Solid-State Electronics 199 (2023) 108497

Contents lists available at ScienceDirect

Solid State Electronics

journal homepage: www.elsevier.com/locate/sse

A physical model for long term data retention characteristics in 3D NAND flash memory

Rashmi Saikia, Souvik Mahapatra *

Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, Maharashtra 400076, India

ARTICLE INFO

The review of this paper was arranged by "Francisco Gamiz"

Keywords:
ABDWT
Data retention
NAND flash
De-trapping
Lateral migration
Program level
Erase-program cycle

ABSTRACT

An Activated Barrier Double Well Thermionic Emission (ABDWT) model is used to simulate long-term Data Retention (DR) in 3D NAND Flash memory cells. The contribution due to only charge De-Trapping (DT) when adjacent cells are at the same charged state and additional contribution due to charge Lateral Migration (LM) when adjacent cells are at different charged states are modeled. The Temperature (T), Program Level (PL), and Erase-Program Cycle (EPC) impacts are studied. The model simulations are verified against the experimental data from various published reports with consistent model parameters. It is shown that at higher temperatures and programming levels, the data retention increases for DT component, whereas LM component increases with the temperature and the lateral electric field.

1. Introduction

Over the last decade, 3D-NAND flash has gained a lot of attention due to its low bit cost and ultra-high storage density, and the ability to overcome the shortcomings of the planar technology. The Silicon Nitride (SiN) based Charge Trap Layer (CTL) has been widely investigated over the floating gate because of its many advantages like ease of processing the continuous 3D structure [1–3]. Fig. 1 (a), (b), shows the bird's eye view of the Bit Cost Scalable (BiCS) 3D NAND flash architecture with a continuous SiN based CTL. Fig. 1 (c) shows the Cell V_T (V_T: threshold voltage) Distribution (CVD) of a Triple Level Cell (TLC) to account for the multiple charge storage levels that are placed in memory to achieve multi-bit capability, at the cost of reduced spacing between levels due to the fundamental limit of maximum charge storage in CTL. Despite the merits of 3D NAND, there are several reliability concerns due to the use of a continuous CTL [4]. Data retention (DR) is a key reliability issue, resulting in charge loss from a given PL and shifting of the CVD towards a lower Program Level (PL). This reduces the spacing between the lower CVD tail of a higher PL and the upper CVD tail of the adjacent lower PL, resulting in Fail Bit Count (FBC) during memory read operation [5]. DR is classified into short and long-term modes [6–10]. During short-term data retention, the charges are lost within a few seconds after programming. For long-term data retention, charge loss occurs for a longer period. Depending upon the mode, multiple charge loss mechanisms often play together making DR complex. Erase program cycle (EPC) further complicates the DR characteristics. Successful separation for these charge mechanisms is necessary to understand the behavior of each component under the different experimental conditions and how it convolutes the total DR characteristics. Therefore, it is crucial to model the individual charge loss mechanism to estimate the accurate lifetime of production-level flash cells. Due to the inherent limitation of long measurement time, long-term DR is measured at an elevated temperature (T) over a relatively shorter time (~hours), and models are used to estimate the same at end-of-use (~years) at lower (~operation) T. Only empirical models are used for the same as of today, with uncertainties in the model parameters (*e.g.*, the use of different Arrhenius T activation energy (E_A) values for different PL) [6–9]. In this work, a physics-based Activated Barrier Double Well Thermionic Emission (ABDWT) modeling framework is introduced to model and explain the long-term data retention characteristics of 3D NAND flash under different experimental conditions with consistent values of model parameters.

