A Stochastic Modeling Framework for NBTI and TDDS in Small Area p-MOSFETs

Anandkrishnan. R, S. Bhagdikar, N. Choudhury, R. Rao, B. Fernandez, A. Chaudhury, N. Parihar and S. Mahapatra

Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India

*Phone: +91-222-572-0408, Email: souvik@ee.iitb.ac.in

***Abstract*— Kinetic Monte Carlo (KMC) simulations are used to simulate the stochastic interface trap generation recovery (ΔV_{IT}) and hole trapping detrapping (ΔV_{HT}) during and after Negative Bias Temperature Instability (NBTI) stress. The simulated mean of threshold voltage shift ($\Delta V_T = \Delta V_{IT} + \Delta V_{HT}$) is verified against continuum simulations and mean of measured data on multiple small area devices. Simulated and measured time constants for steps of Time Dependent Defect Spectroscopy (TDDS) data and step like recovery after NBTI stress are compared and analyzed.**

Keywords— NBTI, HKMG, Kinetic Monte Carlo (KMC), interface trap generation, hole trapping, Reaction-Diffusion (RD) model, Non-Radiative Multi-phonon (NMP) model.

I. Introduction

NBTI is a crucial reliability issue for both large and small area p-channel MOSFETs [1], [2]. Modeling of the ΔV_T time kinetics during and after stress requires understanding of the underlying physical mechanism, which is debated [3], [4]. A continuum model framework having uncorrelated contributions from interface trap generation (ΔV_{IT}), trapping in pre-existing defects (ΔV_{HT}), and generation of bulk insulator defects (ΔV_{OT}) has been used to predict the measured data for different stress bias (V_{GSTR}) and temperature (T) for different technologies [5]-[11]. It uses Reaction Diffusion (RD) model for interface trap generation (ΔN_{IT}), Transient Trap Occupancy Model (TTOM) to calculate trap occupancy and their contribution (ΔV_{IT}), and analytical models for both ΔV_{HT} and ΔV_{OT}. The schematic of continuum framework is shown in Fig.1 [5]. A corresponding stochastic framework and the model prediction of the mean of small area devices have been shown in [12]. The model uses KMC RD [13] but analytical TTOM and ΔV_{HT} (no ΔV_{OT} since it is negligible at ~use bias [6]). However, the consistency of continuum [5] and stochastic [12] versions is debated [3], [14]. Connection between NBTI and TDDS [15] is also debated [3].

In this paper, Non-radiative Multi Phonon (NMP) model [16] is implemented in KMC domain for ΔV_{HT} kinetics. The H_2 lock-in during stress [17] is added to KMC RD for ΔN_{IT} [13]. H_2 diffusivity reduction used in continuum RD for recovery [5] that was questioned in [3] is verified by H_2 hopping and lock-in using stochastic framework. Elastic tunneling based electron capture [10], [18] is used for time constant dispersion in the stochastic TTOM (as macroscopic analyses imply negligible T impact on the TTOM time constants [5]-[11]). Individual and mean ΔV_{IT} and ΔV_{HT} time kinetics are simulated. The mean of measured stress and recovery kinetics is modeled. Continuum simulations are used to verify mean of stochastic results. Time constants measured for NBTI recovery and TDDS steps are compared and explained for different experimental conditions. The connection between KMC and continuum implementations is established.

II. Experimental and Model Framework Details

Gate first High-k Metal Gate (HKMG) planar p-channel MOSFETs with interfacial layer (IL) thickness of ~3Å grown using low T rapid thermal process (RTP) are used. The HfO_2 High-k thickness is 23Å and made by atomic layer deposition process. Refer to [19] for full details. An ultra-fast (UF, 10µs delay) method is used for stress and recovery measurements.

Fig.2 shows a schematic explaining different mechanisms in the KMC RD framework. Gillespie stochastic simulation [20] based propensities for H passivated bond dissociation and re-passivation, H, H_2 diffusion and H_2 lock-in calculate trap generation and passivation at the Si/IL interface, IL bulk, IL/ High-k interface and High-k bulk (all bulk defects are lumped at IL/High-k interface for simplicity) during and after stress. Backend shown in Fig.2 consists of High-k and metal layers. A random fraction of generated traps is selected that go below the Fermi level during recovery as shown in Fig.3. These traps are assigned random time constants (dispersed as per elastic tunneling model [18], refer Fig.4) for electron capture during recovery to implement stochastic TTOM. A stochastic version of the NMP model, Fig.5, with suitable reaction rates between the energy wells is utilized for hole trapping and detrapping. The precursors for trap generation and trapping are randomly distributed in space in a device, Fig.6, and multiple devices are used for KMC simulations. Exponential distribution of a single charge impact, Fig.6, with mean impact obtained from measured data [21] is used for channel percolative impact of random defects, to calculate V_T shift.

