A Kinetic Monte Carlo Study of Retention Time in a POM Molecule-Based Flash Memory

Oves Badami, Toufik Sadi, Fikru Adamu-Lema, Paul Lapham, Dejiang Mu, Vihar Georgiev, Jie Ding, and Asen Asenov

***Abstract*—The modelling of conventional and novel memory devices has gained significant traction in recent years. This is primarily because the need to store an increasingly larger amount of data demands a better understanding of the working of the novel memory devices, to enable faster development of the future technology generations. Furthermore, in-memory computing is also of great interest from the computational perspectives, to overcome the data transfer bottleneck that is prevalent in the von-Neumann architecture. These important factors necessitate the development of comprehensive TCAD simulation tools that can be used for modeling carrier dynamics in the gate oxides of the flash memory cells. In this work, we introduce the kinetic Monte Carlo module that we have developed and integrated within the Nano Electronic Simulation Software (NESS) – to model electronic charge transport in Flash memory type structures. Using the developed module, we perform retention time analysis for a polyoxometalate (POM) molecule-based charge trap flash memory. Our simulation study highlights that retention characteristics for the POM molecules have a unique feature that depends on the properties of the tunneling oxide.**

***Index Terms*—Tunneling, flash memory, polyoxometalate, kinetic Monte Carlo.**

I. Introduction

With the continuing microelectronics progress, the amount of data that is being stored and processed is increasing dramatically. To keep up with the increase in demand for data storage, flash memories have consequently been scaled dramatically and stacked by the semiconductor industry [1]. Moreover, different architectures for flash memories, such as charge trap and nanocrystal-based flash memories, are being investigated. In particular, charge trap based flash memories have gained considerable importance, as they overcome the issue of intercell coupling due to the redistribution of charge observed in the standard floating gate based flash memory [2]. However, conventional charge trap based flash memories suffer from large variability that is inherent due to the amorphous nature of charge-trapping oxide and the randomness of the trap distribution [3]. Self-assembled nano-crystal based flash memories can overcome the variability associated with the amorphous nature of the oxide [4]. However, while fabricating the nanocrystals, it is difficult to control their sizes which will affect adversely the performance of the corresponding memory cells [5], [6], and will increase variability. Hence, a molecule-based charge trap flash memory provides an alternative architecture that overcomes the coupling of problems of the floating gate flash cells, and the variability of the conventional charge trap based flash memory and nanocrystal based flash memory, by localizing the charge carriers in an array of self-assembled molecules. Molecule based memory provide an additional handle to tune the memory characteristics (by tuning the molecular energy levels) as compared to the conventional charge trap based memories wherein the charge levels are randomly distributed in energy [7]. Consequently, molecule-based flash memory have generated a lot of attention recently as possible solution to improve the flash cells variability and reliability issues.

Redox active molecules have an important ability to reliably accept (undergo reduction) and give-up (undergo oxidation) electrons thereby making them a prime candidate for being employed in flash memory cells for storing electrons [8]–[10]. Continuous research in the molecule-based flash memories has allowed the development and fabrication of nanowire based flash memory cells based on gate-all-around nanowire transistors [11]. Redox based memory cells have demonstrated a 10 year retention time which is necessary for commercialization [12]. Researchers have also demonstrated a fast program/erase [12] and stable memory operation over 10^9 cycles [11]. Despite the progress in molecular flash memory technology, the development of a comprehensive physics-based simulator (and methodology) for optimization and path-finding remains unaddressed.

Most of the studies related to the molecule-based flash memories have used organic molecules. However, we have employed We have employed POMs, a type of inorganic molecule: because: 1) they have better compatibility with the CMOS fabrication process as compared to their organic counterparts, as they consist of oxygen atoms similar to SiO_2 [13]; 2) they do not breakdown during the annealing process and hence do

Manuscript received June 3, 2020; accepted August 3, 2020. Date of publication August 20, 2020; date of current version September 28, 2020. This work was supported in part by the EPSRC (UK) under Grants EP/S000224/1 and EP/S001131/1 and in part by NSFC project (Project No.61604105). The review of this article was arranged by Associate Editor Prof. Ashok Srivastava. *(Corresponding author: Jie Ding.)*

Oves Badami is with the Indian Institute of Technology Hyderabad, Sangareddy 502285, India (e-mail: oves.badami@ee.iith.ac.in).

Toufik Sadi is with the Engineered Nanosystems Group, Aalto University, 00076, Finland (e-mail: toufik.sadi@aalto.fi).

Fikru Adamu-Lema, Paul Lapham, Vihar Georgiev, and Asen Asenov are with the School of Engineering, University of Glasgow, Glasgow G12 8QQ, U.K. (e-mail: fikru.adamu-lema@glasgow.ac.uk; p.lapham.1@research.gla.ac.uk; vihar.georgiev@glasgow.ac.uk; Asen.Asenov@glasgow.ac.uk).

Dejiang Mu and Jie Ding are with the Taiyuan University of Technology, Taiyuan, China (e-mail: mdj201809@163.com; dingjie2015@foxmail.com).