2. Overview of DR mechanisms

Charges stored in Charge Trap Layer (CTL) during the programming phase can reduce over time due to multiple processes, thus causing a

* Corresponding author.
E-mail address: souvik@ee.iitb.ac.in (S. Mahapatra).

https://doi.org/10.1016/j.sse.2022.108497

Available online 26 October 2022

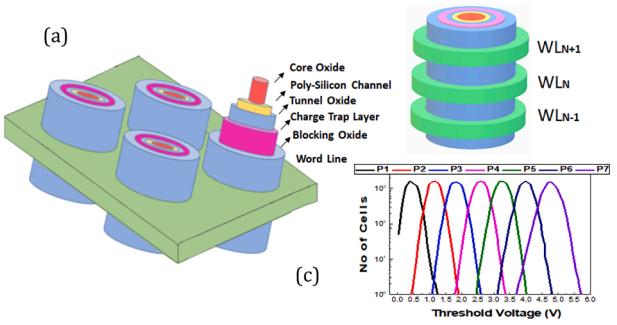

Fig. 1. (a) Schematic of BiCS memory cell with the highlighted layers. of Core Oxide, Poly-Silicon channel, Tunnel Oxide, Charge Trap Nitride Layer, Blocking Oxide, and the metal gate, (b) Schematic showing the stacked word-lines (WL_N) and (c) CVD of a triple-level cell flash from state *P*1 – P7.

reliability issue. Fig. 2 (a) schematically demonstrates the various failure mechanisms during Data Retention (DR) operation and (b) shows the schematic band diagram of the in-cell charge loss components [7]. During the de-trapping (DT) mechanism, the charges trapped in the Tunnel Oxide (TO) trap sites during the Erase Program Cycle (EPC) de-traps due to thermal excitation to the substrate. Generation of traps in TO after multiple EPC can result in charge loss via Trap-Assisted Tunneling (TAT), although this mainly affects Cell Voltage Distribution (CVD) tails and more so after a higher number of EPC steps. As shown in Fig. 2 (c) Lateral Migration (LM) is caused due to charge migration through the continuous CTL throughout the bit-line between the adjacent cells when they are not at the same Program Level (PL). Vertical Redistribution (VR) occurs during the short-term retention when the charge centroid located near the Blocking Oxide (BO) moves vertically within the nitride near the TO interface. The change in the spatial position of the trapped charges rather than the amount of trapped charge is more important and governs a change in the cell V_T (ΔV_T). This work focuses on long-term DR (so VR is ignored) due to DT and LM (TAT is not considered as only the median of the CVD is modeled).

3. Model framework

The Activated Barrier Double Well Thermionic Emission (ABDWT) model is used for calculating the time kinetics of changes in ΔV_T due to de-trapping (DT) and Lateral Migration (LM) (handled separately, total ΔV_T is the sum of individual models). Fig. 3 shows the schematic of the model along with the rate equations. The model has two stable energy states, E_1 and E_2, separated by an energy barrier E_B. The barrier has a Gaussian distribution in energy, with T activated mean (E_{B_MEAN}) and spread (E_{B_SPREAD}) with activation energy E_{A_MEAN} and E_{A_SPREAD} respectively. Once the cell is programmed, charges in the Charge Trap Layer (CTL) are considered to be in the E_1 state. The charges in CTL set up an electric field (E_{OX}) across Tunnel Oxide (TO), Blocking Oxide (BO), and between adjacent cells in the bit-line when their Program Level (PL) is different. DT is caused by E_{OX} across TO (DT across BO is ignored due to its larger thickness), and LM is due to lateral E_{OX} between

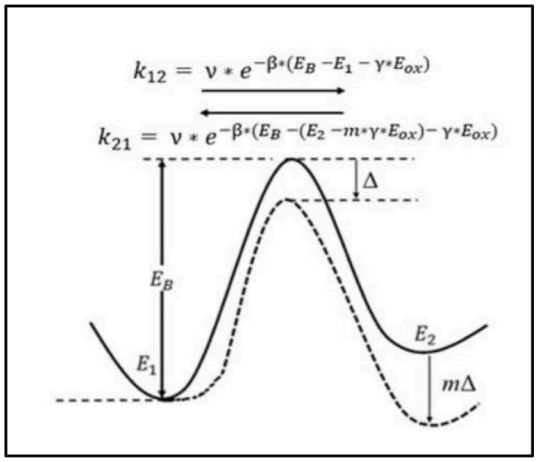

Fig. 3. Schematic of ABDWT model with the rate equations. [11].

