Received 25 September 2018; revised 19 November 2018; accepted 6 December 2018. Date of publication 17 December 2018; date of current version 1 March 2019. The review of this paper was arranged by Editor M. Liu.

Digital Object Identifier 10.1109/JEDS.2018.2886359

Investigation of Retention Noise for 3-D TLC NAND Flash Memory

KUNLIANG WANG [1], GANG DU [1] (Member, IEEE), ZHIYUAN LUN [2], AND XIAOYAN LIU [1] (Member, IEEE)

1 Institute of Microelectronics, Peking University, Beijing 100871, China
2 Solid State Drive Department, Hisilicon Techologies, Co., Ltd., Hangzhou 310052, China

CORRESPONDING AUTHORS: X. Y. LIU AND G. DU (e-mail: xyliu@ime.pku.edu.cn; gangdu@pku.edu.cn)

This work was supported in part by the National Key Research and Development Plan under Grant 2016YFA0202101, in part by the Solid State Drive Department, Hisilicon Techologies, Co., Ltd., Hangzhou, China, in part by the National Natural Science Foundation of China under Grant 61421005, and in part by the National High-Tech Research and Development Program (863 Program) under Grant 2015AA016501.

ABSTRACT In this paper, the retention noise [electron emission statistics (EES)] after program operation of 3-D triple-level program cell (TLC) NAND flash memory is investigated. Three main noise sources, consisting of essential EES (EEES), electron numbers fluctuation, and device parameters fluctuation to broaden the retention V_{th} distributions are comprehensively considered, and the corresponding analytic models are developed. The impact of device parameters fluctuation is relatively larger than EEES and electron numbers fluctuation for our measured 3-D TLC NAND flash memory devices. Using the proposed models, the calculated V_{th} distributions after different data retention times have good agreements with the measurements, which validate our proposed models. This paper provides a method to predict the V_{th} distributions accurately and efficiently, and can help in improving reliability of 3-D TLC and quad-level program cell NAND flash memory.

INDEX TERMS Retention V_{th} distribution, retention noise, 3-D TLC NAND flash memory, reliability.

I. INTRODUCTION

3-D NAND flash memory has been considered as a promising candidate for future memory solutions [1] due to lower cost per bit. At the same time, in order to further increase the storage capacity of NAND flash memory, multi-level program cell (MLC) and triple-level program cell (TLC) technologies are introduced. 3-D quad-level program cell (QLC) NAND flash memory [2] is a promising candidate for lower power operation, and higher density nonvolatile memory. However, data retention errors are the dominant sources of NAND flash memory fail bit counts (FBCs), which is one of the most important challenges to realize multi-bit-per-cell (MB) technologies due to the extremely narrow read V_{th} windows.

Currently, in order to improve the reliability of NAND flash memory, read voltage control technology [3] and error correction codes (ECCs), such as LDPC (Low-Density Parity-Check) and BCH (Bose–Chaudhuri–Hocquenghem) [4] are common used in SSD controller. However, the more powerful ECC mechanisms require the real-like V_{th} distributions after specific operation modes, such as: data retention times and program/erase (P/E) cycles. However, it takes a lot of time to obtain the V_{th} distributions based on the measurements. Modeling the retention V_{th} distribution can make the process more effective and efficient.

In recent years, a lot of work investigates the retention behavior of NAND flash memory. Most of them focus on the single memory cell [5]–[7], and the comprehensively analysis of different physical failure mechanisms [5], [6] and the influence of states of adjacent cells [7] are presented. Some works study on the effects of data retention times, operation temperatures, and P/E cycles on retention characteristics [8]–[11] of NAND flash memory, and the universal retention V_{th} degradation formula is developed. Besides, several modeling approaches have been proposed to predict the retention V_{th} distributions based on array statistical behavior, such as: assuming the retention V_{th} distributions as a mixture of Gaussian [12], [13] and other distribution functions, or fitting the V_{th} distributions by parameter estimation algorithms [14]–[16], or using the experimental results to interpolation or extrapolation [17], and these modeling approaches nicely represent the retention V_{th} distributions. However, the underlying physical mechanism of statistical V_{th} distributions degradation is still lacking.

In this work, the retention noise after program operation of 3-D TLC NAND flash memory is investigated, and the main purpose of this work is to find the relationship between the spread degree of retention V_{th} distribution (electron numbers distribution in nitride layer) and average V_{th} shift (mean of emitted electron numbers). Here, we consider the retention noise as EES (electron emission statistics) [18], which is the reason for the widening of retention V_{th} distributions. Three main noise sources to broaden the retention V_{th} distributions are comprehensively investigated, and the corresponding analytic models are developed. Using the proposed models, the calculated V_{th} distributions after different data retention times have good agreements with the measurements, which validate our proposed models. This work provides a method to predict the retention V_{th} distributions accurately and efficiently, and can help in improving reliability of 3-D TLC and QLC NAND flash memory technologies.

This paper is organized as follows: After the introduction, in Section II, we describe the retention characteristics of 3-D TLC NAND flash memory, and the effects of three main noise sources on retention V_{th} distributions are introduced; In Section III, the corresponding analytic models are developed; Section IV presents the results and discussion, finally, the conclusions are drawn in Section V.

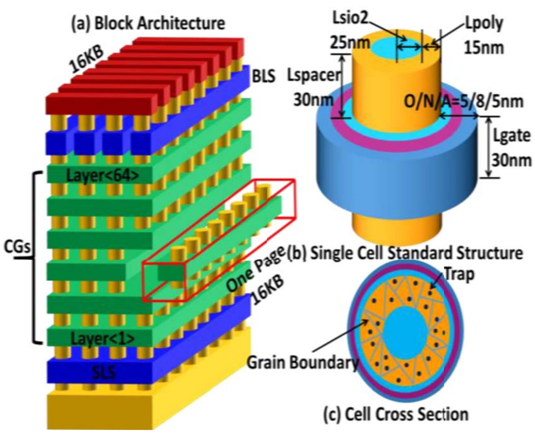

FIGURE 1. (a) Schematic structure of a 3-D TLC NAND flash memory array in a block. (b). Cell structure. (c). Cross section of the single memory cell.

