IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 9 December 2022, accepted 11 January 2023, date of publication 18 January 2023, date of current version 24 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3237967

RESEARCH ARTICLE

# Holistic Optimization of Trap Distribution for Performance/Reliability in 3-D NAND Flash Using Machine Learning

**KIHOON NAM[1], (Member, IEEE), CHANYANG PARK[1], (Member, IEEE), HYEOK YUN[1], (Member, IEEE), JUN-SIK YOON[1], (Member, IEEE), HYUNDONG JANG[1], (Member, IEEE), KYEONGRAE CHO[1], (Member, IEEE), MIN SANG PARK[2], HYUN-CHUL CHOI[3], (Member, IEEE), AND ROCK-HYUN BAEK[1], (Member, IEEE)**

[1]Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea
[2]SK hynix Inc., Icheon-si, Gyeonggi-do 17336, Republic of Korea
[3]Department of Electrical Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea

Corresponding author: Rock-Hyun Baek (rh.baek@postech.ac.kr)

This work was supported in part by the Semiconductor Industry Collaborative Project between POSTECH and SK hynix Inc., and in part by the National Research Foundation of Korea (NRF) through the Korea Government (MSIT) under Grant NRF-2020R1A4A4079777 and Grant NRF-2020M3F3A2A02082436.

**ABSTRACT** A machine learning (ML) method was used to optimize the trap distribution of the charge trap nitride (CTN) to simultaneously improve its performance/reliability (P/R) characteristics, which are tradeoffs in 3-D NAND flash memories. Using an artificial neural network (ANN), we modeled the relationship between trap distributions and P/R characteristics. The ANN was trained using a large experimentally-calibrated technology computer-aided design (TCAD) simulation dataset. The gradient descent method was adapted to optimize the trap distribution, achieving the best P/R characteristics based on the well-trained ANN. Eventually, we found the best trap profile distributed in both space and energy. In particular, the energetic trap distribution had a larger impact on the P/R characteristics than that of the spatial trap distribution. Furthermore, in terms of the P/R characteristics, it was generally preferable to increase all inputs of the energetic trap distribution. However, the acceptor-like trap energy level ($E_{TA}$) and its standard deviation ($\sigma_{EA}$) caused a tradeoff between P/R characteristics; therefore, ML was used to determine their optimal points. The proposed ML method allows the optimization of trap distribution to obtain the best P/R characteristics rapidly and quantitatively. Our findings could be used as a guideline for determining the physical properties of CTN in 3-D NAND flash cells.

**INDEX TERMS** 3D NAND flash, charge trap nitride, device optimization, machine learning, performance, reliability, trap distribution.

## I. INTRODUCTION

NAND flash memory is a typical nonvolatile memory that is widely used in data servers and portable electronic devices. The principal advantage of NAND flash is its low bit cost owing to its high density, which has been facilitated by the conversion of conventional 2-D architectures to 3-D architectures and the adoption of charge trap nitride (CTN) as a storage layer [1], [2], [3], [4]. In particular, CTN has enabled the fine scaling of devices and improved their endurance and reliability compared with conventional floating-gate-based devices [5]. However, as 3-D NAND cells are more stacked and scaled, precise control is required to deposit CTN. In addition, uncertainty in determining the physical

The associate editor coordinating the review of this manuscript and approving it for publication was Cristian Zambelli.


This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

properties of CTN can result in tradeoffs between their erase/program (ERS/PGM) speed and retention characteristics [4], [6], [7], [8].

In our previous study, we only considered the ERS/PGM characteristics with the energetic properties of traps in CTN [9], [10]. However, in this study, we have attempted to solve more complicated problems combined with retention characteristics by optimizing the critical properties of CTN, that is, uneven spatial and energetic trap distributions, by using a deeper neural network. Unlike previous studies, new trap properties were investigated by observing all characteristics simultaneously rather than individually. Previously, the spatial and energetic trap distributions were extracted using the trap spectroscopy by charge injection and sensing (TSCIS) method [11], [12] and the retention model [13], [14]. By optimizing these controllable parameters comprehensively, we first attempted to simultaneously improve the performance, incremental step pulse erasing/programming (ISPE/ISPP) characteristics [15], and the reliability, retention characteristics in the 3-D NAND cell, which exhibit a tradeoff relationship.

Machine learning (ML) in this study can help in determining the optimal values exactly and rapidly. ML has recently been used to predict and optimize nanoscale transistors [16], [17], [18]. An artificial neural network (ANN) modeled the relationship between the trap distributions as input parameters and the absolute values of the threshold voltage shift ($|\Delta V_{th}|$) as the output parameters. The ANN that was trained using a validated simulation dataset determined the best trap distribution exhibiting the best performance/reliability (P/R) characteristics at the same time. Consequently, we proposed using ML methods to rapidly and quantitatively determine the optimal trap distributions. Furthermore, we analyzed sensitive trap parameters to determine the P/R characteristics. The remainder of this paper is organized as follows. Section II describes the methods used to train and optimize the ANN. Section III presents the results of the optimization and trends using ML-based analysis. Finally, the conclusions are presented in Section IV.

## II. SIMULATION STRUCTURE AND METHODS

Fig. 1 shows a part of the 3-D NAND flash string used in the simulation. The cylindrical structure was simulated using Sentaurus technology computer-aided design (TCAD) [19]. Both ends of the string were connected to a bit line (BL) and a source line (SL), and $V_{th}$ was extracted from the BL current vs. gate voltage curves of the selected word line ($WL_{Sel.}$). Unselected WLs ($WL_{Unsel.}$s) were stacked around the $WL_{Sel.}$ to consider adjacent cells. The gate stack consisted of a metal gate, blocking oxide (BOX), charge trap nitride (CTN), bandgap-engineering tunneling oxide (BE-TOX), polysilicon channel, and Macaroni filler. No grain boundaries in the polysilicon channel were assumed to focus on the CTN materials. Drift-diffusion transport, Shockley–Read–Hall recombination, mobility (doping, high-field, and interface-dependent), and Hurkx band-to-band tunneling models were implemented in the simulation. A nonlocal tunneling model was used to describe the movement of charges between the channel and the CTN. In addition, the Poole–Frenkel model was considered, and the trapped charges were calculated in the CTN. The features of the simulation structure were obtained from an actual 3-D NAND cell [9]. In this study, the geometric features of the actual device were fixed only to see the effect of trap distributions in the CTN on the P/R characteristics.

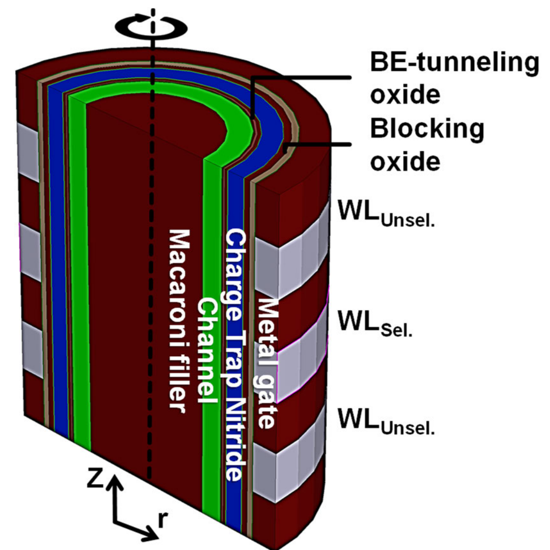
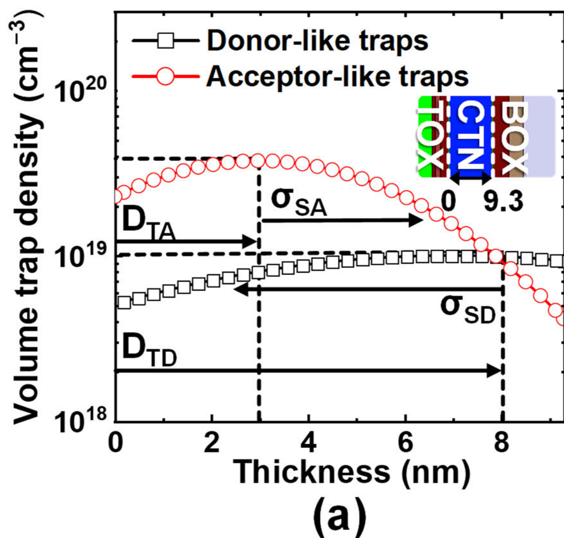



**FIGURE 1.** Schematic of 3-D NAND flash string along the channel and the cell gate stack. The device is simulated in a cylindrical coordinate. $V_{th}$ is extracted from the BL current vs. gate voltage of $WL_{sel.}$.

