, K. Kyung, J. Choi, and K. Kim, "Three-Dimensional 128 Gb MLC Vertical NAND Flash Memory With 24-WL Stacked Layers and 50 MB/s High-Speed Programming," *JSSC*, 2015.

[81] T. Parnell, N. Papandreou, T. Mittelholzer, and H. Pozidis, "Modelling of the Threshold Voltage Distributions of Sub-20nm NAND Flash Memory," in *GLOBECOM*, 2014.

[82] M. Patel, J. S. Kim, and O. Mutlu, "The Reach Profiler (REAPER): Enabling the Mitigation of DRAM Retention Failures via Profiling at Aggressive Conditions,"

in *ISCA*, 2017.
[83] D. A. Patterson, G. Gibson, and R. H. Katz, "A Case for Redundant Arrays of Inexpensive Disks (RAID)," in *SIGMOD*, 1988.
[84] P. Prabhu, A. Akel, L. M. Grupp, S. Y. Wing-Kei, G. E. Suh, E. Kan, and S. Swanson, "Extracting Device Fingerprints from Flash Memory by Exploiting Physical Variations," in *TRUST*, 2011.
[85] M. Qureshi, D. H. Kim, S. Khan, P. Nair, and O. Mutlu, "AVATAR: A Variable-Retention-Time (VRT) Aware Refresh for DRAM Systems," in *DSN*, 2015.
[86] Samsung Electronics Co., Ltd., "Samsung V-NAND Technology," https://www.samsung.com/us/business/oem-solutions/pdfs/V-NAND_technology_WP.pdf, white paper. 2014.
[87] B. Schroeder, R. Lagisetty, and A. Merchant, "Flash Reliability in Production: The Expected and the Unexpected," in *FAST*, 2016.
[88] S. Seabold and J. Perktold, "Statsmodels: Econometric and Statistical Modeling with Python," in *SciPy*, 2010.
[89] K.-D. Suh, B.-H. Suh, Y.-H. Lim, J.-K. Kim, Y.-J. Choi, Y.-N. Koh, S.-S. Lee, S.-C. Suk-Chon, B.-S. Choi, J.-S. Yum *et al.*, "A 3.3 V 32 Mb NAND Flash Memory With Incremental Step Pulse Programming Scheme," *JSSC*, 1995.
[90] TechInsights, Inc., "NAND Flash Memory Roadmap," http://www.techinsights.com/NAND-flash-roadmap/, 2016.
[91] W. Wang, T. Xie, and D. Zhou, "Understanding the Impact of Threshold Voltage on MLC Flash Memory Performance and Reliability," in *ICS*, 2014.
[92] Y. Wang, L. Dong, and R. Mao, "P-Alloc: Process-Variation Tolerant Reliability Management for 3D Charge-Trapping Flash Memory," *TECS*, 2017.
[93] E. H. Wilson, M. Jung, and M. T. Kandemir, "ZombieNAND: Resurrecting Dead NAND Flash for Improved SSD Longevity," in *MASCOTS*, 2014.
[94] Q. Xiong, F. Wu, Z. Lu, Y. Zhu, Y. Zhou, Y. Chu, C. Xie, and P. Huang, "Characterizing 3D Floating Gate NAND Flash," in *SIGMETRICS*, 2017.
[95] Q. Xiong, F. Wu, Z. Lu, Y. Zhu, Y. Zhou, Y. Chu, C. Xie, and P. Huang, "Characterizing 3D Floating Gate NAND Flash: Observations, Analyses, and Implications," *TOS*, 2018.
[96] V. Ye, "The Solution to Bit Error Non-Uniformity of 3D NAND," in *Flash Memory Summit*, 2017.
[97] E. Zhang, W. Wang, C. Zhang, Y. Jin, G. Zhu, Q. Sun, D. W. Zhang, P. Zhou, and F. Xiu, "Tunable Charge-Trap Memory Based on Few-Layer MoS2," *ACS Nano*, 2014.
[98] X. Zhang, J. Li, H. Wang, K. Zhao, and T. Zhang, "Reducing Solid-State Storage Device Write Stress Through Opportunistic In-Place Delta Compression," in *FAST*, 2016.

A APPENDIX

A.1 Write-Induced Errors

We analyze how each type of write-induced error affects the RBER and the threshold voltage distribution of 3D NAND flash memory.

A.1.1 Program Errors. Program errors occur when the data is incorrectly written to the NAND flash memory [4, 9, 11, 79]. Such errors are introduced when multiple programming operations are required to write data to a single cell. For example, in many MLC NAND flash memory devices, *two-step programming* [4, 79] is employed. Two-step programming uses two separate *partial programming* steps to write data to an MLC NAND flash cell. In the first step, the flash controller writes only the LSB to the cell, setting the cell to a temporary voltage state. In the second step, the controller writes the MSB to the cell, but in order to perform this write, the controller must first determine the current voltage state of the cell. This requires reading the partially-programmed data from the cell, during which an error may occur. This error causes the controller to incorrectly set the final voltage state of the cell during the second programming step, and, thus, is called a program error. Prior work [4] shows that program errors occur in state-of-the-art planar MLC NAND flash memory.

Current generations of 3D NAND flash memory use *one-shot programming* [4, 9, 11, 79], which programs *both* the LSB and MSB of a cell at the *same time*. As a result, current 3D NAND flash memory devices do *not* experience program errors. Our measurements in Figure 4 confirm the lack of program errors in 3D NAND flash memory. In an MLC NAND flash memory that has program errors, the threshold voltage distributions of the ER and P1 states have secondary peaks near the P2 and P3 states, respectively [4]. This is because program errors affect only the LSB, since only the LSB is being read during the second programming step. Since there is no second peak in Figure 4, there are no program errors.

Program errors may appear in future 3D NAND flash memory devices. In planar NAND flash memory, two-step programming was introduced when planar MLC NAND flash memory transitioned to the 40 nm manufacturing process technology node, in order to reduce the number of program interference errors [79]. A similar transition may occur in the future to continue scaling the density of 3D NAND flash memory, especially as it becomes increasingly difficult to add more layers into a 3D NAND flash memory chip. Thus, we conclude that today's 3D NAND flash memories do *not* have program errors, but program errors may appear in future generations.

