Comprehensive Simulation of Program, Erase and Retention in Charge Trapping Flash Memories

Abhijeet Paul, Ch. Sridhar, Suny Gedam and S. Mahapatra*
Department of Electrical Engineering, Indian Institute of Technology Bombay - 400076, India. *Email: souvik@ee.iitb.ac.in

Abstract

A simulator is developed for SONOS Flash memories to predict Program (P), Erase (E) and Retention (R) behavior under uniform 1D operation. It provides insight on the impact of trap parameters on P, E and R and can be used to optimize memory stacks.

Introduction and features

SONOS and similar charge trapping based devices [1-6] are attracting much attention as Flash memory candidates as they offer better scalability compared to conventional FG devices. Several reports exist on fabrication and experimental characteristics of these devices [5, 6]. Several attempts were made to develop physical models for electrical characteristics [1, 7-10]. However, no TCAD tool exists till date to simulate the operation of these memories. In this work, we report the development of a device simulator for predicting P, E and R of SONOS cells under 1D operation (Fig.1).

The simulator provides (a) time evolution of electric field in the ONO stack, (b) time evolution and spatial distribution of trapped carriers inside nitride, and (c) overall threshold voltage shift (∆Vt) and time evolution of charge centroid during P-E-R operation. It also provides useful insight into the underlying physical processes and good matching with experimental results, which can be used to obtain values for different physical parameters (effective mass, barrier height, band gap) and trap parameters (distribution profile, capture cross section, energy depth). It can also predict the impact of trap parameters on P-E-R operation and can be used to design ONO stacks for optimizing the performance of SONOS and similar high-K charge trapping based memories.

Simulation flow and physical models

The simulator works as a post-processor to DESSIS [11] (Fig.2). Operating conditions, physical and trap parameters are fed through an input deck and user defined parameter file. At the beginning of P, E or R, the potential across Poly-Si-ONO-substrate is obtained from DESSIS1 simulation (V(G)>0 and V(G)<0 for P and E respectively, V(G)=0 for R). Poly-Si and substrate potentials are kept constant and only the bands in ONO are allowed to change as trapping/de-trapping continues during P, E or R [7]. For each time step, Poisson equation is solved throughout the ONO stack, continuity and SRH trapping/de-trapping statistics [12]-[15] are solved only in the nitride. Top (OxT) and bottom (OxB) oxide layers are considered as pure tunneling barriers (no traps), and electric field in these layers (which changes at each time step) decide conduction and valence band tunneling mechanisms (Fowler-Nordheim, Direct, Modified Fowler-Nordheim) during P/E (Table 1). Only drift is considered as transport mechanism in the nitride during P and E (as electric field is high) [7,9]. Thermal emission [10,18], Poole-Frenkel emission [12,16], trap-to-band tunneling [16-18] and trap-to-trap tunneling [19] are considered for charge loss during R (Table 1). Single energy level, non-interacting traps (having T independent but field dependent capture cross section) are assumed for electrons and holes. Trap generation is not considered in the ONO during P/E operation. Electron and hole effective masses are treated as fitting parameters [7, 20].

Impact of trap profiles

Several trap profiles (along with any linear combination with varying weights) can be specified (Fig.3): (a) Uniform, (b) Gaussian, (c) Inverted Gaussian, (d) Exponential increase from OxT towards OxB, and (e) Exponential decrease from OxT interface towards OxB interface. Trap distribution has relatively smaller impact on program transients (Fig.4a) but significantly impacts erase transients (Fig.4b). Retention loss during R (Fig.4c) shows similar trap profile dependence as in E (Fig.4b). Uniform distribution shows fastest P/E transients. Exponential increase (more traps near OxB) causes relatively fast P but slowest E. Exponential decrease (less traps near OxB) causes slowest P and relatively fast E. When compared to Gaussian, Inverted Gaussian shows faster P and slower E.

Impact of physical and trap parameters

Simulations were done with an Inverted Gaussian profile3. Increase in electron m* (in OxB) results in lower tunneling current (Fig.1a, J2), slower P transient and lower saturation value (Fig.5, LHS). Increase in electron m* (poly-Si) reduces gate injection tunneling current (Fig.1b, J1), and causes faster E transient (Fig.5, RHS). Increase in electron trap depth (φe) in the nitride reduces charge de-trapping, resulting in slightly faster P transient and higher saturation value but drastically slower E transient (Fig.6). Increase in capture cross section

---

1 No nitride charge is assumed for DESSIS seed solution in P mode. A charge (equivalent to that obtained at end of P from the SONOS simulator) is placed at the bottom NO interface for DESSIS seed solution before E and R mode.