All traps were spatially and energetically distributed in the CTN. In addition, each distribution was composed of donor- and acceptor-like traps, which captured the holes and electrons, respectively. Moreover, Gaussian trap distributions of the peak, mean, and standard deviations were assumed.

Fig. 2(a) shows the spatial trap distributions. The spatial trap distribution assumes that the volume trap density is not uniform in the CTN. Each distribution is composed of peak spatial depths ($D_{TD}$ and $D_{TA}$) and standard deviations of the spatial traps ($\sigma_{SD}$ and $\sigma_{SA}$). $\sigma_{SD}$ and $\sigma_{SA}$ contain values that sufficiently reveal the deviation of the volume trap density. A thickness of 0 nm indicates the BE-TOX/CTN interface; the closer trapped charges are to zero, the more they are near the BE-TOX/CTN interface. In other words, the trap has a spatial distribution based on the CTN thickness.

Fig. 2(b) shows the energetic trap distributions. The energetic trap distribution indicates the energetic position of the traps in the bandgap of the CTN. Each distribution is composed of peak trap densities ($N_{TD}$ and $N_{TA}$), energy levels at $N_{TD}$ and $N_{TA}$ ($E_{TD}$ and $E_{TA}$), and standard deviations of the energetic traps ($\sigma_{ED}$ and $\sigma_{EA}$). $E_{TD}$ is obtained from the valence band energy ($E_V$), while $E_{TA}$ is obtained from the conduction band energy ($E_C$). Thus, a value close to zero



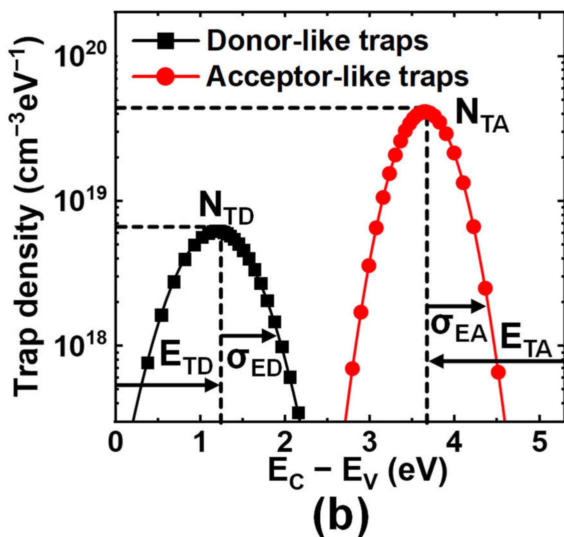
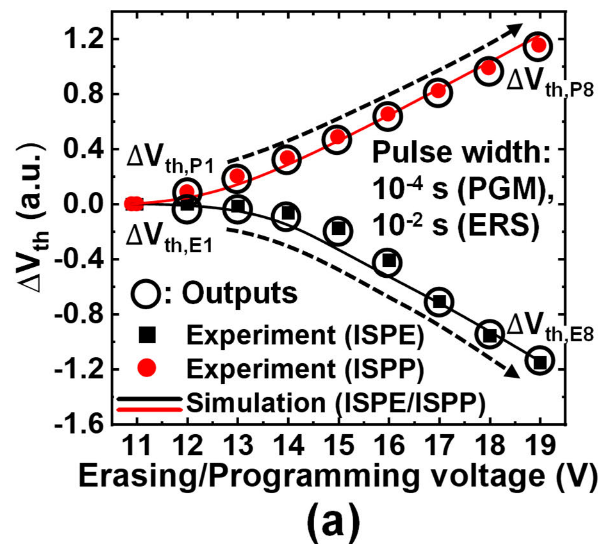




**FIGURE 2. (a) Spatial trap distributions and (b) energetic trap distributions of donor/acceptor-like traps in the CTN. At each thickness point, the volume trap density is determined by the integral of the energetic traps.**



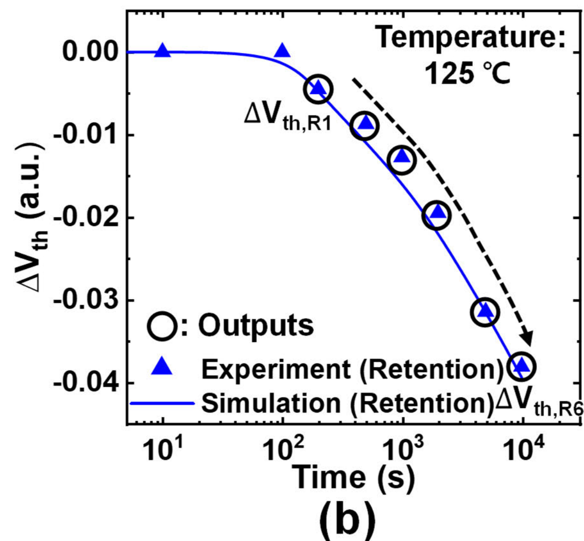




**FIGURE 3. (a) $\Delta V_{th}$ vs. erasing/programming voltage in ISPE/ISPP characteristics and (b) $\Delta V_{th}$ vs. time in retention characteristics. Good agreements are obtained between simulation and experimental data. Moreover, adjacent cells ($WL_{Unsel.}$s) are in the ERS state.**

indicates a shallow trap. Furthermore, the value of the volume trap density in Fig. 2(a) is the integral of the energetic trap distributions at a specific location on the CTN thickness.

Fig. 3(a) shows $\Delta V_{th}$ as a function of the erasing/programming voltage; the ISPE/ISPP schemes. The ISPE/ISPP scheme is a basic ERS/PGM method that rapidly narrows the $V_{th}$ distribution by increasing the operating voltage. Therefore, the higher $|\Delta V_{th}|$, the better the performance at the same operating voltage. The ISPE scheme starts from the PGM state, whereas the ISPP scheme starts from the ERS state. Sixteen $\Delta V_{th}$ values (open circle) are used as outputs at each erasing/programming voltage of 12 V to 19 V.

Fig. 3(b) shows $\Delta V_{th}$ as a function of time. The experiments were conducted at a high temperature (125°C) because $\Delta V_{th}$ barely changed at room temperature. The initial $V_{th}$ was in the PGM state to verify the charge transport, and $WL_{Unsel.}$s were in the ERS state. In terms of reliability, the retention characteristics improves when $|\Delta V_{th}|$ decreases at a given time. Six $\Delta V_{th}$ values (open circle) are used as outputs at each time of $2 \times 10^2$ s to $10^4$ s.