A.1.2 Program/Erase Cycling Errors. A P/E cycling error occurs because of the natural variation of the threshold voltage of cells in each state [14, 69] due to the inaccuracy of each program and erase operation (see Section 2.2). Such inaccuracy during program and erase operations increases as the P/E cycle count increases. To study the impact of P/E cycling errors, we randomly select a flash block within each 3D NAND chip, and wear out the block by programming random data to each page in the block until the block reaches 16K P/E cycles. Using the methodology described in Section 4.1, we obtain the overall RBER and the threshold voltage of each cell at various P/E cycle counts.[10]

Observations. Figure 17 shows how the mean and standard deviation of the threshold voltage distribution of each state change as a function of the P/E cycle count, when we fit our voltage measurements for each state to a Gaussian model. Each subfigure in the top row represents the mean for a different state; each subfigure in the bottom row represents the standard deviation for a different state. The blue dots shows the measured data; each orange line shows a linear trend fitted to the measured data. The x-axis shows the P/E cycle count; the y-axis shows the mean (Figures 17a–17d) or the standard deviation (Figures 17e–17h) of the threshold voltage distribution of each state, in voltage steps. We make four observations from Figure 17. First, the mean and standard deviation of all states increase linearly as the P/E cycle count increases. We fit a line using linear regression, shown as an orange dotted line in each subfigure.[11] Second, the threshold voltage distributions of the ER and P1 states shift to higher voltages, while the distributions of the P2 and P3 states shift to lower voltages, causing the distributions to move closer to the middle of the threshold voltage range. Third, the threshold voltage distributions of all four states become wider (i.e., the standard deviation increases) as the P/E cycle count increases. Since the distributions shift towards the middle of the threshold voltage range and become wider as the P/E cycle count increases, the distributions become closer to each other, which increases the raw bit error rate. Fourth, the magnitude of the threshold voltage shift and the widening of the distributions is much larger for the ER state than it is for the other three states (i.e., P1, P2, P3). Therefore, ER↔P1 errors (i.e., an error that shifts a cell that is originally programmed in the ER state to the P1 state, or vice versa) increase faster than other errors with the P/E cycle count.

Figure 18 shows how the RBER increases as the P/E cycle count increases. The top graph breaks down the errors into which bit (i.e., LSB or MSB) they occur in. The bottom graph breaks down the errors based on how the error changed the cell state due to a shift in the cell threshold voltage. If the error caused *either* the LSB or MSB (but not both) to be read incorrectly, we refer to that error as a single-bit error (ER ↔ P1, P1 ↔ P2, and P2 ↔ P3 in the graph). If *both* the LSB and MSB are read incorrectly as a result of the error, we refer to that error as a multi-bit error. We make four observations from Figure 18. First, both LSB and MSB errors increase as the P/E cycle count increases, following an exponential trend. Second, ER ↔ P1 errors increase at a much faster rate as the P/E cycle count increases, compared to the other types of cell state changes, and ER ↔ P1 errors become the dominant MSB error type when the P/E cycle count reaches 8K P/E cycles (6K is the cross-over point). This is because the electrons trapped in the cell during wearout prevent the cell from being set to very low threshold voltages. As a result, the threshold voltage distribution of the ER state shifts and widens more than the distributions of the other states, as we see in Figure 17. Third, multi-bit errors are less common, but they occur as early as at 1K P/E cycles. Only a large difference between the target and actual threshold voltage

[10] Due to limitations with our experimental testing platform, each data point at a particular P/E cycle count has a retention time of 50 minutes.

[11] For the ER state, a linear fit has a 5.9% higher root mean square error than a power-law fit. However, we choose the linear fit due to its simplicity.

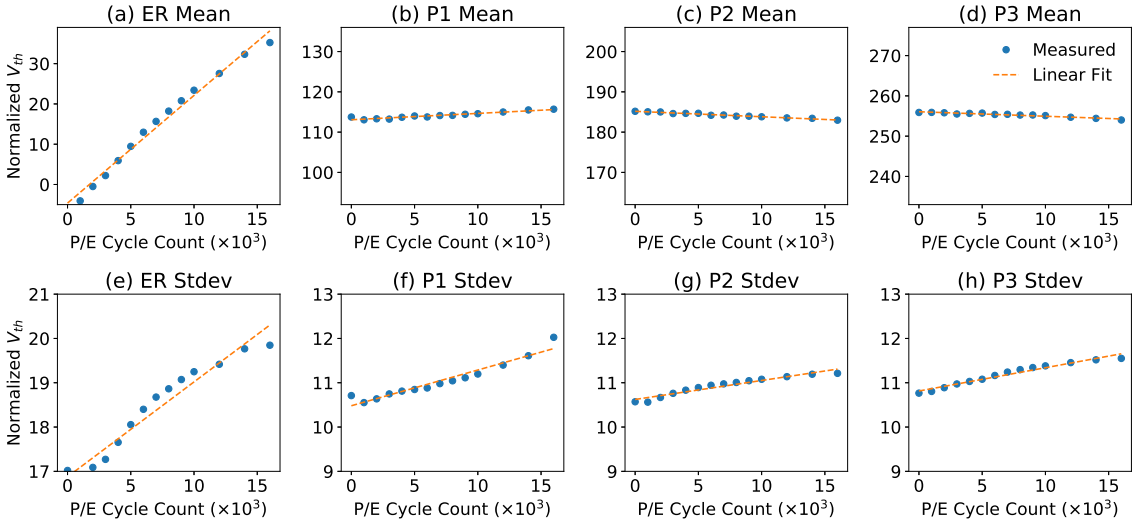


Figure 17: Mean and standard deviation of our Gaussian threshold voltage distribution model of each state, versus P/E cycle count.

can lead to a multi-bit error, which is unlikely to happen. Fourth, MSBs have a 2.1× higher error rate than LSBs, on average across all P/E cycle counts. This is because the flash controller must use two read reference voltages to read a cell's MSB, but needs *only one* read reference voltage to read a cell's LSB.

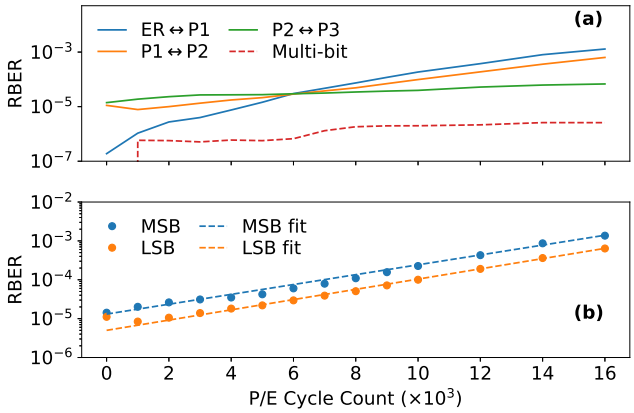


Figure 18: RBER due to P/E cycling errors vs. P/E cycle count.

Figure 19 shows how the optimal read reference voltages change as the P/E cycle count increases. This figure contains three subfigures, each of which shows the optimal voltage for V_a, V_b, and V_c (see Figure 1a). We make two observations from this figure. First, the optimal voltage for V_a increases rapidly as the P/E cycle count increases: after 16K P/E cycles, the voltage goes up by more than 20 voltage steps. Second, the optimal voltages for V_b and V_c remain almost constant as the P/E cycle count increases: neither voltage changes by more than 4 voltage steps after 16K P/E cycles, as expected from the lack of change in P1, P2, and P3 distribution means shown in Figure 17.

