



HAL
open science

1T-NOR Flash memory after endurance degradation: An advanced TCAD simulation

F. Matteo, R. Simola, J. Postel-Pellerin, K. Coulié

► To cite this version:

F. Matteo, R. Simola, J. Postel-Pellerin, K. Coulié. 1T-NOR Flash memory after endurance degradation: An advanced TCAD simulation. *Microelectronics Reliability*, 2022, 138, pp.114621. 10.1016/j.microrel.2022.114621 . hal-03941091

HAL Id: hal-03941091

<https://hal.science/hal-03941091v1>

Submitted on 22 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1T-NOR Flash Memory After Endurance Degradation: An advanced TCAD Simulation

F. Matteo^{a,b}, R. Simola^a, J. Postel-Pellerin^b, K. Coulié^b

^a STMicroelectronics, Rousset, France

^b Aix-Marseille University, CNRS, IM2NP, 13451 Marseille, France

Abstract

In this paper we have performed TCAD simulations of 1T-NOR Flash electrical characteristics after 1 million cycles of Program/Erase (P/E) operations. Thanks to the TCAD simulation, spatial defect distributions have been proposed to explain the endurance degradation. Process simulation was based on a 90 nm node embedded Non-Volatile Memory Technology (eNVM) produced at STMicroelectronics. Hot Carrier Injection (HCI) and advanced tunnelling models were used during Program and Erase, whereas Flash degradation was considered through defects located at Si/SiO₂ interface and inside SiO₂. Obtained results for programming window before and after cycling, as well as consumption current, are in excellent agreement with experiments. Moreover, within this framework it was possible to correctly reproduce the I-V characteristics at the unstressed Flash side after 1 million cycles as previously reported in the literature.

1. Introduction

Charge storage floating gate memory Flash-EEPROM (Electrically Erasable Programmable Read Only Memory), or simply Flash, is arguably one of the most successful Non-Volatile Memory devices that has ever been introduced in the market, with billions of units still processed each year and a compound annual growth rate (CGAR) expected to reach 14.4% until 2028 [1]. Its reliability, which mainly includes endurance and data retention, has been extensively investigated in the last decades [2-4]. In this paper we have used Technology Computer Aided Design (TCAD) to simulate 1T-NOR Flash Programming window closure after 1 million cycles. TCAD is a physics-based numerical modelling approach for accurately simulating both fabrication process and electrical characteristics of microelectronics devices, a tool that has been successfully used for device performance optimization and reliability improvement [5,6,7]. Similar to [8] Flash endurance was modelled through defects; however, in this paper we have adopted an inhomogeneous defects distribution both at Si/SiO₂ interface and in SiO₂ bulk oxide, as well as different types of defects. This approach is consistent with experimental results reported in [4].

2. TCAD simulation framework for fresh Flash

Process simulation was based on a 90nm embedded Non-Volatile Memory technology node. The obtained memory cell is shown in Fig. 1, where n-type doped regions (source, drain, polysilicon control gate and floating gate) are in red, p-type (channel, substrate) in blue, and metallic contacts and silicided areas in grey color.

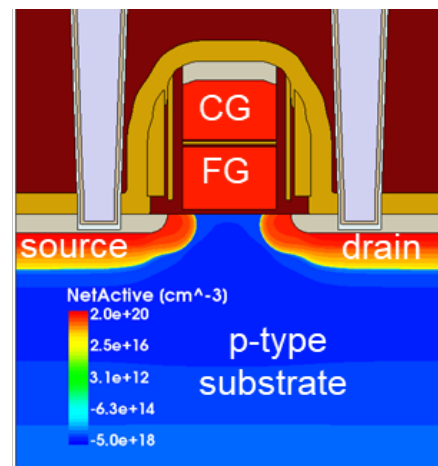


Fig. 1. TCAD process simulated 1T-NOR Flash. (Gate length $L=180$ nm, width $W=90$ nm, oxide $t_{ox}=9$ nm).

Flash gate length is $L=180$ nm and gate oxide

* Corresponding author. franck.matteo@st.com
Tel: +33 (056) 133 6000 ; Fax: +33 (056) 133 6200

thickness $t_{ox}=9$ nm. In this 2D simulation, cell width ($W=90$ nm) was taken into account through a scaling factor. For TCAD calibration we used electrical experimental data. This were obtained using a Keysight B1500A semiconductor device analyzer to apply signals for Quasi-Static Measurements (C-V and Id-Vcg reading characteristics) through Source Monitor Units (SMU), while programming and erasing dynamics were carried out by Waveform Generator Fast Measurement Units (WGFMU in B1530A modules). Endurance tests are performed using Semiconductor Pulse Generator Units (SPGU in B1525A modules), as detailed in a previous work [9].