Under the same bias conditions, the simulation data (lines) were calibrated using the experimental data (symbols), as shown in Fig. 3. They are in good agreement when adjusting the inputs of the trap distributions to calibrated values, as shown in Table 1. Table 1 also summarizes the ranges of the input parameters to be modeled in the artificial neural network (ANN). The inputs of the energetic trap distributions, that is, $N_{TD}$, $E_{TD}$, $\sigma_{ED}$, $N_{TA}$, $E_{TA}$, and $\sigma_{EA}$, were referenced from [13], [14], [20], [21], and [22]. In addition, the inputs of the spatial trap distributions, that is, $D_{TD}$, $\sigma_{SD}$, $D_{TA}$, and $\sigma_{SA}$, were extracted from [12] and [23]. The ranges of each input were sufficiently feasible, and trap parameters could be determined by adjusting the gas flow ratios [24]. Consequently,

**TABLE 1.** Calibrated values and ranges of input parameters.

| Input parameters | Minimum | Maximum | Calibrated |
|---|---|---|---|
| Peak density of donor-like traps, $N_{TD}$ ($cm^{-3}\cdot eV^{-1}$) | $4.00 \times 10^{18}$ | $1.00 \times 10^{19}$ | $6.00 \times 10^{18}$ |
| Energy level of donor-like traps at $N_{TD}$, $E_{TD}$ (eV) | 0.80 | 1.60 | 1.20 |
| Standard deviation of energetic donor-like traps, $\sigma_{ED}$ (eV) | 0.10 | 0.40 | 0.40 |
| Peak spatial depth of donor-like traps, $D_{TD}$ (nm) | 0.00 | 9.30 | 4.70 |
| Standard deviation of spatial donor-like traps, $\sigma_{SD}$ (nm) | 2.00 | 20.00 | 10.00 |
| Peak density of acceptor-like traps, $N_{TA}$ ($cm^{-3}\cdot eV^{-1}$) | $4.00 \times 10^{19}$ | $1.00 \times 10^{20}$ | $5.00 \times 10^{19}$ |
| Energy level of acceptor-like traps at $N_{TA}$, $E_{TA}$ (eV) | 0.80 | 2.20 | 1.64 |
| Standard deviation of energetic acceptor-like traps, $\sigma_{EA}$ (eV) | 0.10 | 0.40 | 0.40 |
| Peak spatial depth of acceptor-like traps, $D_{TA}$ (nm) | 0.00 | 9.30 | 4.70 |
| Standard deviation of spatial acceptor-like traps, $\sigma_{SA}$ (nm) | 2.00 | 20.00 | 10.00 |

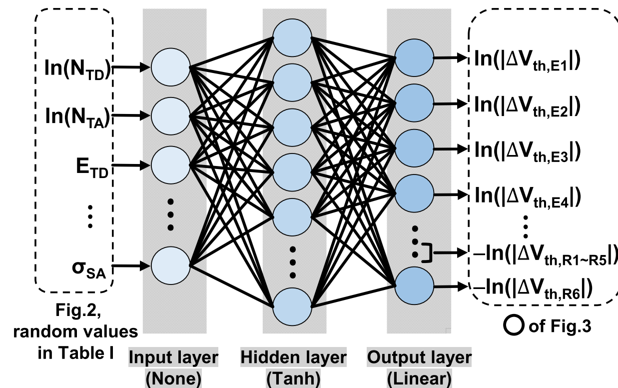



**FIGURE 4.** Schematic of 10 inputs (Fig. 2) for 22 outputs (Fig. 3). MLP was used to train the nonlinear function. After training, an optimization process was performed to find the inputs that made the outputs large by the gradient descent method.

the controllable input parameters were randomized between their minimum and maximum values, having different $|\Delta V_{th}|$ values accordingly.

Fig. 4 shows a schematic of the ANN used in the ML process. The inputs are the 10-feature of the trap distributions, as shown in Fig. 2, and the outputs are the 22-feature of $|\Delta V_{th}|$ (open circle), as shown in Fig. 3. A multilayer perceptron (MLP) was used to model this relationship. The MLP has been widely used to build complex nonlinear models [25]. The feedforward network had one hidden layer with 30 nodes, and its activation function was a hyperbolic tangent. The Levenberg–Marquardt method was adopted for the training algorithm, and the cost function was the mean squared error (MSE). After successful training, the gradient descent method was used as an optimization algorithm in the inverse

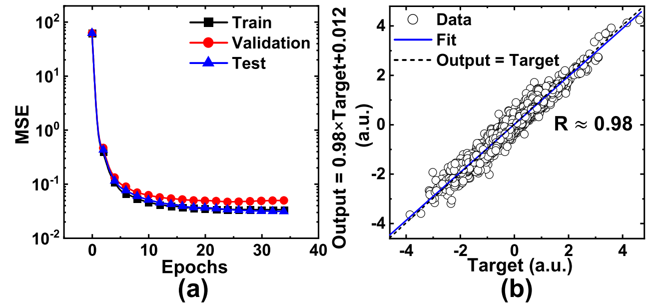



**FIGURE 5.** (a) MSE vs. epochs of train/validation/test sets and (b) a regression plot showing the match between prediction and target in the test sets.

direction. Furthermore, by using ML-based optimization, we can precisely find the specific inputs that are challenging for engineers to find empirically. In the outputs, $|\Delta V_{th}|$ values were used to provide the same direction and were logged to prevent abnormal values (such as $< 0$). However, the $|\Delta V_{th,R1\sim R6}|$ of the retention characteristics were negative because they had different directions for improvement. These results will be discussed in Section III. In the inputs, $N_{TD}$ and $N_{TA}$ were logged, and then all parameters were standardized for rapid and accurate training. For example, it takes 4741 seconds on average to predict one device using TCAD, whereas only 21 seconds using ML. All ML operations were run in MATLAB [26] using a personal desktop computer with AMD Ryzen 7 5800X (8-Core, 3.8 GHz) and 32 GByte RAM.

Table 2 summarizes the Spearman correlations between the inputs and outputs. They have a positive correlation as space changes to red and a negative correlation as it changes to blue. For the ISPE/ISPP characteristics, outputs are good to be larger, but the opposite is true for the retention characteristics. Therefore, correlation values with the same sign are likely to have a tradeoff between the P/R characteristics. For example, when $E_{TA}$ increases, the $|\Delta V_{th,E1\sim E8}|$ of the early phase is likely to decrease (degrade), whereas $|\Delta V_{th,R1\sim R6}|$ is likely to decrease (improve). It can also be observed in the $N_{TA}$ and $\sigma_{EA}$. However, it is difficult to analyze the effect of multiple inputs on multiple outputs with only correlations. Therefore, in this study, we attempted to analyze the complex physical properties of the CTN in a 3-D NAND flash device using ML and suggested the best set of inputs for the tradeoffs.

## III. RESULTS AND DISCUSSION

### A. $|\Delta V_{TH}|$ PREDICTION AND OPTIMIZATION USING ARTIFICIAL NEURAL NETWORK

In this section, we present the results of training and optimization using the ANN. We want $|\Delta V_{th}|$ of ISPE/ISPP characteristics to be larger and $|\Delta V_{th}|$ of retention characteristics to be smaller, but they inherently have tradeoffs. Therefore, the goal of optimization is to find the best set of inputs that make the best P/R characteristics as simultaneously as possible according to an estimated value, which is our evaluation criterion.