Digital Object Identifier 10.1109/TNANO.2020.3016182

not reduce the thermal budget needed for the conventional CMOS fabrication [14]; 3) they can capture multiple electrons and undergo reversible transitions of capturing/emitting electrons [14], providing the possibility of having multi-bit flash memory [15], [16]. The possibility of molecular self-assembly of POM molecules can also reduce significantly the statistical variability. Our first principle calculations have shown that the Lowest Unoccupied Molecular Orbital (LUMO) energy level lies in the bandgap of the silicon thus enabling the POMs to be used as traps in POM based charge trap flash memory [17].

This paper aims to introduce the kinetic Monte Carlo, kMC, module implemented in the Nano-Electronic Simulation Software (NESS), which can be used to simulate different aspects (dynamics of charging and discharging) of flash memory operations. As a practical example, we have used a planar POM molecule-based flash memory, which has been experimentally demonstrated to have a memory characteristic, as a test case [14]. Previous simulation studies have concentrated on the variability (due to random dopant fluctuations and POM distribution) of the threshold voltage and the comparison between bulk and Silicon-on-Insulator architectures [17], [18]. However, a comprehensive study of retention, program and erase time were not reported as it wasn't then possible to perform a time-dependent study. In this simulation study, POM molecules were treated as traps, which when occupied are represented as point charges. The LUMO level of each POM molecule is considered as a trap energy level. In this work, we neglect the change of the energy of the LUMO level due to the occupation by electrons.

The paper is organized as follows. In Section II, we discuss the overall simulation framework of NESS, the kinetic Monte Carlo methodology, and the models that are implemented. In Section III, we present the benchmarking of the implementation in respect of conventional flash memory cell operation as well as the retention time analysis of a POM based flash memory. Finally, in Section IV, we conclude and discuss the future outlook.

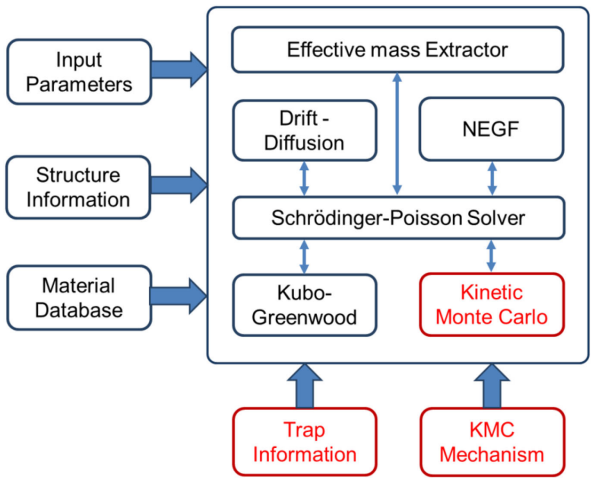

Fig. 1. Illustration of capabilities of NESS simulation engine. The module in discussion for this paper is highlighted in red.

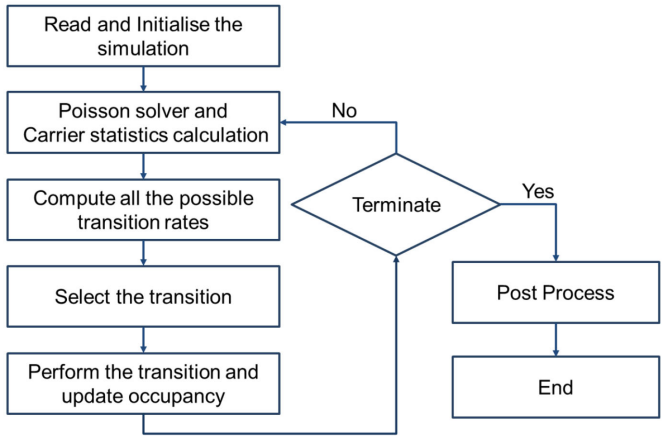

Fig. 2. Flowchart for the kMC methodology as implemented in NESS.

II. Simulation Framework

A. Nano-Electronic Simulation Software - NESS

The Nano-Electronic Simulation Software (NESS) is an in-house simulation engine having a wide range of capabilities as illustrated in Fig. 1. Electron transport can be modelled using a quantum transport model (based on a non-equilibrium Green's function solver capable of accounting for ballistic, phonon and surface roughness scattering) and a classical transport model based on a drift-diffusion solver with different mobility models such as Yamaguchi, Caughey-Thomas and Masetti) module [19]. A Kubo-Greenwood solver allows the accurate calculation of the low field mobility due to the most relevant scattering mechanism including acoustic and optical phonon, and surface roughness scattering [20], [21]. This calculated mobility can be employed in the drift-diffusion solver for a more accurate performance prediction of the MOSFETs. An effective mass extractor module allows the calculation of the parabolic effective masses from the electronic bandstructure obtained using more sophisticated electronic band structure calculation methods, such as the semi-empirical Tight Binding formulation or Density Functional Theory [22]. In addition to these solvers, NESS has a native structure generator that can generate arbitrary semiconductor devices including bulk MOSFETs, nanowires, and stacked nanowires architectures with different cross-section shapes. The structure generator can account for the most relevant sources of statistical variability, including surface roughness, random discrete dopants distribution, interface charges distribution, and metal grain granularity [23].