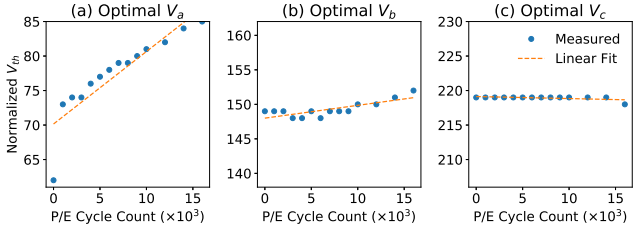


Figure 19: Optimal read reference voltages vs. P/E cycle count.

Insights. To compare the error characteristics of 3D NAND flash memory to that of planar NAND flash memory, we take the equivalent observations on planar NAND flash memory reported by prior works [14, 64, 81], and compare them to our findings for 3D NAND flash memory, which we just described. We find two key differences. First, for 3D NAND flash memory, the threshold voltage distributions for the P2 state and the P3 state shift to *lower* voltages as the P/E cycle count increases. In contrast, for planar NAND flash memory, the distributions of both states shift to *higher* voltages [14, 64, 81]. One possible source of this change is the increased impact of early retention loss with P/E cycle count, which lowers the threshold voltage of cells in higher-voltage states (i.e., P2 and P3) [23]. Second, for 3D NAND flash memory, the *change* in the mean threshold voltage of each state distribution exhibits a linear increase. However, in sub-20 nm planar NAND flash memory, the change in the mean threshold voltage exhibits a *power-law*-based increase with P/E cycle count [64, 81]. In sub-20 nm planar NAND flash memory, the mean threshold voltage of each state distribution increases more rapidly at lower P/E cycle counts than in higher P/E cycle counts, resulting in the power-law-based behavior. However, we note that planar NAND flash memory using an older

manufacturing process technology (e.g., 20–24 nm) exhibits a linear increase with P/E cycle count for the distribution mean [14], just as we observe for 3D NAND flash memory. Thus, there is evidence that when the manufacturing process technology scales below a certain size, the change in the distribution mean transitions from linear behavior to power-law-based behavior with respect to P/E cycle count. As a result, when future 3D NAND flash memory scales down to a sub-20 nm manufacturing process technology node, we might expect that it too will exhibit power-law behavior for the change in the distribution mean. We conclude that the differences we observe between the P/E cycling effect in 3D NAND flash memory and planar NAND flash memory are mainly caused by the use of a significantly different manufacturing process technology node.

A.1.3 Program Interference. When a cell (which we call the *aggressor cell*) is being programmed, cell-to-cell program interference can cause the threshold voltage of nearby flash cells (which we call *victim cells*) to increase unintentionally [15, 16] (see Section 2.2). In 3D NAND flash memory, there are two types of program interference that can occur. The first, *wordline-to-wordline program interference*, affects victim cells along the z-axis of the cell that is programmed (see Figure 3). These victim cells are physically next to the cell that is programmed, and belong to the same bitline (and thus the same flash block). The second, *bitline-to-bitline program interference*, affects victim cells along the x-axis or y-axis of the cell that is programmed. Bitline-to-bitline program interference can affect victim cells in the same wordline (i.e., cells on the y-axis), or it can affect victim cells that belong to other flash blocks (i.e., cells on the x-axis).

To quantitatively analyze the effect of program interference on cell threshold voltage and raw bit error rate, we use the same experimental data that we have for P/E cycling errors (see Section A.1.2). A correlation exists between the amount by which program interference changes the threshold voltage of a victim cell (ΔV_{victim}) and the threshold voltage change of the aggressor cell ($\Delta V_{aggressor}$) [15]. As a result of this *interference correlation*, the threshold voltage of a victim cell is *dependent* on the threshold voltage of the aggressor cell. The strength of this correlation can be quantified as $\frac{\Delta V_{victim}}{\Delta V_{aggressor}}$, which is a property of the NAND device and is largely dependent on the distance between the cells [57]. After programming randomly-generated data to the victim cells and the aggressor cells, we estimate $\Delta V_{aggressor}$ by calculating the threshold voltage difference between the aggressor cell's threshold voltage in its final state and that in the ER state. We estimate ΔV_{victim} by calculating the difference between the victim cell's threshold voltage with and without program interference.[12]

Observations. Figure 20 shows the interference correlation for wordline-to-wordline interference and bitline-to-bitline interference on a victim cell, for aggressor cells of varying distance from the victim cell. For example, the victim cell in BL M, WL N has an interference correlation of 2.7% with the *next wordline* aggressor cell in BL M, WL N+1, which means that, if the threshold voltage of the aggressor cell increases by ΔV, the threshold voltage of the victim cell increases by $0.027\Delta V$ due to wordline-to-wordline program interference. We make two observations from this figure. First, the interference correlation of the *next wordline* aggressor cell (i.e., 2.7%) is over an order of magnitude higher than that of any other aggressor cell, of which the maximum interference correlation is only 0.080% (the *previous wordline* aggressor cell in BL M, WL N-1). Thus, the program interference to the victim cell, is dominated by wordline-to-wordline interference from the *next* wordline. Second, all of the other types of interference have much smaller interference correlation values.

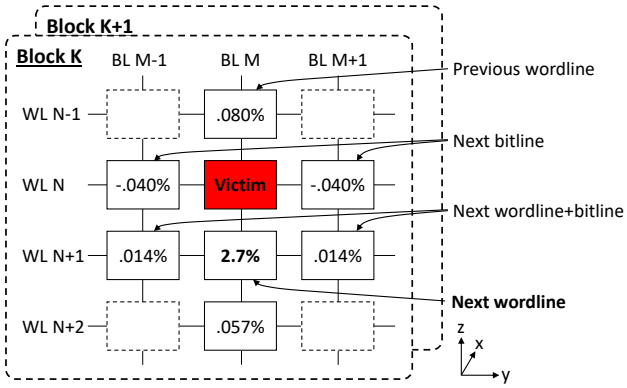


Figure 20: Interference correlation for a victim cell, as a result of programming aggressor cells of varying distances from the victim cell.