**TABLE 2.** Spearman correlation between input and output parameters.

| Input par. | $\|\Delta V_{th,En}\|$ 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | $\|\Delta V_{th,Pn}\|$ 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | $\|\Delta V_{th,Rn}\|$ 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_{TD}$ | -0.01 | -0.01 | -0.01 | -0.00 | -0.00 | 0.00 | 0.01 | -0.01 | 0.48 | 0.48 | 0.48 | 0.45 | 0.42 | 0.37 | 0.31 | 0.25 | -0.13 | -0.13 | -0.13 | -0.13 | -0.14 | -0.14 |
| $E_{TD}$ | 0.00 | 0.01 | 0.02 | 0.02 | 0.02 | 0.01 | 0.01 | 0.01 | 0.07 | 0.06 | 0.06 | 0.08 | 0.08 | 0.07 | 0.06 | 0.05 | -0.41 | -0.43 | -0.44 | -0.45 | -0.47 | -0.48 |
| $\sigma_{ED}$ | -0.02 | -0.02 | -0.02 | -0.01 | 0.00 | 0.01 | 0.02 | -0.02 | 0.73 | 0.72 | 0.71 | 0.66 | 0.61 | 0.53 | 0.44 | 0.36 | -0.08 | -0.10 | -0.11 | -0.12 | -0.13 | -0.15 |
| $D_{TD}$ | -0.03 | -0.04 | -0.03 | -0.02 | -0.01 | -0.00 | -0.01 | -0.03 | -0.05 | -0.07 | -0.07 | -0.08 | -0.08 | -0.07 | -0.07 | -0.06 | -0.02 | -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| $\sigma_{SD}$ | -0.04 | -0.04 | -0.04 | -0.02 | -0.01 | -0.01 | 0.00 | -0.04 | 0.21 | 0.21 | 0.20 | 0.19 | 0.17 | 0.15 | 0.12 | 0.09 | -0.10 | -0.10 | -0.10 | -0.10 | -0.10 | -0.10 |
| $N_{TA}$ | 0.21 | 0.28 | 0.46 | 0.48 | 0.44 | 0.41 | 0.40 | 0.35 | -0.00 | 0.08 | 0.06 | 0.16 | 0.23 | 0.29 | 0.36 | 0.40 | 0.11 | 0.11 | 0.10 | 0.10 | 0.10 | 0.10 |
| $E_{TA}$ | -0.83 | -0.74 | -0.34 | 0.17 | 0.31 | 0.40 | 0.45 | 0.44 | 0.03 | 0.02 | 0.03 | 0.05 | 0.06 | 0.07 | 0.08 | 0.08 | -0.75 | -0.73 | -0.73 | -0.72 | -0.70 | -0.69 |
| $\sigma_{EA}$ | 0.40 | 0.41 | 0.60 | 0.58 | 0.56 | 0.55 | 0.53 | 0.48 | 0.01 | 0.12 | 0.09 | 0.22 | 0.31 | 0.42 | 0.50 | 0.58 | 0.22 | 0.21 | 0.20 | 0.19 | 0.18 | 0.17 |
| $D_{TA}$ | -0.12 | -0.17 | -0.19 | -0.16 | -0.13 | -0.12 | -0.10 | -0.08 | 0.00 | -0.04 | -0.04 | -0.08 | -0.09 | -0.10 | -0.10 | -0.09 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 |
| $\sigma_{SA}$ | 0.16 | 0.23 | 0.28 | 0.24 | 0.23 | 0.21 | 0.20 | 0.18 | 0.04 | 0.10 | 0.10 | 0.14 | 0.16 | 0.17 | 0.18 | 0.19 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |

Fig. 5(a) shows the MSEs of the train/validation/test sets with epochs. After preprocessing, the total number of samples was 2290, and $\Delta V_{th}$ could be extracted from the 57250 BL current vs. gate voltage curves, resulting in the ISPE/ISPP and retention curves of Figs. 1 and 2. The samples were divided into train, validation, and test sets using weights of 70, 15, and 15, respectively. The train sets were used to train the models, and the validation sets were used to measure the performance of models built on the train set. A specific trained model with the best performance was selected using the validation sets. Finally, the test sets were used to measure the expected performance of the model after the validation sets were determined. Furthermore, the smaller the MSE, the better the ANN training. Subsequently, the ANN exhibited the best performance, with an MSE $= 4.5903 \times 10^{-2}$ in 24 epochs after the set validation epochs are over in 34 epochs (10 epochs away).

Fig. 5(b) shows the agreement between the targets and predicted values in the sets. The targets are the existing simulated values, and the corresponding predicted values lie on the y-axis. A regression of the test sets shows an ideal alignment in which R = 0.98 (close to 1). It implies that the trained ANN successfully modeled the nonlinear function between the inputs and outputs. Consequently, a well-trained ANN can predict the correct outputs for random inputs within a finite range.

Figs. 6(a), (b), and (c) show comparisons of the ISPE/ISPP and retention characteristics between the simulated and optimized values. Solid lines denote the results of simulated outputs from the randomized inputs shown in Table 1. The open symbols indicate the improved outputs predicted using the ANN. The loss function for the superior P/R characteristics can be expressed as follows:

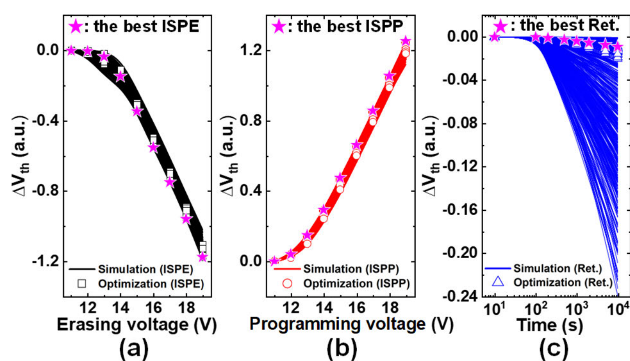



**FIGURE 6.** Comparison of the results between simulation and optimization in (a) the ISPE, (b) ISPP, and (c) retention characteristics. The larger $|\Delta V_{th}|$ in the ISPE/ISPP characteristics and the smaller $|\Delta V_{th}|$ in the retention characteristics, the better the P/R characteristics.

$$loss\left(\hat{y}_j\right) = \exp\left(-\hat{y}_j\right), j = 1, \ldots, 22 \quad (1)$$

where the prediction $\hat{y}_j$is the matrix of $\ln(|\Delta V_{th,E1\sim E8}|)$, $\ln(|\Delta V_{th,P1\sim P8}|)$, and $-\ln(|\Delta V_{th,R1\sim R6}|)$ in order (Fig. 4); thus, the value of feature $j$ is for $j = 1$ to 22. The solution of (1) can be found in the direction where the gradient becomes small. In our case, the gradient decreases as $\hat{y}_j$ increases, according to (1). Therefore, $|\Delta V_{th,E1\sim E8}|$ and $|\Delta V_{th,P1\sim P8}|$ increased, but $|\Delta V_{th,R1\sim R6}|$ decreased because they were made negative based on the previous training. The best ISPE, ISPP, and retention characteristics were determined by only one set of inputs, and the selection method will be discussed in the next paragraph.

