Modeling Channel Length Scaling Impact on NBTI in RMG Si p-FinFETs

Narendra Parihar, Ravi Tiwari and Souvik Mahapatra
Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India
(*Phone: +91-222-572-0408, Fax: +91-222-572-3707, Email: souvik@ee.iitb.ac.in)

***Abstract*— Negative Bias Temperature Instability (NBTI) stress and recovery time kinetics from Replacement Metal Gate (RMG) High-K Metal Gate (HKMG) p-channel FinFETs are measured and modeled. The impact of channel length (L) scaling on shift in threshold voltage (ΔV_T), its power-law time exponent (*n*), Voltage Acceleration Factor (VAF) and Temperature (T) activation (E_A) is analyzed. TCAD and bandstructure calculations are utilized to explain the L dependence of experimental data.**

Keywords— NBTI, FinFET, RMG, trap generation, RD model, TCAD, bandstructure calculation, strain.

I. INTRODUCTION

Negative Bias Temperature Instability (NBTI) remains an important reliability issue in modern p-channel MOSFETs [1]. NBTI causes a gradual buildup of positive charges in the gate insulator and shifts various device parameters such as threshold voltage (ΔV_T), subthreshold slope (ΔSS), trans-conductance (Δg_m), drain current (ΔI_D) etc. over time. Modeling of NBTI time kinetics during and after DC and AC stress is needed to extrapolate measured ΔV_T at short time stress to End of Life (EOL) at use condition. Although NBTI mechanism is debated [2], [3], a framework [4] has achieved the same across different technologies [4]-[8], using uncorrelated contributions from the interface (ΔV_{IT}) and bulk (ΔV_{OT}) oxide trap generation as well as hole trapping (ΔV_{HT}) in preexisting oxide traps. The model framework of [4] is shown in Fig.1. In this, the generation and passivation of traps (ΔN_{IT}) at Si/IL and IL/High-K interfaces of HKMG stack are calculated using the Reaction-Diffusion (RD) model. Transient Trap Occupancy Model (TTOM) calculates their occupancy and contribution (ΔV_{IT}). Empirical models are used for ΔV_{HT} and ΔV_{OT}. Analyses of experimental data by this framework across technologies [4]-[8] have established that the ΔV_{IT} subcomponent dominates EOL degradation at operating conditions. The ΔV_{IT} generation is governed by the breaking of bonds at the channel/IL interface, Fig.2, where inversion layer holes tunnel to and get captured in Si-H bonds and make them weak, the bonds subsequently get broken by thermal process [9]. The ΔV_{IT} magnitude depends on the gate stack Nitridation (N%), channel Germanium percentage (Ge%), and other FEOL processes. The N% in the gate stack and Ge% in the channel influence the tunneling effective mass (m_T), tunneling barrier (φ_B); and hence impacts the generation of ΔV_{IT}, as shown in Fig.2. The impacts of gate stack N% and Ge% in the channel have been studied in [4]-[8]. Layout related mechanical stress impact in FDSOI is analyzed in [7]. In this paper, the impact of channel length (L) scaling on NBTI is studied and modeled in RMG based HKMG Si FinFETs by the same NBTI modeling framework described in [4]. A companion paper [10] models the L dependence of RMG SiGe p-FinFETs. To the best of our knowledge, no modeling is shown till date that quantifies the L dependence of NBTI magnitude and related parameters.

In this work, measured data at different temperature (T) and stress bias (V_{GSTR}) are predicted by the comprehensive model framework of [4] and decomposed into underlying ΔV_{IT}, ΔV_{HT}, and ΔV_{OT} components, each having their unique time, bias and T dependencies. The time evolution of ΔV_T kinetics during and after stress, and the V_{GSTR} and T dependencies of fixed time ΔV_T are modeled for different L. The calibrated model is used for EOL projection of ΔV_T and its components for different L. It is shown that ΔV_T reduces at smaller L, and ΔV_{IT} dominates ΔV_T for all L. Sentaurus Process [11] is used to simulate the mechanical Source/Drain (S/D) stress at different L, while tight binding bandstructure calculations [12] link mechanical stress to ΔN_{IT} by using the model of Fig.2 [9]. The RD framework is incorporated into Sentaurus Device [13], [14] to predict measured ΔN_{IT} time kinetics (from ΔV_T) at different V_{GSTR}, T and L.