B. The Kinetic Monte Carlo Methodology

The overall flow of the kMC solver is shown in Fig. 2. After reading the structural details, the simulation domain is initialized for details such as the number of traps, trap type (neutral or positive when empty), and the initial state of the trap (filled or empty) from the text-based user input file. Following the initialization, the three-dimensional (3D) Poisson's equation is solved in the entire simulation domain. Equilibrium carrier statistics are assumed in the semiconductor region and charge dynamics are self-consistently coupled with the solution of Poisson's equation. To calculate the possible transitions, the initial state of the electron (kinetic energy) is randomly generated according to the Fermi-Dirac distribution.

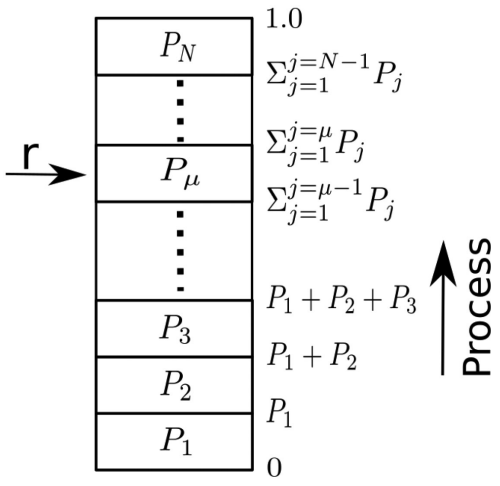

Fig. 3. Illustration of the cumulative probability ladder for different processes. The cumulative probability is also shown.

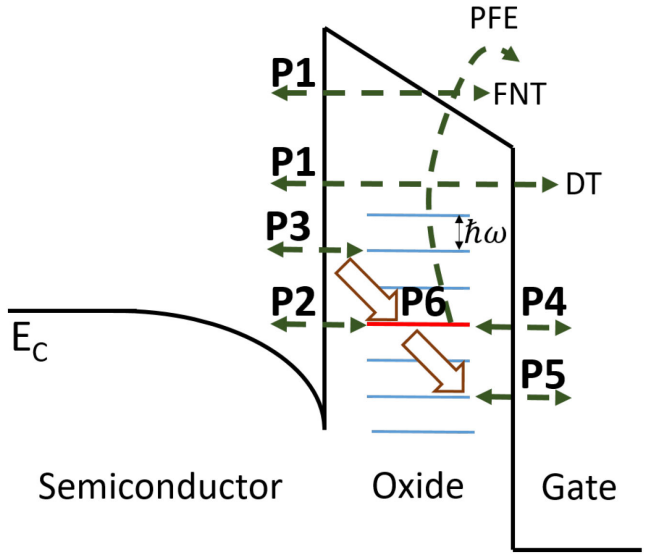

Fig. 4. Illustration of different mechanisms that are implemented in the Kinetic Monte Carlo Module; P1: Tsu-Esaki model, P2 and P4: Elastic capture and emission of electrons, P3 and P5: Inelastic capture and emission of electrons, P6: Poole-Frenkel emission (PFE). The electrons can be captured from and emitted to the electrode (either substrate or gate).

To proceed forward, we compute all the possible transition rates from the electrodes (semi-conductor and gate) to the trap (in our case the POM molecule), for trap-to-trap, from an electrode to another electrode and Poole-Frenkel emission. Using these transition rates, a particular transition, μ, is selected randomly as follows [24]:

$$\frac{\Sigma_{\mu'=1}^{\mu-1} R_{\mu'}}{R_{TOT}} < r \leq \frac{\Sigma_{\mu'=1}^{\mu} R_{\mu'}}{R_{TOT}}, \tag{1}$$

where R_{TOT} is the total transition rate of all the processes and r is the random number having a uniform distribution between 0 and 1. This is illustrated represented in Fig. 3. To make the implementation computationally efficient, we consider tunneling only from the point closest to the trap in the electrodes. This is justified, as the tunneling probability decreases exponentially with the increasing tunneling distance. The time step associated with the transition can be calculated using [24]:

$$\tau = -\frac{ln\,(r')}{R_{TOT}}, \tag{2}$$

where r' is a random number uniformly distributed between 0 and 1. The total time elapsed since the start of the simulation is given by $t = t + \tau$, and by keeping a track of the number of electrons that have crossed the oxide-semiconductor interface, N_{TOT}, the total current can be calculated using the universal formula:

$$I = -\frac{eN_{TOT}}{t}, \tag{3}$$

where t is the total time and e is the elementary electronic charge.

C. Electron Transport Models

In the kMC module, we have implemented a comprehensive set of physics-based models for simulating electron transport in the oxide. These are schematically illustrated in Fig. 4. The developed module also provides flexibility to switch on/off different models. Here we will briefly discuss the implemented transition models:

1) Tsu-Esaki Model: The process of tunneling directly from semiconductor to metal (and vice versa) is accounted for by using the Tsu-Esaki tunneling model. This allows describing the Direct tunneling (DT) and Fowler-Nordheim tunneling (FNT). The tunneling rate can be calculated as [25], [26]

$$R_{TE} = \frac{Am_{si}k_BT}{2\pi^2\hbar^2}\int_{E_c}^{\infty} T_x(E_z) ln\left(\frac{1+\exp\left(\frac{E_{fs}-E_z}{k_BT}\right)}{1+\exp\left(\frac{E_{fg}-E_z}{k_BT}\right)}\right) dE_z, \tag{4}$$

where A is the cross-section area of the device, m_{si} is the effective mass in the semiconductor, k_B is the Boltzmann constant, T is the temperature, E_c is the conduction band edge, $E_{fs(g)}$ is the Fermi-Level in the semiconductor (gate) and $T_x(E_z)$ is the transmission coefficient which is calculated using the Wentzel-Kramers-Brillouin (WKB) approximation. In the bandgap region, we have assumed a parabolic electronic dispersion relationship.