Figure 21 shows how much the threshold voltage of a victim cell shifts (ΔV_{victim}) when a neighboring aggressor cell is programmed to the P3 state, which generates the largest possible program interference. Each curve represents a certain program interference type (i.e., Next WL or Prev WL) and a certain state of the victim cell (V). The curves that have a significant amount of threshold voltage shift (e.g., >6 voltage steps) due to program interference are shown in Figure 21(a); the curves that have a small amount of threshold voltage shift are shown in Figure 21(b). We make three observations from Figure 21. First, the effect of program interference decreases as the P/E cycle count increases (along the x-axis, from left to right). As we discuss in Section A.1.2, electrons trapped in a flash cell due to wearout prevent the cell from returning to the lowest threshold voltage values during an erase operation. As a result, as the P/E cycle count increases, the mean threshold voltage of the ER state increases. This causes $\Delta V_{aggressor}$ to decrease as the P/E cycle count increases, because the starting voltage of the aggressor cell increases but its target voltage after programming remains the same. As we discuss above, the interference correlation (i.e., the ratio between $\Delta V_{aggressor}$ and ΔV_{victim}) is largely a function of the distance between flash cells. Thus, since $\Delta V_{aggressor}$ decreases, ΔV_{victim} also decreases with the P/E cycle count. Second, the amount of program interference induced by an aggressor cell in the next wordline decreases when the victim cell is in a higher-voltage state (Next WL curves in Figure 21a, from top to bottom). This is likely because the voltage difference between the aggressor cell and the victim cell is lower when the victim cell is in a higher-voltage state, reducing the the threshold voltage shift due to program interference. Third, the program interference induced by an aggressor

[12] The cell threshold voltage *without* program interference is obtained by reading the cell *before* the next wordline is programmed.

cell in the previous wordline (Prev WL curves in Figure 21) affects the threshold voltage distribution of only the ER state for a victim cell, but it has little effect on the distributions of the other three states (i.e., P1, P2, P3). This is a result of how programming takes place in NAND flash memory. A program operation can only *increase* the voltage of a cell due to circuit-level limitations. When the aggressor cell in the previous wordline is programmed, the victim cell is already in the ER state, and the victim cell's voltage increases due to program interference. Some time later, the victim cell is programmed. If the target state of the victim cell is P1, P2, or P3, the programming operation needs to further increase the voltage of the cell, and any effects of program interference from the aggressor cell in the previous wordline are eliminated. If, however, the target state of the victim cell is ER, the programming operation does not change the victim cell's voltage, and the effects of program interference from the aggressor cell in the previous wordline remain.

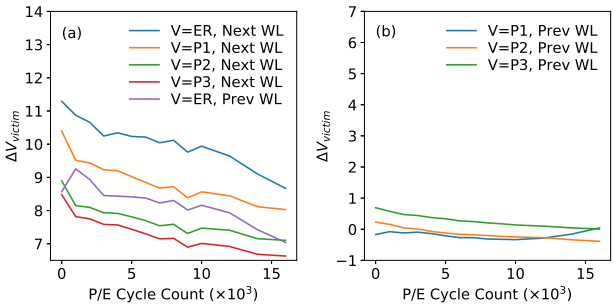


Figure 21: Amount of threshold voltage shift due to program interference vs. P/E cycle count.

Insights. We compare the program interference in 3D NAND flash memory to the program interference observed in planar NAND flash memory, as reported in prior work [15, 16]. We find one major difference. The maximum interference correlation of program interference from a directly-adjacent cell is 40% lower in 3D NAND flash memory (2.7%) than in state-of-the-art (20–24 nm) planar NAND flash memory (4.5% [15]). This is corroborated by findings in prior work [80], which shows that 3D NAND flash memory has 84% lower program interference than 15–19 nm planar NAND flash memory. The lower interference correlation in 3D NAND flash memory is due to the larger manufacturing process technology node (30–50 nm for the chips we test) that it uses compared to state-of-the-art planar NAND flash memory. The amount of interference correlation between neighboring cells is a function of the distance between the cells [57]. In a larger manufacturing process technology node, the flash cells are farther away from each other, causing the interference correlation to decrease. We note that when future 3D NAND flash memory chips use smaller manufacturing process technology nodes, the impact of programming interference will increase, similar to what happened in planar NAND flash memory.

Note that we are the first to compare how the threshold voltage shift caused by program interference changes with the P/E cycle count. As we discuss in our first observation for Figure 21, the program interference effect decreases as the P/E cycle count increases because the increasing effects of wearout reduce the value of $\Delta V_{aggressor}$ during programming. We conclude that the 40% reduction in the program interference effect we observe in 3D NAND flash memory compared to planar NAND flash memory is mainly caused by the difference in manufacturing process technology.

A.2 Early Retention Loss

In this section, we present the results and analysis of retention loss in 3D NAND flash memory in addition to the key findings in Section 4.3. We use the same methodology as described in Section 4.3.

Observations. Figure 22 shows how the mean and the standard deviation of the threshold voltage distribution change with retention time. Each subfigure in the top row shows the mean for a different state; each subfigure in the bottom row show the standard deviation for a different state. The blue dots show the measured data; each orange line shows a linear trend line fitted to the measured data. The x-axis shows the retention time in log scale; the y-axis shows the mean or standard deviation value in voltage steps. We make five observations from this figure. First, the threshold voltage distribution shifts more when the retention time is low. This is the early retention loss phenomenon, which occurs because charge that is trapped near the surface of the charge trap layer is detrapped soon after programming. Second, as the retention time increases, the voltage values of cells in the P1, P2, and P3 states decrease, while the voltage values of cells in the ER state increase. This is because the cells in the ER state have *negative* threshold voltages, and hence they *gain* charge over retention time. Third, the threshold voltage distributions of the ER and P3 states shift faster than the distributions of the P1 and P2 states as the retention time increases. This is because the ER and P3 states have larger voltage differences from the ground than the other states. Fourth, retention loss has little effect on the width of the threshold voltage distribution (i.e., standard deviation values change by less than 1 voltage step after 24 days). This is because the effects of retention loss (i.e., charge leakage) impact cells at a similar rate, causing all of the cells within the threshold voltage distribution to lose a similar amount of voltage. Fifth, the correlation between any distribution parameter (V) and the retention time (t) can be modeled as a linear function (shown by the dotted lines in Figure 22): $V = A \cdot \log(t) + B$. A and B are constants that change based on which parameter V is modeling (i.e., the threshold voltage distribution mean or standard deviation). Prior work shows that planar NAND flash memory has a similar trend for retention loss, even though it uses a different flash cell design. We have already compared and evaluated the differences between 3D NAND and planar NAND flash memory in retention loss speed in Section 4.3, and provided more detail about the linear function that models the threshold voltage distribution parameters in Section 5.2.

