Modeling of Trap Generation in 3-D NAND Charge Trap Flash Memory

Anuj Kumar, Ravi Tiwari, Himanshu Rai, Rashmi Saikia, Arnav Shaurya Bisht and Souvik Mahapatra*
Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India
*Phone: +91-222-572-0408, Email: souvik@iitb.ac.in

2024 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD) | 979-8-3315-1635-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/SISPAD62626.2024.10733029

Abstract: The Reaction-Diffusion-Drift (RDD) framework, implemented in both Sentaurus TCAD and standalone 1-D versions, is used to simulate Program/Erase (P/E) cycling related Tunnel Oxide Trap Generation (TO-TG) in Charge Trap Flash (CTF) NAND. The TCAD version uses NAND string as a device structure and bipolar bias for cycling, but is limited in the P/E cycle count due to computational time overhead. The 1-D version has an identical trap generation model as in TCAD, but it uses unipolar pulse for calibration with TCAD bipolar P/E results and extrapolate to high P/E cycle counts. The simulated TO-TG density is incorporated in the Activated Barrier Double Well Thermionic Emission (ABDWT) model to calculate post-cycling data retention (DR) loss. Experimental DR data are reproduced.

Keywords**:** 3-D NAND, Charge Trap Flash, P/E cycling, data retention, RDD model, ABDWT model.

I. Introduction

Data retention is a key metric of a nonvolatile memory like 3-D CTF NAND, and is caused by in-cell and inter-cell components, as shown in Fig. 1(a) [1]-[3]. It is well known that TO traps play a crucial role for the in-cell component, by triggering mechanisms such as Trap Assisted Tunneling (TAT), Detrapping (DT) and Interface Trap Passivation (ITP), as shown in Fig. 1(b) [1]. P/E cycling increases TO trap density and related DR loss during bake after cycling. While a product in actual use undergoes P/E cycling over its lifetime, this is replicated during qualification through a distributed cycling scheme, where certain % of cycles are done rapidly at a given temperature (T), and the remaining % of cycles are done with inserted delay at the same or a different T [2], [4]. The TAT process is directly related to TO-TG density, and recent results from 3-D CTF NAND show the following behaviors: (a) TG density has a power-law relation with cycle count, with power factor of 0.5, (b) there is no impact of inserted delay between cycles on TG density, and (c) TG density does not change during post-cycling DR bake [2]. However, DT involves detrapping of trapped electrons in generated TO traps, and show impact of inserted delay and retention bake, although the density of traps does not change.

It is important to understand the mechanism governing TO-TG during P/E cycling and develop a suitable modeling framework. The RDD model has been effectively applied in logic devices to explain experimental TG characteristics observed during Bias Temperature Instability (BTI), Stress Induced Leakage Current (SILC) and also Time Dependent Dielectric Breakdown (TDDB) experiments [5]. The logic devices undergo DC or unipolar stress, and the RDD model implementations in deterministic 1-D standalone and 3-D TCAD, and stochastic 3-D standalone have been shown to be equivalent [5]. The 1-D standalone model was also used to simulate distributed P/E cycling using unipolar pulses and assuming TG occurs solely during E step [6]. Recently, the RDD model has also been implemented under bipolar pulses (for P and E) using TCAD, demonstrating suitability for simulating TO-TG in NAND [7]. That implementation however has only shown some basic TO-TG features under a rapid but limited P/E cycle count (#20), and experimental data are not reproduced.