III. Trap Generation, Passivation and Occupancy

Fig.7 locates the generated interface traps during stress for different stress time at first (left) and second (right) interfaces. Fig.8 locates interface traps that remain during recovery after stress for different recovery time at first (left) and second (right) interfaces. The snapshots are from a single simulation, and generated traps are randomly distributed at the first and second interfaces. Fig.9 plots H_2 locations at various instances of stress (left) and recovery (right). H_2 generated from second interface diffuses to 3D space during stress, and back diffuses from 3D space towards second interface and need to search a broken bond for passivation. This causes delayed recovery for cases when broken bonds are few in numbers (*e.g.*, after short time stress or for very long recovery time). Some H_2 can get locked and would never diffuse back, causing lower recovery.

Fig.10 shows the stochastic RD simulated ΔN_{IT} traces from multiple small area devices along with mean during DC stress. Stochastic traces show discrete jumps when traps get created; mean of multiple simulations shows power law dependence at long time with slope n~1/6. Soft-saturation is observed at long

stress time. Fig.11 compares the ΔN_{IT} kinetics obtained from continuum model to the mean ΔN_{IT} obtained using stochastic RD with and without H_2 lock-in. The introduction of H_2 lock-in slightly increases the slope of stochastic mean degradation, but the slope is still n~1/6. Continuum and mean of stochastic simulations closely follow each other. Fig.12 shows stochastic simulated individual and mean recovery kinetics along with continuum RD simulation without diffusivity reduction for short stress and long recovery time. Discrete jumps are seen when a trap gets passivated. The hopping impact is evident as mean of stochastic recovery is slower than continuum model. Fig.13 plots stochastic simulated individual and mean kinetics of recovery without and with lock-in of H_2 and the continuum RD simulation with and without diffusivity reduction after long stress. The hopping effect is smaller at short time recovery after long stress as broken bonds are a plenty. However, lock-in (higher for longer stress) reduces the number of available H_2 for back diffusion and passivation and reduces recovery. Fig.14 plots the stress time dependence of normalized recovery kinetics for mean of KMC recovery simulation after different stress time. The continuum RD model results in ~55% recovery (without diffusivity reduction) for recovery time ~ stress time, which is much faster than that in Fig.14. The results shown in Fig.12 through Fig.14 justify the use of diffusivity reduction in the continuum RD framework [5].

Stochastic TTOM (Fig.3) augments RD model to calculate charge occupancy and actual contribution of generated traps. Time constants are randomly assigned to traps selected for the electron-capture process as per the distribution shown in Fig.4. Fig.15 plots individual traces and mean of ΔV_{IT} recovery from the stochastic TTOM augmented RD model. The continuum result and underlying fast and slower components are shown. The mean of multiple stochastic simulations agrees well with continuum model. The ΔV_{IT} step height for each recovery step (calculated using Fig.6) and plotted in Fig.16 as function of emission time obtained from only stochastic RD and TTOM enabled stochastic RD model. The emission time is distributed from short to long time for TTOM enabled RD model, where the short emission time is due to the electron capture process and long emission time is due to trap passivation process.

IV. Hole Trapping and Detrapping (NMP Model)

The stochastic traces and mean of ΔV_{HT} during and after stress is shown in Fig.17 and Fig.18 respectively. The traces show discrete jumps as traps get filled (Fig.17) and emptied (Fig.18) using the NMP process. Continuum simulations with identical parameters agree with the mean of multiple KMC simulations. NMP model shows early saturation of the stress kinetics when model parameters are adjusted for fast recovery after stress, to remain consistent with experimental data shown later. Fig.19 plots the ΔV_{HT} step height versus emission time extracted from KMC NMP recovery kinetics (for saturating stress parameters). Most of the steps recover within 10s. Mean of multiple KMC (Figs.20, Fig.22) and continuum (Figs.21, Fig.23) simulations for stress (Figs.20, Fig.21) and recovery (Figs.22, Fig.23) show NMP parameters can be adjusted to either obtain fast saturation and early start of recovery or non-saturation with delayed start of recovery. It is not possible to get non-saturated stress and early start of recovery for any choice of NMP parameters. Its implication is discussed next.

