nanomaterials

MDPI

Article

Investigation of Program Efficiency Overshoot in 3D Vertical Channel NAND Flash with Randomly Distributed Traps

Chanyang Park [1], Jun-Sik Yoon [1], Kihoon Nam [1], Hyundong Jang [1], Minsang Park [2] and Rock-Hyun Baek [1,*]

[1] Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea; chrypark@postech.ac.kr (C.P.); junsikyoon@postech.ac.kr (J.-S.Y.); namgee4970@postech.ac.kr (K.N.); hdjang@postech.ac.kr (H.J.)

[2] SK hynix Inc., Icheon 17336, Republic of Korea; minsang.park@sk.com

* Correspondence: rh.baek@postech.ac.kr

Abstract: The incremental step pulse programming slope (ISPP) with random variation was investigated by measuring numerous three−dimensional (3D) NAND flash memory cells with a vertical nanowire channel. We stored multiple bits in a cell with the ISPP scheme and read each cell pulse by pulse. The excessive tunneling from the channel to the storage layer determines the program efficiency overshoot. Then, a broadening of the threshold voltage distribution was observed due to the abnormal program cells. To analyze the randomly varying abnormal program behavior itself, we distinguished between the read variation and over−programming in measurements. Using a 3D Monte−Carlo simulation, which is a probabilistic approach to solve randomness, we clarified the physical origins of over−programming that strongly influence the abnormal program cells in program step voltage, and randomly distributed the trap site in the nitride of a nanoscale 3D NAND string. These causes have concurrent effects, but we divided and analyzed them quantitatively. Our results reveal the origins of the variation and the overshoot in the ISPP, widening the threshold voltage distribution with traps randomly located at the nanoscale. The findings can enhance understanding of random over−programming and help mitigate the most problematic programming obstacles for multiple−bit techniques.

Keywords: abnormal program cell; charge trap nitride; incremental step pulse programming; Monte−Carlo simulation; over−programming; overshoot; program efficiency; 3D NAND flash

Citation: Park, C.; Yoon, J.-S.; Nam, K.; Jang, H.; Park, M.; Baek, R.-H. Investigation of Program Efficiency Overshoot in 3D Vertical Channel NAND Flash with Randomly Distributed Traps. *Nanomaterials* **2023**, *13*, 1451. https://doi.org/10.3390/nano13091451

Academic Editors: Xianbin Li, Wei Zhang and Ming Xu

Received: 18 March 2023
Revised: 19 April 2023
Accepted: 22 April 2023
Published: 24 April 2023

1. Introduction

Three−dimensional (3D) vertical channel NAND flash memory structures [1–4] and n−bit multi−level cell (MLC) operation [5] enable storing 2^n ($n \geq 2$) states in a single cell. The nanostructures consist of cylindrical strings with nanowire thin film transistors vertically stacked, making it easier to store data and improve integration than when the transistors are placed horizontally. Those technologies have successfully helped manufacturers break through the process difficulties of sub−20−nm planar NAND flash. In the planar structure, the bit cost, which refers to the fabrication cost per bit, increases with the cost growth of lithography and etching processes for implementing fine patterns [4]. In the n−bit MLC NAND flash, an incremental step pulse programming (ISPP) algorithm checks whether target cells exceed the program verify (PV) level at every programming pulse. The ISPP scheme controls the tunneling current into the nano−scaled nitride film as intended and narrows the threshold voltage (V_{th}) distribution [6]. The ISPP slope is defined as the derivative V_{th} with respect to the program voltage (V_{pgm}). The ISPP slope is described simply as the threshold voltage difference (ΔV_{th_PLS}) between successive programming pulses (PLS) divided by the program step voltage (V_{step}), i.e., $\Delta V_{th\_PLS}/V_{step}$. Previous studies have attempted to increase the program efficiency or ISPP slope [7–9]. However,

programming (PGM) uniformity or control over the variation in the ISPP slope has become more significant because the V_{th} gap margin between program states becomes smaller, or even overlaps, as the number of bits per cell increases. Therefore, the uniformity and controllability for variation in PGM operation are crucial because higher bit densities require a narrower width of V_{th} distribution (W_{Vth}). This variation is fatal for the penta−(5−bit and 32 PGM states) and the hexa−level (6−bit and 64 PGM states) cell techniques [10,11], which require a narrower W_{Vth} (100–200 mV) than that of the quadruple−level cell [12]. Photonic integrated circuits (PICs) based on silicon and nitride were introduced to overcome the integration limitation of electronic integrated circuits (ICs), changing the signal medium from electricity to light [13,14]. Still, both ICs and PICs require research on the behaviors of subatomic particles such as electrons/traps and photons for ultrahigh density.

Although various factors cause abnormal program cells (APCs) in the high−density n−bit MLC, quantitatively identifying each cause is difficult. Thus, memory companies generally depend on defense algorithms, such as Bose–Chaudhuri–Hocquenghem [15], and low−density parity−check codes [16] instead of accurately identifying APCs. A simulation−based study analyzed the random behavior of APCs [17], but a huge amount of hardware data is needed to describe their randomness. A previous study has also investigated the change in the number of traps in nitride. It is well known that variations in the number of trapped electrons lead to an expansion of the V_{th} distribution [18]. In this study, via 3D Monte−Carlo simulation (MCS), the V_{th} value for the randomly distributed single trap charges was quantitatively confirmed toward the z−axis and radial directions, even when the same number of electrons were trapped. We reflected the capacitance ratio between the tunneling oxide and blocking oxide, including bisecting nitride films that were attached to each other. We also considered the energy band for the different trap positions. Then, 3D MCS was used to predict the probabilistic tunneling outcomes, the V_{th} distribution, and the overshoot in the ISPP slope. Moreover, the V_{th} distributions created by the superposition of the over−programming (O−PGM) and read variation were decomposed and analyzed.

Meanwhile, previous studies analyzed the parasitic effects on the V_{th} distribution and proposed solutions [19–21]. They investigated two−dimensional (2D) planar devices with different cell geometries, bias conditions, and cell characteristics. The solutions increase the reading and programming time for use in 3D structures and complicate the circuitry. Compared to planar floating gate devices, 3D NAND flash memory based on charge trap nitride (CTN) can mitigate the floating gate coupling effect [12,22] and upbuild the integration by stacking pairs of wordline (WL) in the upward direction. However, 3D NAND flash suffers from obstacles, such as an additional coupling component after conversion from 2D to 3D [23], as well as large variations in the electrical properties of the cells. These problems are mainly due to geometric effects [24] that do not exist in 2D structures. The previous studies analyzed the pristine or programmed states of devices. However, we decomposed the noisy factors and analyzed how the program procedure was influenced by the geometric effect in the tapered channel pillar [25]. In other words, we examined cell states both within the tunneling process and after the ISPP. The previous research quantitatively calculated how the random variation, which is caused by the geometric effect and the number of preoccupied charges, occurs during the tunneling process [25]. In this work, the extraction methodology was used and further developed to physically analyze how the V_{th} varies depending on the number of traps and the location when the number of traps is controlled based on the measured average ISPP slope. In particular, we randomized the trap position on a nanoscale using 3D MCS and confirmed the APC behavior with numerous measured cells to support the simulated data.