II. RETENTION CHARACTERISTICS

The mainstream 3-D TLC NAND flash memory array structure, cell structure, and the cross section of single memory cell are shown in Figure 1. Figure 2 (a) shows the physical degeneration mechanism of a single NAND flash memory cell for data retention operation. Here, N is the electron numbers in nitride layer after ISPP program operation, the electron charge tends to escape from the nitride layer by

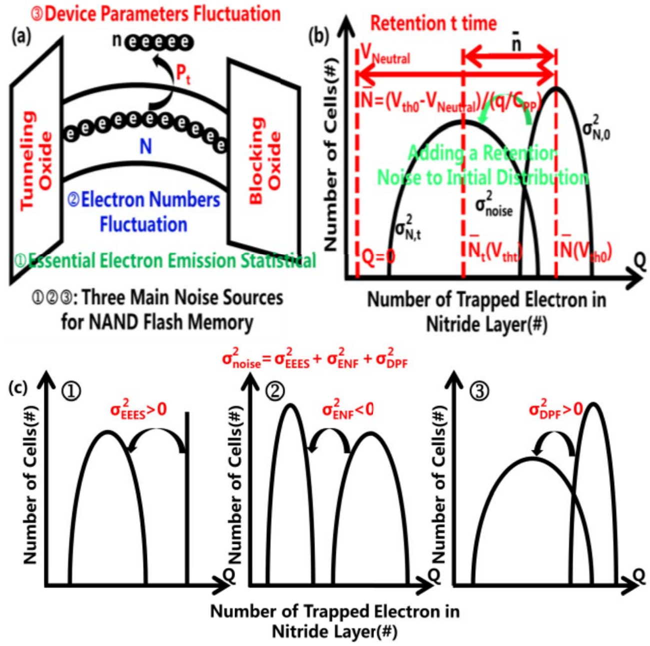

FIGURE 2. (a) Physical degradation mechanism of a single NAND flash memory cell for data retention operation. Each electron charge has the emission probability P_t to escape from the nitride layer. (b) The evolution of retention V_{th} distribution, just like adding a retention noise to the initial program V_{th} distribution due to electron emission statistics (EES). (c) Three main noise sources for NAND flash memory, consisting of essential electron emission statistics (EEES), electron numbers fluctuation (ENF), and device parameters fluctuation (DPF) are comprehensively considered.

the electron emission probability P_t during data retention, and n is emitted electron numbers from nitride layer after retention t time, which causes the NAND flash memory cell V_{th} degradation as eq. (1):

$$\Delta V_{th} = -q \cdot n/C_{PP} \tag{1}$$

Here, q is electronic charge, and C_{PP} is the control-gate to nitride layer coupling capacitance. Figure 2 (b) plots the evolution of retention V_{th} distribution (electron numbers distribution), just like adding a retention noise ($\bar{n}$ and σ^2_{noise} are the mean and variance of retention noise respectively) to the initial program V_{th} distribution (initial electron numbers distribution) due to the statistical emission of electrons (EES). Here, $\overline{N}$ and $\sigma^2_{N,0}$ are the mean and variance of initial electron numbers distribution respectively, $\overline{N_t}$ and $\sigma^2_{N,t}$ are the mean and variance of electron numbers distribution after retention t time respectively, and $\bar{n}$ is the mean of emitted electron numbers after retention t time.

Figure 3 shows the typical measured retention V_{th} distributions for a page (16KB) memory cells of 3-D TLC NAND flash memory, and they shift in negative direction after 4.8h data retention time at 85°C. Here, it can be noted that the x axis (V_{th}) only represents the offsets relative to the default read voltages (A, B, C, D, E, F, and G), and it is difficult to obtain the real voltage values due to the default read voltages were fixed in the SSD controller. Therefore, $V_{Neutral}$ (average V_{th} for neutral V_{th} distribution) is defined

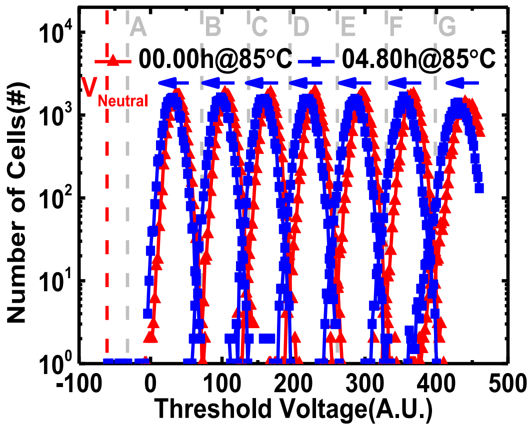

FIGURE 3. Typical measured retention V_{th} distributions for a page (16KB) memory cells, and they shift in negative direction after 4.8h data retention time at 85°C.