2) Inelastic Tunneling: The injection of an electron from the terminal (substrate or gate) into the trap (and vice versa) can occur inelastically, i.e. while tunneling the electron may lose or gain energy through the exchange of phonons. The transition rate for electron tunnelling from an electrode to a trap, R_{ED}^{IE}, and from a trap to an electrode, R_{DE}^{IE}, when phonons with energy $\hbar\omega$ (ω is the frequency of the phonon and $\hbar$ is the reduced Planck's constant), are absorbed/emitted to complete the transition are given by [27]

$$R_{ED}^{IE} = \frac{(4\pi)^2 r_D^3}{\hbar E_g}\frac{e^2\hbar^2F^2}{2m_i} N_{DOS}(E) f_{FD}(E) T_x(E)$$
$$\times L_p(2S\,(f_{BE}(f_{BE}+1))^{0.5}) e^{-\frac{p\hbar\omega}{k_BT}},$$

$$R_{DE}^{IE} = \frac{(4\pi)^2 r_D^3}{\hbar E_g} \frac{e^2\hbar^2F^2}{2m_i} N_{DOS}(E)(1-f_{FD}(E))T_x(E) \times L_p(2S\left(f_{BE}(f_{BE}+1)\right)^{0.5}) e^{-\frac{p\hbar\omega}{k_BT}}, \quad (5)$$

where p is the number of phonons involved in the transition, E is the energy of the electron in the electrode (generated randomly according to the Fermi-Dirac distribution), $N_{DOS}(E)$ is the density of states, $f_{FD}(E)$ is the Fermi-Dirac distribution function, f_{BE} is the value of the Bose-Einstein distribution function for phonon energy ($\hbar\omega$) F is the electric field, E_g is the oxide bandgap, m_i is the effective mass in the insulator, S is the Huang-Rhys factor, r_D is the localization radius of the electron in a trap, with trap-depth E_D. $L_p(z)$ is the multi-phonon transition probability [27] and it is given by:

$$L_p(z) = \left(\frac{f_{BE}+1}{f_{BE}}\right)^{0.5p} \exp\left(-S(2f_{BE}+1)\right) I_p(z), \quad (6)$$

where $I_p(z)$ is the modified Bessel function. For the conditions when either electrode to trap or trap to electrode need to emit phonons for the transition to occur, the exponential term in eq. 5 is set to unity [28].

3) Elastic Tunneling: The elastic tunneling model is used to model the tunneling phenomenon when the electron that is generated in the terminal aligns with the trap energy level. This necessity of alignment makes this process a special case. The elastic electrode-to-trap (R_{ED}^E) and trap-to-electrode (R_{DE}^E) tunneling rate can be modelled by [28]

$$R_{ED}^E = \left(\frac{m_e}{m_i}\right)^{2.5} \left(\frac{8E^{1.5}}{3\hbar\sqrt{E_D}}\right) f_{FD}(E)T_x(E)$$
$$R_{DE}^E = \left(\frac{m_e}{m_i}\right)^{2.5} \left(\frac{8E^{1.5}}{3\hbar\sqrt{E_D}}\right) (1-f_{FD}(E))T_x(E) \quad (7)$$

where m_e is the electron mass in the terminal (substrate/gate), E is the energy of the electron in the electrode (generated randomly according to the Fermi-Dirac distribution).

4) Trap-to-Trap Tunneling: Electron transport from one trap to another is modeled through electron hopping. The tunneling rate from a trap located at $\mathbf{r_1}$ to a trap at $\mathbf{r_2}$ and separated by energy ΔE is calculated as [29]

$$R_{DD} = f_0 \exp\left(\frac{-2r_{12}}{r_D}\right) f_1(1-f_2) \begin{cases} 1, & \Delta \mathrm{E} \le 0 \\ \exp(-\frac{\Delta E}{k_BT}) & \Delta E > 0, \end{cases} \quad (8)$$

where r_{12} is the magnitude of the distance between the two traps and r_D is localization length – depending on the trap depth E_D (trap energy with respect to the conduction band edge of the oxide) and oxide effective mass m_i – given by $\hbar/\sqrt{2m_iE_D}$. f_0 is the vibration frequency of the atoms ($\approx 10^{13}$ sec^{-1}), f_1 and f_2 are the occupancies (between 0 and 1) of the traps located at r_1 and r_2, respectively. The energy difference between the traps is discretized in the units of the phonon energy ($\hbar\omega$).

5) Poole-Frenkel Emission: In addition to the already discussed mechanisms, an electron that is trapped can also transit to the electrode through the Poole-Frenkel emission process.