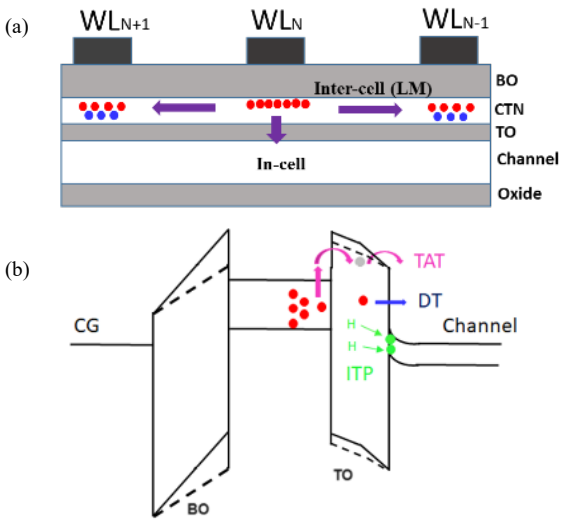

Fig.1. (a) Schematic diagram showing the In-cell and Inter-cell charge loss mechanisms, (b) Energy band diagram showing long-term in-cell components. Reproduced from [1].

In this work, enhancement is done to the TCAD setup to increase the P/E cycle count (to #100) without any runtime penalty; TO-TG is simulated in both P and E steps by the Capture Emission Depassivation (CED) and Multi-State Configuration (MSC) modules of the Sentaurus Device [8]. The 1-D standalone model is calibrated with TCAD, and is used to simulate higher P/E cycle count (>10K) relevant for Triple Level Cell (TLC) applications; only unipolar pulses are used. The unipolar high cycle data are also verified by using TCAD based effective DC simulation. The simulated TO-TG density (N_0) is then used as an input into ABDWT model, to simulate DR loss during bake after cycling under Solid Pattern (SP) tests [3]. The TAT and DT contributions are isolated (DT and ITP are difficult to isolate and hence lumped), and experimental data are reproduced for various measurement conditions.

Authorized licensed use limited to: TU Wien Bibliothek. Downloaded on November 07,2024 at 14:07:23 UTC from IEEE Xplore. Restrictions apply.

II. Trap Generation During Cycling

Fig.2 shows the 2-D cross sectional cut of a 3-D NAND string used for TCAD simulation, with 3 consecutive Word Lines (WL's) and Select FETs; the structure is simulated using Sentaurus Process [9]. The biasing scheme during P and E are shown, with electron and hole density (in poly-Si channel) respectively. Fig.2 also shows the RDD model. Channel carrier density (n or p), electric field (E_{OX}) and T breaks Hydrogen (H) passivated bonds at the channel poly-Si/TO interface, H is released and defects are created (X-). Other defects (Y-, Z-) are created by diffusing H in the TO bulk (T activated), the resulting H_2 molecule and ionic (H_2^+, OH^-) species diffuse and/or drift along the NAND string and control the cycle count dependence of TG (power law slope, ion dominated case result in 0.5 slope [5]-[7]). Note, TO has a sandwiched ONO stack, and the N layer impedes diffusion of H related species, and hence, unlike logic, the diffusion/drift happens along the direction of NAND string (via the bottom O and poly-Si layers).

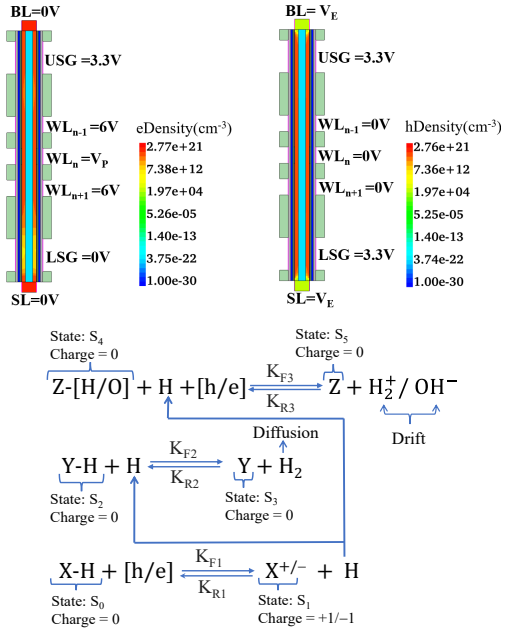

Fig.2. Sentaurus process [9] generated 2-D cross section of 3-D NAND CTF string, showing 3 consecutive word lines and select FETs. The P/E biasing schemes and electron and hole density in the channel poly-Si layer are shown. Also shown are the chemical equations for RDD model. The first reaction (generation of H) is solved at the poly-Si/TO interface, while second (generation of H_2) and third (generation of H_2^+ and OH^-) reactions are solved at a suitable interface defined in the middle of the bottom O layer of the ONO TO [7]. P pulse ($+V_G$) results in electrons in the channel and eventual generation of H_2 and OH^-, while E pulse ($-V_G$) results in holes in the channel and eventual generation of H_2 and H_2^+.