V. Analysis of Experimental Data

The mean of measured ΔV_T time kinetics during stress is predicted using ΔV_{IT} and ΔV_{HT} from continuum (identical to stochastic mean) simulations, see Fig.24 (a). Simulated ΔV_{HT} saturates fast and recovers quickly, as non-saturated ΔV_{HT} also results in delayed recovery that cannot predict experimental data. Fig.24 (b) shows mean stress kinetics for different V_{GSTR} along with prediction by continuum model. The contribution by ΔV_{HT} is smaller at higher V_{GSTR} and longer time, and ΔV_T is dominated by ΔV_{IT}. The fast recovery of ΔV_{HT} is verified by using TDDS measurements [15]. In TDDS measurements, the device is stressed for shorter stress time (100ms) and at low T so that the degradation is dominated by ΔV_{HT}, as can be seen in Fig.24 (a). Fig.25 shows the individual recovery traces and mean for two different devices under TDDS measurements. The step height as a function of emission time extracted from the recovery traces of Fig.25 is shown in Fig.26. Most of the steps recover within ~10s and hence verifies the shorter time constant associated with ΔV_{HT} recovery. Unlike [15], TDDS in this and other [6] devices has widely distributed random time steps, reminiscent of a purely stochastic process. Fig.27 shows the continuum framework prediction of the mean of recovery kinetics measured in multiple small area devices under various stress cases. Fig.27 (b) has relatively higher ΔV_{HT} compared to Fig.27 (a) and hence shows faster recovery; while Fig.27 (c) has relatively higher ΔV_{IT} compared to Fig.27 (a) and hence shows slower recovery. The step heights versus emission time extracted from Fig.27 are shown in Fig.28. Most of the traps recover within ~10s for Fig.28 (b), the ΔV_{HT}-dominated case. Fig.29 shows the step height as a function of emission time simulated using the stochastic framework (KMC RD + TTOM + KMC NMP) for cases when ΔV_{HT} dominates ΔV_T (top) and when ΔV_{IT} dominates ΔV_T (bottom). When ΔV_{HT} dominates ΔV_T, most of the steps recover within ~10s, consistent with experiments (Fig.26 and Fig.28 (b)). When ΔV_{IT} dominates ΔV_T, the emission time is widely distributed from short to long time, consistent with experimental data in Fig.28 (a, c).

VI. Conclusion

Measured step-like ΔV_T stress recovery kinetics and mean of multiple small area measurements are captured by a KMC stochastic framework. Simulated ΔV_{IT} evolves in time with power law kinetics for mean, while ΔV_{HT} saturates at longer time stress. ΔV_{IT} recovery has widely distributed time constant due to electron capture (short) and trap passivation (long). ΔV_{HT} has short time constant for NBTI recovery and TDDS. Mean of stochastic simulations is found to be consistent with the continuum model.

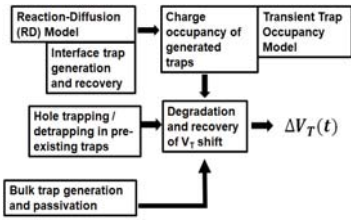


Fig.1. Schematic of a comprehensive NBTI modeling framework consisting of uncorrelted interface trap generation (ΔV_{IT}), hole trapping (ΔV_{HT}) and bulk trap generation (ΔV_{OT}) components.

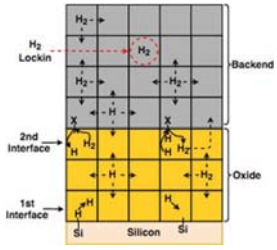


Fig.2. Schematic of the stochastic RD model for H passivated bond dissociation and re-passivation at first and second interfaces and diffusion of H and H_2.

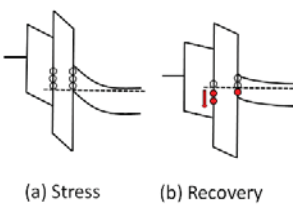


Fig.3: The energy band diagram of PMOS during NBTI (a) stress and (b) recovery. Solid circle represents the donor traps after electron capture.