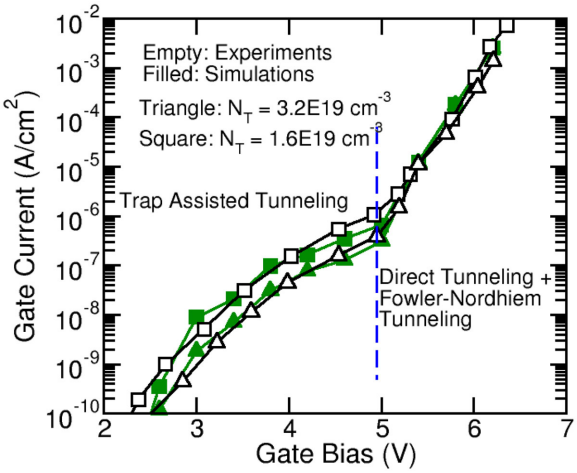

Fig. 5. Comparison between the simulations and experimental data for gate current for different trap density. All the other parameters, other than the trap density, were kept the same. Experimental data are from Ref. [30].

The transition rate, R_{PF}, can be calculated using [28]:

$$R_{PF} = f_0 \exp\left(\frac{-E_D}{k_BT}\right)\left(\frac{1}{\beta^2}(1+(\beta-1)\exp(\beta))+0.5\right), \quad (9)$$

where $\beta = \sqrt{e^3F/(\pi\epsilon_0\epsilon_{opt}k_B^2\,T^2)}$, F is the electric field, and ϵ_{opt} is the optical permittivity. We assume that as soon as the electron is emitted from the trap it is collected by the gate electrode.

III. Results

A. Comparison With Experiments

To benchmark the models and their implementation, Fig. 5 compares the experimental data from Ref. [30] with the simulations. For the simulations, we have used a MOS capacitor with an oxide thickness of 5 nm as specified in the experimental data. We also assumed that the traps are uniformly distributed in the gate oxide. The trap energy is assumed to be in the range of 2.3 eV - 3.1 eV which again matches well with the values reported in Ref. [31]. The relaxation energy ($E_{REL} = S\hbar\omega$) extracted from the calibration process was assumed to be 0.6 eV. The trap density used for the simulations was 3.2×10^{19} cm^{-3} and 1.6×10^{19} cm^{-3}. The disagreement between the extracted trap density values in this work and those reported in Ref. [30] is attributed to the different simulation models employed in both studies. However, the trap densities reported in this work agree well with those extracted from the data by other research groups employing a similar simulation model [32]. The results show a good agreement between the simulations and experimental data, thus validating our implementation.

B. POM Flash Memory Cell Simulation

Fig. 6(a) shows the schematic diagram of a flash memory cell with the POM molecule layer which stores electrons. It also highlights the important parts (tunnelling oxide, POM layer, and control oxide). The inset shows the schematic diagram

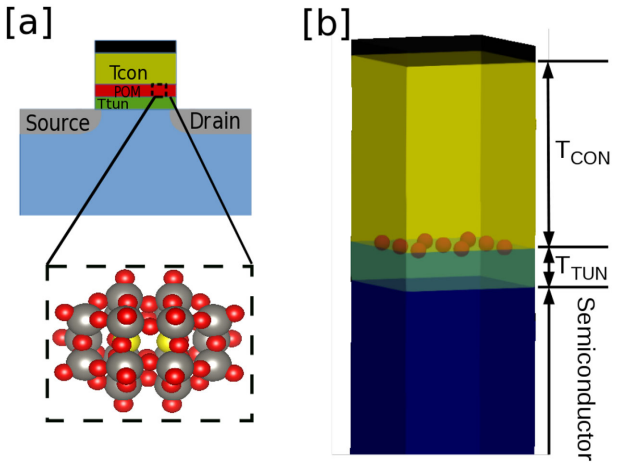

Fig. 6. (a) Schematic of a molecule (in this case POM) based charge trap flash memory cell. (b) A POM ($W_{18}O_{54}(SO_3)_2$) molecule (without counter-cations) is also illustrated in the inset (yellow sphere - S, grey sphere - W and red sphere - O) (b) MOS capacitor used in the simulations. In this work, we have considered a 3 × 3 arrangement of the POM molecules.

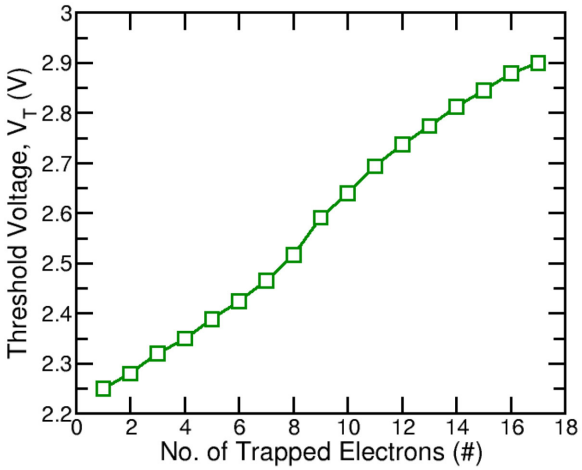

Fig. 7. Threshold voltage as a function of the number of trapped charges. As expected the threshold voltage increases linearly with the trapped electrons.

of the POM molecule used in this work. Fig. 6(b) shows the simulation domain (a MOS capacitor) used in this work. The distance between any two neighboring POM molecules was set to be 3 nm. The LUMO level for the POM was set to be 3.67 eV below the conduction band edge of the SiO_2 as extracted from the density functional theory simulations [17].

Fig. 7 shows the impact of the number of electrons trapped in the POM layer on the threshold voltage. In this work, we define the threshold voltage as a gate bias at which the inversion charge density (in the center of the device) becomes 10% of the p-doping density (10^{18} cm^{-3}).