Fig.3 shows the TCAD simulated atomic (H), molecular (H_2) and ionic (H_2^+ and OH^-) species and TO-TG profiles along the NAND string after 100 P/E cycles. Only N_{TO2} is plotted, however, N_{TO1} and N_{TO2} have identical density as H released from channel/TO interface generates bulk TO traps. Note that N_{TO1} and N_{TO2} are themselves the sum of traps created during the P and E pulses, N_{TO1} is triggered by electrons and holes in the channel respectively, while subsequent released H generates N_{TO2}.

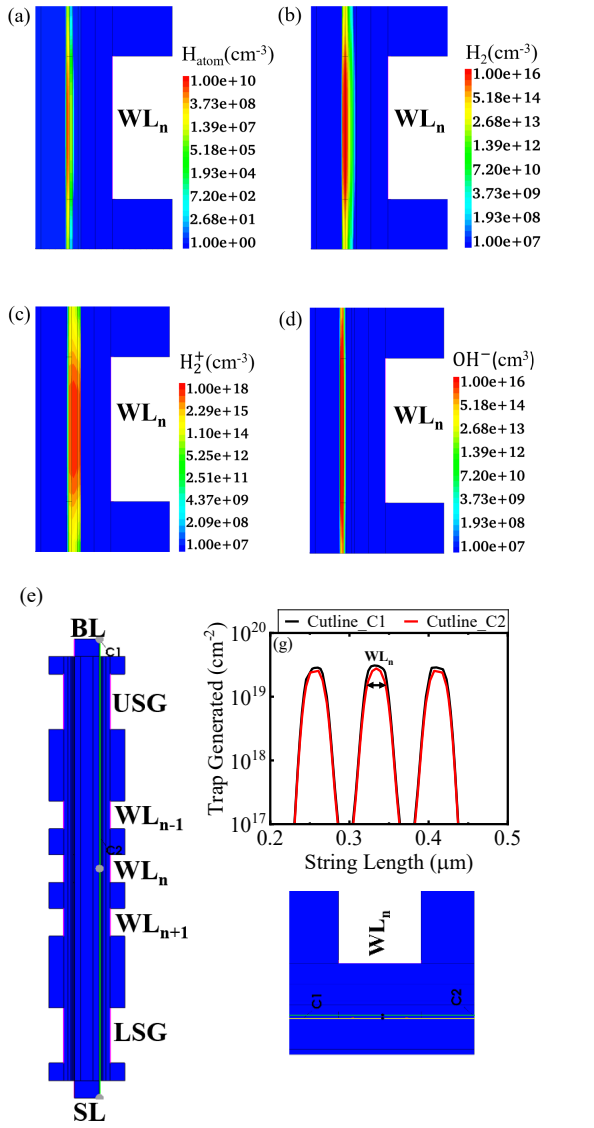

Fig.3. Sentaurus device [8] implemented RDD model simulation results, showing generated (a) H, (b) H_2, (c) H_2^+, (d) OH^- and (e) TO-TG (inside TO) in the NAND string (the target cell and neighboring intercell regions are shown for clarity in (f)). RDD model invokes the Capture Emission Depassivation (CED) and Multi-State-Configuration (MSC) modules of Sentaurus device as discussed in [5], [7], [8].