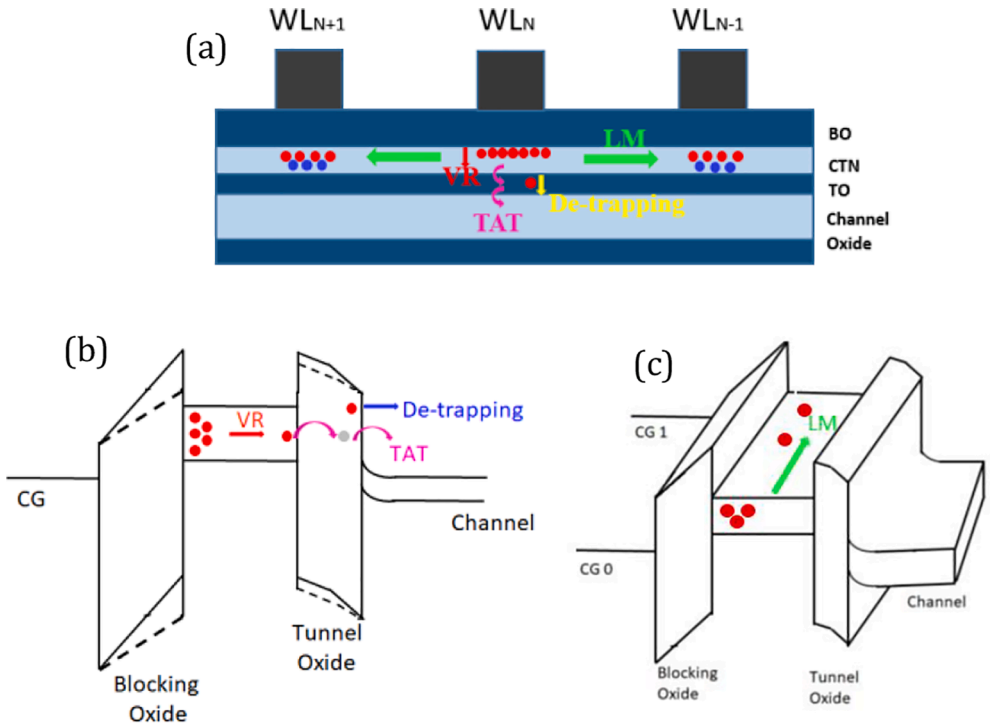

Fig. 2. (a) Cross- sectional explanation of VR, LM, de-trapping, and TAT mechanism. Schematic band diagram of charge loss during short- term retention, (b) vertical charge loss (solid line-before, dash line-after), (c) lateral charge loss. re-created as per [6].

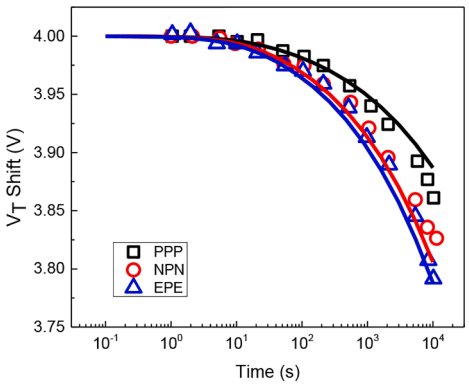

Fig. 4. Modeling of V_T shifts versus retention time under different adjacent cell states (neutral (N), erase (E) and program (P)) at 90 °C. Experimental data from [12]. Symbol-data, Line-Model.

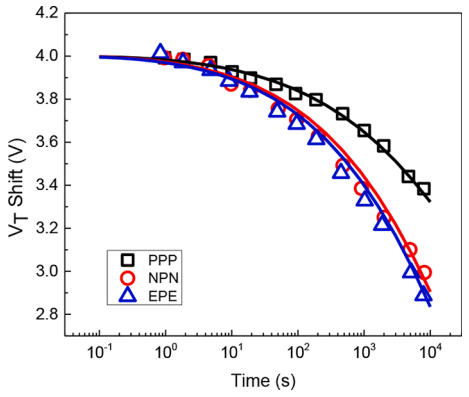

Fig. 5. Modeling of V_T shifts versus retention time under different adjacent cell states (neutral (N), erase (E) and program (P)) at 200 °C. Experimental data from [12]. Symbol-data, Line-Model.