Figure 23 shows how the RBER increases with retention time for a block that has endured 10K P/E cycles. The top graph breaks down the errors according to the change in cell state as a result of the errors; the bottom graph breaks down the errors into MSB and LSB page errors. We make two observations from Figure 23, in addition to our observations in Section 4.3. First, retention errors are dominated by P2 ↔ P3 errors, because the threshold voltage distribution of the P3 state not only shifts more but also widens

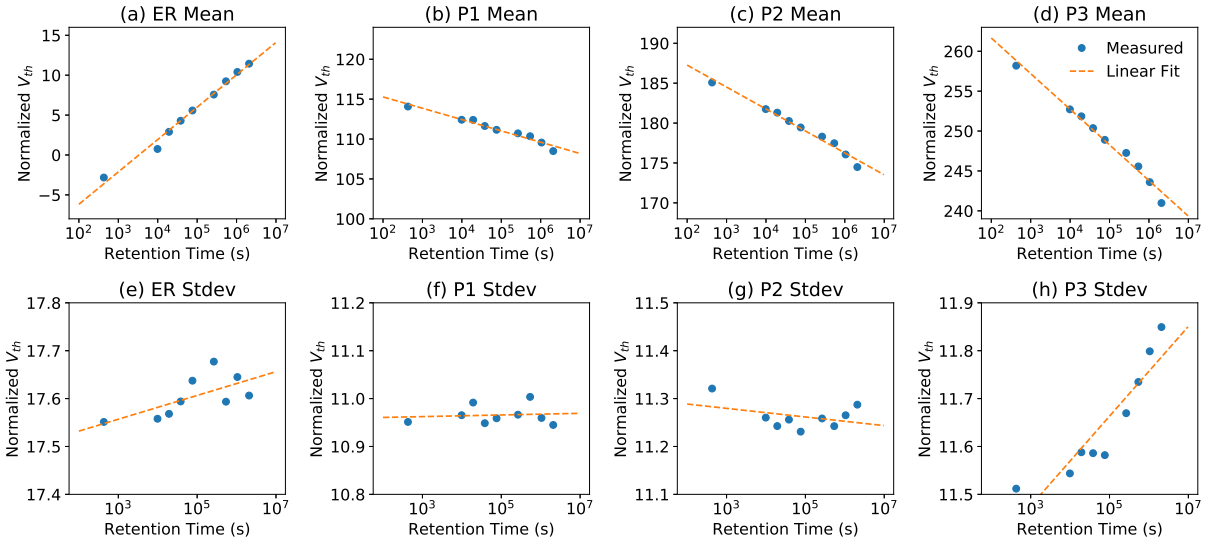


Figure 22: Mean and standard deviation of our Gaussian threshold voltage distribution model of each state, versus retention time.

more with retention time than the distributions of the other states (see Figure 22). Although the distribution of the ER state also shifts significantly, there are fewer ER ↔ P1 errors to begin with. Second, the MSB error rate increases faster than the LSB error rate as the retention time increases. This is because as the distributions of both the ER and P3 states shift more than those of the P1 and P2 states, cells in the ER and P3 states are more likely to have errors. These errors (ER ↔ P1 and P2 ↔ P3) affect the MSB of the cell.

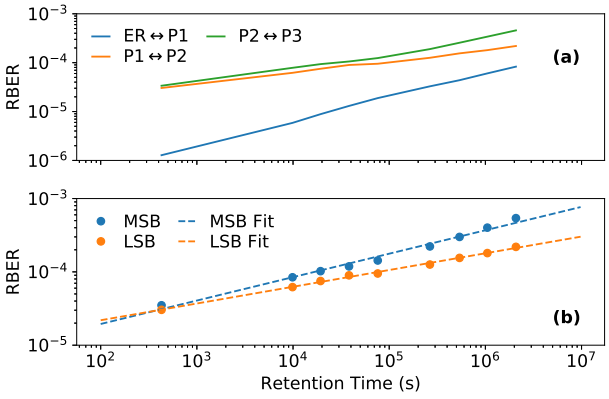


Figure 23: RBER vs. retention time, broken down by (a) the state transition of each flash cell, and (b) MSB or LSB page.

Insights. We compare the errors due to retention loss in 3D NAND flash memory to those in planar NAND flash memory, as reported in prior work [6, 7, 69]. We find another major difference in 3D NAND flash memory in terms of threshold voltage distribution, in addition to those discussed in Section 4.3. We find that the retention loss phenomenon we observe in 3D NAND flash memory (1) shifts the threshold voltage distributions of the P1, P2 and P3 states lower, and (2) has little effect on the width of the distribution of each state. In contrast, the retention loss phenomenon observed in planar NAND flash memory (1) does *not* shift the P1 and P2 state distributions by much, and (2) increases the width of each state's distribution *significantly* [6]. This indicates that a mechanism that adjusts the optimal read reference voltage to the threshold voltage shift caused by retention loss can be more effective on 3D NAND flash memory than on planar NAND flash memory, because the distributions shift by a greater amount (indicating a greater need for voltage adjustment) with a smaller amount of overlap between two threshold voltage distributions (reducing the number of read errors when the optimal read reference voltage is used). We conclude that, due to the early retention loss phenomenon we observe in 3D NAND flash memory, the threshold voltage of a flash cell changes quickly within several hours after programming, leading to significant changes in RBER and optimal read reference voltage values.

A.3 Read-Induced Errors

In this section, we analyze how each type of read-induced error affects the RBER and the threshold voltage distribution of 3D NAND flash memory.

A.3.1 Read Errors. A read error is a type of read-induced error where two reads to a flash cell may return different data values if the read reference voltage used to read the cell is close to the cell's threshold voltage [24, 29, 42] (see Section 2.2). A read error adds uncertainty to the outcome of *every* read operation performed by the SSD controller. However, despite the potential for widespread impact, read errors are *not* well-studied by prior work.

To quantify read errors, we use the data we collected in Section 4.3. For each cell, we see if the *actual* read outcome (i.e., the bit value output by the flash controller after a read operation)

matches the *expected* read outcome (i.e., the value that the read should have returned based on the current voltage of the flash cell). We determine the expected read outcome by comparing V_{ref} with V_{th} (i.e., we expect to read 1 if $V_{th} < V_{ref}$, because V_{ref} is high enough that it should turn on the cell). We obtain V_{th} by combining the outcomes of multiple reads when sweeping the read reference voltage, thus we expect that the combined output eliminates the impact of read errors and is thus accurate. We say that a read error occurs if the actual read outcome and the expected read outcome do *not* match.

Observations. Figure 24 shows how the read error rate changes as a function of the *read offset* (i.e., $V_{ref} - V_{th}$). We observe that, as the absolute value of the read offset increases, the read error rate decreases exponentially. This is likely because when V_{ref} is closer to V_{th} (i.e., when $V_{ref} - V_{th}$ has a smaller absolute value), the amount of noise (i.e., voltage fluctuations) in the sense amplifier increases exponentially [24, 29]. The larger amount of noise increases the likelihood that the sense amplifier incorrectly detects whether the cell turns on, which leads to a larger probability that a read error occurs.

