Carrier Energy-Dependent Capture Model for Improved CTF TCAD Simulations

Bokyeom Kim[1*], Uihui Kwon[1], Woon Ih Choi[1], HyeYoung Kwon[1], and Dae Sin Kim[1]
[1]Computational Science and Engineering Team, Samsung Electronics, Hwasung-si, Gyeonggi-do, Republic of Korea
*Email Address: bk777.kim@samsung.com

***Abstract*— Kinetic description of carrier transport & trapping across a charge trap layer (CTL) is essential for the accurate modeling of 3D-NAND charge-trap flash devices. In many conventional studies, the immediate thermalization and trapping is assumed, which result in the over-estimation of programming speed. To tackle this issue, several approaches have been proposed to reflect the energy-dependent capture probability of the tunneling-injected carrier. However, there were limitations in improving the predictability of programming behavior in the high-kinetic-energy regime. Through in-depth CTF modeling studies, we could find out that it is important to accurately trace the energy loss of electron through inelastic scattering across CTL for the accurate fitting across the entire bias region. Our new model demonstrates superior calibration capability compared to the conventional approaches without modifying material parameter depending on program voltage. The over-estimation of trapped charges in conventional approaches is mainly caused by the uncaptured electrons having high-energy injected over high thermal barrier in multiphonon emission theory. Our model mitigates well programming slope overestimation and removes unphysical behavior observed in the injection-energy-independent relaxation model, which deepens the understanding of charge trapping dynamics and improves the predictability for process optimization for next-generation NAND flash technologies.**

Keywords—CTF, Capture, TCAD, energy relaxation

I. Introduction

Charge-trap flash (CTF) memory, particularly in 3D-NAND architectures, has become a cornerstone of modern nonvolatile storage, prompting continued efforts to improve the physical understanding of the atomistic origin of CTF behavior [1]. A detailed understanding of carrier transport within the charge trap layer (CTL), as well as the spatial distribution of trapped charge, is essential for accurately predicting program/erase and retention characteristics. However, many conventional studies assume instantaneous thermalization and trapping, which leads to an overestimation of programming speed [2,3]. In reality, the electrons are injected into the CTL with high energy and undergo energy loss through scattering within the CTL before being captured. To mitigate these inaccuracies, various models have been introduced to account for the energy dependence of carrier capture in the CTL. These include approaches such as modulating the capture probability as a function of electron kinetic energy [4,5], and employing an empirical Gaussian distribution to define a net generation rate within the CTL [3].

In this study, in addition to conventional approaches, we demonstrate that accurately tracking the energy loss of electron due to inelastic scattering within the CTL is crucial for the precise calibration of CTF devices. We incorporate a key physical mechanism neglected in previous works: the injection-energy-dependent relaxation length of high-energy electrons. This effectively prevents both the overestimation of the ISPP (Incremental Step Pulse Programming) slope and unphysical programming curves observed in the injection-energy-independent relaxation model.

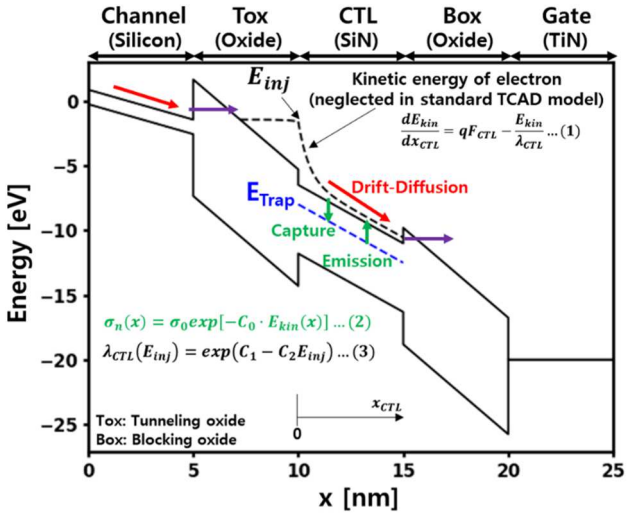

Fig.1 Schematic of TONOS stack illustrated with energy band diagrams, where key physical mechanisms are highlighted using colored arrows. To indicate the energy of the electron injected into the conduction band, the calculated kinetic energy was plotted above the conduction band edge as a dashed line.