In Fig.2, a C-V characteristic on a gate oxide large ($1 \times 10^5 \mu m^2$) plate capacitor is reported. The very good agreement between simulation and experiments indicates that oxide thickness, poly doping, channel doping profile are correctly modeled.

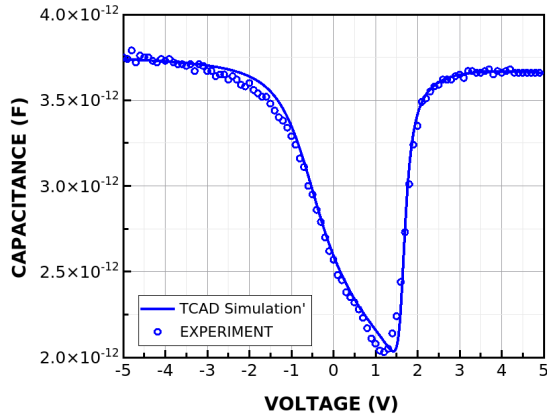


Figure 2: C-V characteristic on gate oxide plate capacitor (area: $1 \times 10^5 \mu m^2$). Good agreement between experiments and TCAD indicates process flow was correctly simulated.

For carriers transport in the channel, the so-called hydrodynamic model [10-11] was adopted since it accounts for energy transport of carriers, and it allows for non-equilibrium between hot electrons and the substrate. The physical mechanism for Flash programming is Channel Hot Injection (CHI) and was simulated using the Lucky electron model [12] for which gate current I_G is given by Eq.1 [11]:

$$I_G = \iint J_n(x, y) \cdot P_s \cdot P_{ins} \cdot \left(\int_{E_B}^{\infty} P_\varepsilon P_r d\varepsilon \right) dx dy \quad (1)$$

and

$$P_s(y) = \exp(-y/\lambda) \quad (2)$$

$$P_\varepsilon(\varepsilon) = \frac{1}{\lambda F_{eff}} \exp\left(-\frac{\varepsilon}{\lambda F_{eff}}\right) \quad (3)$$

$$P_r(\varepsilon) = \frac{1}{2\lambda_r} \left(1 - \sqrt{\frac{E_B}{\varepsilon}} \right) \quad (4)$$

where P_s is the probability that the electron will travel a distance y to the interface without losing any energy, $P_\varepsilon d\varepsilon$ in Eq. 2 the probability that the electron has energy between ε and $\varepsilon + d\varepsilon$, P_{ins} in Eq. 1 is the probability of scattering in the image force potential well, and P_r in Eq. 4 is the probability that the electron will be redirected, F_{eff} is the effective electric field in the channel. In this approach, the energy distribution of the hot electrons is assumed to be Maxwellian, and the main physical parameters are the height of the Si/SiO₂ barrier (E_B), the electron scattering mean free path in the semiconductor λ and the redirection mean free path λ_r .

For Erase operation, the physical mechanism is Fowler-Nordheim tunneling [13] through the oxide and is described by a current density J_{FN} given by Eq.5;

$$J_{FN} = AF^2 \exp(-B/F) \quad (5)$$

where F is the insulator electric field at the interface and A , B are physical related parameters. In this article, we use a slightly more general model, the so called Non-Local Tunneling, based on WKB approximation and described in [11][14].

In Fig.3 we report the Id-Vcg characteristics after a programming pulse ($V_{cg}=10$ V, $V_d=4.2$ V, $t=5 \mu s$) and after an erase pulse ($V_{cg}= -17$ V, $t= 0.5$ ms), as well as the TCAD simulated ones.

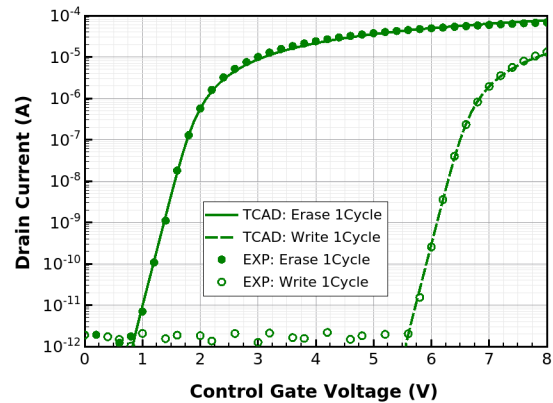


Figure 3: Flash programmed and erased characteristics, after 1 cycle (fresh cell).