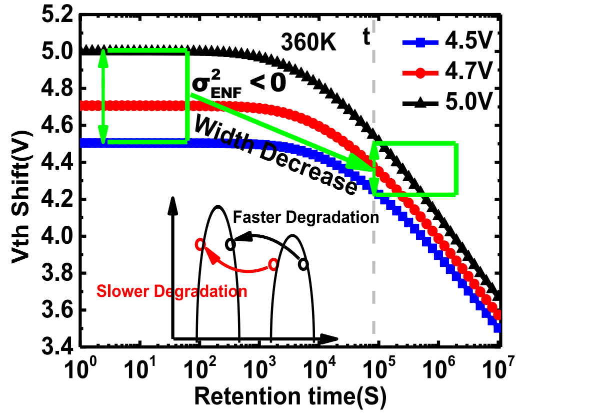

FIGURE 4. Simulated single cell V_{th} shift with different initial V_{th} 4.5V, 4.7V, and 5.0V, and the retention V_{th} distribution width decreases gradually, just like adding a negative variance noise to initial program V_{th} distribution.

in this work as shown in Figure 3 to calculate the mean $\overline{N}$ of initial electron numbers distribution in nitride layer after ISPP program as eq. (2) and Figure 2 (b), which can be calibrated in combination with the experimental results. Here, $V_{th0} - V_{Neutral}$ is the average V_{th} offset duo to the electron injection during the ISPP program.

$$\overline{N} = (V_{th0} - V_{Neutral})/(q/C_{PP}) \quad (2)$$

In addition, from Figure 3 we can find that the retention V_{th} distributions are not only shifted but also broadened, which severely degrades the reliability of NAND flash memory due to the narrow read V_{th} windows. The reason for the widening of retention V_{th} distribution is the electron emission statistics (EES), which is influenced by three main noise sources, consisting of essential electron emission statistics (EEES), electron numbers fluctuation (ENF), and device parameters fluctuation (DPF) as shown in Figure 2 (a) and Figure 2 (c). For essential electron emission statistics, the 'essential' means that even though all the memory cells of a page were in the same initial V_{th} after ISPP program, their V_{th} after data retention operation would be statistical distributed due to the granular nature of electron [19], [20] as shown in Figure 2 (c): ①.

Electron numbers fluctuation is also a physical effect to affect the retention V_{th} distribution. As we all known, ISPP is a very effective method to control a tight V_{th} distribution precisely. Ideally case, a uniform program V_{th} distribution between V_{PV} (program-verify) and $V_{PV} + V_{Step}$ can be obtained by ISPP algorithm [21]. Especially, program V_{th} distribution can be distorted by ISPP noise [19], [20], WL-WL interference [22], and RTN effect of tunneling oxide and Poly-Si [23], [24], thus, the electron numbers in the nitride layer for a page memory cells are variable.

Figure 4 shows the simulated single cell V_{th} shift with different initial V_{th} 4.5V, 4.7V, and 5.0V, respectively, where a self-consistent simulator (takes into consideration the tunneling processes, charge trapping/de-trapping mechanisms, and drift-diffusion transport within the storage layer) is used to simulate the program, erase, retention, and read operations as calibrated and verified in our previous works [25]–[27]. From the figure we can find that the retention V_{th} distribution width decreases gradually, this is because of larger degradation rate for larger initial electron numbers in nitride layer, just like adding a negative variance noise ($\sigma^2_{ENF} < 0$) to initial program V_{th} distribution.

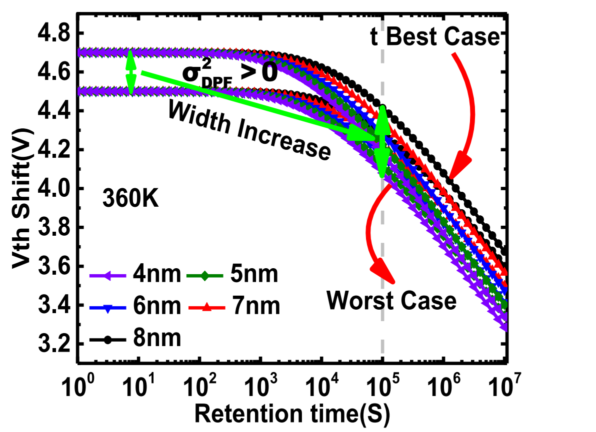

FIGURE 5. Simulated single cell V_{th} shift with different tunneling oxide layer thickness for different initial V_{th} 4.5V and 4.7V, respectively. The V_{th} distribution width increases gradually due to device parameters fluctuation, and the device parameters fluctuation plays an opposite effect compared with the electron numbers fluctuation.

In addition, device parameters fluctuation is a major reason to spread the retention V_{th} distribution. Figure 5 shows the simulated single cell V_{th} shift with different tunneling oxide layer thickness (extreme case) for different initial V_{th} 4.5V and 4.7V, respectively. From this figure we can find that the device parameters fluctuation lead to the V_{th} distribution width increases ($\sigma^2_{DPF} > 0$) due to the faster degradation cells always faster and the slower degradation cells always slower. The effect of electron numbers fluctuation ($\sigma^2_{ENF} < 0$) plays an opposite effect compared with the EEES ($\sigma^2_{EEES} > 0$)

and device parameters fluctuation ($\sigma^2_{DPF} > 0$) as shown in Figure 1 (c).

Considering the effects of EEES, electron numbers fluctuation, and device parameters fluctuation (σ^2_{EEES}, σ^2_{ENF}, and σ^2_{DPF}), the variance of retention noise σ^2_{noise} can be written as following:

$$\sigma^2_{noise} = \sigma^2_{EEES} + \sigma^2_{ENF} + \sigma^2_{DPF} \tag{3}$$

Table 1 lists the main parameters used in this work.

TABLE 1. Main parameters used in this work.

Parameters	Meaning
N	Initial electron numbers in nitride layer
N_t	Electron numbers in nitride layer after retention t time
n	Emitted electron numbers after retention t time
$\overline{N}$	Mean of initial electron numbers distribution
$\overline{N_t}$	Mean of electron numbers distribution after retention t time
$\overline{n}$	Mean of emitted electron numbers
V_{th0}	Initial average V_{th} after ISPP program operation
V_{tht}	Average V_{th} after retention t time
$\sigma^2_{N,0}$	Variance of initial electron numbers distribution
$\sigma^2_{N,t}$	Variance of electron numbers distribution after retention t time
q	Electronic charge
$V_{Neutral}$	Average V_{th} for neutral V_{th} distribution
C_{PP}	Control-gate to nitride layer coupling capacitance, which can be extracted from our previous work [28]
P_t	Electron emission probability during data retention
$P(n)$	Probability of emitted n electrons for a page memory cells
N_m	Maximum storage electron numbers in nitride layer
σ^2_{noise}	Variance of retention noise
σ^2_{EEES}	Variance of retention noise due to EEES
σ^2_{ENF}	Variance of retention noise due to ENF
σ^2_{DPF}	Variance of retention noise due to DPF