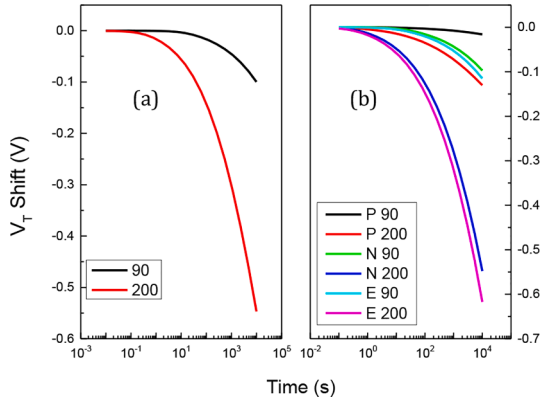

Fig. 6. Component time kinetics (a) de-trapping and (b) LM component at 90 °C and 200 °C for PPP, NPN, and EPE conditions.

adjacent cells. E_{OX} lowers E_B (by Δ) and E_2 (by m*Δ), $\Delta = \gamma * E_{OX}$, and γ is the field acceleration factor. The lowering of E_B triggers over-the-barrier thermionic emission and charge transfer from state E_1 to state E_2; reduction in charges for state E_1 is related to DR. N_0 is a pre-factor related to initial trap density, β is thermal energy (1/kT), and υ is the attempt-to-escape frequency. E_1 is the reference level, all energies are w. r.t E_1. The value of the attempt to escape frequency is taken as $1*10^{13}$ per second. See [11] for further details on the model.

Table 1
ABDWT model parameters for de-trapping and lateral migration for Fig. 4 and Fig. 5.

Parameters	DT	LM
E_2 (eV)	0.12	0.12
E_{B_MEAN} (eV)	2.8	2.8
E_{B_SPREAD} (eV)	0.3	0.3
E_{A_MEAN} (eV)	0.013	0.013
E_{A_SPREAD} (eV)	0.0047	0.0047
γ (C. cm)	8.5e-9	8.5e-9
m (no unit)	8	8
N_0(/cm^2)	2e13	2e13

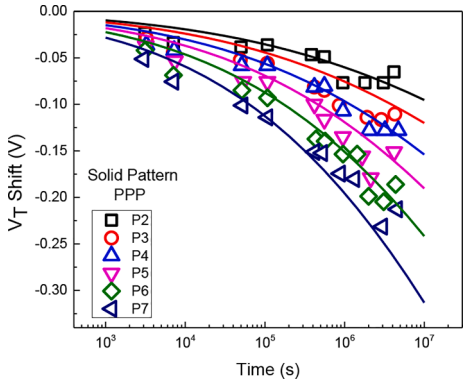

Fig. 7. Modeling of the V_T shift during solid patterns (PPP) at 85 °C with cell state changing from *P*2 state to P7 state. Experimental data from [13]. Symbol-data, Line-Model.

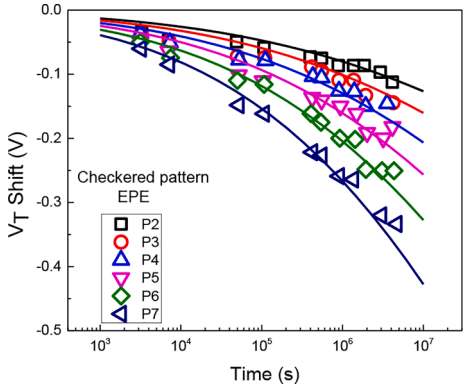

Fig. 8. Modeling of the V_T shift in checkered patterns (EPE) at 85 °C for target cell state changing from *P*2 state to P7 state. Experimental data from [13]. Symbol-data, Line-Model.