Fig.4 plots the P/E cycle dependence of traps triggered by (a) electrons and (b) holes, at the poly-Si/TO interface (N_{TO1}) and inside TO bulk (N_{TO2}), only 20 cycles are shown for clarity. The electron related traps are generated during the P pulse (10μs) and heals during the E pulse (1ms), and the hole related traps show the opposite behavior, for N_{TO1}. The corresponding secondary traps (N_{TO2}, from subsequent reactions) show a steady growth. The sum of electron and hole traps shows equal density for N_{TO1} and N_{TO2} (this is expected as H released from first reaction creates traps out of subsequent reactions), and a steady growth with cycle count, Fig.4 (c). It should be noted that traps are calculated at two interfaces in TCAD purely due to implementation considerations, and in general, N_{TO1} should be at or near the poly-Si/TO interface and N_{TO2} spread across the bottom O

Authorized licensed use limited to: TU Wien Bibliothek. Downloaded on November 07,2024 at 14:07:23 UTC from IEEE Xplore. Restrictions apply.

layer of ONO TO, and respectively resulting in DR loss via the DT and TAT mechanisms. Fig. 4(d) shows reproduction of TCAD bipolar P/E simulated TO-TG density with 1-D standalone unipolar simulation and TCAD-DC simulation. Identical results are obtained, since RDD model results in identical TO-TG trap time kinetics under DC, unipolar and bipolar pulses.

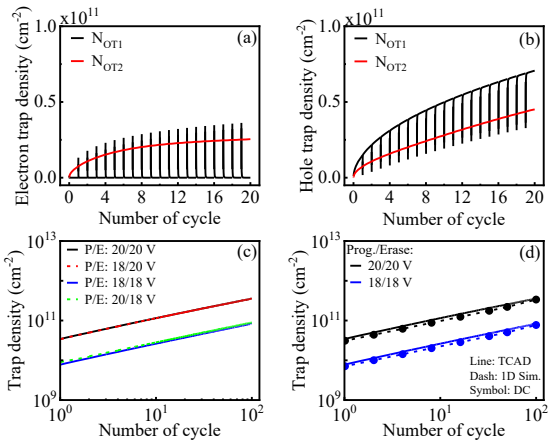

Fig.4. TCAD simulated P/E cycling related traps linked to channel (a) electrons and (b) holes, and (c) Total trap density at various P/E biases. (d) Total trap density during P/E cycling is compared with unipolar pulse and DC. P of 10µs and E pulse of 1ms is used throughout the work.

Fig.5 plots the TO-TG density (a, b) versus P/E cycling count during P/E cycling at (a) various (+/–) V_{WL} (wordline bias) but fixed T and (b) various T but fixed (+/–) V_{WL}, and also (c, d) during retention bake after cycling (at same T as cycling). The simulations are done by using 1-D standalone under unipolar pulse till 10K cycles, after calibrating with TCAD under bipolar pulse till 100 cycles, for all V_{WL} and T conditions. Furthermore, TCAD DC simulations are also used (for effective time related to 10K P/E cycles), showing identical results as pulsed (unipolar or bipolar) cases. This implies thar TCAD DC can be directly calibrated to TCAD bipolar till 100 P/E cycles and extrapolated to get TO-TG at higher cycle counts. The retention bake is done after 10K cycles, by using the standalone model. TO-TG density from both TCAD and standalone implementations show power law dependence on cycle count, with 0.5 slope, which holds for all V_{WL} and T. Moreover, no change in TO-TG density is observed during post-cycling DR bake. These simulation results are consistent with experimental features [2]. Note that the absence of any variation during DR bake is also verified by using TCAD, but after 100 cycles (not explicitly shown for brevity).

III. Data Retention After Cycling

The ABDWT model framework for DR bake is shown in Fig.6 [3]. It involves energy levels E_1 (representing a trap) and E_2 (representing the channel) separated by an energy barrier, with mean E_{BM} and spread E_{BS} that are T activated. The internal oxide electric field lowers E_{BM} and E_2 during DR bake, which triggers an over-the barrier transfer from E_1 to E_2, resulting in occupancy changes for E_1 (loss) and E_2 (gain). Two independent ABDWT models are used for DR loss, each for TAT and DT+ITP. E_1 is a trap in storage nitride for TAT (programmed state determines N_0), and that in the TO for DT (N_0 is related to trap density).