III. MODELS

A. ESSENTIAL ELECTRON EMISSION STATISTICS

Retention characteristics of NAND flash memory can be understood as a series of electron emission events over time. One ultimate reason to broaden the retention V_{th} distribution of NAND flash memory for data retention is EEES. Assuming all the memory cells of a page have same device parameters and electron numbers N in nitride layer after ISPP program, we can model the electron emission events from the nitride layer during retention operation as mutually independent events, and it can be described as the binominal distribution as eq. (4):

$$P(n) = C_N^n \cdot P_t^n \cdot (1 - P_t)^{N-n} \tag{4}$$

Here, N is the electron numbers in nitride layer of NAND flash memory after ISPP program for a page memory cells, P_t is the average electron emission probability between time 0 (program finished) and time t (data retention operation), n is the emitted electrons numbers from nitride layer for a page memory cells, and P(n) is the probability of emitted n electrons for a page memory cells. Thus, its mean $\overline{n}$ (can be obtained from measurements) and the variance σ^2_{EEES} can be written as eq. (5) and (6):

$$\begin{aligned}\overline{n} &= \sum_{n=0}^{N} n \cdot P(n) = \sum_{n=0}^{N} n \cdot C_N^n \cdot P_t^n \cdot (1 - P_t)^{N-n} \\ &= \sum_{n=1}^{N} n \cdot \frac{N!}{n!(N-n)!} \cdot P_t^n \cdot (1 - P_t)^{N-n} \\ &= N \cdot P_t \cdot \sum_{n=1}^{N} \frac{(N-1)!}{(n-1)!(N-n)!} \cdot P_t^{n-1} \\ &\quad \times (1 - P_t)^{(N-1)-(n-1)} \\ &= N \cdot P_t \end{aligned} \tag{5}$$

$$\begin{aligned}\sigma^2_{EEES} &= \sum_{n=1}^{N} n^2 \cdot C_N^n \cdot P_t^n \cdot (1 - P_t)^{N-n} - \overline{n}^2 \\ &= \sum_{n=1}^{N} [n \cdot (n-1) + n] \cdot C_N^n \cdot P_t^n \cdot (1 - P_t)^{N-n} - \overline{n}^2 \\ &= N \cdot (N-1) \cdot P_t^2 + N \cdot P_t - (N \cdot P_t)^2 \\ &= N \cdot P_t \cdot (1 - P_t) = \overline{n} \cdot (1 - P_t) \end{aligned} \tag{6}$$

Then, we can obtain the relationship between the variance of retention noise σ^2_{noise} and the mean $\overline{n}$ of emitted electron numbers due to effect of essential electron emission statistics (EEES) can be expressed as eq. (7):

$$\sigma^2_{EEES} = \overline{n} \cdot (1 - \overline{n}/N) = \overline{n} \cdot (1 - \overline{n}/\overline{N}) \tag{7}$$

B. ELECTRON NUMBERS FLUCTUATION

Electron numbers fluctuation is also a physical effect to affect the retention V_{th} distribution. The variance of retention noise can be expressed as eq. (8):

$$\sigma^2_{ENF} = \sigma^2_{N,t} - \sigma^2_{N,0} \tag{8}$$

According to the definition of variance, the variance of initial electron numbers distribution $\sigma^2_{N,0}$ can be expressed as eq. (9):

$$\sigma^2_{N,0} = \sum_{N=0}^{N_m} (N - \overline{N})^2 \cdot P(N) \tag{9}$$

Here, N_m is the maximum storage electron numbers in nitride layer, and P(N) is the probability of storage N electrons in nitride layer after ISPP program. Then, we need to calculate the variance $\sigma^2_{N,t}$ of electron numbers fluctuation after retention t time.

To take into account the statistical dispersion of N, the probability function of electron emission events for NAND flash memory can be written as eq. (10):

$$P(n) = \sum_{N=0}^{N_m} C_N^n \cdot P_t^n \cdot (1 - P_t)^{N-n} \cdot P(N) \tag{10}$$

Therefore, the mean $\overline{n}$ of emitted electron numbers as eq. (5) can be rewritten as eq. (11) [29]:

$$\overline{n} = \sum_{n=0}^{N} n \cdot \sum_{N=0}^{N_m} C_N^n \cdot P_t^n \cdot (1 - P_t)^{N-n} \cdot P(N) = \overline{N} \cdot P_t \tag{11}$$

From eq. (11) we can find that the average electron emission probability P_t can be approximately expressed as eq. (12):

$$P_t = \bar{n}/\bar{N} \tag{12}$$

Therefore, N_t can be written as eq. (13), namely, the number of residual electrons is equal to the number of initial electrons minus the number of emitted electrons during data retention.

$$N_t = N - N \cdot P_t = N \cdot (1 - P_t) \tag{13}$$

Combining eq. (9) and (13), we can obtain $\sigma^2_{N,t}$ as eq. (14) and (15):

$$\sigma^2_{N,t} = \sum_{N_t=0}^{N_m} (N_t - \overline{N_t})^2 \cdot P(N_t) \tag{14}$$

$$\begin{aligned}\sigma^2_{N,t} &= \sum_{N=0}^{N_m} ((1 - P_t) \cdot N - (1 - P_t) \cdot \bar{N})^2 \cdot P(N) \\ &= (1 - P_t)^2 \cdot \sigma^2_{N,0} = (1 - \bar{n}/\bar{N})^2 \cdot \sigma^2_{N,0}\end{aligned} \tag{15}$$