4. Model validation

Fig. 4 and Fig. 5 show experimental Data Retention (DR) characteristics from [12], at 2 different DR bake T, and for 3 different DR conditions at each T: the target cell is fixed at 4 V, while adjacent cells along the bit-line are at 3 V (PPP, all three cells are programmed), 0 V (NPN, adjacent cells are neutral, the target cell is programmed), and –1.5 V (EPE, adjacent cells are erased, the target cell is programmed), along with model calculation with underlying DT and LM components shown in Fig. 6.

In PPP, the LM has the least contribution as lateral E_{OX} between the target and adjacent cell is small, and DT due to E_{OX} across TO dominates. As the neighboring cell V_T reduces to neutral (N) and erase (E), LM due to lateral E_{OX} between the target and the adjacent cells increases but DT stays fixed. Both components get activated at higher T.

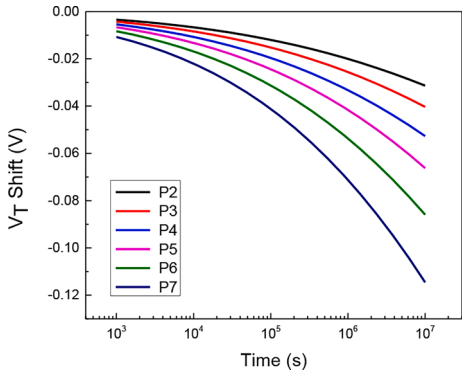

Fig. 9. Time kinetics of LM component during the checkered condition for *P*2 state to P7 state.

Table 2
ABDWT model parameters for de-trapping and lateral migration for Fig. 7 and Fig. 8.

Parameters	DT	LM
E_2 (eV)	0.12	0.16
E_{B_MEAN} (eV)	2.8	2.8
E_{B_SPREAD} (eV)	0.3	0.3
E_{A_MEAN} (eV)	0.013	0.013
E_{A_SPREAD} (eV)	0.0047	0.0047
γ (C. cm)	1.5e-9	1.5e-9
m (no unit)	7	7
N_0(/cm^2)	4e13	4e13

Parameters are listed in Table 1, identical values are used for both DT and LM to model different experimental conditions.

Fig. 7 and Fig. 8 show experimental DR characteristics from [13], for solid (PPP, all three cells programmed) and checkered (EPE, adjacent cells are erased, the target is programmed) patterns, with target cell at different PL (P2 to P7), along with model calculation. In this case, DR experiments are done after the cells are subjected to 300 EPC. LM contribution is negligible for the solid pattern as all three cells (target and two adjacent) are programmed at the same PL. However, it gets triggered in a checkered pattern because the adjacent cells are fixed at an erased state but the target cell is programmed at a different PL. As shown in Fig. 9, as the difference between the target cell V_T and the adjacent cell V_T increases the LM component also increases. The DT contribution increases for higher PL, however for a particular PL, it is kept fixed between solid and checkered patterns as the target cell voltage remains the same in both experiments(*i.e.*, the DT model of Fig. 7 is also used for Fig. 8). Parameters are listed in Table 2.

Except for E_2, identical values are used for all parameters to model DT and LM for different PL. We speculate that the E_2 value is different between DT and LM due to the charge accumulated in the inter-cell regions during EPC in these experiments. Some of the parameters are slightly different between the two sources of data and are presumably due to structural and material properties being different between different production lines.

5. Conclusion

Long-term Data Retention is successfully modeled with De-trapping (DT) and Lateral Migration (LM) contributions calculated using the ABDWT framework. DR gets accelerated at a higher temperature (T), from a higher program level (PL). The Checkered programming pattern compared to the solid programming pattern shows higher DR due to the additional component LM along with the DT which also shows positive temperature activation. DR also increases as the difference in PL between the adjacent cells and the target cell increases as the lateral field increases the LM component, and all aspects can be handled by the ABDWT model framework. Importantly, only DT affects the DR in the solid pattern configuration, and an identical DT contribution is used for a given PL between solid and checkered patterns. Therefore, this physics-based framework can be used to determine end-of-life DR during the product qualification phase.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Proper reference has been provided for the data.