In this study, we measured V_{th} distributions in wafer (WF) and its packaged chip that are mass−produced NAND flash. Among the several factors widening the V_{th} distribution during the ISPP, the O−PGM inducing an ISPP slope of >1 was quantitatively separated. Furthermore, causal factors for the overshoot in the program efficiency and ISPP slope were analyzed using calibrated 3D MCS.

2. Materials and Methods

2.1. Classification of Abnormal Program Cell Components

2.1.1. Causal Factors Broadening the Threshold Voltage Distribution

To identify the causal factors broadening the V_{th} distribution, we experimented with a chip and WF of a 3D U−shaped stacked memory array transistor (SMArT) structure [1,3], as shown in Figure 1. V_{th} and ΔV_{th_PLS} were read for approximately 70,000–80,000 cells in each step of ISPP and a page unit. The nanowire thin film transistor cell of 3D NAND flash memory comprises metal−oxide−nitride−oxide−polysilicon. Moreover, it is penetrated by a macaroni oxide filler to mitigate the grain boundary effect, as shown in the inset of Figure 1. The basic memory operations act as a page or block unit, and the 3D NAND array is a stacked structure of multiple nanoscale thin film transistors with multiple materials. The stacked cells organize a string connected to a sourceline (SL) that supplies charges and a bitline (BL) that transmits/receives signals. They operate like a source and drain in a logic device, and the BL current is transferred to a page buffer, and the analog signal of the current, or V_{th}, is converted into digital data, or bit.

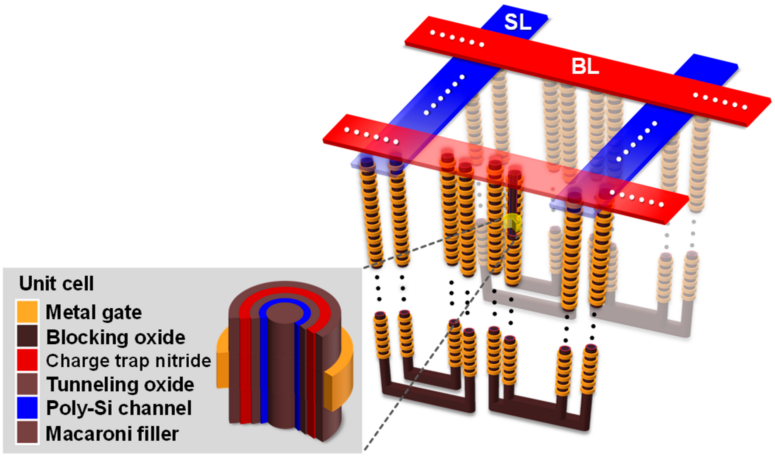

Figure 1. Schematic of a 3D U−shaped SMArT structure (inset: a nanowire thin film transistor cell composed of multiple materials based on charge trap nitride as storage layer). The channel strings are connected to the sourceline (SL) and the bitline (BL), which are the source of electrons, and a signal line connected to the page buffer, which is a circuit that converts the analog channel current into digital data (bits), respectively.

Considering the PV level as the stop line for programming, cells may persist to cross the level and span the larger V_{th} shift than V_{step}, compared to the ideally programmed cells with $W_{Vth} = V_{step}$. The V_{th} distributions stretch on both sides simultaneously due to factors such as under−programming (U−PGM) [26] due to the negative overshoot of ISPP slope or charge loss, background pattern dependency (BPD) [19], cell−to−cell interference (CTCI) [20,23], read variation, and intrinsic O−PGM. We experimented with the target cell after defining the data pattern of the adjacent cells with five PGM/erase cycles, the initialization procedure in Figure 2a. Furthermore, we used the ISPP scheme to store multiple bits in cells, as in Figure 2b. After the initialization condition of erasing the adjacent cells, we experimented with the page unit test. Compared to the planar arrays, 3D NAND flash arrays have a thicker pitch between layers and the nonmetallic storage layer, as shown in Figure 2c. Thus, BPD and CTCI can be ignored in our measurement conditions.

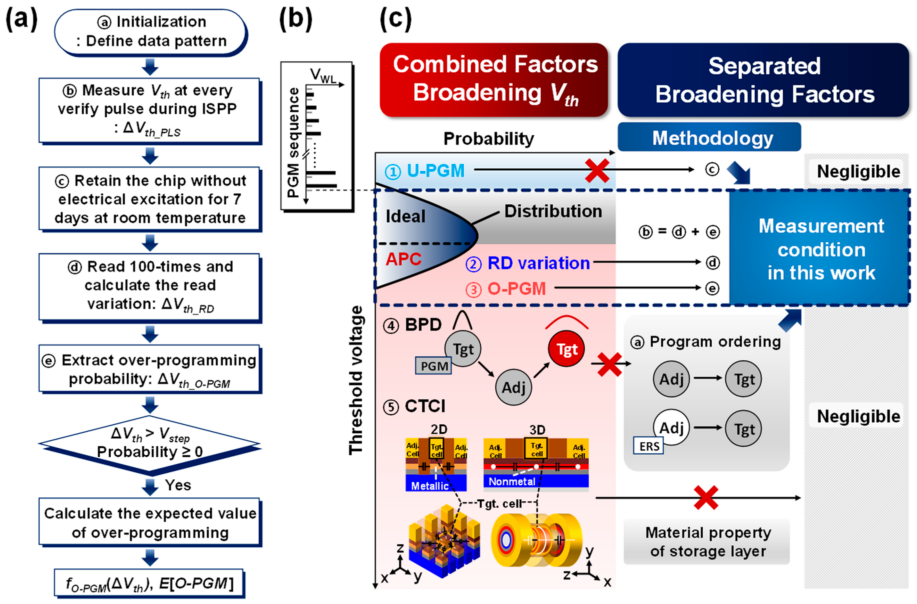

Figure 2. (**a**) The flow chart of the experimental procedure to analyze the program efficiency overshoot in a 3D NAND flash memory chip. (**b**) Wordline voltage (V_{WL}) versus PGM sequence during ISPP scheme. (**c**) The combined factors broadening V_{th} (① under−programming ② read variation ③ intrinsic over−programming ④ background pattern dependency from adjacent (Adj.) cells to a target (Tgt.) cell ⑤ cell−to−cell interference) and the separated broadening factors with the controlled measurement condition in this work. The methodology (ⓐ–ⓔ) to exclude ①, ④, and ⑤ corresponds to (**a**).

Furthermore, U−PGM is neglected because its probability is less than 0.3% in our measurement conditions. We define the APC as the cell programmed to a higher V_{th} than V_{step} from the PV level. For n−bit MLC, in which the number of program states becomes two to the power of n, the APC invades the next or higher program state. When the invasion of APC into the V_{th} of a higher program state exceeds the error correction threshold in the defense algorithms [15,16], the cells overlapping V_{th} distributions fail to be read. Because APC is harmful for the V_{th} distribution that stores multiple bits, identifying the causes of APC is essential. Interestingly, APC can be triggered during both the reading and PGM operations. When the target cells are verified incorrectly during, immediately after, or long after the PGM operation, the cells can be read as larger than the data that are actually stored (read variation), or they can be more programmed (intrinsic O−PGM). We analyzed the read variation in the V_{th} distribution in the next section. Moreover, Figure 2a shows the experimental procedure in this work, and Figure 2c represents the methodology corresponding to the procedure.