Then, we can obtain the relationship between the variance of retention noise σ^2_{noise} and the mean $\bar{n}$ of emitted electron numbers due to effect of electron numbers fluctuation (ENF) can be expressed as eq. (16):

$$\sigma^2_{ENF} = (1 - \bar{n}/\bar{N})^2 \cdot \sigma^2_{N,0} - \sigma^2_{N,0} < 0. \tag{16}$$

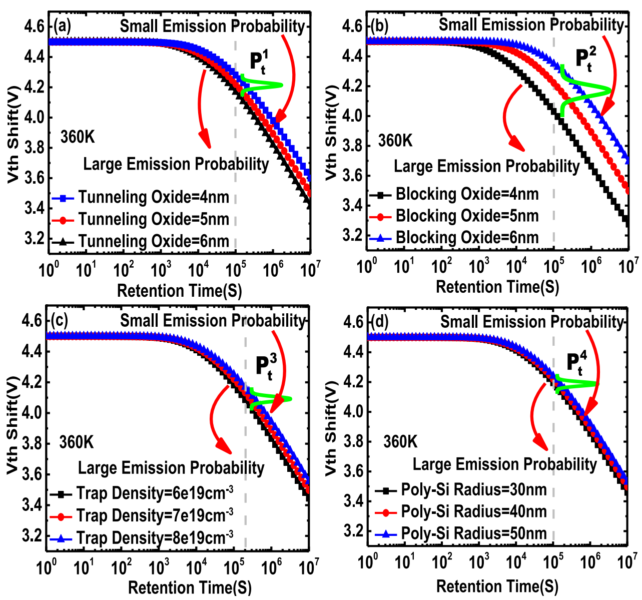

FIGURE 6. Simulated retention V_{th} shift with same initial V_{th} (4.5V) at different memory cell device parameters (a) tunneling oxide layer thicknesses, (b) blocking oxide layer thicknesses, (c) trap density of nitride layer, and (d) Poly-Si channel radiuses, respectively. The adopted standard memory device parameters: Lgate/Lspacer=30/30nm, Poly-Si channel radius = 40nm, and O/N/A=5/8/5nm.

C. DEVICE PARAMETERS FLUCTUATION

Device parameters fluctuation is a major reason to spread the retention V_{th} distribution. Figure 6 show the simulated retention V_{th} shift with same initial V_{th} (4.5V) considering the variability of (a) tunneling oxide layer thicknesses, (b) blocking oxide layer thicknesses, (c) trap density of nitride layer, and (d) Poly-Si channel radiuses, respectively. From the figures we can find that different device parameters lead to different electron emission probabilities (P^1_t, P^2_t, P^3_t, P^4_t). In this work, we use Gaussian distributions to describe the effect of different device parameters fluctuation as shown in the Figure 6 (a)-(d) due to device parameters fluctuation tends to have a Gaussian distribution [30]–[34].

However, Figure 6 (a)-(d) only shows the effect of single device parameter fluctuation. When coupling all the effects of different device parameters variability (more than 4) sources of NAND flash memory, P_t can be understand as the unified electron emission probability, which can be obtained from a specific function operation, and its probability density function as eq. (17):

$$f(P_t) = \frac{1}{\sqrt{2\pi \cdot w \cdot \overline{P_t}}} e^{-\frac{(P_t - \overline{P_t})^2}{2 \cdot w \cdot \overline{P_t}}} \tag{17}$$

Here, $\overline{P_t}$ is the mean of electron emission probability, and w is a calibrated parameter for different 2-D and 3-D NAND flash memory technologies, which reflects the difference of electron emission capability due to device parameters fluctuation.

The mean $\bar{n}$ of emitted electron numbers for data retention operation after the effect of device parameters fluctuation can be written as eq. (18):

$$\bar{n} = N \cdot \overline{P_t} \tag{18}$$

The probability P(n) of emitted n electrons for data retention operation due to device parameters fluctuation can be written as eq. (19):

$$P(n) = \frac{1}{\sqrt{2\pi \cdot w \cdot \bar{n}}} e^{-\frac{(n - \bar{n})^2}{2 \cdot w \cdot \bar{n}}} \tag{19}$$

Then, we can obtain the relationship between the variance of retention noise σ^2_{noise} and the mean $\bar{n}$ of emitted electron numbers due to device parameters fluctuation (DPF) can be expressed as eq. (20):

$$\sigma^2_{DPF} = w \cdot \bar{n} \tag{20}$$

Finally, the relationship between the variance of retention noise σ^2_{noise} and the mean $\bar{n}$ of emitted electron numbers can be written as eq. (21):

$$\begin{aligned}\sigma^2_{noise} &= \sigma^2_{EEES} + \sigma^2_{ENF} + \sigma^2_{DPF} \\ &= \bar{n} \cdot (1 - \frac{\bar{n}}{\bar{N}}) + w \cdot \bar{n} + \sigma^2_{N,0} \cdot \left(\left(1 - \frac{\bar{n}}{\bar{N}}\right)^2 - 1\right).\end{aligned} \tag{21}$$

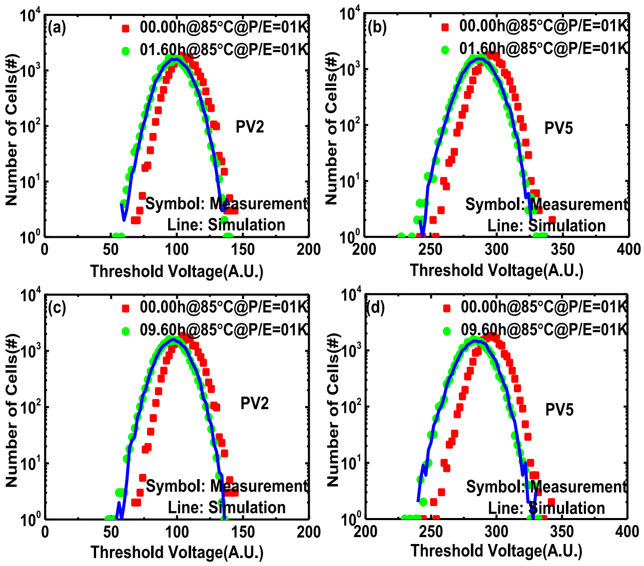

FIGURE 7. Measured V_{th} distributions for (a) PV2 and (b) PV5 after 1.6h data retention time at 85°C, which can be used to calibrated parameters w and $V_{Neutral}$. Simulated V_{th} distributions for (c) PV2 and (d) PV5 after 9.6h data retention time at 85°C also have good agreements with the measurements.