This excellent agreement between experiments and simulations was obtained by fine-tuning the floating

gate capacitive coupling ratio, a mandatory step for a 2D simulated device, and by calibrating the two injection models.

3. Flash Degradation Simulation

The two mechanisms involved in Flash operation are channel hot electrons injection into the floating gate during writing and Fowler-Nordheim tunneling, during erase. Both are responsible for oxide degradation leading to programming window closure, and both have been extensively studied in the past. Concerning high field Fowler-Nordheim induced degradation, the breaking of Si-H bonds at Si/SiO₂ interface during electrical stress leads to the generation of interface traps [15]. Moreover, the generation of bulk oxide traps which results from Fowler-Nordheim stress favors charge trapping and induces leakage current (SILC) [16]. Finally, positive charges buildup in the oxide due to Anode Hole Injection (AHI) seems first to occur, while at high injection levels negative charge trapping prevails and reduces the efficiency of tunneling current injection [17-18]. It must be emphasized that degradation during Flash erase operation is evenly distributed along SiO₂ gate oxide and Si/SiO₂ interface. On the contrary, channel hot electron (CHE) generated defects are mainly located at or near drain region, towards which they have been accelerated by an intense lateral electric field. Such defects include Si-SiO₂ interface states, bulk oxide traps (for holes and electrons) and near interface oxide traps [19]. Based on these considerations, the following paradigm for explaining programming window closure in terms of defects distribution (Fig. 4 a)).

- Defects are generated in the oxide, at the Si-SiO₂ interface and in the silicon substrate.
- Defects are not evenly distributed but are more concentrated near the drain side.
- Electrons are trapped in the oxide during P/E operations.
- Holes are trapped during erase operation, through anode hole injection (AHI) mechanism (there are more trapped holes drain side due to traps generation by hot carriers).

Each trap distribution implemented in the TCAD simulation is uniform. However, we consider different types and concentrations and thus the global distribution in the structure is not uniform. Moreover, to take into account the asymmetric degradation due to the hot carrier injection through the oxide, we distinguish two regions, the Source side and Drain

side (illustrated in Fig.4 b)).

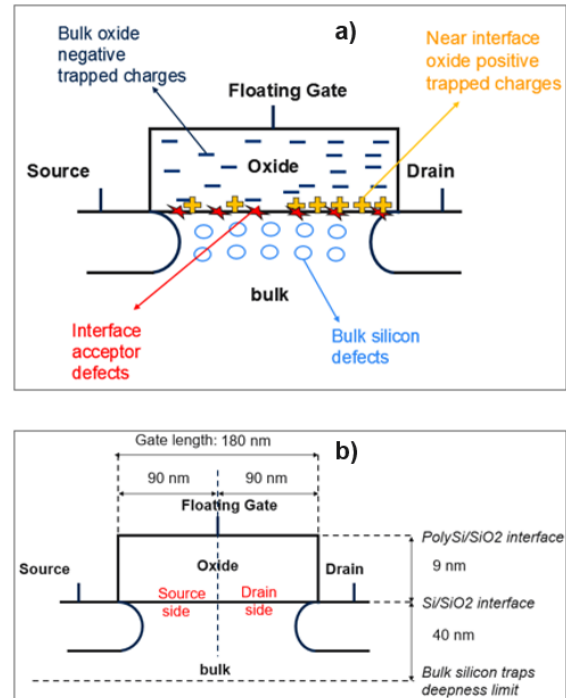


Figure 4: a) Post Cycling Defects Distribution. b) Traps Geometric specifications

In Fig.5, we show the post cycling programming window at 1 million cycles which is a high value for this technology (typically between 10⁵ and 10⁶). The threshold voltage shifts of both states are due to programming and erasing efficiency reduction and defects electrostatic influence on the read operation (also called static aging).