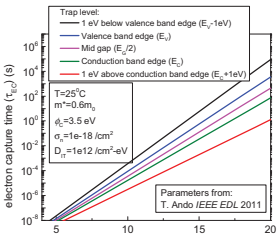


Fig.4. TTOM time constant distribution as a function of oxide depth for different trap energy levels.

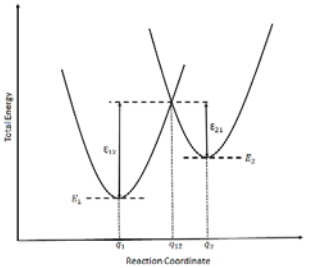


Fig.5. Schematic of the Two Well Non-Radiative Multiphonon model.

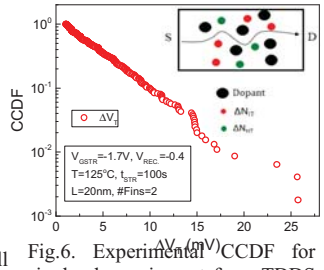
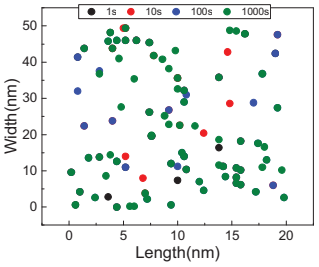


Fig.6. Experimental CCDF for single charge impact from TDDS measurement.



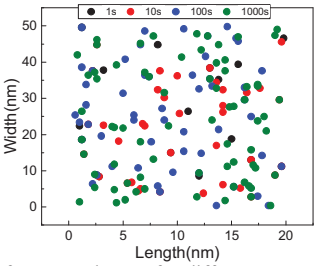
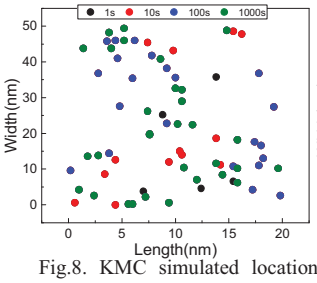


Fig.7. KMC simulated location of generated traps for different stress time at first (left) and second (right) interface.



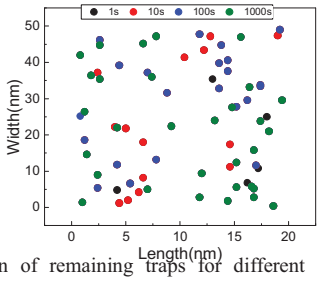


Fig.8. KMC simulated location of remaining traps for different recovery time at first (left) and second (right) interface.

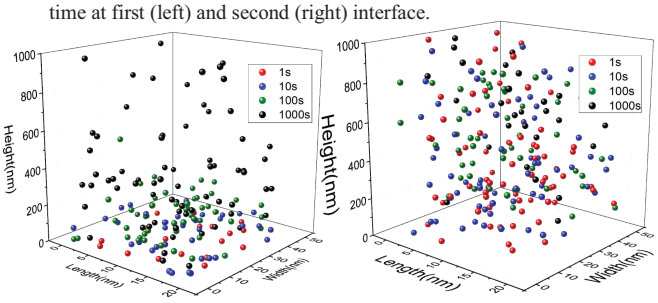


Fig.9. KMC simulated 3 dimensional H_2 profile for different stress time (left) and recovery time (right).

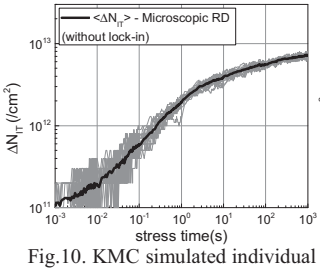


Fig.10. KMC simulated individual and mean ΔV_{IT} stress kinetics. Long time degradation shows 1/6 power law time dependence.

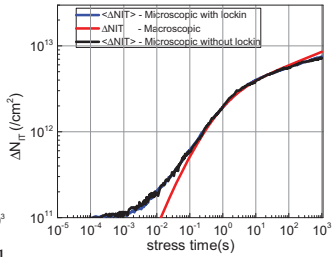
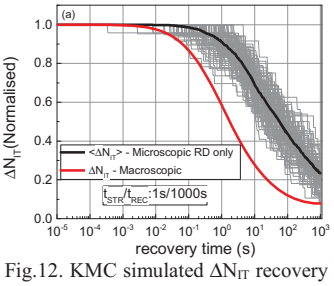


Fig.11. KMC simulated ΔV_{IT} stress kinetics with H_2 lock-in. Continuum result is also shown.