IV. RESULTS AND DISCUSSION

After model the three main noise sources of NAND flash memory as mentioned above, parameter w and $V_{Neutral}$ can be calibrated in combination with the measured retention V_{th} distributions (3-D TLC NAND flash memory) for PV2 and PV5 after 1.6h data retention time at 85°C as shown in Figure 7 (a) and (b). Then, we obtain that w=12 and $V_{Neutral} = -30$ here. And we can find that the effect of device parameters fluctuation is relatively larger than EEES and electron numbers fluctuation for our measured 3-D TLC NAND flash memory devices. Using the calibrated parameter w and $V_{Neutral}$, the simulated retention V_{th} distributions for PV2 and PV5 after 9.6h data retention time at 85°C also have good agreements with the experimental results are shown in Figure 7 (c) and (d).

Figure 8 shows the calculated mean $\overline{N}$ of initial electron numbers distributions, mean $\overline{N_t}$ of electron numbers distributions after 4.8h data retention time, and the mean $\overline{n}$ of emitted electron numbers after 4.8h data retention time at 85°C for different PV level. Here, we only show the results from PV1 to PV6 due to the incompleteness of measured V_{th} distribution for PV7.

After obtaining the parameter w, $V_{Neutral}$, and mean $\overline{n}$ of emitted electron numbers no matter what data retention time and PV level, according to eq. (21), we can obtain the variance of retention noise, which reflect the spread degree of retention V_{th} distribution. Figure 9 and Figure 10 show the simulated V_{th} distributions versus measured one after 1.6h and 9.6h data retention time at 85°C, and good agreements are obtained, which validate our proposed analytic models. Especially, this models can be applied to different 2-D and 3-D NAND flash memory technologies and application scenarios, thus, this work provides a method to predict the V_{th} distributions accurately and efficiently, which can be used for read voltage optimization [35], [36]

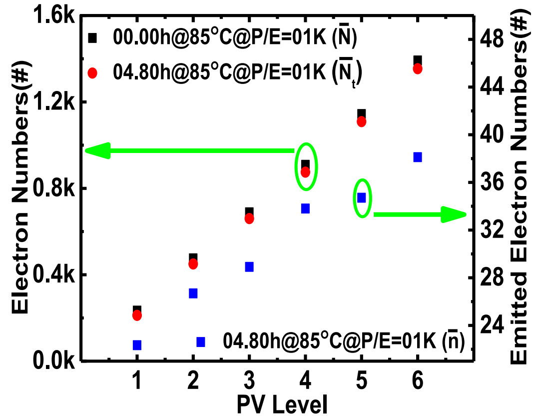

FIGURE 8. Calculated mean $\overline{N}$ of initial electron numbers distribution, mean $\overline{N_t}$ of electron numbers distribution after 4.8h data retention time, and the mean $\overline{n}$ of emitted electron numbers after 4.8h data retention time at 85°C.

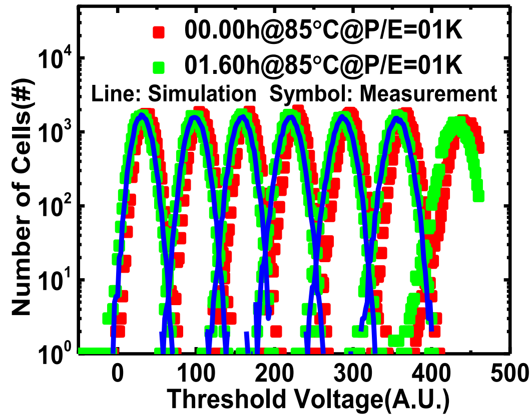

FIGURE 9. Simulated retention V_{th} distributions versus measured one after 1.6h data retention time at 85°C baking temperature for different PV level.

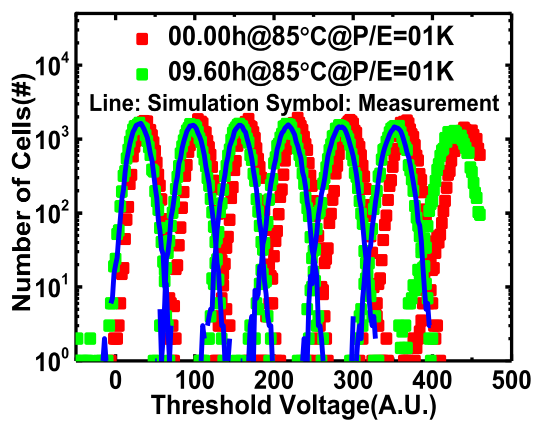

FIGURE 10. Simulated retention V_{th} distributions versus measured one after 9.6h data retention time at 85°C baking temperature for different PV level.

and design for more powerful ECC decoding algorithm to improve the reliability of NAND flash memory.

V. CONCLUSION

In this paper, the retention noise after program operation of 3-D TLC NAND flash memory is investigated. Three main noise sources of NAND flash memory to broaden the retention V_{th} distribution for data retention are investigated, and the corresponding analytic models are developed. Using the proposed models, the simulated V_{th} distributions after different data retention times have good agreements with the experimental results, and it can be applied to different 2-D and 3-D NAND flash memory technologies and application scenarios. This work provides a method to predict the V_{th} distributions accurately and efficiently, which can be used for read voltage optimization and design for more powerful ECC decoding algorithm to improve the reliability of NAND flash memory.