Fig. 7 shows the estimated values of optimization are usually greater than those of conventional simulation. We

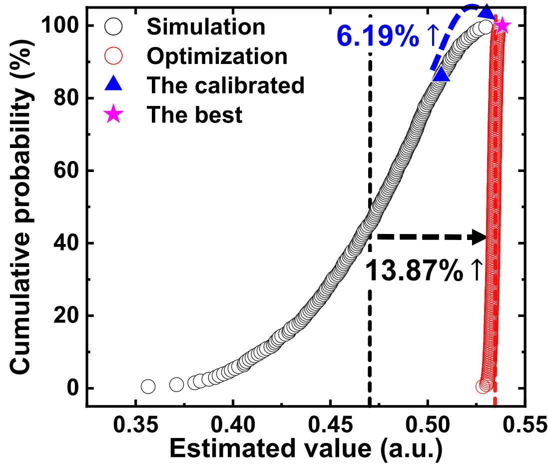



**FIGURE 7.** Cumulative probability of the estimated value of (2), the sum of enhancement for each characteristic between simulation and optimization. The estimated values of optimization are larger than those of simulation, which means the former is better than the latter. From this perspective, the largest estimated value means the best result.

initially set up randomized inputs and performed 2000 iterations using the well-trained ANN to optimize the trap distributions. In addition, we calculated the estimated values for 2000 samples using the following equation:

$$Estimated\ value = \frac{1}{16}\sum_{j=1}^{16}\hat{y}_j - \frac{1}{6}\sum_{j=17}^{22}\hat{y}_j \quad (2)$$

The estimated value of simulation was calculated as its own value, not $\hat{y}_j$. Equation (2) is our evaluation criterion. In (2), the former indicates the $|\Delta V_{th,E1\sim E8}|$ and $|\Delta V_{th,P1\sim P8}|$ of ISPE/ISPP characteristics, and the latter indicates the $|\Delta V_{th,R1\sim R6}|$ of retention characteristics; the larger the former and the smaller the latter, the better the P/R characteristics. Therefore, the largest estimated value among the optimizations was selected as the best P/R characteristics, which is the final goal of optimization process. As a result, the best result increased by 6.19% from the calibrated one. Also, the average of cumulative distribution function for optimization increased by 13.87% from the simulated one. These values are sufficiently large considering the small size of the retention characteristics and the fast processing speed (just 40 minutes). Finally, we analyzed the best set of inputs that made the best P/R characteristics in the following section. Furthermore, the sensitive inputs that significantly affected the outputs were also investigated in detail.

### *B. ANALYSIS OF PERFORMANCE/RELIABILITY BASED ON MACHINE LEARNING APPROACH*

Figs. 8(a) and (b) show the qualitative direction of the trap distributions for the best P/R characteristics and the best donor/acceptor-like traps showing trap profiles with the largest estimated value.

Fig. 8(a) shows the best spatial trap distributions compared with the calibrated distributions. If the volume trap density was large near the BE-TOX/CTN interface, we had a question of which one was larger, the amount of charge tunneled from the channel in the ISPE/ISPP characteristics or the amount of charge leaked to the channel in the retention characteristics. As a result of ML, the best distributions have larger volume trap densities with wide $\sigma_{SD}$ and narrow $\sigma_{SA}$, and then $D_{TD}$ and $D_{TA}$ are closer to the BE-TOX/CTN interface than the calibrated distributions. In general, these results suggest that the spatial trap distributions should be large and close to the BE-TOX/CTN interface to improve the P/R characteristics.

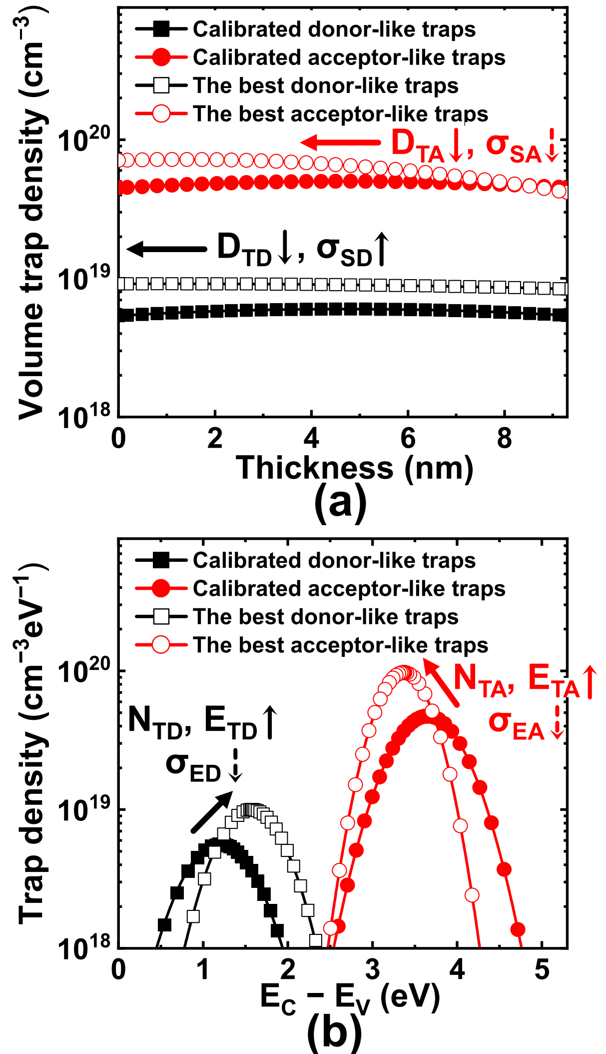



**FIGURE 8.** Comparison of trap profiles between calibrated and the best ones in (a) the spatial trap distributions and (b) the energetic trap distributions. The best P/R characteristics mainly come from Increasing volume trap density as $N_{TD}$ and $N_{TA}$ increase and having deep energy levels in the traps as $E_{TD}$ and $E_{TA}$ increase.

Fig. 8(b) shows the best energetic trap distributions compared with the calibrated distributions. As a result of ML, the best distributions have larger $N_{TD}$ and $N_{TA}$ and deeper $E_{TD}$ and $E_{TA}$ values, but $\sigma_{ED}$ and $\sigma_{EA}$ decrease slightly compared with the calibrated values. These results suggest that the large trap densities and deep energy levels of the energetic trap distributions can improve the P/R characteristics, generally.

**TABLE 3.** Values and dominant factors for the best trap distributions.

| Input parameters | Calibrated | The best | Dominant factor for the best |
|---|---|---|---|
| $N_{TD}$ (cm$^{-3}$·eV$^{-1}$) | 6.00 ×10$^{18}$ | 1.00 ×10$^{19}$ | Available trap sites ↑ |
| $E_{TD}$ (eV) | 1.20 | 1.57 | Attempt-to-escape factor ↓ |
| $\sigma_{ED}$ (eV) | 0.40 | 0.36 | Shallow traps ↓ |
| $D_{TD}$ (nm) | 4.70 | 0.99 | Tunneling probability (ERS/PGM efficiency) ↑ |
| $\sigma_{SD}$ (nm) | 10.00 | 20.00 | Tunneling probability ↑ > Charge leakage ↑ |
| $N_{TA}$ (cm$^{-3}$·eV$^{-1}$) | 5.00 ×10$^{19}$ | 9.69 ×10$^{19}$ | Available trap sites ↑ |
| $E_{TA}$ (eV) | 1.64 | 1.92 | Attempt-to-escape factor ↓ |
| $\sigma_{EA}$ (eV) | 0.40 | 0.30 | Shallow traps ↓ |
| $D_{TA}$ (nm) | 4.70 | 1.36 | Tunneling probability (ERS/PGM efficiency) ↑ |
| $\sigma_{SA}$ (nm) | 10.00 | 7.60 | Charge leakage (Retention) ↓ |

The values of the best inputs are listed in Table 3. Also, the dominant factors for the best are briefly summarized. Especially, some inputs are close to their minimum/maximum values, whereas others have ambiguous values, such as $E_{TA}$ and $\sigma_{EA}$. It means that all inputs should not necessarily be their minimum or maximum for the best. Furthermore, the difference among the optimizations is insignificant (Fig. 6), indicating that several optimized inputs can also improve the P/R characteristics because their results would not differ much from the best results. Therefore, we analyzed the box plot for the optimized inputs to get the sensitivity of each input parameter, as shown in Fig. 9.