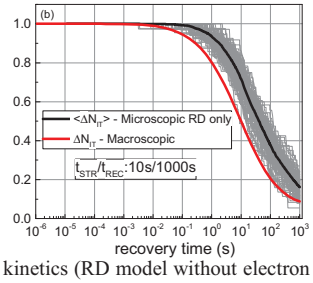
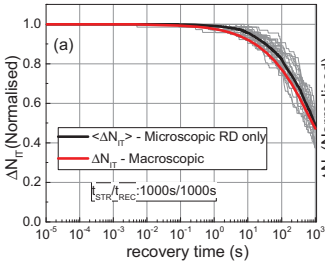


Fig.12. KMC simulated ΔN_{IT} recovery kinetics (RD model without electron capture) and continuum simulation without diffusivity slow down effect for stress time of (a) 1s and (b) 10. Lock in is not considered.



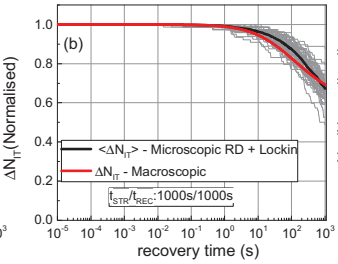


Fig.13. KMC simulated ΔN_{IT} recovery kinetics (RD model without electron capture) (a) without H_2 lock-in and continuum simulation without diffusivity slow down effect (b) with H_2 lock-in and continuum simulation with diffusivity slow down effect.

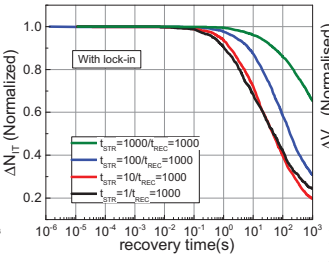


Fig.14. KMC simulated normalized recovery kinetics after different stress time.

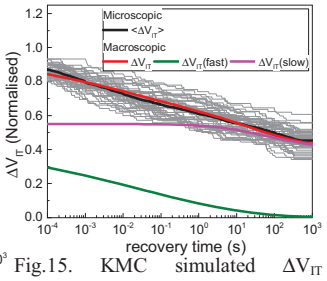


Fig.15. KMC simulated ΔV_{IT} recovery kinetics showing individual traces and mean. Continuum simulation result with fast and slow subcomponent is shown.

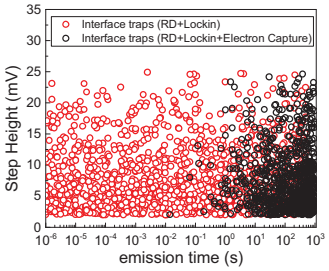


Fig.16. KMC simulated step height as a function of emission time for interface trap recovery using trap passivation, without and with electron capture effect.

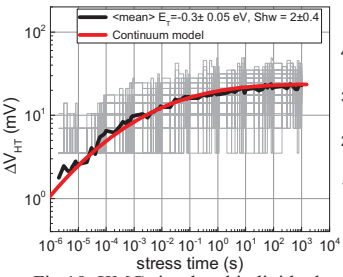


Fig.19. KMC simulated individual and mean ΔV_{HT} stress kinetics. Continuum result is shown.

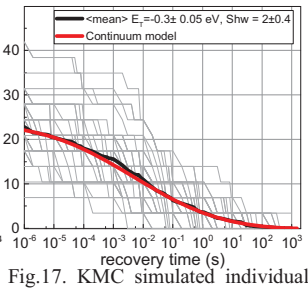


Fig.17. KMC simulated individual and mean ΔV_{HT} recovery kinetics. Continuum result is shown.

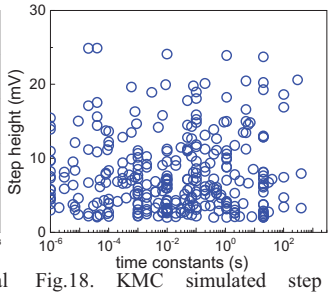


Fig.18. KMC simulated step height as a function of step time for ΔV_{HT} (from NMP).