II. Methods

We investigate the programming dynamics of a CTF device composed of a TiN-Oxide-Nitride-Oxide-Silicon (TONOS) stack. A schematic of the simulations is shown in Fig. 1. The simulation framework and parameters are adopted from previous studies [3–5]. The drift-diffusion, Poisson, and Shockley–Read–Hall (SRH) equations are solved self-consistently with trapping and tunneling mechanisms incorporated within the SRH framework [3,5]. As illustrated in Fig.1, electrons with kinetic energies exceeding 1 eV is injected into CTL, resulting in two primary effects: (1) a decrease in capture probability, and (2) an increase in scattering. The following section provides a detailed explanation of how effects (1) and (2) are incorporated into conventional TCAD simulations.

A. Decrease in Capture Probability

The multiphonon emission (MPE) theory for non-radiative carrier capture describes the transition from a free

state at the conduction band minimum (CBM) to a bound state at the trap level [6]. However, in CTF device, free electrons with kinetic energies exceeding 1 eV are injected into CTL. As the electron energy increases, the thermal barrier (E_B) for the transition from high-energy free states to the bound state also increases (Fig.2) [7]. Consequently, the capture cross section (σ_c), which represents the capture probability, decreases exponentially with E_B ($\sigma_c \propto exp(-E_B/k_BT)$).

To account for the reduced capture probability of high-energy electrons, we first calculate the kinetic energy of electrons (E_{kin}) injected into CTL using

$$\frac{dE_{kin}}{dx_{CTL}} = qF_{CTL} - \frac{E_{kin}}{\lambda_{CTL}} \dots (1)$$

where x_{CTL} is a spatial coordinate starting from tunneling oxide (Tox)/CTL interface [4-5]. Injection energy (E_{inj}), i.e., $E_{kin}(0)$, is calculated by subtracting conduction band energy at the Tox/CTL interface from that at the substrate/Tox interface. λ_{CTL} denotes the energy relaxation length, and F_{CTL} represents the electric field along the CTL. The capture cross section is modeled as the following formulation to decay exponentially with increasing E_{kin}

$$\sigma_c = \sigma_0 \, exp\big(-C_0 \cdot E_{kin}(x_{CTL})\big) \dots (2)$$

where C_0 is the decay factor, and σ_0 denotes the capture cross section for the transition from the CBM state to the trap state.

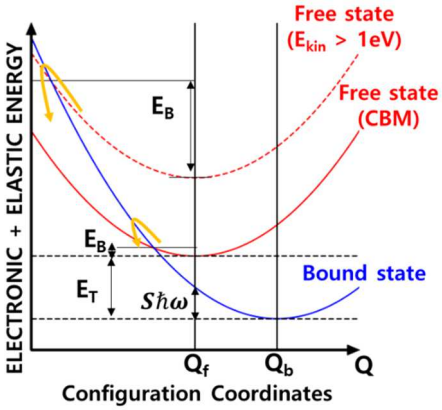

Fig.2 Schematic illustration of non-radiative electron capture process based on the MPE theory [6-7]. Electron capture is described with the transition from a free state to a bound state at the trap level. In CTF device, free electrons with kinetic energies exceeding 1 eV are injected into CTL, resulting in higher thermal barriers (E_B) for the transition. Q_f and Q_b denote configurations at free and bound state, respectively. E_T is trap depth, $S\hbar\omega$ is the reorganization energy of trap.