2 Simulations were done with non-optimized trap parameters, only to verify the relative impact of trap distribution profiles.

3 Closest to the trap distribution obtained experimentally.
increases trapping efficiency and hence results in faster P speed and higher saturation value but reduces E speed (Fig.7). Increase in attempt to escape frequency [8, 9] increases probability of electron detrapping resulting in lower P speed but faster E speed (Fig.9). R transients show similar (to E) dependence on physical parameters, i.e., higher charge loss for lower capture c.s., higher attempt to escape frequency and lower trap depth (not explicitly shown in this paper).

**Prediction of experimental results**

Fig.9 shows experimental and simulated P-E transients for two (different nitride thickness) n-type SONOS devices under different P/E bias. Fig.10 shows experimental and simulated R characteristics at different T. The simulator predicts wide range of experimental P-E-R results with good accuracy, and provides trap profiles (Fig.11) and physical parameters values (Table 2). Both devices show identical, highly non-uniform trap profiles (much higher trap density near OxT-N interface than N-OxB interface and nitride center), and is consistent with that obtained by pure experimental method [23]. These devices have deposited top oxide and hence larger trap density at and near OxT-N interface.

Fig.12 shows the time evolution of trapped electrons in nitride during P/E operation. Hole current is found to be very small during both P/E (larger valence band offset). Therefore, electron injection followed by trapping and electron de-trapping followed by ejection are respectively responsible for $\Delta V_T$ shift during P and E. Figure 13(a) shows the variation of bottom and top oxide electric fields during P/E. During P, electron trapping (negative charge) in nitride results in OxB field to decrease and OxT field to increase in time. For E, electron detrapping (less negative charge) in nitride results in OxT field to increase and OxB field to decrease in time. Fig. 13(b) shows the tunneling currents in top and bottom oxide during P. Decreasing OxB field results in a gradual (continuous) change in the tunneling regime from FN to DT, while increasing OxT field increases FN tunneling current [7]. Fig.13 is a good indicator that internal modules of the simulator are working as per expectation.

**Optimization of trap profile**

The simulator can be used to determine ideal trap profile for P-E-R optimization. As shown in Fig.14, pushing the trap distribution towards OxB would increase P speed, but also increase E speed and R loss. Changing trap density in bulk of nitride (keeping density near OxT and OxB constant) does not have a major impact on P-E-R. Therefore, process steps that impact the quality of N-OxB or OxT-N interfaces would play a major role on P-E-R optimization of SONOS devices.

**Summary and Outlook**

To summarize, a post-processor simulator is developed to predict program, erase and retention behavior of SONOS cells. It can predict P-E-R characteristics over a wide range of experimental conditions and provide information about trap distribution, trap parameters and other physical parameters. The simulator can be used for material characterization and optimization during stack development of SONOS and similar charge trapping devices.

**Acknowledgment**

J. Vasi, D.K. Sharma (IIT Bombay) for useful suggestions. Nitesh Jain (presently at UT, Austin) for help during early development. E. Murakami of Renesas Technologies (Japan) for devices and support.

**References**

Fig 1: Band schematic of ONO stack under three modes of operation (a) programming (b) erase and (c) retention. In Fig. 1(a) depending on the band bending tunneling currents J2 and J3 can be DT, FN or MFN while J1 and J4 can be DT or FN. Similarly in Fig. 1(b) J1 and J4 can be DT, FN or MFN while J2 and J3 can be DT or FN. In Fig. 1(c) charge loss processes are (1) Band to band tunneling, (2) trap to band, (3) Poole-Frenkel emission, (4) Thermal emission and (5) Trap to trap tunneling. OxT is top-oxide, N is Si$_3$N$_4$ and OxB is bottom-oxide.

Table 1: System equations solved in the simulator.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Stack</th>
<th>Equations</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-E</td>
<td>Oxide</td>
<td>Poisson and DT or FN or MFN</td>
<td>[7, 9, 21, 22]</td>
</tr>
<tr>
<td></td>
<td>Nitride</td>
<td>Poisson, SRH Continuity</td>
<td>[7, 9, 12]</td>
</tr>
<tr>
<td>R</td>
<td>Nitride</td>
<td>Trap-trap, Trap-band Thermal emission Poole Frenkel Poisson</td>
<td>[10, 17, 18, 19]</td>
</tr>
</tbody>
</table>

Fig 2: Flowchart showing the steps of simulation flow.