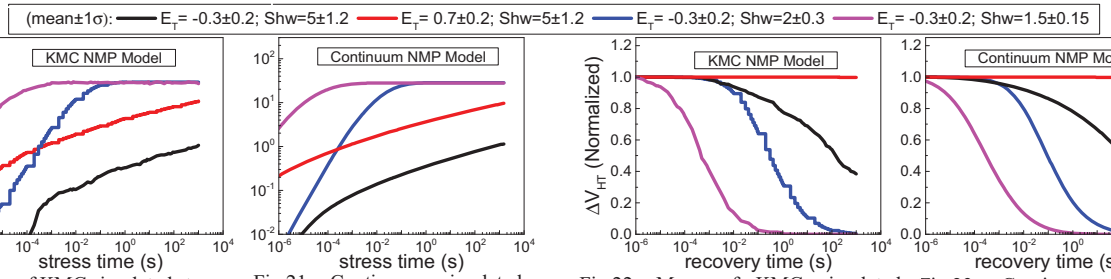


Fig.20. Mean of KMC simulated stress kinetics for various NMP parameters (E_t and Shw), which impact slope of long time degradation.

Fig.21. Continuum simulated stress kinetics for variation in NMP model parameters as in Fig.21.

Fig.22. Mean of KMC simulated recovery kinetics for variation in NMP parameters (as Fig.21), which impact recovery kinetics (fast or delayed).

Fig.23. Continuum simulated recovery kinetics for variation in NMP model parameters as in Fig.21.

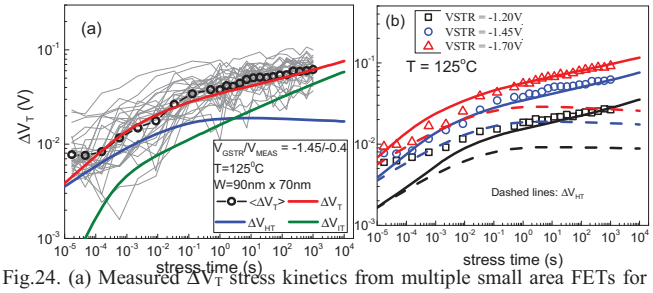


Fig.24. (a) Measured ΔV_T stress kinetics from multiple small area FETs for a fixed V_{GSTR} and T, their mean; and model prediction of mean with different subcomponents. (b) Mean of measured ΔV_T stress kinetics in small area FETs for multiple V_{GSTR} and T, and model prediction with different subcomponents.

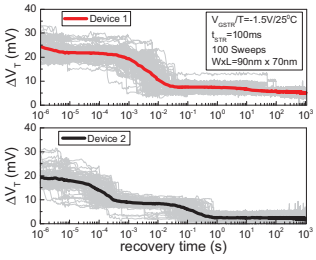


Fig.25. Measured ΔV_T recovery kinetics for subsequent TDDS sweeps in 2 similar small area FETs and their mean.

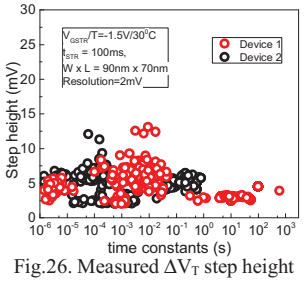


Fig.26. Measured ΔV_T step height as a function of step time from recovery kinetics for TDDS in 2 similar small area FETs.

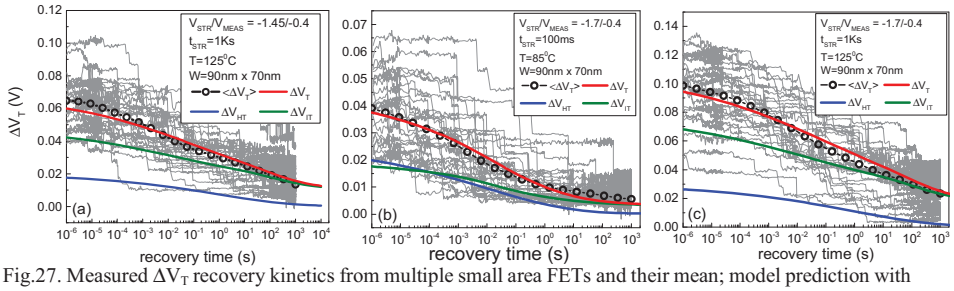


Fig.27. Measured ΔV_T recovery kinetics from multiple small area FETs and their mean; model prediction with subcomponents for (a) moderate V_{GSTR}/T (reference case), (b) high V_{GSTR} but for shorter stress time (ΔV_{HT} dominated), and (c) high V_{GSTR} but for longer stress time (ΔV_{IT} dominated).