B. Increase in Scattering

For electron energies below 10 eV, electrons injected with higher energy lose their energy over shorter distances due to increased scattering [8]. This phenomenon has been observed in silicon nitride [9]. To capture this behavior, the energy relaxation length in the nitride, λ_{CTL}, is modeled as a decreasing function of the injected electron energy.

$$\lambda_{CTL}(E_{inj}) = exp(C_1 - C_2 E_{inj}) \; or \; C_1 E_{inj}^{-C_2} \dots (3)$$

where C_1 and C_2 are fitting parameters [10].

III. Results & Discussion

We performed simulations on two structures: Silicon-Nitride-Al_2O_3-Nitride-Oxide-Silicon (SNANOS) and TONOS stack. In SNANOS stack, the superior calibration capability of our model is proved. In TONOS stack, we investigate the impact of injection-energy-dependent relaxation on CTF programming (PGM) simulation and explore the applicability of ISPP schedule optimization by analyzing detailed physical quantities such as trapped charge density.

A. Calibration Without Changing Material Constants Depending on V_{PGM}

First, we simulated the SNANOS stack using the conventional TCAD model without considering kinetic energy. As shown in Fig. 3(a), the conventional TCAD model utilizes constant capture cross section which fails to reflect the abrupt reduction in σ_c at higher electron kinetic energies, resulting in an overly rapid increase in threshold voltage (V_T) compared to the experimental data—leading to PGM slope overestimation.

To reduce the discrepancy with experimental results, conventional TCAD model have to adjust intrinsic material constants such as electron mobility of CTL ($\mu_{n,CTL}$), tunneling mass (m_{tun}), and trap density (N_T) to fit the measured value of V_T [11,12]. For example, the electron mobility was set to a value greater than 1 $cm^2/V \cdot s$, because a higher mobility reduces the amount of charge captured by traps [11].

However, the experimentally measured mobility of amorphous silicon nitride (a-SiN) is on the order of 10^{-5} $cm^2/V \cdot s$ or even lower, whereas values around 1 $cm^2/V \cdot s$ are typical for a-Si [13-15]. In the case of CTF device, the apparent reduction in charge trapping observed in the experiment, as compared to the simulation, is not due to more than a 10^5-fold increase in electron mobility of a-SiN, but rather due to high energy electrons which fail to be captured because of a large thermal barrier E_B, as explained by MPE theory [6-7].

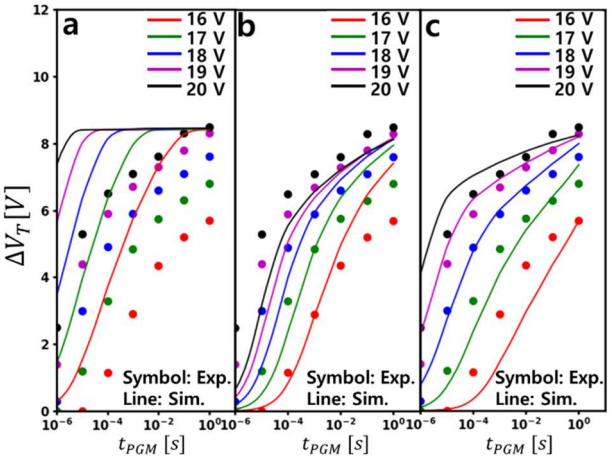

Fig.3 PGM curves for the SNANOS stack simulated by utilizing (a) conventional TCAD model, (b) injection-energy-independent relaxation model, and (c) injection-energy-dependent relaxation model

As shown in Fig. 3(b), even when incorporating the reduction in σ_c due to kinetic energy of electron while using a mobility value on the order of $10^{-5}\ cm^2/V \cdot s$, results in a decrease in the slope of PGM curves. However, this approach fails to reproduce the measured data in the high-bias regime (i.e., high-kinetic-energy regime), as it does not account for the relaxation length dependence on the injection energy.