Fig 3: P-E-R transients simulated for various profiles: (1) Uniform, (2) Gaussian, (3) Exp. Dec., (4) Exp. Inc. (5) Inverted Gaussian, for 3.8nm/6nm/5nm (OxT/N/OxB).

Fig 4: Programming AV_T for various profile (Fig. 3) is shown in (a) at Vg = 12 V. Normalized erase AV_T is shown in (b) at Vg = -12 V. Uniform profile shows maximum AV_T during programming. Also it gives the fastest erase. Erase transients show a significant dependence on trap-profile. Retention shows similar nature as erase but after a longer time. Normalization has been done with AV_T max. after programming.

Fig 5: Effect of electron effective mass $m^*$ on P-E transients. (Simulated at Vg = 12V).

Fig 6: Effect of electron trap depth $\phi_{tr} (n,p)$ in N on P-E transients. (Simulated at Vg = 12V.)

Fig 7: Effect of capture cross-section $\sigma_{n,p}$ in N on P-E transients. (Simulated at Vg = 12V.)
**Fig. 8:** Effect of attempt to escape frequency \((\nu_a)\) in \(N\) on P-E transients. (Simulated at \(|V_g| = 12V\))

**Fig. 9:** Simulated and experimental P/E transients for two stacks (1) 5.8nm/6nm/5nm (2) 5.8nm/6nm/5nm at different gate biases. Thicker stack provides higher \(\Delta V_T\) since amount of charge trapped is higher. Devices were initially programmed at 11V for 10s.

**Fig. 10:** Simulated and experimental retention curves for the stack 5.8nm/6nm/5nm at three temperatures \((T)\); 27°C, 70°C and 120°C. The initial \(V_T\) is 4.7 V for all the three retention curves. Faster retention loss is observed at higher \(T\).

**Fig. 11:** Estimated trap profile obtained after matching in two stacks. Profiles are similar with higher trap density at OxT-N interface compared to N-OxB side. This is due to poor interface quality at OxT-N interface where both materials are deposited.

**Fig. 12:** Temporal evolution of electrons in \(N\). Figure shows, erase is due to back-tunneling of electrons which decreases \(\Delta V_T\).

**Table 2:** Extracted physical parameter values

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Units</th>
<th>Value</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma_{\text{ox}})</td>
<td>cm(^{-2})</td>
<td>2.2 x 10(^{-14})</td>
<td>±0.2</td>
</tr>
<tr>
<td>(V_{\text{ox}}) (PF)</td>
<td>Hz x 10(^{14})</td>
<td>1.3 ±0.1</td>
<td></td>
</tr>
<tr>
<td>(V_{\text{ox}}) (TB)</td>
<td>Hz x 10(^{14})</td>
<td>1.3 ±0.1</td>
<td></td>
</tr>
<tr>
<td>(\Phi_{\text{nit}})</td>
<td>eV</td>
<td>1.46 ±0.02</td>
<td></td>
</tr>
<tr>
<td>(\Phi_{\text{ox}})</td>
<td>eV</td>
<td>1.44 ±0.02</td>
<td></td>
</tr>
<tr>
<td>(m_{\text{eff}})</td>
<td>Unit less</td>
<td>0.463 ±0.01</td>
<td></td>
</tr>
<tr>
<td>(m_{\text{bulk}})</td>
<td>Unit less</td>
<td>0.52 ±0.03</td>
<td></td>
</tr>
<tr>
<td>(m_{\text{ox}})</td>
<td>Unit less</td>
<td>0.466 ±0.01</td>
<td></td>
</tr>
<tr>
<td>(\Phi_{\text{ohm}})</td>
<td>eV</td>
<td>3.12 ±0.03</td>
<td></td>
</tr>
<tr>
<td>(\Phi_{\text{ohm}})</td>
<td>eV</td>
<td>4.0 ±0.2</td>
<td></td>
</tr>
<tr>
<td>(\Phi_{\text{ohm}})</td>
<td>eV</td>
<td>3.1 ±0.1</td>
<td></td>
</tr>
<tr>
<td>(\Phi_{\text{ohm}})</td>
<td>eV</td>
<td>4.0 ±0.2</td>
<td></td>
</tr>
<tr>
<td>(\Phi_{\text{ohm}})</td>
<td>eV</td>
<td>1.05 ±0.05</td>
<td></td>
</tr>
<tr>
<td>(\Phi_{\text{ohm}})</td>
<td>eV</td>
<td>1.85 ±0.05</td>
<td></td>
</tr>
</tbody>
</table>