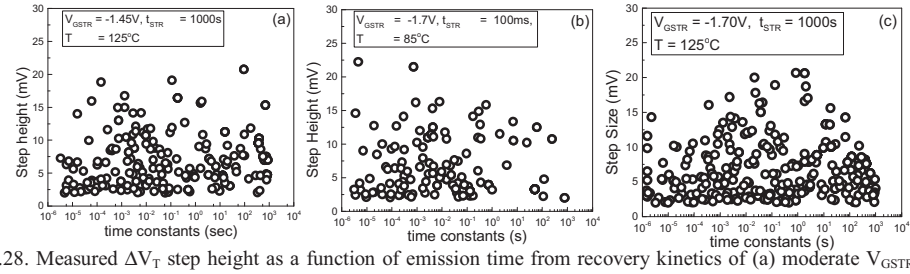


Fig.28. Measured ΔV_T step height as a function of emission time from recovery kinetics of (a) moderate V_{GSTR}/T (Fig.27 (a)), (b) high V_{GSTR} but for shorter stress time (Fig.27 (b)), and (c) high V_{GSTR} but for longer stress time (Fig.27 (c)).

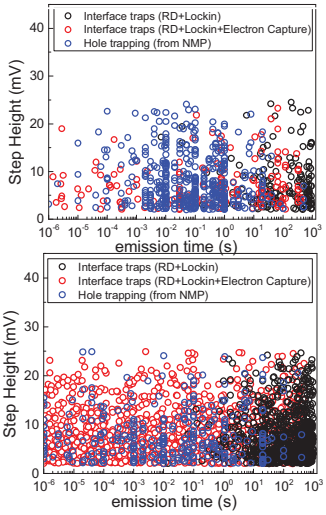


Fig.29. KMC simulated step height as a function of emission time for ΔV_{HT} and ΔV_{IT} for the case of ΔV_{HT} dominated stress (top) and ΔV_{IT} dominated stress (bottom).

REFERENCES

[1] S. Ramey, A. Ashutosh, C. Auth, J. Clifford, M. Hattendorf, J. Hicks, R. James, A. Rahman, V. Sharma, A. St Amour, and C. Wiegand, "Intrinsic transistor reliability improvements from 22nm tri-gate technology," *IEEE Int. Reliab. Phys. Symp. Proc.*, pp. 4–8, 2013.

[2] A. Kerber and P. Srinivasan, "Impact of stress mode on stochastic BTI in scaled MG/HK CMOS devices," *IEEE Electron Device Lett.*, vol. 35, no. 4, pp. 431–433, Apr. 2014.

[3] J. H. Stathis, S. Mahapatra, and T. Grasser, "Controversial issues in negative bias temperature instability," *Microelectron. Reliab.*, vol. 81, no. November 2017, pp. 244–251, 2018.

[4] S. Mahapatra and Narendra Parihar, "A review of NBTI mechanisms and models," *Microelectronics Reliability*, Volume 81, 2018, Pages 127-135, ISSN 0026-2714.

[5] N. Parihar, N. Goel, S. Mukhopadhyay and S. Mahapatra, "BTI Analysis Tool-Modeling of NBTI DC, AC Stress and Recovery Time Kinetics, Nitrogen Impact, and EOL Estimation," in *IEEE Trans. Electron Devices*, vol. 65, no. 2, pp. 392-403, Feb.2018.

[6] N. Parihar, U. Sharma, R. G. Southwick, M. Wang, J. H. Stathis and S. Mahapatra, "Ultrafast Measurements and Physical Modeling of NBTI Stress and Recovery in RMG FinFETs Under Diverse DC–AC Experimental Conditions," in *IEEE Transactions on Electron Devices*, vol. 65, no. 1, pp. 23-30, Jan. 2018.N. Parihar, IEDM, p.7.3, 2018.

[7] N. Parihar, R. G. Southwick, M. Wang, J. H. Stathis and S. Mahapatra, "Modeling of NBTI Kinetics in RMG Si and SiGe FinFETs, Part-II: AC Stress and Recovery," *IEEE Trans. Electron Devices*, vol. 65, no. 5, pp. 1699-1706, May 2018.