REFERENCES

[1] H. Tanaka *et al.*, "Bit cost scalable technology with punch and plug process for ultra high density flash memory," in *VLSI Dig. Tech. Papers*, Kyoto, Japan, 2007, pp. 178–179, doi: 10.1109/VLSIT.2007.4339708.

[2] S. Lee *et al.*, "A 1Tb 4b/cell 64-stacked-WL 3D NAND flash memory with 12MB/s program throughput," in *IEEE ISSCC Dig. Tech. Papers*, San Francisco, CA, USA, 2018, pp. 340–341, doi: 10.1109/ISSCC.2018.8310323.

[3] A. Kobayashi, T. Tokutomi, and K. Takeuchi, "Versatile TLC NAND flash memory control to reduce read disturb errors by 85% and extend read cycles by 6.7-times of read-hot and cold data for cloud data centers," in *Proc. VLSI Circuits*, Honolulu, HI, USA, 2016, pp. 1–2, doi: 10.1109/VLSIC.2016.7573505.

[4] S. Tanakamaru, Y. Yanagihara, and K. Takeuchi, "Error-prediction LDPC and error-recovery schemes for highly reliable solid-state drives (SSDs)," *IEEE J. Solid-State Circuits*, vol. 48, no. 11, pp. 2920–2933, Nov. 2013, doi: 10.1109/JSSC.2013.2280078.

[5] K. Lee, M. Kang, Y. Hwang, and H. Shin, "Accurate lifetime estimation of sub-20-nm NAND flash memory," *IEEE Trans. Electron Devices*, vol. 63, no. 2, pp. 659–666, Feb. 2016, doi: 10.1109/TED.2015.2509004.

[6] K. Lee and H. Shin, "Investigation of retention characteristics for trap-assisted tunneling mechanism in sub 20-nm NAND flash memory," *IEEE Trans. Device Mater. Rel.*, vol. 17, no. 4, pp. 758–762, Dec. 2017, doi: 10.1109/TDMR.2017.2772046.

[7] H.-J. Kang *et al.*, "Comprehensive analysis of retention characteristics in 3-D NAND flash memory cells with tube-type poly-Si channel structure," in *VLSI Dig. Tech. Papers*, Kyoto, Japan, 2015, pp. 182–183, doi: 10.1109/VLSIT.2015.7223670.

[8] C. M. Compagnoni *et al.*, "Investigation of the threshold voltage instability after distributed cycling in nanoscale NAND flash memory arrays," in *IRPS Tech. Dig.*, Anaheim, CA, USA, 2010, pp. 604–610, doi: 10.1109/IRPS.2010.5488762.

[9] C. Miccoli, C. M. Compagnoni, S. Beltrami, A. S. Spinelli, and A. Visconti, "Threshold-voltage instability due to damage recovery in nanoscale NAND flash memories," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2406–2414, Aug. 2011, doi: 10.1109/TED.2011.2150751.

[10] D. Resnati, G. Nicosia, G. M. Paolucci, A. Visconti, and C. M. Compagnoni, "Cycling-induced charge trapping/detrapping in flash memories—Part I: Experimental evidence," *IEEE Trans. Electron Devices*, vol. 63, no. 12, pp. 4753–4760, Dec. 2016, doi: 10.1109/TED.2016.2617888.

[11] D. Resnati, G. Nicosia, G. M. Paolucci, A. Visconti, and C. M. Compagnoni, "Cycling-induced charge trapping/detrapping in flash memories—Part II: Modeling," *IEEE Trans. Electron Devices*, vol. 63, no. 12, pp. 4761–4768, Dec. 2016, doi: 10.1109/TED.2016.2617890.

[12] G. Q. Dong, Y. Y. Pan, N. Xie, C. Varanasi, and T. Zhang, "Estimating information-theoretical NAND flash memory storage capacity and its implication to memory system design space exploration," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 9, pp. 1705–1714, Sep. 2012, doi: 10.1109/TVLSI.2011.2160747.

[13] G. Q. Dong, Y. Y. Pan, and T. Zhang, "Using lifetime-aware progressive programming to improve SLC NAND flash memory write endurance," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 22, no. 6, pp. 1270–1280, Jun. 2014, doi: 10.1109/TVLSI.2013.2267753.

[14] T. Parnell, N. Papandreou, T. Mittelholzer, and H. Pozidis, "Modelling of the threshold voltage distributions of sub-20nm NAND flash memory," in *Proc. IEEE Glob. Commun. Conf.*, Austin, TX, USA, 2014, pp. 2351–2356, doi: 10.1109/GLOCOM.2014.7037159.

[15] D.-H. Lee and W. Sung, "Estimation of NAND flash memory threshold voltage distribution for optimum soft-decision error correction," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 440–449, Jan. 2013, doi: 10.1109/TSP.2012.2222399.

[16] D.-H. Lee and W. Sung, "Decision directed estimation of threshold voltage distribution in NAND flash memory," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 919–927, Feb. 2014, doi: 10.1109/TSP.2013.2295056.

[17] H. B. Li, "Modeling of threshold voltage distribution in NAND flash memory: A Monte Carlo method," *IEEE Trans. Electron Devices*, vol. 63, no. 9, pp. 3527–3532, Sep. 2016, doi: 10.1109/TED.2016.2593913.

[18] C. Miccoli *et al.*, "Impact of neutral threshold-voltage spread and electron-emission statistics on data retention of nanoscale NAND flash," *IEEE Electron Device Lett.*, vol. 31, no. 11, pp. 1202–1204, Nov. 2010, doi: 10.1109/LED.2010.2069082.

[19] C. M. Compagnoni, R. Gusmeroli, A. S. Spinelli, and A. Visconti, "Analytical model for the electron-injection statistics during programming of nanoscale NAND flash memories," *IEEE Trans. Electron Devices*, vol. 55, no. 11, pp. 3192–3199, Nov. 2008, doi: 10.1109/TED.2008.2003332.