2.1.2. Read Variation Effect on Expansion of Threshold Voltage Distribution

To confirm the read variation effect on the expansion of V_{th} distribution, we ran the chip for at least seven days at room temperature, as in Figure 2a. Thus, retention error caused by material properties in nitride and oxide thin films while reading, including short−term retention within a few seconds [27,28], no longer occurred noticeably (Figure 2c). Then, the widths (W_{RD}s) of the read variation distribution (ΔV_{th_RD}) were obtained from the difference between the left and right tail bits of the read variation distribution at a 0.5% probability with repeatedly reading one hundred times, similar to previous research [25]. The W_{RD}s in the chip and WF are represented in Figure 3a.

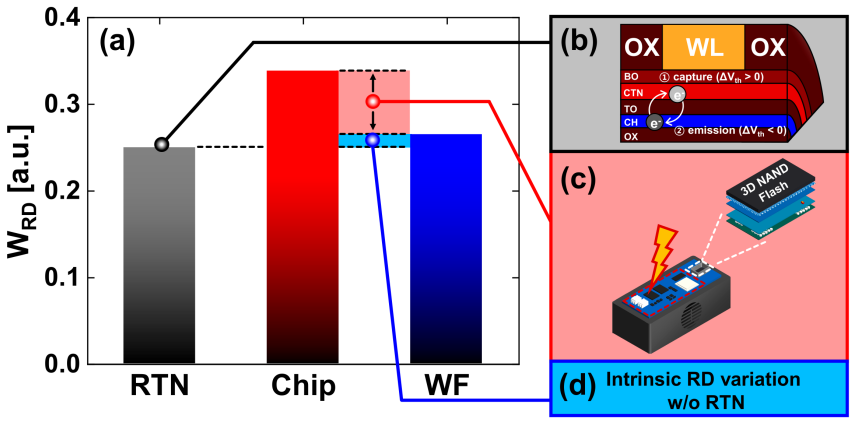

Figure 3. (**a**) Distribution width of read variation in chip and WF obtained by reading iteratively seven days after completion of programming. (**b**) RTN behavior with charge capture and emission. (**c**) Noise from the peripheral circuitry in the chip−level equipment. (**d**) Measured intrinsic read variation in WF except for RTN.

The W_{RD} of the chip is larger than that of the WF because the chip was mounted on a board with an additional peripheral circuit, which aided the experiments using numerous cells. However, the W_{RD} of the chip and WF were fixed regardless of the different V_{step}s and locations, as in the previous study [25].

Random telegraph noise (RTN) can be observed in the discrete fluctuations between upper and lower boundaries in V_{th} and channel current at tens of nanometer nodes [29–32], similar to the structures in Figure 3b,d [33,34]. However, the read variation measured from the entire page is generated by undesirable effects, such as RTN, carrier transport through random grain boundaries in the polycrystalline material [35,36], common source line noise [37], and noise from the peripheral circuitry to support chip measurements [25]. In a previous study [38], the causes of the V_{th} distribution widening were classified into program error and RTN. RTN was responsible for U−PGM, and the study [39] confirmed that it was programmed below the PV level. However, we considered that RTN also affects regions programmed above V_{step} from the PV level because the variation of V_{th} in the read operation occurred even after the program was completed. In particular, reflecting the read variation by the RTN in the previous studies [18,29–34,39], and despite the different optimal measurement conditions of RTN [31], we verified that the measured W_{RD}s of the chip and WF in Figure 3a,c,d are larger than the W_{RD} of RTN in Figure 3a,b. Hence, we confirmed that non−RTN factors influenced the read operation in Figure 3c,d.

2.2. Characterization of Over−Programming in Abnormal Program Cell

After eliminating the read variation from the V_{th} distribution as in the experimental procedure in Figure 2a, we can determine how the PGM operation itself affects APC [25]. In Figure 4, n indicates the ordinal number for the respective V_{step}s. We show experiments of ΔV_{th_PLS} from the (n − 1)th to nth PGM pulse to analyze the difference in V_{th}. In Figure 4, ΔV_{th_PLS} is the V_{th} difference between two consecutive PGM pulses, whereas ΔV_{th_RD} in Figures 3 and 4 is a variation in V_{th} during repetitive read operations.

The ΔV_{th_PLS} distribution of 70,000–80,000 cells was measured based on V_{step}. We split V_{step} to observe different trapping occurrences. In other words, a higher V_{step} will result in more trapping occurrences and a greater probability of O−PGM.

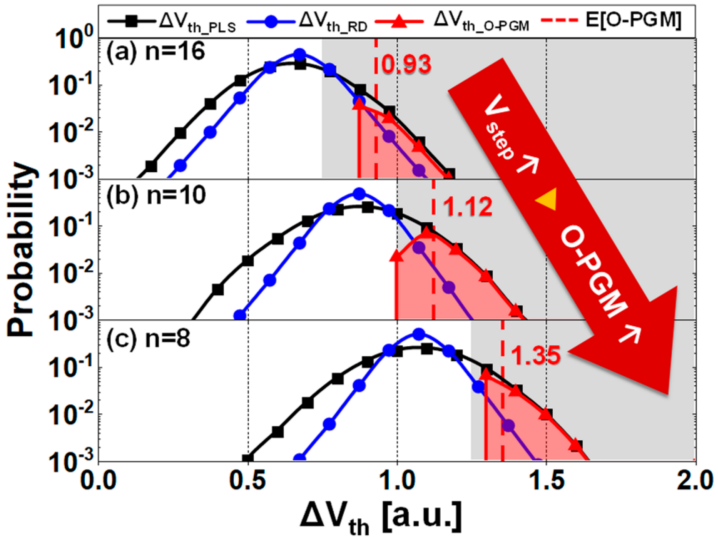

Figure 4. ΔV_{th_PLS} distribution (black lines) from (n − 1)th programming pulse to nth programming pulse during the ISPP, ΔV_{th_RD} by read variation (blue lines), ΔV_{th_O-PGM} by O−PGM (red lines), and the conditional expected values of O−PGM (red dashed lines) to (**a**) V_{step} = 0.75 a.u., (**b**) V_{step} = 1 a.u., and (**c**) V_{step} = 1.25 a.u., respectively; grey regions represent abnormal program cell regimes.

3. Results and Discussion

3.1. Extraction of Over−Programming and Dependence on the Program Step Voltage

Assume that read variation is the unique cause of APC. Then, ΔV_{th_PLS} should be distributed as read variation (blue line) centered on the V_{step} × ISPP slope in Figure 4. However, the real ΔV_{th_PLS} (black line), was more widely distributed than a read variation (blue line), which indicates that O−PGM in the cell itself is also responsible for APC.