Fig. 8 shows the threshold voltage as a function of the time for the retention time analysis. In the simulations, we have assumed that the POM molecules are doubly charged (2 × reduced) at time $t = 0$ sec. We have set the Huang-Rhys parameters, S, to 6 and the phonon energy, $\hbar\omega$ to 0.06 eV, which are consistent with the values extracted by calibrating the tunneling current to the experimental data [31]. Subsequently, the electrons from the

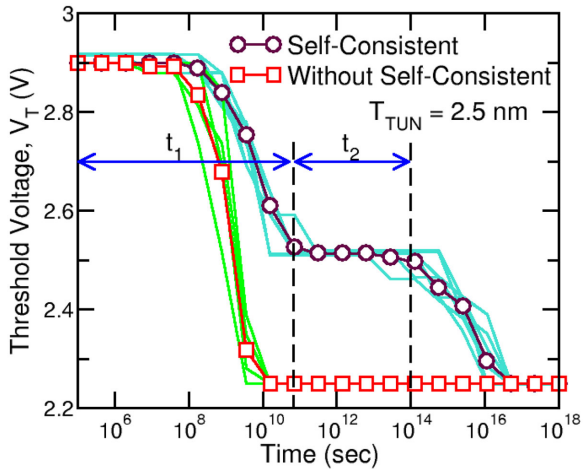

Fig. 8. Evolution of the threshold voltage as a function of time calculated using the self-consistent (Poisson's equation and carrier statistics) and non-self consistent solution.

POM molecules tunnel through the oxide layer leading to a reduction in the threshold voltage. Fig. 8 compares the threshold voltage extracted from the simulations, when the Poisson's equation and equilibrium carrier statistics are solved self-consistently, and when the self-consistency is not considered. Since the electron dynamics are modeled using a statistical method, the inherent variability (solid lines) and the mean (line+symbol) of the threshold voltage is shown in Fig. 8. It can be seen that when the self-consistency is considered there is a plateau-like feature that is not present otherwise. This feature occurs because at the start each POM stores two electrons (2 × reduces), and all the POM molecules are significantly more likely to emit one electron each and settle in 1 × reduced state (then there is a reduction in the tunneling rate after each POM has emitted an electron) and then subsequently emit the second electron. The electric field near the POM molecule which has 2 trapped electrons(2 × reduced) is more than the case when there is one trapped electron (1× reduced). Hence, the inelastic defect to electrode tunneling rate is more for a 2× reduced POM than the 1× reduced. Consequently, all POMs initially tend to lose one electron and settle into 1× reduced state. This reduction in the transition rates is not observed in the case where self-consistency is not considered, as the electrostatic potential is not modified to reflect the changes in the number of electrons trapped in the POM molecules.

Since the tunneling oxide is amorphous, a wide range of values of important oxide parameters have been extracted from the experimental data [31]–[33]. Hence, in Fig. 9, we compare the impact of the relevant oxide parameters ($E_{REL} = S\hbar\omega$) on the retention characteristics of the device. The tunneling oxide thickness was set to 1.5 nm. The results indicate that the unique plateau-like feature that is observed for $S\hbar\omega = 0.36$ eV vanishes for $S\hbar\omega = 1.8$ eV. This highlights the multi-bit storage capability of the POM molecules based flash memory which depends strongly on the oxide properties or the fabrication process. To understand this behavior, we plot the multi-phonon transition probability in Fig. 10. Due to the interaction between

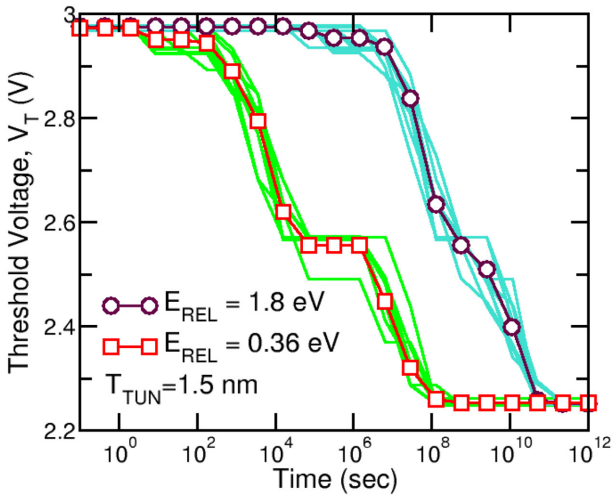

Fig. 9. Impact of the oxide parameters (relaxation energy, E_{REL}) on the threshold voltage evolution. The mean of the simulations results (solid lines) is indicated by line+symbol.