Fig. 9 shows the standardized values of all inputs are drawn in a box plot because of the different feature sizes. The box size of the energetic trap distributions ($N_{TD}$, $E_{TD}$, $\sigma_{ED}$, $N_{TA}$, $E_{TA}$, and $\sigma_{EA}$) is smaller than that of the spatial trap distributions ($D_{TD}$, $\sigma_{SD}$, $D_{TA}$, and $\sigma_{SA}$). The median lines are close to the maximum, implying that the sensitivity of the energetic trap distributions is greater than that of the spatial trap distributions. This result could be expected from the relatively small correlation between the spatial trap distributions and $|\Delta V_{th}|$ (Table 2), and it was verified through the ML-based analysis. In addition, the boxes for $N_{TD}$ and $N_{TA}$ contain the maximum. We confirmed that as $N_{TD}$ and $N_{TA}$ increased, $|\Delta V_{th,E1\sim E8}|$ and $|\Delta V_{th,P1\sim P8}|$ increased because of the many available trap sites, but $|\Delta V_{th,R1\sim R8}|$ also increased with the increasing electric field in the CTN [22], [27]. However, the maximum values of $N_{TD}$ and $N_{TA}$ are preferred because the degree of increase in $|\Delta V_{th,E1\sim E8}|$ and $|\Delta V_{th,P1\sim P8}|$ are greater than that of $|\Delta V_{th,R1\sim R8}|$.

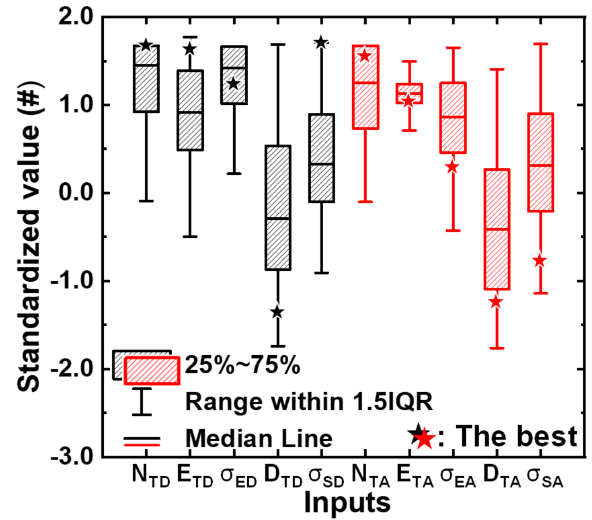



**FIGURE 9.** Box plot of 10 inputs that results in the optimized outputs. They are standardized because of different feature sizes. The star symbol indicates the standardized best inputs, as listed in Table 3.

Fig. 10 shows the output trends for the remaining four inputs of the energetic trap distributions. Each of the remaining four inputs was split, leaving the other inputs fixed at their calibrated values. Additionally, $|\Delta V_{th,E2}|$, $|\Delta V_{th,E8}|$, $|\Delta V_{th,P8}|$, and $|\Delta V_{th,R6}|$, representative outputs of each characteristic, were used for the analysis. As $E_{TD}$ increased, $|\Delta V_{th,E8}|$ and $|\Delta V_{th,P8}|$ remained almost the same, whereas $|\Delta V_{th,R6}|$ decreased and then was saturated, as shown in Fig. 10(a). This is because deeper energy reduces the attempt-to-escape factor for traps [28].

In Fig. 10(b), as $\sigma_{ED}$ increased, $|\Delta V_{th,E8}|$ remained the same, $|\Delta V_{th,P8}|$ increased slightly, and $|\Delta V_{th,R6}|$ decreased after $\sigma_{ED}$ = 0.16 eV. Initially, $|\Delta V_{th,R6}|$ increased slightly due to the shallow hole traps in the $WL_{Sel.}$ region, but after $\sigma_{ED}$ = 0.16 eV, $|\Delta V_{th,R6}|$ decreased because the holes already filled in the gate space were rather reduced due to the increase in the shallow hole traps. Therefore, the holes that invade the $WL_{Sel.}$ region decreased, positively affecting the retention characteristics.

Fig. 10(c) shows the results of increasing $E_{TA}$. We found that the early ($|\Delta V_{th,E2}|$) and late ($|\Delta V_{th,E8}|$) phases of the ISPE exhibited different trends. In the early phase, the escape of trapped electrons begins first, and so the lower $E_{TA}$ results in the higher ERS efficiency [29]. However, this phenomenon was reversed in the late phase because of the same reason for increasing $E_{TD}$. In addition, as the $E_{TA}$ increased, $|\Delta V_{th,R6}|$ decreased rapidly but started to saturate. Therefore, when considering the tradeoffs between the ISPE and retention characteristics, the ML determined the best value of $E_{TA}$ to be 1.92 eV (Table 3).

Fig. 10(d) shows the results of increasing $\sigma_{EA}$. The tradeoff between the P/R characteristics is clearly demonstrated. As $\sigma_{EA}$ increased, the integral of the energetic trap distributions also increased, resulting in an increase in $|\Delta V_{th,E8}|$ and

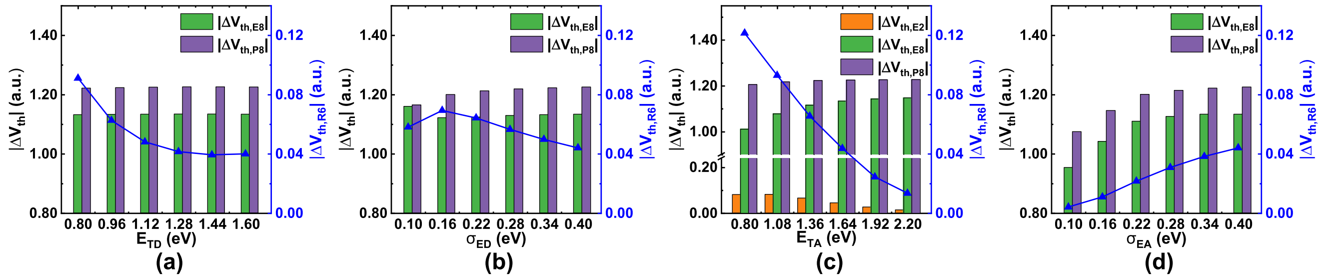



**FIGURE 10. Trends of P/R characteristics with (a) $E_{TD}$, (b) $\sigma_{ED}$, (c) $E_{TA}$, and (d) $\sigma_{EA}$. There is a clear tradeoff for P/R characteristics in $E_{TA}$ and $\sigma_{EA}$.**

$|\Delta V_{th,P8}|$. However, $|\Delta V_{th,R6}|$ increased because the shallow electron traps increase and then escape easily. Therefore, when considering the tradeoff between the ISPE/ISPP and retention characteristics, the ML determined the best value of $\sigma_{EA}$ to be 0.30 eV (Table 3).

## IV. CONCLUSION

ML-based analysis was used to determine the optimal trap distributions of CTN for improving the P/R characteristics simultaneously in 3-D NAND flash memory. The ANN could model the relationship between the trap distributions and $|\Delta V_{th}|$ and was trained using an experimentally-calibrated TCAD simulation dataset. The trained ANN exhibited high accuracy and predicted specific inputs for the superior P/R characteristics. As a result, we found the values of the best inputs using our evaluation criterion and analyzed the trends of each sensitive input parameter. In general, the influence of the energetic trap distributions on $|\Delta V_{th}|$ was greater than that of the spatial trap distributions. Moreover, increasing all inputs of the energetic trap distributions proved to be better for the P/R characteristics. However, it was found that $E_{TA}$ and $\sigma_{EA}$ should not be maximized because they cause a tradeoff in the P/R characteristics. Therefore, even if the actual process was difficult, ML could determine their optimal values. Therefore, this study is helpful in determining the complex physical properties of CTN in the 3-D NAND flash.