After determining the value of n when the program efficiency begins to saturate, which is when most APCs manifest, the ΔV_{th_PLS} distributions were obtained and are shown in Figure 4. The ordinal numbers for V_{step} = 0.75, 1, and 1.25 a.u. are 16, 10, and 8, respectively. In Figure 4a–c, the number of programming pulses decreases as V_{step} increases because a higher V_{step} requires fewer pulses to reach a similar PV level. The extraction method of O−PGM from the APC can be described as follows [25], and it produces the subtraction ΔV_{th_RD} from ΔV_{th_PLS} on the condition in Figure 2a. The intrinsic O−PGM (ΔV_{th_O-PGM}) can be obtained through this method.

The extracted O−PGM is expressed as a distribution rather than a single value, so a quantitative index is required to compare how much O−PGM occurred. Therefore, we should define the probability of O−PGM ($f_{O-PGM}(\Delta V_{th})$) and calculate its conditional expected values ($E[O-PGM]$), which is similar to calculating the conditional expected value in the previous study [25].

We defined the probability functions for ΔV_{th}s by PGM pulse ($f_{PLS}(\Delta V_{th})$), read variation ($f_{RD}(\Delta V_{th})$), and O−PGM ($f_{O-PGM}(\Delta V_{th})$). The $f_{RD}(\Delta V_{th})$ should be translated to the peak of $f_{PLS}(\Delta V_{th})$, where most cells are distributed on ΔV_{th_PLS} to extract $f_{O-PGM}(\Delta V_{th})$. Therefore, $g_{RD}(\Delta V_{th})$, the probability of read variation centered on V_{step}×ISPP slope can be described as the following equation:

$$g_{RD}(\Delta V_{th}) = f_{RD}\left(\Delta V_{th} - V_{step} \times ISPP\ slope\right) \quad (1)$$

The probability of O−PGM can be obtained by the following equation:

$$f_{O-PGM}(\Delta V_{th}) = f_{PLS}(\Delta V_{th}) - g_{RD}(\Delta V_{th}), \quad (2)$$

where $\Delta V_{th} > V_{step}$ and the probability is nonnegative. The extraction procedure is the same as calculating the probability of intersection with the complement of read variation from the total probability of ΔV_{th_PLS}.

Finally, we can obtain a quantitative value for the $f_{O-PGM}(\Delta V_{th})$ that considers the conditional probability so that the summation of the probability is equal to one where $\Delta V_{th} > V_{step}$. We used the conditional expected value of $f_{O-PGM}(\Delta V_{th})$ ($E[O-PGM]$) as the following equation:

$$E[O-PGM] = \frac{\int_{V_{step}}^{\infty} \Delta V_{th} f_{O-PGM}(\Delta V_{th}) d(\Delta V_{th})}{\int_{V_{step}}^{\infty} f_{O-PGM}(\Delta V_{th}) d(\Delta V_{th})}, \quad (3)$$

where the APC is in the region of $\Delta V_{th} > V_{step}$, as previously defined in Section 2.1.1 and Figure 2a.

Then, we obtained the values of $E[O-PGM]$ = 0.93, 1.12, and 1.35 a.u. for V_{step} = 0.75, 1, and 1.25 a.u., respectively. Comparing the $E[O-PGM]$, the O−PGM contributes to APC by increasing as V_{step} increases. The influence of the PGM itself was confirmed by removing the read variation component independent of V_{step}s. For the intrinsic O−PGM to be ideal, a higher activation energy for detrapping is required than the read variation prone to capture/emission described in Figure 3. It means the O−PGM cells have experienced stronger Fowler–Nordheim tunneling than the process design in ISPP. Furthermore, the measured ΔV_{th} was affected by the trap characteristics, such as trap locations and energy levels rather than the fixed trap fluctuation that could easily be detrapped/trapped or captured/emitted in the short term. This is analyzed in the next section.

3.2. In−Depth Analysis of Over−Programming

The number of traps can be controlled by changing V_{step} in the measurement. Moreover, we simulated V_{th} shifts, i.e., O−PGM, in the simulation without the read variation from the repetitive capture/emission and the external noise in Figure 3. MCS is used for the probabilistic approach to phenomena in which the unpredictable and inseparable variables are parametrized [17]. In particular, MCS can implement random and nonuniform trap behaviors due to the variation in nanotechnology manufacturing on the deca−nanometer scale. In this study, 3D MCS obtained the probabilistic distributions determined by the states of traps in the 3D nanowire vertical channel NAND flash structure, such as the number of traps, their locations, and energy levels. They are material properties of nanoscale films that are difficult to observe by measurement. We confirmed the random ISPP slopes in a previous work [25]. Further developed, the program sequence from the beginning to end, as compared to the previous study [25], is shown in Figures 2a and 5a. The measured ISPP slopes (average ~0.8) from the 70,000–80,000 samples in Figure 5a were calibrated using 3D MCS, which randomly changes the number of traps, their locations, and their energy levels, shown in Figure 5b. The accurate number of traps cannot be determined due to randomization, and the change in the trap number causes the overshoot and the undershoot of program efficiency. Thus, we chose the range for shooting and the reference value based on the measured program efficiency. To realize an average ISPP slope of 0.8, 190–265 nitride traps in a cell were occupied by electrons or a single PGM pulse (Figure 5b).

The number of traps was calculated for a nanowire FET device, including fringing capacitances [40]. The fringing capacitances are the extrinsic parasitic capacitances between the gate and other conductive materials beyond the effective channel length, such as the WLs of adjacent cells, the channel regions of the intercell, and the channel below the gate conducted by the fringing field. In other words, the average value of the ISPP slope in 3D MCS was calibrated to match the measurement, as shown in the triangles in Figure 5. APCs had ISPP slopes above 1 and even up to 2. The slopes indicate that the effect of the programming itself was larger than the read variation. In Figure 5, we focused on the cells entering the V_{th} region from PV + V_{step} to PV + 2 × V_{step}, defined as the APC region. Although the dimension of the gate stack and the pitch between the materials are designed at the manufacturing engineering to produce an average ISPP slope, program efficiency overshoot occurs in realistic operation. The Gaussian distribution of the injection probability has been previously defined as the Poisson statistics and V_{th} distribution by

RTN in a floating gate device [18]. We proceeded with the analysis of 3D structures based on CTN, the latest process technology, by extending the analysis performed in previous studies based on the floating gate device [12,18,22,30]. Despite the similar V_{step} and lower ISPP slope (average ~0.8 compared to the floating gate with 1), the measured V_{th} distribution is wider than the one described by the Poisson statistics and RTN [18].