References

[1] Micheloni R, Aritome S, Crippa L. Array Architectures for 3-D NAND Flash Memories. Proc IEEE 2017;105(9):1634–49. https://doi.org/10.1109/JPROC.2017.2697000.
[2] Micheloni R, Crippa L, Zambelli C, Olivo P. Architectural and Integration Options for 3D NAND Flash Memories. Computers 2017;6:27. https://doi.org/10.3390/computers6030027.
[3] Monzio Compagnoni C, Spinelli AS. Reliability of NAND Flash Arrays: A Review of What the 2-D–to–3-D Transition Meant. IEEE Trans Electron Devices 2019;66(11):4504–16. https://doi.org/10.1109/TED.2019.2917785.
[4] Raghunathan S. (Invited) 3D-NAND Reliability: Review of key mechanisms and mitigations,". In: 2020 4th IEEE Electron Devices Technology & Manufacturing Conference (EDTM); 2020. p. 1–4. https://doi.org/10.1109/EDTM47692.2020.9117872.
[5] Cai Y, Luo Y, Haratsch EF, Mai K, Mutlu O. Data retention in MLC NAND flash memory: Characterization, optimization, and recovery. In: 2015 IEEE 21st International Symposium on High Performance Computer Architecture (HPCA); 2015. p. 551–63. https://doi.org/10.1109/HPCA.2015.7056062.
[6] Woo C, et al. Modeling of Charge Loss Mechanisms during the Short Term Retention Operation in 3-D NAND Flash Memories. Symp VLSI Technol 2019;2019:T214–5. https://doi.org/10.23919/VLSIT.2019.8776579.
[7] Kim S, et al. Separation of Lateral Migration Components by Hole During the Short-Term Retention Operation in 3-D NAND Flash Memories. IEEE Trans Electron Devices 2020;67(6):2645–7. https://doi.org/10.1109/TED.2020.2989734.
[8] Kim S, Lee K, Woo C, Hwang Y, Shin H. Analysis of Failure Mechanisms During the Long-Term Retention Operation in 3-D NAND Flash Memories. IEEE Trans Electron Devices 2020;67(12):5472–8. https://doi.org/10.1109/TED.2020.3028349.
[9] Woo C, et al. Modeling of Charge Failure Mechanisms during the Short Term Retention Depending on Program/Erase Cycle Counts in 3-D NAND Flash Memories. In: 2020 IEEE International Reliability Physics Symposium (IRPS); 2020. p. 1–6. https://doi.org/10.1109/IRPS45951.2020.9129306.
[10] Choi B, et al. Comprehensive evaluation of early retention (fast charge loss within a few seconds) characteristics in tube-type 3-D NAND flash memory. IEEE Symp VLSI Technol 2016;2016:1–2. https://doi.org/10.1109/VLSIT.2016.7573385.
[11] Choudhury N, Parihar N, Goel N, Thirunavukkarasu A, Mahapatra S. Modeling of DC - AC NBTI Stress - Recovery Time Kinetics in P-Channel Planar Bulk and FDSOI MOSFETs and FinFETs. IEEE J Electron Devices Soc 2020;8:1281–8. https://doi.org/10.1109/JEDS.2020.3023803.
[12] Kang H-J, et al. Comprehensive analysis of retention characteristics in 3-D NAND flash memory cells with tube-type poly-Si channel structure. In: 2015 Symposium on VLSI Technology (VLSI Technology); 2015. T182–3. https://doi.org/10.1109/VLSIT.2015.7223670.
[13] Mizoguchi K, Kotaki S, Deguchi Y, Takeuchi K. Lateral charge migration suppression of 3D-NAND flash by vth nearing for near data computing. pp. 19.2.1-19.2.4. In: 2017 IEEE International Electron Devices Meeting (IEDM); 2017. https://doi.org/10.1109/IEDM.2017.8268420.