II. EXPERIMENTAL

RMG HKMG p-FinFETs from a commercial technology at 14nm node is used. Time evolution of ΔV_T is measured using ultrafast (UF) method with 10μs delay. The measurements are performed at different V_{GSTR} and T and for different channel lengths (L=160nm, 80nm, 20nm). Sentaurus TCAD is used for mechanical stress and NBTI kinetics calculations. The impact of strain on band structure is calculated using [12].

III. MODELING OF STRESS AND RECOVERY

Fig.3 and Fig.4, respectively, show the modeling of time kinetics during and after DC stress. The model subcomponents and overall prediction for a fixed V_{GSTR} and T (Fig.3 (a) and Fig.4 (a)), and model prediction for multiple V_{GSTR} and T (Fig. 3 (b) and Fig.4 (b)) are illustrated. ΔV_{IT} shows power law time dependence with time exponent of ~1/6 for longer-time stress. During recovery, a fraction of ΔV_{IT} recovers fast by capturing electrons from the substrate and the remaining ΔV_{IT} recovers slowly by trap re-passivation [4]. ΔV_{HT} saturates at long stress time and recovers fast. ΔV_{OT} has power law time dependence for stress with exponent ~1/3 and recovers slowly. The model prediction of time kinetics during and after Mode-B AC stress is shown in Fig.5 and Fig.6, respectively. The subcomponents and overall prediction for a fixed V_{GSTR} and T (Fig.5 (a) and Fig.6 (a)), and prediction for multiple V_{GSTR} and T (Fig.5 (b) and Fig.6 (b)) are shown. ΔV_{HT} is negligible; ΔV_{IT} and ΔV_{OT}

show identical time kinetics as DC for Mode-B AC stress. The recovery after AC stress is delayed since only the slower ΔV_{IT} (as electron capture happens in the last half cycle before onset of measurement) and slower ΔV_{OT} components contribute [3]. Fig.7 shows the model prediction of fixed time Mode-B AC degradation versus duty cycle (Fig.7 (a)) and frequency (Fig.7 (b)); frequency independence is observed.

Such analyses are done for all L. Fig.8 shows the model prediction of the time kinetics during and after DC stress for different L at a fixed V_{GSTR} and T. The degradation magnitude reduces while the time exponent (n) increases with reduction in L. Fig.9 through Fig.11 show the model prediction of V_{GSTR} dependence of fixed time ΔV_T for different L. Subcomponents and overall prediction at a fixed T (left), and overall prediction at different T (right) are shown. The VAF of ΔV_{IT} increases at lower L; the VAFs of ΔV_{HT} and ΔV_{OT} are independent of L. Fig.12 through Fig.14 show the prediction of T dependence of fixed time ΔV_T for different L. Subcomponents and overall prediction at a fixed V_{GSTR} (left), as well as overall prediction at different V_{GSTR} (right) are shown. The T activation of all the subcomponents is independent of L. The overall ΔV_T reduces, while its VAF (Fig.9-Fig.11) and E_A (Fig.12-Fig.14) increase as L is reduced. All subcomponents reduce at scaled L, due to slight thickening of IL and change of oxide field (E_{OX}); ΔV_{IT} reduction is larger than others due to mechanical stress impact as discussed later. The relative ΔV_{OT} contribution increases at lower L across V_{GSTR} and T, which is evident at shorter time stress (Fig.15) and EOL under use bias (Fig.16), although the overall ΔV_T is always dominated by ΔV_{IT}. This increases the VAF and E_A of overall ΔV_T, as these parameters are higher for the ΔV_{OT} subcomponent. VAF of ΔV_{IT} also increases with the reduction in L, as discussed later, and therefore contributes in increasing the VAF of overall ΔV_T. Higher fractional ΔV_{OT} is also responsible for higher n of overall ΔV_T at lower L, due to its higher n (~1/3) than that of ΔV_{IT} (n~1/6). ΔV_{IT} dominates at different V_{GSTR} and T, unless both V_{GSTR} and T are large when ΔV_{OT} also equally contributes (due to its high VAF and E_A). As expected for a matured process, ΔV_{HT} is low and has much less impact for different stress conditions and L.