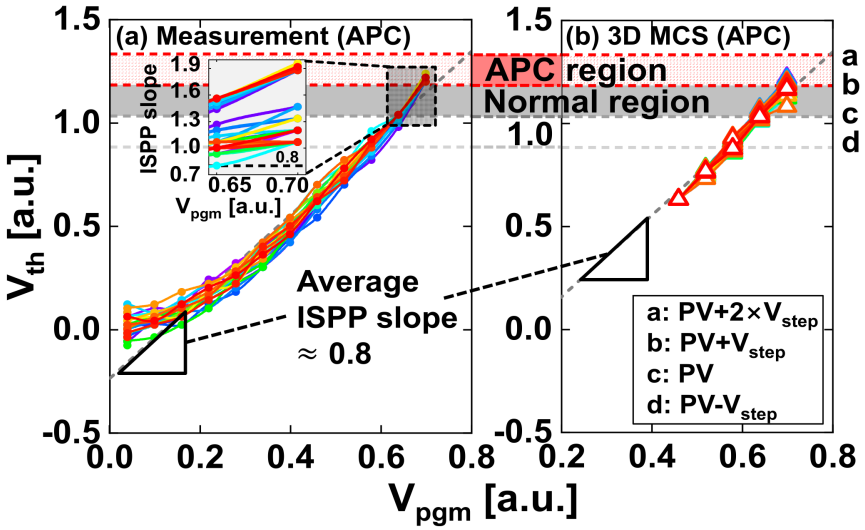

Figure 5. V_{th} versus V_{pgm} of APCs in (**a**) measurement and (**b**) 3D MCS. The inset of (**a**) is the overshoot in the ISPP slope for APCs near the PV level. Different colored lines in (**a**,**b**) represent the curves in the rightmost APCs of the ΔV_{th_PLS} distribution. The measured average ISPP slope was calibrated using 3D MCS considering fringing capacitances. The average ISPP slopes in measurement and 3D MCS are indicated by triangles, respectively. The cells in the V_{th} region from a (PV + 2 × V_{step}) to b (PV + V_{step}) are APCs, the cells from b (PV + V_{step}) to c (PV) are normal, and the cells from c (PV) to d (PV − V_{step}) are before passing the programming reference, c (PV).

There are more APCs for each PGM state of an n−bit MLC in Figure 5a, but only the rightmost APCs of the ΔV_{th_PLS} distribution are displayed to distinguish them visually due to their many overlaps. The rightmost APCs exhibit the worst case of O−PGM.

Figure 6 displays (a) the energy band diagram of a simulated polysilicon channel and (b) the contour plot showing that ΔV_{th_PLS} concentrates around the top of the barrier. We assume that material parameters, such as maximum trap concentration, maximum trapping distance from the tunneling oxide for electrons, and the blocking oxide for holes and capture cross−section in the device physics of the simulation, do not change with the trap location. These device parameters affect the Poisson statistics and represent the injection probability for Fowler–Nordheim tunneling [18]. Splitting the parameters confirms the influence of the electron trapping process and the injection probability. However, the objective of the 3D MCS was to confirm the V_{th} to the positions of the trapped electrons due to the infinitesimal positioning of the traps in the nanoscale film rather than changing the probability of capturing the electrons. Therefore, since we analyzed 3D MCS with electrons trapped, we can analyze the effect of the trapped position of electrons on V_{th}, as shown in Figure 6b.

For a given number of traps that have randomly distributed trap sites in every PGM pulse, the nitride traps near the top of the barrier of the channel induce a greater V_{th} shift than traps at other locations. Furthermore, at the nitride/tunneling oxide interface trap, the barrier height based on the energy level of the SL increases significantly. In particular, the barrier height increases the most when the interface trap is near the top of the barrier. Thus, V_{th} increases, as shown in Figure 6.

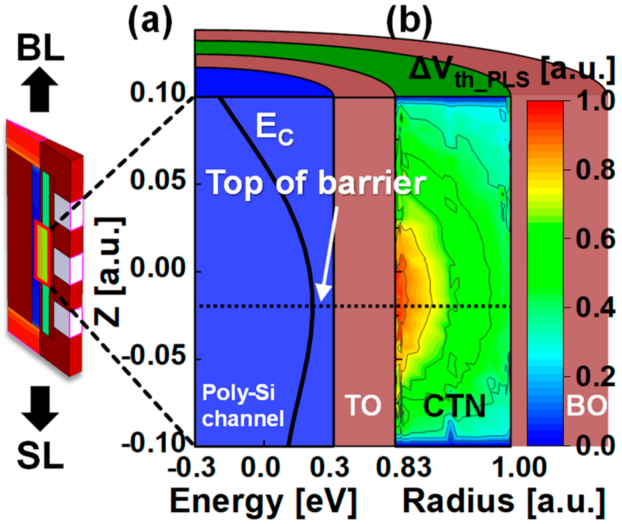

Figure 6. (**a**) Band diagram of the poly−Si channel in the read operation after PGM operation and (**b**) contour plot with the magnitude of ΔV_{th_PLS} highlighted in different colors at different nitride traps in a target cell in the string.

Figure 7 shows the energy band diagram of a polysilicon channel enlarged by the channel length of the target cell, and the inset of Figure 7a shows the energy band diagram of a channel string. Trap #2 is located at the same z−coordinate as the top of barrier in the channel. Figure 7b shows the dependence of barrier height on the trapping position from trap sites #1–#5. In particular, trap #2 has the largest barrier height in the channel and can be expected to have the largest V_{th}, as shown in Figure 6. Positions #1 and #5, far from the top of the barrier, are located approximately at the barrier height of cells without the trap.

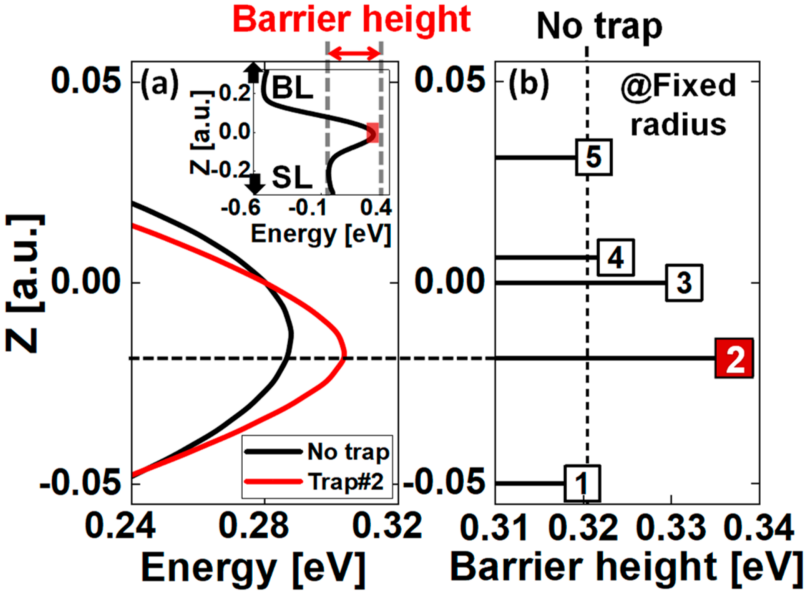

Figure 7. (**a**) Band diagram of a channel enlarged by the channel length of the target cell in an inset, with and without a single trap (or before and after programming) at trap #2 in (**b**), (**b**) change in barrier height with respect to the position of single trap numbered from 1 to 5.