## ACKNOWLEDGMENT

The EDA tool was supported by the IC Design Education Center (IDEC).

## REFERENCES

[1] J. Jang, H.-S. Kim, W. Cho, H. Cho, J. Kim, S. I. Shim, J.-H. Jeong, B.-K. Son, D. W. Kim, and J.-J. Shim, "Vertical cell array using TCAT (terabit cell array transistor) technology for ultra high density NAND flash memory," in *Proc. Symp. VLSI Technol.*, Kyoto, Japan, Jun. 2009, pp. 192–193.

[2] H. Tanaka, M. Kido, K. Yahashi, M. Oomura, R. Katsumata, M. Kito, Y. Fukuzumi, M. Sato, Y. Nagata, Y. Matsuoka, Y. Iwata, H. Aochi, and A. Nitayama, "Bit cost scalable technology with punch and plug process for ultra high density flash memory," in *Proc. IEEE Symp. VLSI Technol.*, Kyoto, Japan, Jun. 2007, pp. 14–15, doi: 10.1109/VLSIT.2007.4339708.

[3] R. Katsumata, M. Kito, Y. Fukuzumi, M. Kido, H. Tanaka, Y. Komori, M. Ishiduki, J. Matsunami, T. Fujiwara, and Y. Nagata, "Pipe-shaped BiCS flash memory with 16 stacked layers and multi-level-cell operation for ultra high density storage devices," in *Proc. Symp. VLSI Technol.*, Kyoto, Japan, Jun. 2009, pp. 136–137.

[4] E.-S. Choi and S.-K. Park, "Device considerations for high density and highly reliable 3D NAND flash cell in near future," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2012, pp. 9.4.1–9.4.4, doi: 10.1109/IEDM.2012.6479011.

[5] W. Liu, F. Wu, M. Zhang, Y. Wang, Z. Lu, X. Lu, and C. Xie, "Characterizing the reliability and threshold voltage shifting of 3D charge trap NAND flash," in *Proc. Design, Autom. Test Eur. Conf. Exhib. (DATE)*, Florence, Italy, Mar. 2019, pp. 312–315, doi: 10.23919/DATE.2019.8714941.

[6] C. Sandhya, U. Ganguly, K. Singh, C. Olsen, S. Seutter, G. Conti, K. Ahmed, N. Krishna, J. Vasi, and S. Mahapatra, "The effect of band gap engineering of the nitride storage node on performance and reliability of charge trap flash," in *Proc. 15th Int. Symp. Phys. Failure Anal. Integr. Circuits*, Singapore, Jul. 2008, pp. 1–7, doi: 10.1109/IPFA.2008.4588192.

[7] G. Van Den Bosch, A. Furnemont, M. B. Zahid, R. Degraeve, L. Breuil, A. Cacciato, A. Rothschild, C. Olsen, U. Ganguly, and J. Van Houdt, "Nitride engineering for improved erase performance and retention of TANOS NAND flash memory," in *Proc. Joint Non-Volatile Semiconductor Memory Workshop Int. Conf. Memory Technol. Design*, Opio, France, 2008, pp. 128–129, doi: 10.1109/NVSMW.2008.45.

[8] E.-S. Choi, H.-S. Yoo, K.-H. Park, S.-J. Kim, J.-R. Ahn, M.-S. Lee, Y.-O. Hong, S.-G. Kim, J.-C. Om, M.-S. Joo, S.-H. Pyi, S.-S. Lee, S.-K. Lee, and G.-H. Bae, "Modeling and characterization of program/erasure speed and retention of TiN-gate MANOS (Si-oxide-$SiN_x$-$Al_2O_3$-metal gate) cells for NAND flash memory," in *Proc. 22nd IEEE Non-Volatile Semiconductor Memory Workshop*, Aug. 2007, pp. 83–84, doi: 10.1109/NVSMW.2007.4290591.

[9] K. Nam, C. Park, J.-S. Yoon, H. Jang, M. S. Park, J. Sim, and R.-H. Baek, "Origin of incremental step pulse programming (ISPP) slope degradation in charge trap nitride based multi-layer 3D NAND flash," *Solid-State Electron.*, vol. 175, Jan. 2021, Art. no. 107930, doi: 10.1016/J.SSE.2020.107930.

[10] K. Nam, C. Park, J.-S. Yoon, H. Yun, H. Jang, K. Cho, H.-J. Kang, M.-S. Park, J. Sim, H.-C. Choi, and R.-H. Baek, "Optimal energetic-trap distribution of nano-scaled charge trap nitride for wider $V_{th}$ window in 3D NAND flash using a machine-learning method," *Nanomaterials*, vol. 12, no. 11, p. 1808, May 2022, doi: 10.3390/NANO12111808.

[11] R. Degraeve, M. Cho, B. Govoreanu, B. Kaczer, M. B. Zahid, J. Van Houdt, M. Jurczak, and G. Groeseneken, "Trap spectroscopy by charge injection and sensing (TSCIS): A quantitative electrical technique for studying defects in dielectric stacks," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2008, pp. 1–4, doi: 10.1109/IEDM.2008.4796812.

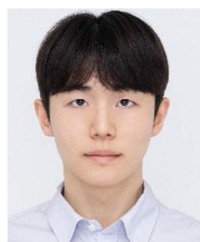
[12] A. Suhane, A. Arreghini, R. Degraeve, G. Van den bosch, L. Breuil, M. B. Zahid, M. Jurczak, K. De Meyer, and J. Van Houdt, "Validation of retention modeling as a trap-profiling technique for SiN-based charge-trapping memories," *IEEE Electron Device Lett.*, vol. 31, no. 1, pp. 77–79, Jan. 2010, doi: 10.1109/LED.2009.2035718.

[13] H.-J. Kang, N. Choi, S.-M. Joe, J.-H. Seo, E. Choi, S.-K. Park, B.-G. Park, and J.-H. Lee, "Comprehensive analysis of retention characteristics in 3-D NAND flash memory cells with tube-type poly-Si channel structure," in *Proc. Symp. VLSI Technol.*, Kyoto, Japan, Jun. 2015, pp. T182–T183, doi: 10.1109/VLSIT.2015.7223670.

[14] Y. Yang and M. H. White, "Charge retention of scaled SONOS nonvolatile memory devices at elevated temperatures," *Solid-State Electron.*, vol. 44, no. 6, pp. 949–958, Jan. 2000, doi: 10.1016/S0038-1101(00)00012-5.

[15] K.-D. Suh, B.-H. Suh, Y.-H. Lim, J.-K. Kim, Y.-J. Choi, Y.-N. Koh, S.-S. Lee, S.-C. Kwon, B.-S. Choi, and J.-S. Yum, "A 3.3 V 32 Mb NAND flash memory with incremental step pulse programming scheme," *IEEE J. Solid-State Circuits*, vol. 30, no. 11, pp. 1149–1156, Nov. 1995, doi: 10.1109/4.475701.