Defect concentrations were implemented to fit the post cycling programming window. First, the concentration of silicon channel defect is chosen to fit experimental subthreshold slope which is exclusively dependent on the channel carrier mobility. Negative trapped charges in the bulk oxide increase the threshold voltage by static aging and also decrease the Fowler-Nordheim injection efficiency [18]. Thus, the erase state after stress is simulated by adding that kind of defects following the paradigm that the concentration must be higher on drain side (due to hot carrier degradation). Finally, for the programmed state, we considered that the channel hot electron injection depends on the channel mobility (already considered in the first step) and Si/SiO₂ interface states at source side as suggested in [20]. Indeed, interface states at source side reduce the electron quantities available for the hot carrier injection. During write operation, interface states at drain side

are masked by the high voltage applied on drain contact. Even though acceptor interface states decrease the hot carrier injection, they also induce a strong static aging, shifting the threshold voltage to higher values. To balance this static aging and get a good agreement, we added holes trapped next to the Si/SiO₂ interface. These holes are typically generated by anode hole injection with a higher concentration at drain side due to hot carrier injection.

In Table 1, we have reported the parameter set that was used in all simulations of the degraded Flash.

defect type	source side	drain side
SiO ₂ bulk (cm ⁻³)	-2.5×10 ¹⁸	-5.0×10 ¹⁸
SiO ₂ near interface (cm ⁻²)	+1.0×10 ¹²	+5.8×10 ¹²
Si-SiO ₂ interface (cm ⁻² /eV)	1.8×10 ¹²	1.8×10 ¹²
Silicon channel (cm ⁻³)	2.0×10 ¹⁸	2.0×10 ¹⁸

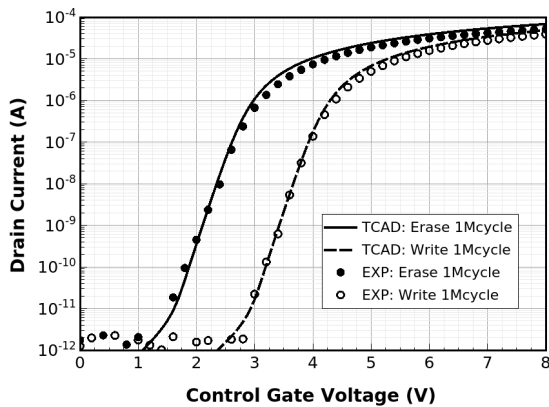


Figure 5: Flash programmed and erased characteristics, after 1 million cycles (degraded cell).

Thanks to the TCAD simulation, as shown in Fig. 6, it is possible to separate the two programming window closure causes, the static aging and the injection efficiency loss due to defects.

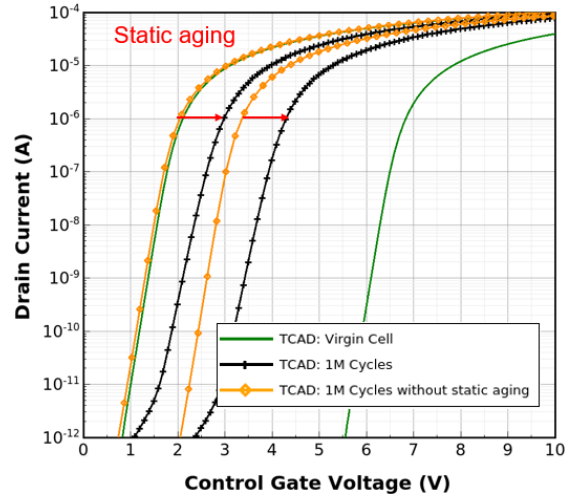


Figure 6: Flash programmed and erased characteristics after 1 million cycles, static aging and injection efficiency loss are decorrelated.

For the erased state, the injection efficiency loss is very low, so the V_t shift of 1V is due to static aging. This result is in agreement with a previous work in which negative trapped charges were implemented in the tunnel oxide of an EEPROM to simulate the programming window closure during an endurance test [21]. Indeed, for an equivalent charge density in the EEPROM tunnel oxide, the injection efficiency loss by Fowler-Nordheim is also negligible. For the programmed state, we also observe the static aging of 1V but the preponderant phenomenon is the hot carrier injection efficiency loss. This result is highlighted on programming kinetics, that is time-evolution of threshold voltage (extracted at constant drain current of 1 μA) during programming, shown in Fig.7.

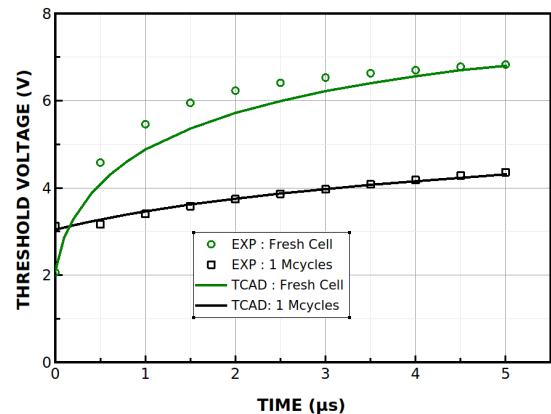


Figure 7: Programming kinetics of 1T-NOR Flash, before and after 1 million cycles (degraded cell).