By contrast, our proposed model, which explicitly accounts for the injection energy, captures the physical trends across the entire bias range in good agreement with experimental observations (Fig. 3c). We have found that our model exhibits high calibration capability, as it can reproduce experimental results by utilizing the energy dynamics of electrons without requiring any modification of material parameters such as $\mu_{n,CTL}$, m_{tun}, and N_T.

B. Injection-energy-dependent relaxation length

In TONOS stack, to investigate the impact of injection-energy-dependent relaxation length in detail, we simulate two cases: (1) an injection-energy-independent relaxation length and (2) an injection-energy-dependent relaxation length. In both cases, the programming slope is reduced compared to the conventional TCAD (Fig. 4a, 4c). However, when an injection-energy-independent relaxation length is used, programming becomes less efficient as programming bias (V_{PGM}) increases. This unphysical behavior, inconsistent with experimental observations, has also been observed in prior study using an injection-energy-independent relaxation length (Fig. 4(b) in [4]). This is because, at higher injection energies, electrons traverse the CTL without sufficient energy loss, leading to reduced capture probability during transport (Fig. 4b). In contrast, in our model, electrons injected at higher V_{PGM} lose energy more rapidly near the Tox/CTL interface and got more effectively trapped (Fig. 4d).

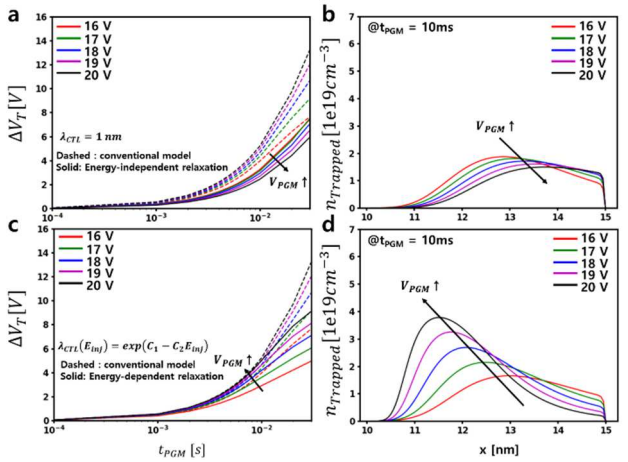

Fig.4 PGM curves and the trapped charge concentration at 10ms calculated for the TONOS stack: (a,b) injection-energy-independent model: (a) PGM curve, (b) trapped charge concentration; (c,d) injection-energy-dependent relaxation model: (c) PGM curve, (d) trapped charge concentration.

Through our study, we have identified the physical origin of the Gaussian-shaped distribution [3]. When a 1 ms pulse is applied, the centroid of trapped charge varies with the applied bias (Fig. 5). At lower biases, electrons experience less scattering and are captured farther from the Tox/CTL interface. The left side of the Gaussian corresponds to the exponentially increasing capture probability with respect to electron energy, while the right side reflects the exponential decrease in free electrons due to trapping at trap sites. This behavior was effectively modeled by injecting free electrons through a Gaussian-shaped net generation rate [3].

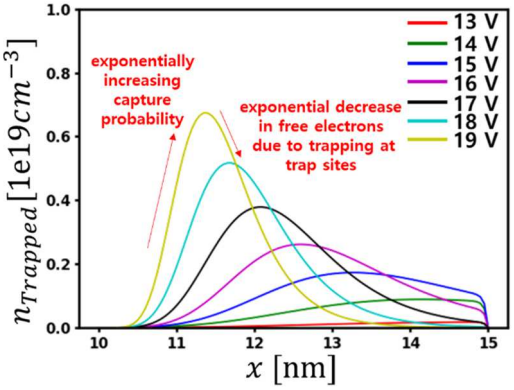

Fig.5 Trapped charge concentration after 1 ms program pulses at different bias voltages (13 V to 19 V).