[16] H.-C. Choi, H. Yun, J.-S. Yoon, and R.-H. Baek, "Neural approach for modeling and optimizing Si-MOSFET manufacturing," *IEEE Access*, vol. 8, pp. 159351–159370, 2020, doi: 10.1109/ACCESS.2020.3019933.
[17] H. Yun, J.-S. Yoon, J. Jeong, S. Lee, H.-C. Choi, and R.-H. Baek, "Neural network based design optimization of 14-nm node fully-depleted SOI FET for SoC and 3DIC applications," *IEEE J. Electron Devices Soc.*, vol. 8, pp. 1272–1280, 2020, doi: 10.1109/JEDS.2020.3022367.
[18] J.-S. Yoon, S. Lee, H. Yun, and R.-H. Baek, "Digital/analog performance optimization of vertical nanowire FETs using machine learning," *IEEE Access*, vol. 9, pp. 29071–29077, 2021, doi: 10.1109/ACCESS.2021.3059475.
[19] *Sentaurus Device User Guide, Version: O-2018.06*, Synopsys, Mountain View, CA, USA, 2018.
[20] V. A. Gritsenko, S. S. Nekrashevich, V. V. Vasilev, and A. V. Shaposhnikov, "Electronic structure of memory traps in silicon nitride," *Microelectron. Eng.*, vol. 86, nos. 7–9, pp. 1866–1869, Jul. 2009, doi: 10.1016/J.MEE.2009.03.093.
[21] T. Ishida, Y. Okuyama, and R. Yamada, "Characterization of charge traps in metal-oxide-nitride-oxide-semiconductor (MONOS) structures for embedded flash memories," in *Proc. IEEE Int. Rel. Phys. Symp. Proc.*, San Jose, CA, USA, Mar. 2006, pp. 516–522, doi: 10.1109/RELPHY.2006.251272.
[22] Y. Wang and M. H. White, "An analytical retention model for SONOS nonvolatile memory devices in the excess electron state," *Solid-State Electron.*, vol. 49, no. 1, pp. 97–107, Jan. 2005, doi: 10.1016/J.SSE.2004.06.009.
[23] A. Arreghini, N. Akil, F. Driussi, D. Esseni, L. Selmi, and M. J. van Duuren, "Long term charge retention dynamics of SONOS cells," *Solid-State Electron.*, vol. 52, no. 9, pp. 1460–1466, Sep. 2008, doi: 10.1016/J.SSE.2008.04.016.
[24] T. H. Kim, I. H. Park, J. D. Lee, H. C. Shin, and B. G. Park, "Electron trap density distribution of Si-rich silicon nitride extracted using the modified negative charge decay model of silicon-oxide-nitride-oxide-silicon structure at elevated temperatures," *Appl. Phys. Lett.*, vol. 89, no. 6, Aug. 2006, Art. no. 063508, doi: 10.1063/1.2335619.
[25] A. Pinkus, "Approximation theory of the MLP model in neural networks," *Acta Numer.*, vol. 8, pp. 143–195, Jan. 1999, doi: 10.1017/S0962492900002919.
[26] *MATLAB User Manual, Version R2021b*, Mathworks, Inc., Natick, MA, USA, 2021.
[27] S. J. Wrazien, Y. Zhao, J. D. Krayer, and M. H. White, "Characterization of SONOS oxynitride nonvolatile semiconductor memory devices," *Solid-State Electron.*, vol. 47, no. 5, pp. 885–891, May 2003, doi: 10.1016/S0038-1101(02)00448-3.
[28] A. V. Vishnyakov, Y. N. Novikov, V. A. Gritsenko, and K. A. Nasyrov, "The charge transport mechanism in silicon nitride: Multi-phonon trap ionization," *Solid-State Electron.*, vol. 53, no. 3, pp. 251–255, Mar. 2009, doi: 10.1016/J.SSE.2008.07.005.
[29] A. Padovani, A. Arreghini, L. Vandelli, L. Larcher, G. V. Bosch, P. Pavan, and J. V. Houdt, "A comprehensive understanding of the erase of TANOS memories through charge separation experiments and simulations," *IEEE Trans. Electron Devices*, vol. 58, no. 9, pp. 3147–3155, Sep. 2011, doi: 10.1109/TED.2011.2159722.



**KIHOON NAM** (Member, IEEE) received the B.S. degree in electrical and computer engineering from Ajou University, Suwon, Republic of Korea, in 2019, and the M.S. degree in electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea, in 2021, where he is currently pursuing the Ph.D. degree.

His research interest includes operation and structure of vertical NAND memory (3D NAND flash and gate-all-around FET).

**CHANYANG PARK** (Member, IEEE) received the B.S. degree in electrical and electronics engineering from Chung-Ang University, Seoul, Republic of Korea, in 2018, and the M.S. degree in electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea, in 2020, where he is currently pursuing the Ph.D. degree.

His research interests include experimental characterization and simulation of 3D vertical NAND flash memory devices.

**HYEOK YUN** (Member, IEEE) received the B.S. degree in electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea, in 2019, where he is currently pursuing the integrated M.S. and Ph.D. degree in electrical engineering.

His research interests include variability of multi-gate field-effect transistors (FinFETs, nanowire FETs, and nanosheet FETs) and machine learning.

**JUN-SIK YOON** (Member, IEEE) received the B.S. degree in electrical engineering and the Ph.D. degree in creative IT engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea, in 2012 and 2016, respectively.

From 2016 to 2018, he was a Postdoctoral Research Fellow at POSTECH. Since 2019, he has been a Research Assistant Professor of electrical engineering with POSTECH. His research interests include characterization and simulation of advanced nanoscale devices (fin, gate-all-around, tunneling, and nanosheet FETs) and applications (chemical sensor and solar cell).

**HYUNDONG JANG** (Member, IEEE) received the B.S. degree in electrical engineering from Yeungnam University, Gyeongsan, Republic of Korea, in 2018, and the M.S. degree in electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea, in 2020, where he is currently pursuing the Ph.D. degree.

His research interests include characterization and simulation of 3D vertical NAND flash memory devices and machine learning.

**KYEONGRAE CHO** (Member, IEEE) received the B.S. degree in electrical engineering from Hongik University, Seoul, Republic of Korea, in 2020, and the M.S. degree in electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea, in 2022, where he is currently pursuing the Ph.D. degree.

His research interests include operation and structure of vertical NAND flash memory devices and machine learning.

**MIN SANG PARK** received the B.S. and Ph.D. degrees in electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea, in 2005 and 2011, respectively. Since 2011, he has been a Technical Leader with the NAND Flash Device Technology Team, SK hynix Inc., Icheon, Republic of Korea.

**HYUN-CHUL CHOI** (Member, IEEE) received the B.S. degree in electronic and electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 2002, and the M.S. and Ph.D. degrees in electronic and electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea, in 2004 and 2011, respectively.

From 2011 to 2014, he was a Research Scientist with the Multimedia Development Laboratory, Daum Communications. From 2014 to 2020, he was an Assistant Professor with the Department of Electronic Engineering, Yeungnam University, Gyeongsan, Republic of Korea. Since 2020, he has been an Associate Professor with Yeungnam University. His research interests include multimedia retrieval, computational photography, 3D photogrammetry, object tracking, and recognition based on machine learning and neural networks.

**ROCK-HYUN BAEK** (Member, IEEE) received the B.S. degree in electrical engineering from Korea University, in 2004, and the M.S. and Ph.D. degrees in electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea, in 2006 and 2011, respectively.

From 2011 to 2015, he was a Postdoctoral Researcher and a Technical Engineer at SEMATECH, Albany, NY, USA. From 2015 to 2017, he was a Senior Device Engineer at the Samsung Research and Development Center (Pathfinding TEAM), Republic of Korea. From 2011 to 2017, he was an Assistant Professor of electrical engineering with POSTECH, where he is currently an Associate Professor. His research interests include advanced logic devices (fin, gate-all-around, nanosheet FETs, and vertical FET), materials (Si, SiGe, and Ge), memory (3D-NAND and DRAM peri), 3DIC (TSV and M3D), and unit circuit characterization based on electrical measurement, TCAD, and machine learning technique.