IV. Physical Mechanism

Si/IL interfacial Si-H bond dissociation (ΔN_{IT} creation) is due to tunneling (m_T, φ_B) and capture of inversion holes into stretched (due to polarization) covalent bonds and subsequent T activation [9]. The reaction rate (k_F) depends on 4 parameters [4], [9]: pre-factor (k_{FIT}), T independent oxide field (E_{OX}) acceleration (Γ_0), polarization (α) and T activation (E_{AKF1}), Fig.2. The parameters α and E_{AKF1} are independent of L scaling. Both k_{FIT} and Γ_0 depend on m_T and φ_B and change with L due to mechanical stress in the channel. TCAD process simulation is used to quantify the impact of SiGe S/D epitaxial stressors on mechanical channel stress, and is shown in Fig.17. Stress profiles at the middle of the fin (surface orientation of <110>) show that the component along the channel direction (Stress-ZZ) dominates, which is compressive in nature. Due to higher stress near the S/D edges, the compressive stress at the middle of the fin increases significantly at lower L; see Fig.18. Bandstructure calculations for <110> surface that corresponds to fin sidewall orientation suggest an increase in m_T along the perpendicular (tunneling) direction due to increase in uniaxial compressive stress (UCS) as L is reduced, with no significant impact on φ_B, as shown in Fig.19. Note that this is a feature of the <110> surface and would impact FinFETs. Planar devices having <100> surface would be impacted differently [10]. The m_T values obtained in this work are consistent with the values reported elsewhere [15]. Fig.20 shows the impact of uniaxial compressive strain on the parameters k_{FIT} and Γ_0. Note, three orders change in k_{FIT} would result in an order change in ΔV_{IT} [4]; k_{FIT} reduces and Γ_0 increases with increase in strain (due to L reduction). Fig.21 compares the k_{FIT} and Γ_0 for different L as obtained using theoretical calculations and prediction of measured data (after component decomposition). Consistency of theoretical and experimental values is observed. Lower k_{FIT} results in relatively higher reduction in ΔV_{IT} component than the others as L is scaled.

The RD model is incorporated in Sentaurus Device. Fig.22 through Fig.24 compare the TCAD prediction and measured stress time kinetics (left) and the V_{GSTR} dependence (right) of ΔN_{IT} for different L; ΔN_{IT} is extracted from model prediction of measured data as discussed above. TCAD can accurately predict the time kinetics and T dependence of VAF at different L with only 2 process dependent parameters (k_{FIT} and Γ_0) that are obtained using strain and bandstructure calculations.

V. Implication

Although this work uses a large variation in L to highlight the effect, the impact caused by IL scaling would be absent for changes in L relevant for scaling of core devices. Increase in stress with L scaling is very drastic at L of interest. Resultant impact on ΔV_{IT} would impact the effect of L scaling on EOL ΔV_T (since ΔV_{OT} at use bias is negligible). Some NBTI relief is expected due to mechanical stress as L is scaled for <110> surface dominated FinFETs. However, the impact of L scaling would be different in planar or GAA nano sheets with <100> dominated surface, which needs to be investigated.

VI. Conclusion

Overall ΔV_T and its underlying subcomponents reduce at smaller L. Increase in mechanical strain at smaller L reduces the pre-factor but increases VAF of ΔV_{IT} by increasing the out of <110> plane tunneling m_T, which controls the Si-H bond dissociation at the fin sidewalls. A slightly higher IL thickness at smaller L reduces E_{OX} and all underlying subcomponents. The relative contribution of ΔV_{IT} reduces, while that of ΔV_{OT} increases and ΔV_{HT} (negligible) remains constant at lower L, although ΔV_{IT} dominates overall ΔV_T at low (~ use) V_{GSTR} for short time stress and at EOL for all L. Higher relative ΔV_{OT} (having higher n, VAF and E_A) increases n, VAF and E_A of overall ΔV_T as L is reduced. Higher VAF of ΔV_{IT} at lower L also contributes in increasing the VAF of overall ΔV_T.

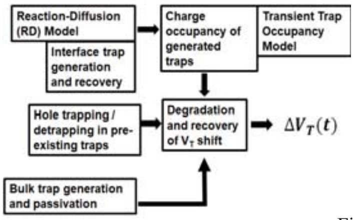

Fig.1. Schematic of a comprehensive NBTI modeling framework consisting of uncorrelated contributions from interface trap generation (ΔV_{IT}), hole trapping (ΔV_{HT}), and bulk trap generation (ΔV_{OT}).