[20] G. Molas *et al.*, "Degradation of floating-gate memory reliability by few electron phenomena," *IEEE Trans. Electron Devices*, vol. 53, no. 10, pp. 2610–2619, Oct. 2006, doi: 10.1109/TED.2006.882284.

[21] K.-T. Park *et al.*, "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*, vol. 43, no. 4, pp. 919–928, Apr. 2008, doi: 10.1109/JSSC.2008.917558.

[22] J.-D. Lee, S.-H. Hur, and J.-D. Choi, "Effects of floating-gate interference on NAND flash memory cell operation," *IEEE Electron Device Lett.*, vol. 23, no. 5, pp. 264–266, May 2002, doi: 10.1109/55.998871.

[23] D. Resnati *et al.*, "Characterization and modeling of temperature effects in 3-D NAND flash arrays—Part I: Polysilicon-induced variability," *IEEE Trans. Electron Devices*, vol. 65, no. 8, pp. 3199–3206, Aug. 2018, doi: 10.1109/TED.2018.2838524.

[24] G. Nicosia *et al.*, "Characterization and modeling of temperature effects in 3-D NAND flash arrays—Part II: Random telegraph noise," *IEEE Trans. Electron Devices*, vol. 65, no. 8, pp. 3207–3213, Aug. 2018, doi: 10.1109/TED.2018.2839904.

[25] Z. Y. Lun *et al.*, "Simulation on endurance characteristic of charge trapping memory," in *Proc. IEEE SISPAD*, Glasgow, U.K., 2013, pp. 92–295, doi: 10.1109/SISPAD.2013.6650632.

[26] Z. Y. Lun *et al.*, "Investigation of retention behavior for 3D charge trapping NAND flash memory by 2D self-consistent simulation," in *Proc. IEEE SISPAD*, Yokohama, Japan, 2014, pp. 141–144, doi: 10.1109/SISPAD.2014.6931583.

[27] Z. Y. Lun, G. Du, K. Zhao, X. Liu, and Y. Wang, "A two-dimensional simulation method for investigating charge transport behavior in 3-D charge trapping memory," *Sci. China Inf. Sci.*, vol. 59, no. 12, pp. 192–201, Dec. 2016, doi: 10.1007/s11432-015-5475-7.

[28] K. Wang, G. Du, Z. Y. Lun, W. Y. Chen, and X. Y. Liu, "Modeling of program Vth distribution for 3-D TLC NAND flash memory," *Sci. China Inf. Sci.*, to be published, doi: 10.1007/s11432-018-9490-1.

[29] Y.-Y. Chiu, M. Aoki, M. Yano, and R. Shirota, "Impact of string pattern on the threshold-voltage spread of program-inhibited cell in NAND flash," *IEEE J. Electron Devices Soc.*, vol. 4, no. 4, pp. 174–178, Jul. 2016, doi: 10.1109/JEDS.2016.2565820.

[30] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1837–1852, Sep. 2003, doi: 10.1109/TED.2003.815862.
[31] A. Cathignol *et al.*, "Quantitative evaluation of statistical variability sources in a 45-nm technological node LP N-MOSFET," *IEEE Electron Device Lett.*, vol. 29, no. 6, pp. 609–611, Jun. 2008, doi: 10.1109/LED.2008.922978.
[32] A. Spessot *et al.*, "Compact modeling of variability effects in nanoscale NAND flash memories," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2302–2309, Aug. 2011, doi: 10.1109/TED.2011.2147319.
[33] D. Nagy *et al.*, "FinFET versus gate-all-around nanowire FET: Performance, scaling, and variability," *IEEE J. Electron Devices Soc.*, vol. 6, pp. 332–340, Mar. 2018, doi: 10.1109/JEDS.2018.2804383.
[34] D. Logoteta, N. Cavassilas, A. Cresti, M. G. Pala, and M. Bescond, "Impact of the gate and insulator geometrical model on the static performance and variability of ultrascaled silicon nanowire FETs," *IEEE Trans. Electron Devices*, vol. 65, no. 2, pp. 424–430, Feb. 2018, doi: 10.1109/TED.2017.2785123.
[35] B. Peleato, R. Agarwal, J. M. Cioff, M. Qin, and P. H. Siegel, "Adaptive read thresholds for NAND flash," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3069–3081, Sep. 2015, doi: 10.1109/TCOMM.2015.2453413.
[36] C. A. Aslam, Y. L. Guan, and K. Cai, "Read and write voltage signal optimization for multi-level-cell (MLC) NAND flash memory," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1613–1623, Apr. 2016, doi: 10.1109/TCOMM.2016.2533498.

KUNLIANG WANG received the B.S. degree in microelectronics from Beijing Jiaotong University, Beijing, China, in 2015. He is currently pursuing the Ph.D. degree with the Institute of Microelectronics, Peking University, Beijing. His research interest is modeling of reliability in 3-D charge trapping NAND flash memory.

GANG DU received the B.S. and Ph.D. degrees in microelectronics from Peking University, Beijing, China, in 1998 and 2002, respectively, where he is currently a Professor with the Institute of Microelectronics. His research interests include Monte Carlo simulation method for nanoscale devices, carrier quasi-ballistic transport effect, and novel structure MOSFET modeling.

ZHIYUAN LUN received the B.S. degree from Jiangnan University, Wuxi, China, in 2010 and the Ph.D. degree from Peking University, Beijing, China, in 2016. He is currently a Research Engineer with Solid State Drive Department, Hisilicon Technologies, Company, Ltd., Hangzhou, China.

XIAOYAN LIU (M'01) received the B.S., M.S., and Ph.D. degrees in microelectronics from Peking University, Beijing, China, in 1988, 1991, and 2001, respectively, where she is currently a Professor with the Institute of Microelectronics. Her research interests include nanoscale device physics, device simulation, and nanoscale device modeling.