Figure 8 shows the increase in W_{PLS} as the interface trap increases within a 1 Å range from the tunneling oxide to the nitride. Figure 8b shows that W_{PLS} increases by 0.015–0.031 a.u., and Figure 8c shows that the ratio of the O−PGM to the total ΔV_{th_PLS} distribution increases by 0.17–1.23% as the ratio of the single interface traps to the total

traps in the nitride increases. Datapoints showed that the ratio of the interface traps to the total traps in the absence of interface traps is zero. These data points are omitted in Figure 8b,c because we adopted a semi−logarithmic scale. The traps in the interfacial region and near the intercell migrate more in the lateral direction than in the noninterface region, resulting in a wider left side of the V_{th} distribution. Because the interface traps are closer to the tunneling oxide than the noninterface traps, V_{th} is calculated as higher, causing a widening of the right side of the V_{th} distribution. In addition, the interface traps screen the channel charges from the gate bias and the channel current reduces; then, the V_{th} becomes larger than bulk nitride traps. Thus, as the number of single interface traps increases, the ΔV_{th_PLS} distribution widens. Moreover, as shown in the contour spacing of Figure 6b, the difference in ΔV_{th_PLS} according to the distance from the nitride/tunneling oxide interface is largest at the top of the barrier. The difference decreases as it moves away from the top of the barrier in the z−direction. These results indicate that, although the number of traps is the same, the location of the traps affects the likelihood of a V_{th} shift occurring. Overall, in the nitride region, the quality of the nitride/tunneling oxide interface is an essential factor in reducing the O−PGM and the extent of the V_{th} distribution. In particular, more defects at and near the interface (nonideal storage) will manifest in the scale−down process because the mitigation treatment for them is more difficult due to the thermal budget for the multiple materials. To make matters worse, their proportion to total traps, including bulk nitride traps (ideal storage), increases. More O−PGM will therefore take place, and the V_{th} will broaden, as in Figure 8.

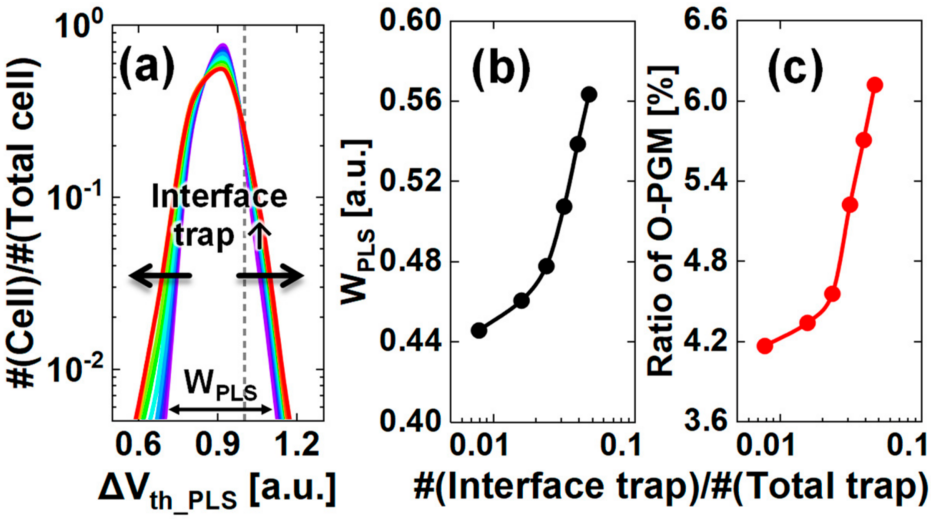

Figure 8. (**a**) Increase in ΔV_{th_PLS} distribution width with an increasing number of nitride/tunneling oxide interface traps within a 1 Å range. The lines in (**a**) are rainbow−colored from purple to red as the interface trap increases. The distribution broadens in the direction of the arrow as the interface trap increases. The dashed line indicates V_{step} values obtained experimentally. (**b**) Width of the ΔV_{th_PLS} distribution (W_{PLS}) and (**c**) increasing ratio of O−PGM to the number of nitride/tunneling oxide interface traps per total number of traps in the nitride region.

We utilized the method to separate the read variation and O−PGM from the V_{th} [25]. Predictive 3D modeling of V_{th} distribution in gate-all-around cylindrical nanowire devices, including parasitic capacitances [40], which requires a large amount of measurement data and quantitatively confirms the influence of each on APCs. As technology nodes evolve and the number of bits stored per cell increases, the V_{th} gap margin between adjacent program states becomes smaller. Thus, data failures will increase because the overlapped data between the program states exceed the error correction threshold. Furthermore, 3D MCS implements the uncertainty and randomness of traps in 3D nanotechnology manufacturing with shrinking trends. In the next−generation process technology, such as the biconcave nitride [2], designing where to bend can be applied to find the optimal point in minimizing

the program efficiency overshoot and reducing process cost. Utilizing the study of trap properties inside and at the boundary between the thin film materials to ensure data reliability will help determine the priority criteria for the PGM and the read strategies to be properly modified.

4. Conclusions

The factors that widen the V_{th} distribution or attribute to the abnormal program cells were quantitatively classified and confirmed by experimental tests from the mass−produced chip and wafer. Furthermore, the intrinsic over−programming from the entire abnormal program behaviors was obtained by the extraction method, then analyzed using 3D MCS based on the electronic properties of the nanoscale thin film, such as the number of traps, their locations, and energy. As traps are closer to the top of the barrier in the conduction band and interface, it is more critical for over−programming despite having the same number of traps as in the nitride volume of the target cell. These material properties of the nanostructure can enhance the understanding of random over−programming and help determine an approach to mitigate the most troublesome program problem in an ISPP scheme for the n−bit MLC technique. We analyzed the impact of the read variation and over−programming on abnormal program cells. Furthermore, the analysis of the failures in verifying operations between program pulses during the ISPP, which is the situation when the programming for cells is successful but redundant program pulses are applied, can be advanced in future research. As a result, the width of the V_{th} distribution can become smaller and the V_{th} gap margin wider.

Author Contributions: Conceptualization, C.P., H.J., M.P. and R.-H.B.; data curation, C.P. and J.-S.Y.; formal analysis, C.P. and M.P.; funding acquisition, R.-H.B.; investigation, C.P.; methodology, C.P., M.P. and R.-H.B.; project administration, R.-H.B.; resources: J.-S.Y., M.P. and R.-H.B.; software, C.P., J.-S.Y. and K.N.; supervision: R.-H.B.; validation, C.P., J.-S.Y., M.P. and R.-H.B.; visualization: C.P.; writing—original draft, C.P.; writing—review and editing, J.-S.Y., K.N., H.J., M.P. and R.-H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Semiconductor Industry Collaborative Project between POSTECH and SK hynix Inc.; the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT), grant number NRF-2020R1A4A4079777; Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT), grant number 2019-0-01906, Artificial Intelligence Graduate School Program (POSTECH); the Ministry of Trade, Industry & Energy (MOTIE) and Korea Semiconductor Research Consortium (KSRC) support program for the development of the future semiconductor devices, grant number 20020265; BK 21 FOUR program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data can be made available upon request from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Choi, E.-S.; Park, S.-K. Device considerations for high density and highly reliable 3D NAND flash cell in near future. In Proceedings of the IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 10–13 December 2012.
2. Jang, J.; Kim, H.-S.; Cho, W.; Cho, H.; Kim, J.; Shim, S.I.; Jang, Y.; Jeong, J.-H.; Son, B.-K.; Kim, D.W.; et al. Vertical cell array using TCAT (Terabit Cell Array Transistor) technology for ultra high density NAND flash memory. In Proceedings of the IEEE Symposium on VLSI Technology, Kyoto, Japan, 15–17 June 2009.
3. Katsumata, R.; Kito, M.; Fukuzumi, Y.; Kido, M.; Tanaka, H.; Komori, Y.; Ishiduki, M.; Matsunami, J.; Fujiwara, T.; Nagata, Y.; et al. Pipe-shaped BiCS flash memory with 16 stacked layers and multi-level-cell operation for ultra high density storage devices. In Proceedings of the IEEE Symposium on VLSI Technology, Kyoto, Japan, 15–17 June 2009.
4. Tanaka, H.; Kido, M.; Yahashi, K.; Oomura, M.; Katsumata, R.; Kito, M.; Fukuzumi, Y.; Sato, M.; Nagata, Y.; Matsuoka, Y.; et al. Bit cost scalable technology with punch and plug process for ultra high density flash memory. In Proceedings of the IEEE Symposium on VLSI Technology, Kyoto, Japan, 12–14 June 2007.