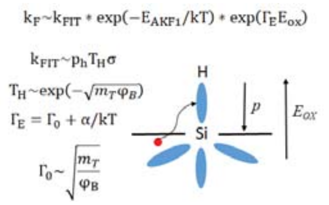

Fig.2. Schematic of Si-H bond dissociation at the channel/IL interface. Inversion layer holes (density p_H) tunnel into (mass m_T, barrier φ_B) interfacial Si-H bonds aided by the oxide electric field (E_{OX}). The bond is broken by thermal excitation with activation E_{AKF1}.

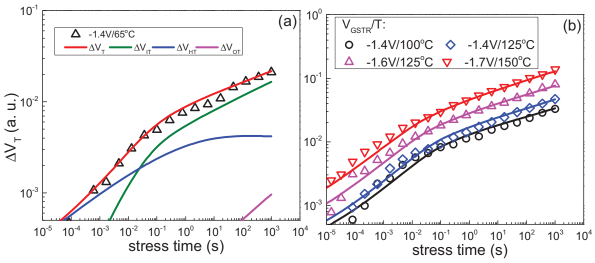

Fig.3. Time evolution of UF measured ΔV_T for DC stress (a) with model prediction and different subcomponents at fixed V_{GSTR} and T, and (b) overall prediction at multiple V_{GSTR} and T.

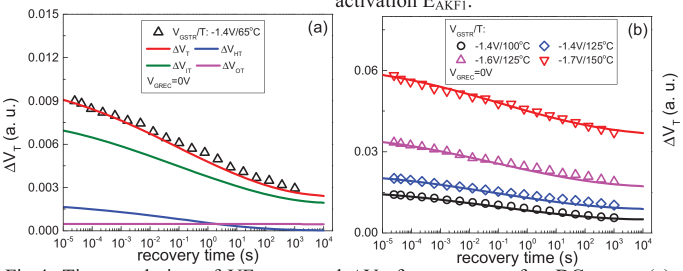

Fig.4. Time evolution of UF measured ΔV_T for recovery after DC stress (a) with model prediction and different subcomponents at fixed V_{GSTR} and T, and (b) overall prediction at multiple V_{GSTR} and T.

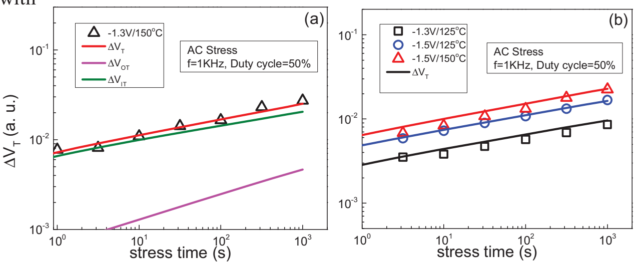

Fig.5. Time evolution of UF measured ΔV_T for AC stress (a) with model prediction and different subcomponents at fixed V_{GSTR} and T, and (b) overall prediction at multiple V_{GSTR} and T.

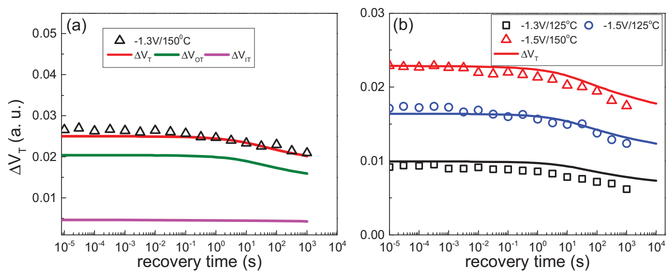

Fig.6. Time evolution of UF measured ΔV_T for recovery after AC stress (a) with model prediction and different subcomponents at fixed V_{GSTR} and T, and (b) overall prediction at multiple V_{GSTR} and T.

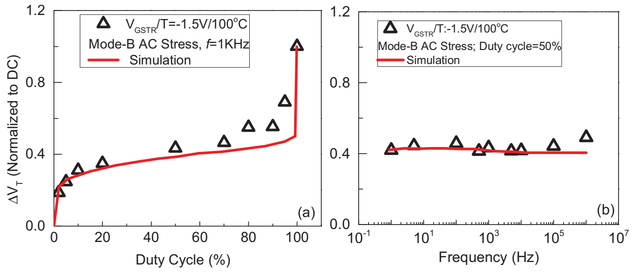

Fig.7. Model prediction of fixed time Mode-B AC degradation as a function of (a) duty cycle (for a fixed frequency) and (b) frequency (for a fixed duty cycle).