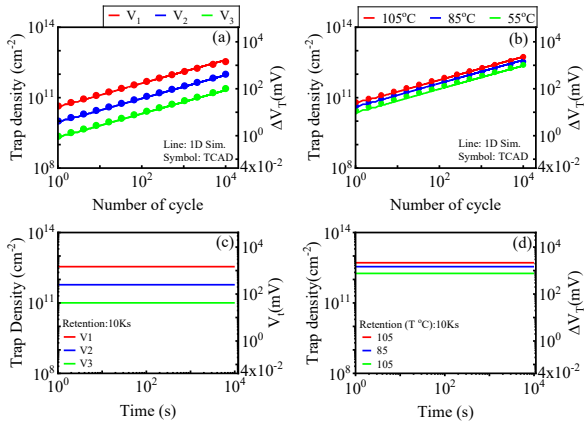

Fig.5. Simulated TO-TG density (a, b) during and (c, d) after P/E cycling at different P/E V_{WL} and T. Cycling is done with standalone model with effective unipolar pulses (shown above). TO-TG density with equivalence TCAD DC simulation is also shown.

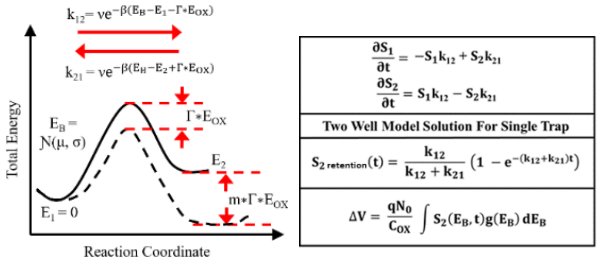

Fig.6. Schematic of ABDWT model which is incorporated in the retention framework [3]. The rate equations for the model are also shown.

Fig. 7(a) shows the simulated TO-TG density versus P/E cycles (TCAD bipolar extended using standalone unipolar) used as N_0 to model the DT component of DR loss (shown later in Fig.8). The TAT component is modeled by using a related change in E_{BM}, shown in Fig. 7(b). Other than these two, all other ABDWT model parameters are kept fixed to model DR loss experiments, refer to [3] for details.

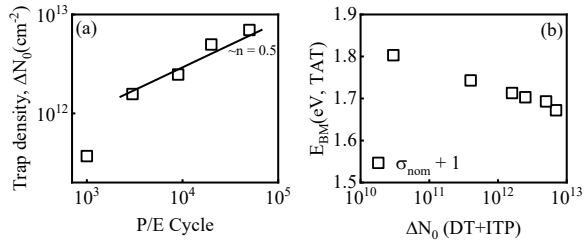

Fig.7. (a) TCAD simulated trap density (ΔN_0) used in DT+ITP component of DR loss (b) correlation of E_{BM} to ΔN_0 used in TAT component of DR loss using the ABDWT model framework [3].

Authorized licensed use limited to: TU Wien Bibliothek. Downloaded on November 07,2024 at 14:07:23 UTC from IEEE Xplore. Restrictions apply.

Fig.8 shows the time kinetics of (a) DT+ITP and (b) TAT subcomponents of DR loss after different P/E cycle counts from ABDWT model. These subcomponents are added to obtain DR loss under SP test (all wordlines are at the same programmed state), and compared to experimental data as shown in Fig.9. The model can successfully reproduce data after different P/E cycles, thereby validating the combined RDD-ABDWT model.