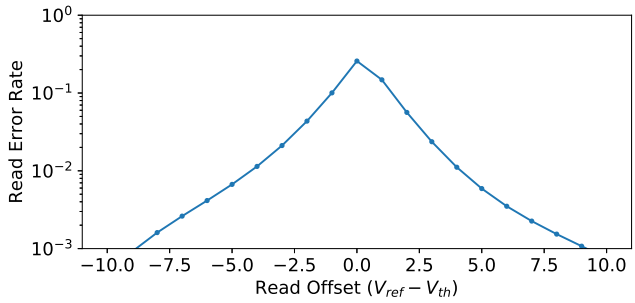


Figure 24: Read error rate vs. read offset ($V_{ref} - V_{th}$).

Figure 25 shows the correlation between the read error rate and the total RBER in a flash page. We observe that the read error rate is linearly correlated with the overall RBER. This is because, when the RBER is high, the threshold voltage distributions of neighboring states overlap with each other by a greater amount. This causes a larger number of cells to be close to the read reference voltage value, increasing the probability that a read error occurs (see Figure 24).

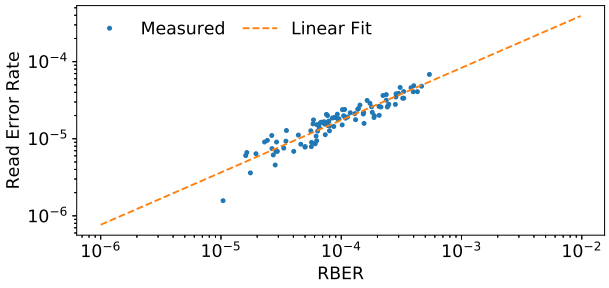


Figure 25: Relationship between the read error rate and the RBER.

Insights. We are the first to discover and quantify the extent of read errors, and to show the correlation of these errors with the RBER and with the read reference voltage. We conclude that read errors are correlated with the read offset (i.e., $V_{ref} - V_{th}$) and the overall RBER of the flash page.

A.3.2 Read Disturb Errors. Read disturb errors occur when a read operation to one page in a flash block may introduce errors in *other, unread* pages in the same block [5, 76] (see Section 2.2). Read disturb errors are caused by the high pass-through voltage applied to cells in the unread pages.

To characterize read disturb errors, we first randomly select 11 flash blocks and wear out each block to 10K P/E cycles by repeatedly erasing and programming pseudorandomly generated data into each page of each block. Then, we program pseudorandomly-generated data to each page of each flash block. To minimize the impact of other errors, especially retention errors due to early retention loss, we wait until the data has a 2-day retention time before inducing read disturb. This ensures that, according to our results in Section 4.3, after 2 days, retention loss has slowed down and can only shift the threshold voltage by at most 1 voltage step during the relatively short characterization process (~9 h). To induce read disturb in the flash block, we repeatedly read from a wordline within the block for up to 900K times (i.e., up to 900K read disturbs). During this process, to characterize the read disturb effect, we obtain the RBER and threshold voltage distribution at ten different read disturb counts from 0 to 900K.

Observations. Figure 26 shows how the mean and standard deviation of the threshold voltage distribution change with read disturb count. Each subfigure in the top row shows the mean for a different state; each subfigure in the bottom row shows the standard deviation for a different state. The blue dots shows the measured data; each orange line shows a linear trend line fitted to the measured data. The x-axis shows the P/E cycle count; the y-axis shows the distribution parameters in voltage steps. We make three observations from this figure. First, the read disturb effect increases the mean threshold voltage of the ER state significantly, by ~8 voltage steps after 900K read disturbs. In contrast, the mean threshold voltages of the programmed states change by only a small amount (<3 voltage steps). The increase in the mean threshold voltage is lower for a higher V_{th} state. This is because the impact of read disturb is correlated with the difference between the pass-through voltage (see Section 2.1) and the threshold voltage of a cell. When the difference is larger (i.e., when the threshold voltage of a cell is lower), the impact of read disturb increases. In fact, we observe that the threshold voltage distribution of the P3 state even shifts to slightly *lower* voltage values during the experiment, because read disturb has little effect on cells in the P3 state, and the impact of retention loss dominates. Second, the distribution width of each state (i.e., standard deviation) *decreases* slightly as the read disturb count increases, by <0.2 voltage steps after 900K read disturbs. Third, the change in each distribution parameter can be modeled as a linear function of the read disturb count (as shown by the orange dotted lines). This shows that read disturb in 3D NAND flash memory follows a similar linear trend as that observed in planar NAND flash memory by prior work [5].

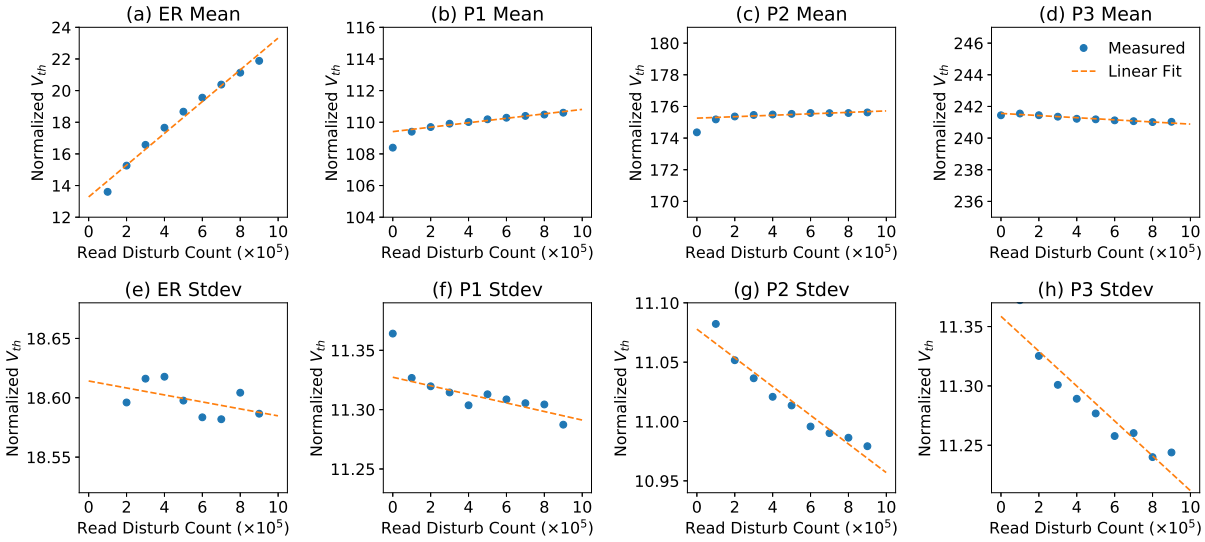


Figure 26: Mean and standard deviation of threshold voltage distribution of each state, vs. read disturb count.