5. Lee, S.; Lee, Y.-T.; Han, W.-K.; Kim, D.-H.; Kim, M.-S.; Moon, S.-H.; Cho, H.C.; Lee, J.-W.; Byeon, D.-S.; Lim, Y.-H.; et al. A 3.3 V 4 Gb four-level NAND flash memory with 90 nm CMOS technology. In Proceedings of the IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, USA, 15–19 February 2004.
6. Suh, K.-D.; Suh, B.-H.; Lim, Y.-H.; Kim, J.-K.; Choi, Y.-J.; Koh, Y.-N.; Lee, S.-S.; Kwon, S.-C.; Choi, B.-S.; Yum, J.-S.; et al. A 3.3 V 32 Mb NAND flash memory with incremental step pulse programming scheme. *IEEE J. Solid-State Circuits* **1995**, *30*, 1149–1156.
7. Chen, W.-C.; Lue, H.-T.; Hsiao, Y.-H.; Hsu, T.-H.; Lin, X.-W.; Lu, C.-Y. Charge storage efficiency (CSE) effect in modeling the incremental step pulse programming (ISPP) in charge-trapping 3D NAND flash devices. In Proceedings of the IEEE IEDM, Washington, DC, USA, 7–9 December 2015.
8. Kim, Y.; Seo, J.Y.; Lee, S.-H.; Park, B.-G. A new programming method to alleviate the program speed variation in three-dimensional stacked array NAND flash memory. *J. Semicond. Technol. Sci.* **2014**, *14*, 566–571. [CrossRef]
9. Lue, H.-T.; Hsu, T.-H.; Wang, S.-Y.; Lai, E.-K.; Hsieh, K.-Y.; Liu, R.; Lu, C.-Y. Study of incremental step pulse programming (ISPP) and STI edge effect of BE-SONOS NAND flash. In Proceedings of the IEEE International Reliability Physics Symposium (IRPS), Phoenix, AZ, USA, 27 April–1 May 2008.
10. Fujiwara, M.; Morooka, T.; Nagashima, S.; Kato, T.; Fukuda, N.; Kariya, N.; Ogura, T.; Kurusu, T.; Shimada, Y.; Ishikawa, T.; et al. 3D semicircular flash memory cell: Novel split-gate technology to boost bit density. In Proceedings of the IEEE IEDM, San Francisco, CA, USA, 7–11 December 2019.
11. Aiba, Y.; Tanaka, H.; Maeda, T.; Sawa, K.; Kikushima, F.; Miura, M.; Fujisawa, T.; Matsuo, M.; Sanuki, T. Cryogenic operation of 3D flash memory for new applications and bit cost scaling with 6-bit per cell (HLC) and beyond. In Proceedings of the IEEE Electron Devices Technology and Manufacturing Conference (EDTM), Chengdu, China, 8–11 April 2021.
12. Shibata, N.; Maejima, H.; Isobe, K.; Iwasa, K.; Nakagawa, M.; Fujiu, M.; Shimizu, T.; Honma, M.; Hoshi, S.; Kawaai, T.; et al. A 70 nm 16 Gb 16-level-cell NAND flash memory. *IEEE J. Solid-State Circuits* **2008**, *43*, 929–937. [CrossRef]
13. Xiang, C.; Jin, W.; Bowers, J.E. Silicon nitride passive and active photonic integrated circuits: Trends and prospects. *Photon. Res.* **2022**, *10*, A82–A96. [CrossRef]
14. Dai, D.; Liang, D.; Cheben, P. Next-generation silicon photonics: Introduction. *Photon. Res.* **2022**, *10*, NGSP1–NGSP3. [CrossRef]
15. Massey, J.L. Shift-register synthesis and BCH decoding. *IEEE. Trans. Inf. Theory* **1969**, *15*, 122–127. [CrossRef]
16. Gallager, R.G. Low-density parity-check codes. *IRE Trans. Inf. Theory* **1962**, *8*, 21–28.
17. Amoroso, S.M.; Maconi, A.; Mauri, A.; Compagnoni, C.M.; Greco, E.; Camozzi, E.; Vigano, S.; Tessariol, P.; Ghetti, A.; Spinelli, A.S.; et al. 3D Monte Carlo simulation of the programming dynamics and their statistical variability in nanoscale charge-trap memories. In Proceedings of the IEEE IEDM, San Francisco, CA, USA, 6–8 December 2010.
18. Compagnoni, C.M.; Spinelli, A.S.; Gusmeroli, R.; Beltrami, S.; Ghetti, A.; Visconti, A. Ultimate accuracy for the NAND flash program algorithm due to the electron injection statistics. *IEEE Trans. Electron Devices* **2008**, *55*, 2695–2702. [CrossRef]
19. Jung, T.-S.; Choi, Y.-J.; Suh, K.-D.; Suh, B.-H.; Kim, J.-K.; Lim, Y.-H.; Koh, Y.-N.; Park, J.-W.; Lee, K.-J.; Park, J.-H.; et al. A 117-mm^2 3.3-V only 128-Mb multilevel NAND flash memory for mass storage applications. *IEEE J. Solid-State Circuits* **1996**, *31*, 1575–1583. [CrossRef]
20. Park, K.T.; Kang, M.; Kim, D.; Hwang, S.-W.; Choi, B.Y.; Lee, Y.-T.; Kim, C.; Kim, K. A zeroing cell-to-cell interference page architecture with temporary LSB storing and parallel MSB program scheme for MLC NAND flash memories. *IEEE J. Solid-State Circuits* **2008**, *43*, 919–928. [CrossRef]
21. Jung, T.-S.; Choi, Y.-J.; Suh, K.-D.; Suh, B.-H.; Kim, J.-K.; Lim, Y.-H.; Koh, Y.-N.; Park, J.-W.; Lee, K.-J.; Park, J.-H.; et al. A 3.3 V 128 Mb multi-level NAND flash memory for mass storage applications. In Proceedings of the IEEE ISSCC, San Francisco, CA, USA, 10 February 1996.
22. Lee, J.-D.; Hur, S.-H.; Choi, J.-D. Effects of floating-gate interference on NAND flash memory cell operation. *IEEE Electron Device Lett.* **2002**, *23*, 264–266.
23. Shim, K.-S.; Choi, E.-S.; Jung, S.-W.; Kim, S.-H.; Yoo, H.-S.; Jeon, K.-S.; Joo, H.-S.; Oh, J.-S.; Jang, Y.-S.; Park, K.-J.; et al. Inherent issues and challenges of program disturbance of 3D NAND flash cell. In Proceedings of the IEEE International Memory Workshop (IMW), Milan, Italy, 20–23 May 2012.
24. Kim, K.T.; An, S.W.; Jung, H.S.; Yoo, K.-H.; Kim, T.W. The effects of taper-angle on the electrical characteristics of vertical NAND flash memories. *IEEE Electron Device Lett.* **2017**, *38*, 1375–1378. [CrossRef]
25. Park, C.; Yoon, J.-S.; Nam, K.; Jang, H.; Park, M.S.; Baek, R.-H. Quantitative analysis of irregular channel shape effects on charge-trapping efficiency using massive 3D NAND data. *Mater. Sci. Semicond. Process.* **2023**, *157*, 107333. [CrossRef]
26. Mizoguchi, K.; Takahashi, T.; Aritome, S.; Takeuchi, K. Data-retention characteristics comparison of 2D and 3D TLC NAND flash memories. In Proceedings of the IEEE IMW, Monterey, CA, USA, 14–17 May 2017.
27. Woo, C.; Lee, M.; Kim, S.; Park, J.; Choi, G.-B.; Seo, M.-S.; Noh, K.H.; Kang, M.; Shin, H. Modeling of charge loss mechanisms during the short term retention operation in 3-D NAND flash memories. In Proceedings of the IEEE Symposium on VLSI Technology, Kyoto, Japan, 9–14 June 2019.
28. Choi, B.; Jang, S.H.; Yoon, J.; Lee, J.; Jeon, M.; Lee, Y.; Han, J.; Lee, J.; Kim, D.M.; Kim, D.H.; et al. Comprehensive evaluation of early retention (fast charge loss within a few seconds) characteristics in tube-type 3-D NAND flash memory. In Proceedings of the IEEE Symposium on VLSI Technology, Honolulu, HI, USA, 14–16 June 2016.