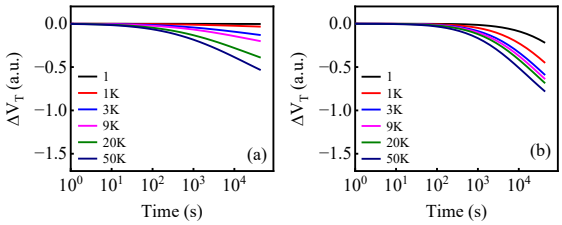

Fig.8. DR loss time kinetics corresponding to (a) DT+ITP and (b) TAT subcomponents after different P/E cycle counts. TO-TG from RDD model is used in ABDWT model as (a) N_0 and (b) E_{BM} (see Fig.7).

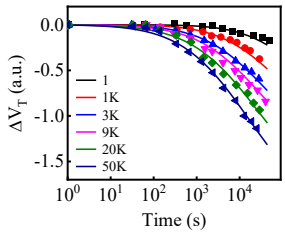

Fig.9. Measured (symbols) and modeled (lines) DR loss kinetics under SP test from P7 state, for different P/E cycle counts.

IV. Conclusion

TCAD implemented RDD model is used to calculate TO -TG due to bipolar P/E cycling up to 100 cycle counts. The standalone RDD model under unipolar cycling is calibrated with TCAD and extended to 10K (or higher) cycle counts relevant for TLC applications. DC RDD simulation from TCAD with an effective time is used to re-verify unipolar results. All cases (TCAD bipolar pulse, standalone unipolar pulse and TCAD DC) results in power-law dependence of TO-TG on cycle count with 0.5 slope, and TO-TG density does not change during the retention bake, consistent with experimental data. The TO-TG density is used in ABDWT model to calculate DR loss during retention bake after P/E cycling, and can well reproduce measured data.

Acknowledgement

The authors would like to thank Mohit Bajaj of Synopsys India Pvt. Ltd. for help with TCAD setup, and industrial partner for measured data.

References

[1] C. Woo et al., “Modeling of Charge Failure Mechanisms during the Short Term Retention Depending on Program/Erase Cycle Counts in 3-D NAND Flash Memories,” IEEE Int. Rel. Phys. Symp. (IRPS), 2020, pp. 1-6.

[2] G. Nicosia, N. Righetti and Y. Dong, "Distributed Cycling in Charge Trap-Based 3D NAND Arrays: Model and Qualification Tests Implications,” IEEE Int. Mem. Works. (IMW), 2023, pp. 1-4.

[3] R. Saikia, H. Rai and S. Mahapatra, "Modeling of Post-Cycling Retention Bake in 3-D CTF TLC NAND Arrays," IEEE Int. Rel. Phys. Symp. (IRPS), 2024, pp. 1-6.

[4] N. Mielke, H. P. Belgal, A. Fazio, Q. Meng and N. Righos, "Recovery Effects in the Distributed Cycling of Flash Memories," IEEE Int. Rel. Phys. Symp. (IRPS), 2006, pp. 1-6.

[5] S. Mahapatra et al., "A Generic Trap Generation Framework for MOSFET Reliability—Part I: Gate Only Stress–BTI, SILC, and TDDB," in IEEE Trans. Electron Devices, vol. 71, no. 1, pp. 114-125, 2024.

[6] T. Samadder, S. Kumar, K. Thakor and S. Mahapatra, "A Theoretical Framework for Trap Generation and Passivation in NAND Flash Tunnel Oxide During Distributed Cycling and Retention Bake," IEEE Int. Rel. Phys. Symp. (IRPS), 2024, pp. 1-6.

[7] A. Kumar, R. Tiwari, M. Bajaj, D. Dolgos, L. Smith and S. Mahapatra, "Study of Trap Generation in NAND Flash Tunnel Oxide using TCAD," IEEE Electron Devices Tech. and Manf. (EDTM), 2024, pp. 1-3.

[8] Sentaurus Device user guide, Version T-2022.03.

[9] Sentaurus Process user guide, Version T-2022.03.

Authorized licensed use limited to: TU Wien Bibliothek. Downloaded on November 07,2024 at 14:07:23 UTC from IEEE Xplore. Restrictions apply.