Figure 27 plots how RBER increases with read disturb count for a flash block that has endured 10K P/E cycles. The top graph breaks down the errors according to the change in cell state as a result of the errors; the bottom graph breaks down the errors into MSB and LSB errors. We make three observations from Figure 27. First, ER↔P1 errors increase significantly with read disturb count, whereas P1↔P2 and P2↔P3 errors do not. This is because the ER state threshold voltage distribution shifts significantly with read disturb count (see Figure 26), reducing the threshold voltage difference between the ER and P1 states. Second, MSB errors increase much faster than LSB errors with read disturb count because ER↔P1 errors are a type of MSB error, and they increase significantly with read disturb count. Third, the increase in RBER with read disturb count follows a linear trend (as shown by the dotted line in Figure 27b), which is similar to the observation made for planar NAND flash memory by prior work [5].

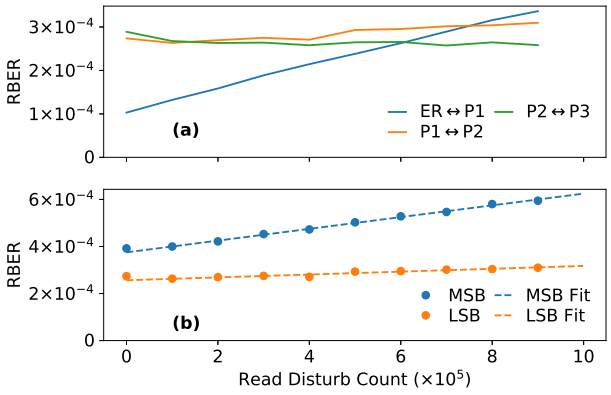


Figure 27: RBER vs. read disturb count, broken down by (a) the state transition of each flash cell, and (b) MSB or LSB page.

Figure 28 shows how the optimal read reference voltages change with read disturb count. The three subfigures show the optimal voltages for V_a, V_b, and V_c. We make two observations from this figure. First, the optimal voltages for V_b and V_c change by relatively little as the read disturb count increases (<3 voltage steps after 900K read disturbs), whereas the optimal V_a changes more with the read disturb count. This is because read disturb causes the threshold voltage distributions of lower-voltage states to change by a greater amount, which requires the read reference voltages separating the lower-voltage states (e.g., V_a) to change more. Second, the increase in the optimal V_a follows a linear trend with read disturb count, because the ER state threshold voltage distribution shifts linearly (as we see from Figure 26).

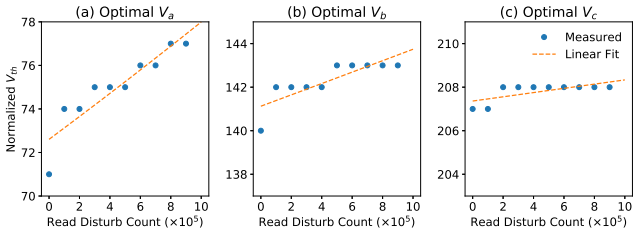


Figure 28: Optimal read reference voltages vs. read disturb count.

Insights. We compare the read disturb effect that we observe in 3D NAND flash memory to that observed in planar NAND flash memory by prior work [5]. We make the observation that, although RBER increases linearly with read disturb count in both 3D NAND and planar NAND flash memory, the slope of the increase (i.e., the sensitivity of the RBER to read disturb) at 10K P/E cycles is 96.7% *lower* in 3D NAND flash memory than that in planar NAND flash memory [5]. We believe that this difference in the sensitivity to read

disturb effect is due to the use of a larger process technology node (30–40 nm) in current 3D NAND flash memory. The comparable planar NAND flash memory results from prior work are collected on 20–24 nm planar NAND flash memory devices [5]. We expect the read disturb effect in 3D NAND flash memory to increase in the future as the process technology node size shrinks. We conclude that the 96.7% reduction in the read disturb effect we observe in 3D NAND flash memory compared to planar NAND flash memory is mainly caused by the difference in manufacturing process technology nodes of the two types of NAND flash memories.

A.4 Layer-to-Layer Process Variation

In this section, we present new results and analyses of the layer-to-layer process variation phenomenon in 3D NAND flash memory, in addition to the key findings we already presented in Section 4.2. We use the same methodology as we describe in Section 4.2.

Figure 29 shows how the threshold voltage distribution mean and standard deviation of each state changes with layer number, for a flash block that has endured 10K P/E cycles. Each subfigure in the top row shows the mean for a different state; each subfigure in the bottom row shows the standard deviation for a different state. We make two observations from this figure. First, the ER state threshold voltage increases by as much as 25 voltage steps as the layer number changes, while the mean threshold voltages of the other three states do not vary by much. This is because the threshold voltage of a cell in ER state is set after an erase operation, and the value it is set to is a function of manufacturing process variation and of wearout. In contrast, the threshold voltage of a cell in one of the other states (P1, P2, or P3) is set to a *fixed* target voltage value *regardless* of process variation [3, 69, 89, 91] (see Section 2.1). Since only the voltage of the ER state is affected by layer-to-layer process variation, only one of the read reference voltages, V_a, changes with the layer number, as we already observed in Figure 6. Second, the distribution widths of ER and P1 states (i.e., their standard deviations) increase in the top layers, and decrease in the bottom layers. This pattern is similar to the pattern of how the RBER changes with layer number, which we show in Figure 5 (Section 4.2). A wider threshold voltage distribution increases the overlap of neighboring distributions, leading to more errors in the top layer. However, the distribution widths of the P2 and P3 states mainly decrease as layer number increases. Unfortunately, we are unable to completely explain why mean threshold voltage and distribution width change differently with layer number for different states because we do not have exact circuit-level information about layer-to-layer process variation.

We conclude that layer-to-layer process variation significantly impacts the threshold voltage distribution and leads to large variations in RBER and optimal read reference voltages across layers.

A.5 Bitline-to-Bitline Process Variation

We perform an analysis of the variation of RBER and threshold voltage distribution along the y-axis (i.e., across groups of bitlines) for a flash block that has endured 10K P/E cycles. We use a similar methodology to our layer-to-layer process variation experiments (see Section 4.2).