Our model provides a more physical interpretation of the experimental observation that ISPP exhibits faster programming speed than decremental SPP (DSPP) or constant pulse schemes [16] (Fig. 6a). In the case of ISPP, electrons are initially trapped farther from the Tox/CTL interface, resulting in a smaller impact on the Tox electric field and the tunneling flux (Fig. 6b). In contrast, for constant pulse and DSPP, electrons are initially trapped closer to the Tox/CTL interface, which reduces the Tox electric field and suppresses the tunneling flux in the initial stage (Fig. 6c-d).

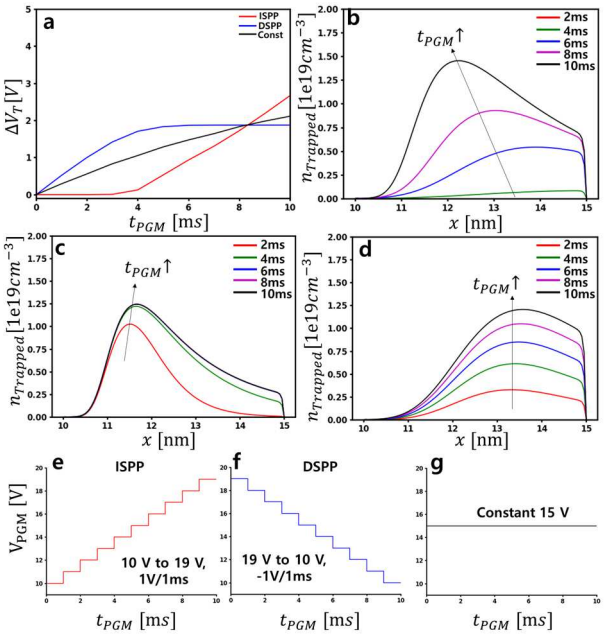

Fig.6 (a) Comparison of ISPP, DSPP, and constant PGM curves. Insets show the corresponding pulse schedules. Time evolution of trapped charge concentration for (b) ISPP, (c) DSPP, and (d) constant PGM conditions. (e) incremental, (f), decremental, and (g) constant pulse schemes are illustrated.

Using the conventional TCAD model for ISPP scheme optimization mislead optimization directions. We test three ISPP schedules: (ISPP1) a fast-ramping over a wide voltage range, (ISPP2) a slow-ramping over a narrow voltage range, and (ISPP3) a moderate-ramping over an intermediate voltage range. ISPP3 appears optimal under the conventional model, whereas ISPP1 is found to be optimal with our proposed model (Fig. 7). In our model, ISPP1 is best schedule, as it promotes charge trapping over the broadest spatial extent of the CTL. However, such an analysis is not possible with conventional TCAD models, where electrons lose energy instantaneously and become trapped immediately at the Tox/CTL interface.

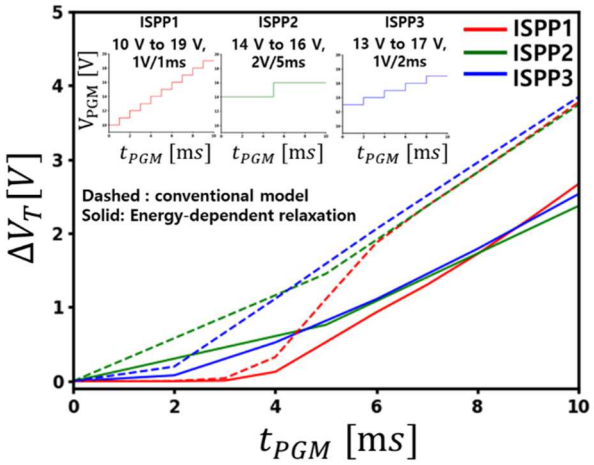

Fig.7 Comparison of PGM curves for three different ISPP schedules. Insets show the corresponding pulse schedules. The optimal ISPP schedule differs from that predicted by the conventional model.