On fresh cell, most of the charge is injected during the

first $\sim 2 \mu\text{s}$, whereas threshold voltage evolution is nearly linear with only a 1 V variation on degraded cell. Similar behavior is observed for drain current consumption in Fig.8. For the fresh cell, a substantial gate current injection leads very quickly to a highly negatively charged floating gate, which in turn induces a drain current reduction. For the cycled cell, small current injection (almost linear as in Fig.7) only slightly reduces the drain current.

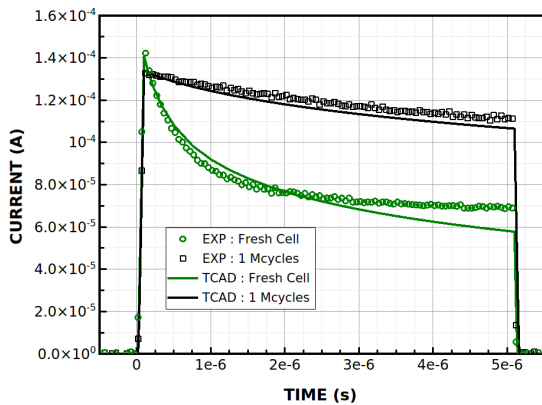


Figure 8: Drain current consumption for the fresh and degraded cell.

Finally, in Fig.9 we show programmed I-V cell characteristics for stressed and unstressed side after 1M cycles, as in [4]. The unstressed side I-V was obtained performing one programming pulse at source side ($V_s=4.2 \text{ V}$, $V_d=0 \text{ V}$, $V_g=10 \text{ V}$) after 1 million cycles.

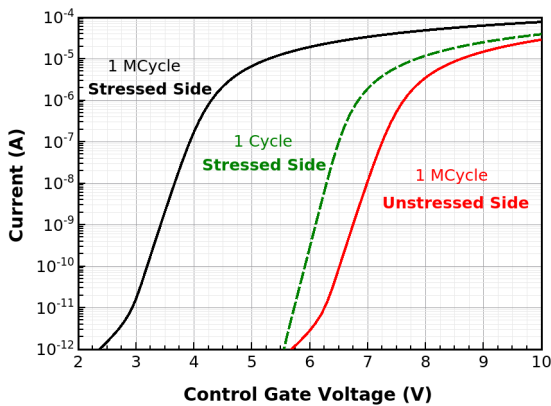


Figure 9: programmed state read of stressed side vs unstressed side.

The V_t shift is only due to the electrostatic defect influence and is in qualitative agreement with experiments reported in [4]. Regarding the drain current consumption, our simulation is also qualitatively in agreement with [4] (Fig.10).

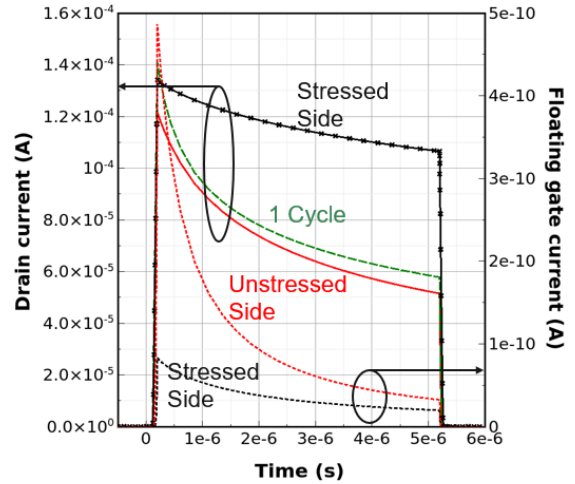


Figure 10: Drain current consumption (Solid line) and Floating gate current (Dots) stressed side vs unstressed side

Concerning the programming pulse at source side (unstressed side), the current consumption follows a similar slope than the 1 cycle one but shifted to a lower value. It means that hot carrier injection is still efficient. Indeed, during the programming pulse at source side, acceptor interface traps (negatively charged) on the low voltage side ($V_d=0\text{V}$) which normally reduce the hot carrier injection, are compensated by the high concentration of trapped hole next to the Si/SiO₂. The shift to lower current values is due to a higher erased threshold voltage when the cell has been cycled. However, while initial erased state is the same, stressed side has a higher drain current peak than the unstressed side. Due to the higher hot carrier injection efficiency at the unstressed side, the injection peak (floating gate current) is higher than the one at the stressed side.