Figure 30 shows how the threshold voltage distribution mean and standard deviation of each state changes with layer number, for a flash block that has endured 10K P/E cycles. Each subfigure in the top row shows the mean for a different state; each subfigure in the bottom row shows the standard deviation for a different state. Note that we normalize the number of bitlines from 0 to 100, by multiplying the actual bitline number with a constant, to maintain the anonymity of the chip vendors. We make two observations from this figure. First, the variations in mean threshold voltage and the distribution width (i.e., standard deviation) are much smaller

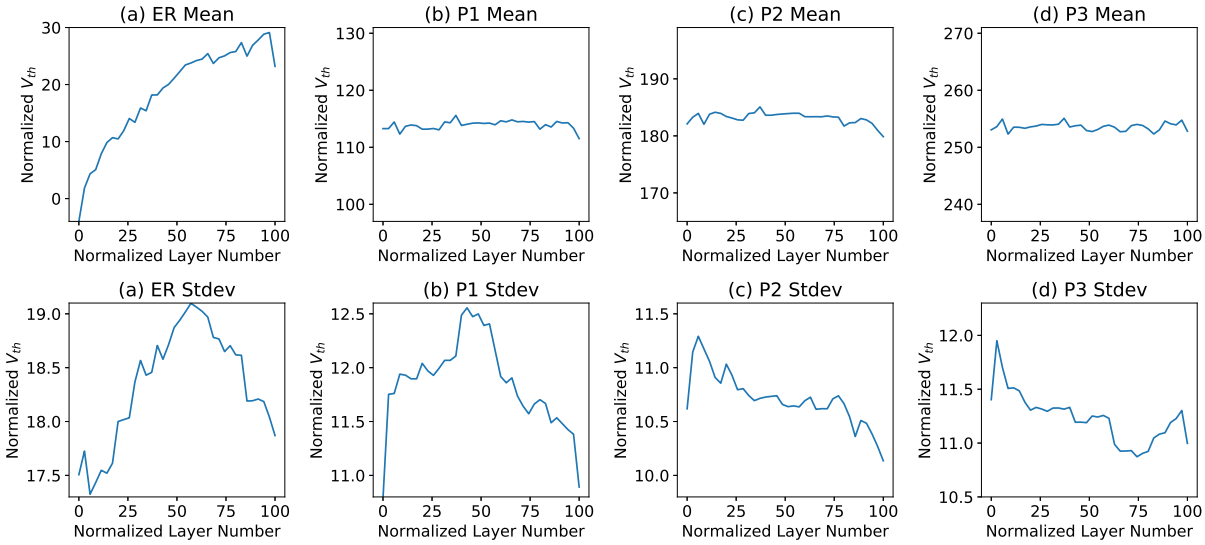


Figure 29: Mean and standard deviation of our Gaussian threshold voltage distribution model of each state, versus layer number.

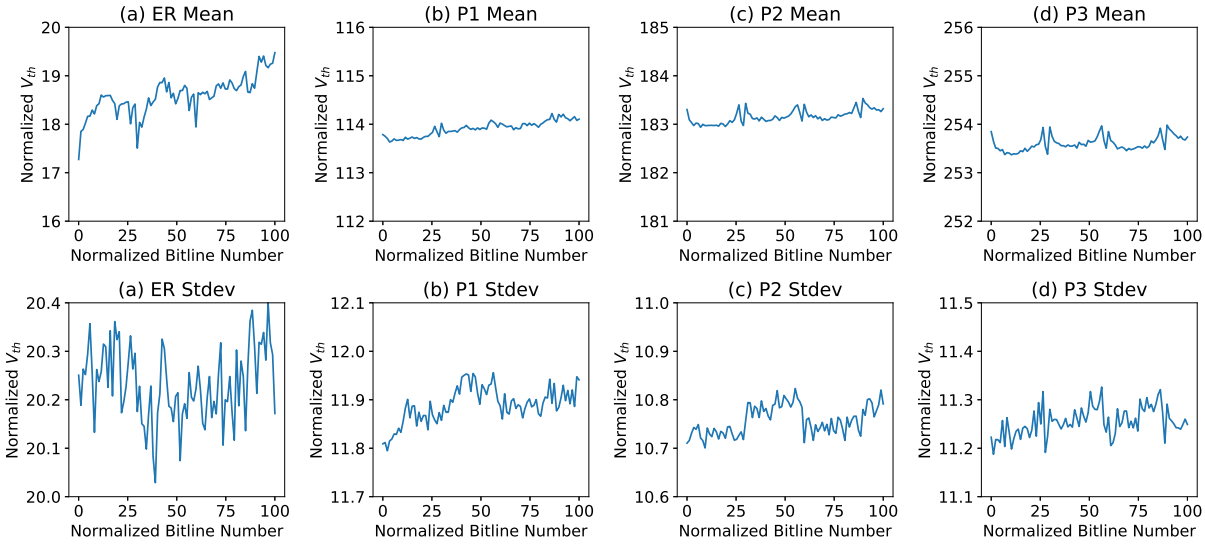


Figure 30: Mean and standard deviation of our Gaussian threshold voltage distribution model of each state, versus bitline number.

in this figure compared to that observed in Figure 29 for layer-to-layer variation (Appendix A.4). This indicates that bitline-to-bitline process variation is much smaller compared to layer-to-layer process variation in 3D NAND flash memory. Second, we observe that the pattern of the mean threshold voltage repeats periodically, for every 25 bitlines. We believe that this indicates a repetitive architecture in the way that the 3D NAND flash memory chip is organized (for example, each block may be made up of four arrays of flash cells that are connected together). Unfortunately, we cannot completely explain this behavior without access to circuit-level design information that is proprietary to NAND flash memory vendors.

Figures 31 and 32 show how the RBER and optimal read reference voltages change with bitline number, for a flash block that has endured 10K P/E cycles. We observe that neither RBER nor the optimal read reference voltages change by much across bitlines. This indicates that the changes that we observe in Figure 30 may not be significant enough to lead to variation in the reliability of different bitlines. We conclude that bitline-to-bitline process variation is much smaller than layer-to-layer process variation in 3D NAND flash memory.

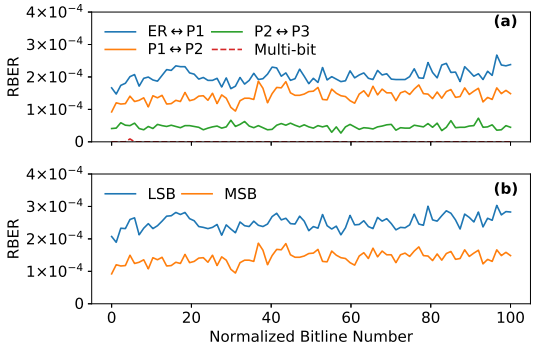


Figure 31: RBER vs. bitline number, broken down by (a) the state transition of each flash cell, and (b) MSB or LSB page.

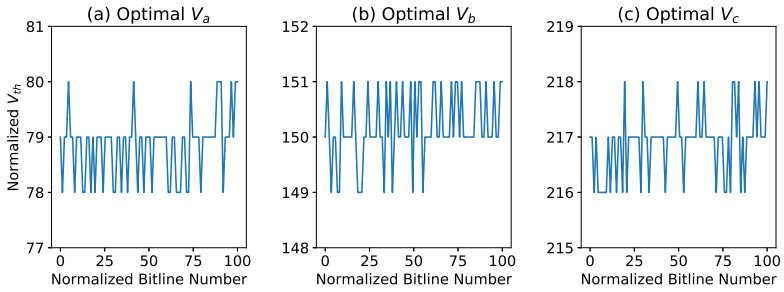


Figure 32: Optimal read reference voltages vs. bitline number.