[8] N. Parihar, R. G. Southwick, M. Wang, J. H. Stathis and S. Mahapatra, "Modeling of NBTI Kinetics in RMG Si and SiGe FinFETs, Part-I: DC Stress and Recovery," *IEEE Trans. Electron Devices*, vol. 65, no. 5, pp. 1699-1706, May 2018.

[9] N. Parihar, R. G. Southwick, M. Wang, J. H. Stathis and S. Mahapatra, "Modeling of NBTI Kinetics in RMG Si and SiGe FinFETs, Part-II: AC Stress and Recovery," *IEEE Trans. Electron Devices*, vol. 65, no. 5, pp. 1699-1706, May 2018.

[10] N. Parihar and S. Mahapatra, "Prediction of NBTI stress and recovery time kinetics in Si capped SiGe p-MOSFETs," in *Proc. Int. Rel. Phys. Symp.*, 2018, pp. P-TX.5-1-P-TX.5-7.

[11] V. Huard *et al*., "Key parameters driving transistor degradation in advanced strained SiGe channels," in *Proc. Int. Rel. Phys. Symp*, 2018, pp. P-TX.4-1-P-TX.4-6.

[12] A. Chaudhary, B. Fernandez, N. Parihar, and S. Mahapatra, "Consistency of the Two Component Composite Modeling Framework for NBTI in Large and Small Area p-MOSFETs," *IEEE Trans. Electron Devices*, vol. 64, no. 1, pp. 256–263, 2017.

[13] T. Naphade, N. Goel, P. R. Nair, and S. Mahapatra, "Investigation of stochastic implementation of reaction diffusion (RD) models for NBTI related interface trap generation," *IEEE Int. Reliab. Phys. Symp. Proc.*, pp. 1–11, 2013.

[14] T. Grasser, K. Rott, H. Reisinger, M. Waltl, F. Schanovsky, and B. Kaczer, "NBTI in nanoscale MOSFETs-the ultimate modeling benchmark," *IEEE Trans. Electron Devices*, vol. 61, no. 11, 2014.

[15] T. Grasser, H. Reisinger, P. J. Wagner, F. Schanovsky, W. Goes, and B. Kaczer, "The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability," *IEEE Int. Reliab. Phys. Symp. Proc.*, pp. 16–25, 2010.

[16] G. Rzepa, J. Franco, A. Subirats, M. Jech, A. Chasin, A. Grill, M. Waltl, T. Knobloch, B. Stampfer, T. Chiarella, N. Horiguchi, L. A. Ragnarsson, D. Linten, B. Kaczer, and T. Grasser, "Efficient physical defect model applied to PBTI in high-κ stacks," *IEEE Int. Reliab. Phys. Symp. Proc.*, p. XT11.1-XT11.6, 2017.

[17] S. Rangan, N. Mielke, and E. C. C. Yeh, "Universal recovery behavior of negative bias temperature instability [PMOSFETs]," *IEEE Int. Electron Devices Meet. 2003*, p. 14.3.1-14.3.4, 2003.

[18] T. L. Tewksbury and H. Lee, "Characterization, Modeling, and Minimization of Transient Threshold Voltage Shifts in MOSFET's," *IEEE J. Solid-State Circuits*, vol. 29, no. 3, pp. 239–252, 1994.

[19] K. Joshi, S. Hung, S. Mukhopadhyay, V. Chaudhary, N. Nanaware, B. Rajamohnan, T. Sato, M. Bevan, A. Wei, A. Noori, B. Mc.Dougal, C. Ni, G. Saheli, C. Lazik, P. Liu, D. Chu, L. Date, S. Datta, A. Brand, J. Swenberg, and S. Mahapatra, "HKMG process impact on N, P BTI: Role of thermal IL scaling, IL/HK integration and post HK nitridation," *IEEE Int. Reliab. Phys. Symp. Proc.*, pp. 1–10, 2013.

[20] Radek Erban, S. Jonathan Chapman,Philip and K. Maini," A practical guide to stochastic simulations of reaction-diffusion processes," (arxiv.org/abs/0704.1908).

[21] B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, R. Degraeve, L. A. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, "Origin of NBTI variability in deeply scaled pFETs," *IEEE Int. Reliab. Phys. Symp. Proc.*, pp. 26–32, 2010.