4. Conclusion

We have performed TCAD process and electrical 2D simulations of 1T-NOR Flash device. Using available advanced models for carrier transport, channel hot injection and tunneling, we were able to successfully simulate the Flash programming window, its programming kinetics and consumption current. Flash degradation after 1 million cycles was taken into account by a non-uniform defects distribution located at Si-SiO₂ interface and inside SiO₂. This heuristic approach led to a good agreement with experiments, including I-V characteristics of degraded cell probed at the unstressed side.

References

- [1] NOR Flash Market to Grow at a CAGR of 14.4% to reach US\$ 6,069.5 million from 2021 to 2028, <https://www.theinsightpartners.com/reports/nor-flash-market>
- [2] Aritome S. et al., "Reliability issues of Flash memory cells, Proceedings of the IEEE," vol. 81, no. 5, pp. 776-788, 1993.
- [3] Yang H. et al., "Reliability issues and models of sub-90nm NAND Flash memory cells," IEEE International Conference on Solid-State and Integrated Circuit Technology, 2006.
- [4] Della Marca V. et al., "Impact of endurance degradation on the programming efficiency and the energy consumption of NOR flash memories," Micro. Reliab., vol. 54, pp. 2262-2265, 2014.
- [5] <https://www.synopsys.com/silicon/tcad.html>
- [6] Jun Young Lim et al. "Analysis of Intrinsic Charge Loss Mechanisms for Nanoscale NAND Flash Memory," IEEE Transactions on Device and Materials Reliability, vol. 15, no. 3, pp. 319-325, 2015
- [7] Pfäffli et al. "TCAD for Reliability," Microelectronics Reliability, vol. 52, pp. 1761-1768, 2012.
- [8] Postel-Pellerin J. et al., "Simulation of the programming efficiency and the energy consumption of Flash memories during endurance degradation," IEEE International Semiconductor Conference (CAS), pp. 101-104, 2016.
- [9] V. Della Marca et al. "Quantitative correlation between Flash and equivalent transistor for endurance electrical parameters extraction", Microelectronics Reliability, vol. 88-90, p.159-163, 2018.
- [10] Grasser T. et al., "A review of hydrodynamic and energy-transport models for semiconductor device simulation," Proceedings of the IEEE, vol. 91, no. 2, pp. 251-274, 2003.
- [11] Sentaurus™ Device User Guide Version Q-2019.12, Dec 2019.
- [12] Hasnat K. et al., "A Pseudo-Lucky Electron Model for Simulation of Electron Gate Current in Submicron NMOSFET's, IEEE Transactions on Electron Devices," vol. 43, no. 8, pp. 1264-1273, 1996.
- [13] Lezlinger M. and Snow E. H., "Fowler-Nordheim tunneling into thermally grown SiO₂," J. Appl. Phys., vol. 40, no. 1, pp. 278-283, 1969.
- [14] Jeong M. et al., "Comparison of Raised and Schottky Source/Drain MOSFETs Using a Novel Tunneling Contact Model, IEEE IEDM Technical Digest," pp. 733-736, 1998.
- [15] Schroder D. K. and Babcock J. A., "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," J. Appl. Phys, vol. 94, no. 1, pp. 1-18, 2003.
- [16] Regnier A. et al., "EEPROM Compact Model with SILC Simulation Capability," IEEE Annual Non-Volatile Memory Technology Symposium, 2006.
- [17] Papadas C. et al., "Model for Programming Window Degradation in FLOTOX EEPROM cells," IEEE Electron Device Letters, vol. 13, no. 2, pp. 89 - 91, 1992.
- [18] Ku P. S. and Schroder D. K., "Charges trapped throughout the oxide and their impact on the Fowler-Nordheim current in MOS devices, IEEE Transactions on Electron Devices," vol. 41, no. 9, pp. 1669 - 1672, 1994.
- [19] Doyle B., "The Hot-Carrier Effect," in ULSI Devices, edited by C. Y. Chang, S. M. Sze, pp. 275-281, 2000.
- [20] Torrente G. "Investigation of degradation mechanisms and related performance concerns in 40nm NOR Flash memories," Thesis, pp. 153-158, 2017
- [21] Matteo F. et al., "Simulation of state of the art EEPROM programming window closure during endurance degradation," IEEE Design & Technology of Integrated System, 2021