IV. Conclusion

In this work, we developed a novel carrier energy-dependent capture model that overcomes the limitations of existing approaches, including programming slope overestimation and unphysical programming curves. The reduced charge trapping observed experimentally, compared to the simulation, is due to the reduced capture probability of high-energy electrons, attributed to the large thermal barrier, consistent with multiphonon emission theory. In addition, we have found the physical origin of the empirically used Gaussian function. Compared to conventional models, our approach shows potential for more accurate optimization of ISPP schedules.

References

[1] W. I. Choi, et al., “Switchable chemical-bond reorganization for the stable charge trapping in amorphous silicon nitride”, Adv. Mater., vol. 36, no. 9, p. 2308054, 2024.
[2] E. Nowak et al., "Intrinsic fluctuations in vertical NAND flash memories," in Symp. VLSI Technol., pp. 21–22, 2012.
[3] F. Schanovsky et al., “Modeling the Operation of Charge Trap Flash Memory–Part I: The Importance of Carrier Energy Relaxation”, IEEE Trans. Electron Devices, vol. 71. no. 1, pp. 547-553, 2023.
[4] M. Hogyoku et al., “TCAD simulation for capture/emission of carriers by traps in SiN: trap-assisted tunneling model extended for capture of carriers injected via Fowler–Nordheim tunneling”, Jpn. J. Appl. Phys., vol. 61, p. SC1087, 2022.
[5] S. Spiga et al., “Experimental and simulation study of the program efficiency of HfO2 based charge trapping memories”, 2010 Proceedings of the ESSDRC, pp. 408–411, 2010.
[6] C. H. Henry et al., “Nonradiative capture and recombination by multiphonon emission in GaAs and GaP,” Phys. Rev. B, vol. 15, no. 2, pp. 989–1016, 1977.
[7] A. Palma et al., “Electric field dependence of the electron capture cross section of neutral traps in SiO_2,” J. Electrochem. Soc., vol. 143, no. 8, pp. 2687 – 2693, 1996.
[8] M. P. Seah et al., “Quantitative electron spectroscopy of surfaces: A standard data base for electron inelastic mean free paths in solids”, Surf. Interface Anal., vol. 1, no. 1, pp. 2–11, 1979.
[9] V. I. Garmash et al., “Experimental determination of the energy dependence of electron inelastic mean free path in silicon oxide and silicon nitride,” J. Surf. Invest. X-ray, Synchrotron Neutron Tech., vol. 10, pp. 767–770, 2016.
[10] T. Tomita et al., “Energy relaxation length for ballistic electron transport in SiO_2,” Phys. Status Solidi B, vol. 204, no. 1, pp. 129 – 132, 1997.
[11] F. Schanovsky et al., “A TCAD compatible SONOS trapping layer model for accurate programming dynamics,” in Proc. IEEE Int. Mem. Workshop (IMW), 2021, pp. 1–4.
[12] F. Driussi et al., “Simulation study of the trapping properties of HfO_2-based charge-trap memory cells,” IEEE Trans. Electron Devices, vol. 61, no. 6, pp. 2056 – 2063, 2014.
[13] İ. Ay et al., “Steady-state and transient photoconductivity in hydrogenated amorphous silicon nitride films,” Sol. Energy Mater. Sol. Cells, vol. 80, no. 2, pp. 209–216, 2003.
[14] T. Güngör et al., “Drift mobility measurements in a-SiNx:H,” J. Non-Cryst. Solids, vol. 282, no. 2–3, pp. 197–202, 2001.
[15] S. A. Awan et al., “Electrical conduction processes in silicon nitride thin films prepared by RF magnetron sputtering using nitrogen gas,” Thin Solid Films, vol. 355, pp. 456–460, 1999.
[16] G. J. Hemink et al., “Fast and accurate programming method for multi-level NAND EEPROMs”, in Symp. VLSI Technol., Dig. Tech. Papers, pp. 129–130, 1995