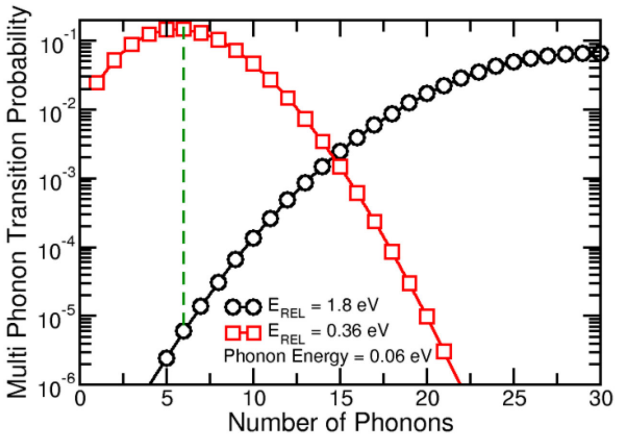

Fig. 10. Multi-phonon transition probability, $L_p(z)$, as a function of the number of phonons for the systems in 9.

the electrostatics (accounting for the electron trapped in the POM) and the position of the LUMO level of the POM molecule it turns out that the number of phonons needed to mediate the transition is about 7 (phonon energy was set to 0.06 eV in both the cases). As can be seen from Fig. 10, the multi-phonon transition probability, $L_p(z)$, for the $E_{REL} = 1.8$ eV is significantly smaller than that obtained at $E_{REL} = 0.36$ eV around the number of phonons needed to mediate the transition. This reduction in the multiphonon transition probability reduces the electron tunneling rate and hence smears out of plateau characteristic that is observed in the case for $E_{REL} = 0.36$ eV.

IV. Conclusion

In this paper, we have implemented a new versatile kinetic Monte Carlo module in NESS to simulate the electron transport through the amorphous oxide based on a comprehensive set of physics-based models. We have validated the implementation by computing the gate leakage current and comparing it with the experimental data reported in the literature. We have further performed a retention time analysis of POM molecule-based flash memory cells. Our simulation results suggest that for the retention time a unique plateau is observed which suggests a possible multi-level memory implementation. Further in-depth analysis of this plateau highlighted that it depends on the relaxation energy which is a material parameter for the oxide. The developed module would also allow us to perform a thorough study of different aspects of the flash memory operation including programming and erase time analysis in the presence of different sources of variability which will be a part of our future works.

References

[1] K. Park *et al.*, "Three-dimensional 128 Gb MLC vertical NAND flash memory with 24-WL stacked layers and 50 MB/s high-speed programming," *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 204–213, Jan. 2015.

[2] J. Lee, S. Hur, and J.-H. 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.

[3] P. Kumar, P. Nair, R. Sharma, S. Kamohara, and S. Mahapatra, "Lateral profiling of trapped charge in SONOS flash EEPROMs programmed using CHE injection," *IEEE Trans. Electron Devices*, vol. 53, no. 4, pp. 698–705, Apr. 2006.

[4] S. Tiwari, F. Rana, K. Chan, H. Hanafi, W. Chan, and D. Buchanan, "Volatile and non-volatile memories in silicon with nano-crystal storage," in *Proc. Int. Electron. Devices Meet.*, Dec. 1995, pp. 521–524.

[5] J. Shaw, T. Hou, H. Raza, and E. C. Kan, "Statistical metrology of metal nanocrystal memories with 3-D finite-element analysis," *IEEE Trans. Electron Devices*, vol. 56, no. 8, pp. 1729–1735, Aug. 2009.

[6] L. Perniola *et al.*, "Modeling of the programming Window distribution in multinanocrystals memories," *IEEE Trans. Nanotechnol.*, vol. 2, no. 4, pp. 277–284, Dec. 2003.

[7] S. Pookpanratana *et al.*, "Non-volatile memory devices with redox-active diruthenium molecular compound," *J. Phys.: Condensed Matter*, vol. 28, no. 9, Feb. 2016, Art. no. 094009.

[8] X. Chen, Y. Zhou, V. A. L. Roy, and S.-T. Han, "Evolutionary metal oxide clusters for novel applications: Toward high-density data storage in nonvolatile memories," *Adv. Mater.*, vol. 30, no. 3, 2018, Art. no. 1703950. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.201703950

[9] C. Li *et al.*, "Multilevel memory based on molecular devices," *Appl. Phys. Lett.*, vol. 84, no. 11, pp. 1949–1951, 2004. [Online]. Available: https://doi.org/10.1063/1.1667615

[10] H. Zhu and Q. Li, "Novel molecular non-volatile memory: Application of redox-active molecules," *Appl. Sci.*, vol. 6, no. 1, 2016. [Online]. Available: https://www.mdpi.com/2076-3417/6/1/7

[11] H. Zhu *et al.*, "Redox-active molecular nanowire flash memory for high-endurance and high-density nonvolatile memory applications," *ACS Appl. Mater. Interfaces*, vol. 7, no. 49, pp. 27 306–27 313, 2015, pMID: 26600234. [Online]. Available: https://doi.org/10.1021/acsami.5b08517

[12] K. Jiang *et al.*, "Nonvolatile memory based on redox-active ruthenium molecular monolayers," *Appl. Phys. Lett.*, vol. 115, no. 16, 2019, Art. no. 162102. [Online]. Available: https://doi.org/10.1063/1.5108675

[13] V. P. Georgiev, S. M. Amoroso, L. Vilá-Nadal, C. Busche, L. Cronin, and A. Asenov, "FDSOI molecular flash cell with reduced variability for low power flash applications," in *Proc. 44th Eur. Solid State Device Res. Conf.*, Sep. 2014, pp. 353–356.