29. Nicosia, G.; Mannara, A.; Resnati, D.; Paolucci, G.M.; Tessariol, P.; Spinelli, A.S.; Lacaita, A.L.; Goda, A.; Compagnoni, C.M. Characterization and modeling of temperature effects in 3-D NAND flash arrays—Part II: Random telegraph noise. *IEEE Trans. Electron Devices* **2018**, *65*, 3207–3213. [CrossRef]
30. Kim, S.; Lee, M.; Choi, G.-B.; Lee, J.; Lee, Y.; Cho, M.; Ahn, K.-O.; Kim, J. RTS noise reduction of 1Y-nm floating gate NAND flash memory using process optimization. In Proceedings of the IEEE IRPS, Monterey, CA, USA, 19–23 April 2015.
31. Jeong, M.-K.; Joe, S.-M.; Seo, C.-S.; Han, K.-R.; Choi, E.; Park, S.-K.; Lee, J.-H. Analysis of random telegraph noise and low frequency noise properties in 3-d stacked NAND flash memory with tube-type poly-Si channel structure. In Proceedings of the IEEE Symposium on VLSI Technology, Honolulu, HI, USA, 12–14 June 2012.
32. Kurata, H.; Otsuga, K.; Kotabe, A.; Kajiyama, S.; Osabe, T.; Sasago, Y.; Narumi, S.; Tokami, K.; Kamohara, S.; Tsuchiya, O. The impact of random telegraph signals on the scaling of multilevel flash memories. In Proceedings of the IEEE Symposium on VLSI Circuits, Honolulu, HI, USA, 15–17 June 2006.
33. Hsieh, C.-C.; Lue, H.-T.; Hsu, T.-H.; Du, P.-Y.; Chiang, K.-H.; Lu, C.-Y. A Monte Carlo simulation method to predict large-density NAND product memory window from small-array test element group (TEG) verified on a 3D NAND flash test chip. In Proceedings of the Symposium on VLSI Technology, Honolulu, HI, USA, 14–16 June 2016.
34. Jia, X.; Jin, L.; Zhou, W.; Lu, J.; Amoroso, S.M.; Brown, A.R.; Lee, K.-H.; Asenov, P.; Lin, X.-W.; Liu, H.; et al. Investigation of random telegraph noise under different programmed cell V_t levels in charge trap based 3D NAND flash. *IEEE Electron Device Lett.* **2022**, *43*, 878–881. [CrossRef]
35. Resnati, D.; Mannara, A.; Nicosia, G.; Paolucci, G.M.; Tessariol, P.; Spinelli, A.S.; Lacaita, A.L.; Compagnoni, C.M. Characterization and modeling of temperature effects in 3-D NAND flash arrays—Part I: Polysilicon-induced variability. *IEEE Trans. Electron Devices* **2018**, *65*, 3199–3206. [CrossRef]
36. Kang, H.-J.; Jeong, M.-K.; Joe, S.-M.; Seo, J.-H.; Park, S.-K.; Jin, S.H.; Park, B.-G.; Lee, J.-H. Effect of traps on transient bit-line current behavior in word-line stacked NAND flash memory with poly-Si body. In Proceedings of the IEEE Symposium on VLSI Technology, Honolulu, HI, USA, 9–12 June 2014.
37. Byeon, D.-S.; Lee, S.-S.; Lim, Y.-H.; Kang, D.; Han, W.-K.; Kim, D.-H.; Suh, K.-D. A comparison between 63 nm 8 Gb and 90 nm 4Gb multi-level cell NAND flash memory for mass storage application. In Proceedings of the IEEE Asian Solid-State Circuits Conference, Hsinchu, Taiwan, 1–3 November 2005.
38. Paolucci, G.M.; Compagnoni, C.M.; Spinelli, A.S.; Lacaita, A.L.; Goda, A. Fitting cells into a narrow V_T interval: Physical constraints along the lifetime of an extremely scaled NAND flash memory array. *IEEE Trans. Electron Devices* **2015**, *62*, 1491–1497. [CrossRef]
39. Joe, S.-M.; Yi, J.-H.; Park, S.-K.; Kwon, H.-I.; Lee, J.-H. Position-dependent threshold-voltage variation by random telegraph noise in NAND Flash memory strings. *IEEE Electron Device Lett.* **2010**, *31*, 635–637.
40. Zou, J.; Xu, Q.; Lou, J.; Wang, R.; Huang, R.; Wang, Y. Predictive 3-D modeling of parasitic gate capacitance in gate-all-around cylindrical silicon nanowire MOSFETs. *IEEE Trans. Electron Devices* **2011**, *58*, 3379–3387.