[14] C. Busche *et al.*, "Design and fabrication of memory devices based on nanoscale polyoxometalate clusters," *Nature*, vol. 515, pp. 545–549, 2014. [Online]. Available: https://doi.org/10.1038/nature13951

[15] H. Zhu *et al.*, "Non-volatile memory with self-assembled ferrocene charge trapping layer," *Appl. Phys. Lett.*, vol. 103, no. 5, 2013, Art. no. 053102, 2013. [Online]. Available: https://doi.org/10.1063/1.4817009

[16] X. Chen *et al.*, "Polyoxometalates-modulated reduced graphene oxide flash memory with ambipolar trapping as bidirectional artificial synapse," *Adv. Electron. Mater.*, vol. 4, no. 12, 2018, Art. no. 1800444. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/aelm.201800444

[17] V. P. Georgiev, S. Markov, L. Vilá-Nadal, C. Busche, L. Cronin, and A. Asenov, "Optimization and evaluation of variability in the programming window of a flash cell with molecular metal–oxide storage," *IEEE Trans. Electron Devices*, vol. 61, no. 6, pp. 2019–2026, Jun. 2014.

[18] V. P. Georgiev *et al.*, "Comparison between bulk and FDSOI POM flash cell: A multiscale simulation study," *IEEE Trans. Electron Devices*, vol. 62, no. 2, pp. 680–684, Feb. 2015.

[19] S. Berrada *et al.*, "Ness: New flexible nano-electronic simulation software," in *Proc. Int. Conf. Simul. Semicond. Processes Devices*, Sep. 2018, pp. 22–25.

[20] C. Medina-Bailon *et al.*, "Mobility of circular and elliptical Si nanowire transistors using a multi-subband 1D formalism," *IEEE Electron Device Lett.*, vol. 40, no. 10, pp. 1571–1574, Oct. 2019.

[21] T. Sadi *et al.*, "Simulation of the impact of ionized impurity scattering on the total mobility in si nanowire transistors," *Materials*, vol. 12, no. 1, 2019. [Online]. Available: https://www.mdpi.com/1996-1944/12/1/124

[22] O. Badami *et al.*, "Comprehensive study of cross-section dependent effective masses for silicon based gate-all-around transistors," *Appl. Sci.*, vol. 9, no. 9, 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/9/1895

[23] J. Lee *et al.*, "Variability predictions for the next technology generations of n-type SixGe1-x Nanowire MOSFETs," *Micromachines*, vol. 9, no. 12, 2018. [Online]. Available: https://www.mdpi.com/2072-666X/9/12/643

[24] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Rev. Mod. Phys.*, vol. 55, pp. 645–705, Jul. 1983. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.55.645

[25] G. Jegert, A. Kersch, W. Weinreich, U. Schröder, and P. Lugli, "Modeling of leakage currents in high- dielectrics: Three-dimensional approach via kinetic Monte Carlo," *Appl. Phys. Lett.*, vol. 96, no. 6, 2010, Art. no. 062113. [Online]. Available: https://doi.org/10.1063/1.3310065

[26] R. Tsu and L. Esaki, "Tunneling in a finite superlattice," *Appl. Phys. Lett.*, vol. 22, no. 11, pp. 562–564, 1973. [Online]. Available: https://doi.org/10.1063/1.1654509

[27] M. Herrmann and A. Schenk, "Field and high-temperature dependence of the long term charge loss in erasable programmable read only memories: Measurements and modeling," *J. Appl. Phys.*, vol. 77, no. 9, pp. 4522–4540, 1995. [Online]. Available: https://doi.org/10.1063/1.359414

[28] G. C. Jegert, "Modeling of leakage currents in high-κ dielectrics," Ph.D. dissertation, Technische Universität München, Munich, Germany, 2012.

[29] T. Sadi, A. Mehonic, L. Montesi, M. Buckwell, A. Kenyon, and A. Asenov, "Investigation of resistance switching in SiOxRRAM cells using a 3D multi-scale kinetic Monte Carlo simulator," *J. Phys.: Condens. Matter*, vol. 30, no. 8, Feb. 2018, Art. no. 084005. [Online]. Available: https://doi.org/10.1088percent2F1361-648xpercent2Faaa7c1

[30] B. Ricco, G. Gozzi, and M. Lanzoni, "Modeling and simulation of stress-induced leakage current in ultrathin SiO films," *IEEE Trans. Electron Devices*, vol. 45, no. 7, pp. 1554–1560, Jul. 1998.

[31] L. Vandelli, A. Padovani, L. Larcher, R. G. Southwick, W. B. Knowlton, and G. Bersuker, "A physical model of the temperature dependence of the current throughSiO_2/HfO_2stacks," *IEEE Trans. Electron Devices*, vol. 58, no. 9, pp. 2878–2887, Sep. 2011.

[32] F. Jiménez-Molinos, A. Palma, F. Gamiz, J. Banqueri, and J. A. López-Villanueva, "Physical model for trap-assisted inelastic tunneling in metal-oxide-semiconductor structures," *J. Appl. Phys.*, vol. 90, no. 7, pp. 3396–3404, 2001. [Online]. Available: https://doi.org/10.1063/1.1398603

[33] A. Gehring, F. Jiménez-Molinos, H. Kosina, A. Palma, F. Gámiz, and S. Selberherr, "Modeling of retention time degradation due to inelastic trap-assisted tunneling in eeprom devices," *Microelectron. Rel.*, vol. 43, no. 9, pp. 1495–1500, 2003, 14th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0026271403002658