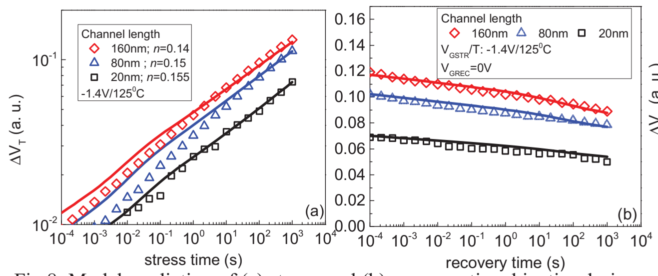

Fig.8. Model prediction of (a) stress, and (b) recovery time kinetics during DC stress for different channel lengths at a fixed V_{GSTR} and T.

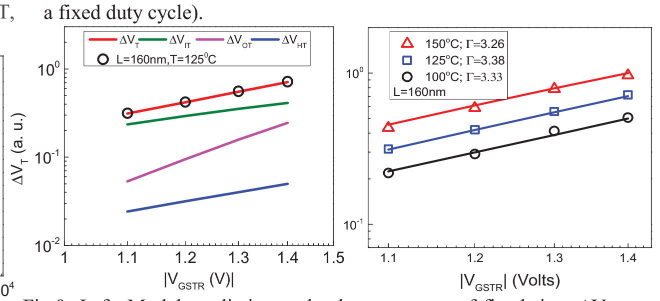

Fig.9. Left: Model prediction and subcomponents of fixed time ΔV_T as a function of V_{GSTR} for a fixed T. Right: Overall model prediction of fixed time ΔV_T as a function of V_{GSTR} for different T. Channel length of 160nm is used.

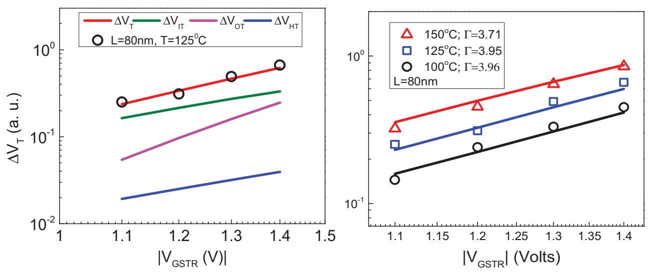

Fig.10. Left: Model prediction and subcomponents of fixed time ΔV_T as a function of V_{GSTR} for a fixed T. Right: Overall model prediction of fixed time ΔV_T as a function of V_{GSTR} for different T. Channel length of 80nm is used.

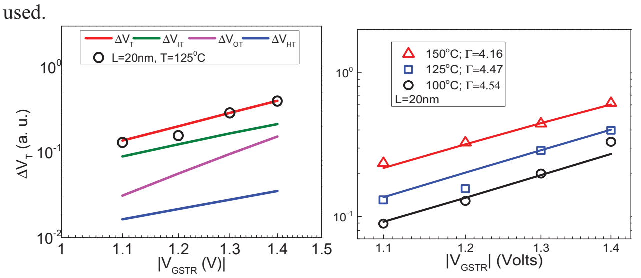

Fig.11. Left: Model prediction and subcomponents of fixed time ΔV_T as a function of V_{GSTR} for a fixed T. Right: Overall model prediction of fixed time ΔV_T as a function of V_{GSTR} for different T. Channel length of 20nm is used.

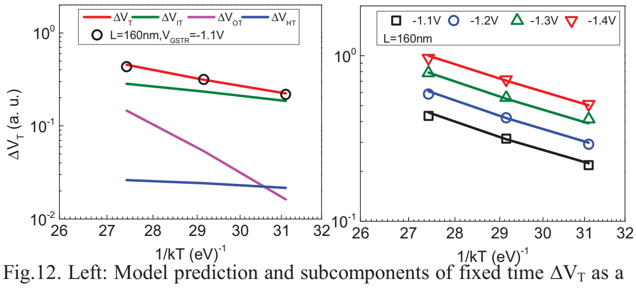

Fig.12. Left: Model prediction and subcomponents of fixed time ΔV_T as a function of T for a fixed V_{GSTR}. Right: Overall model prediction of fixed time ΔV_T as a function of T for different V_{GSTR}. Channel length of 160nm is used.

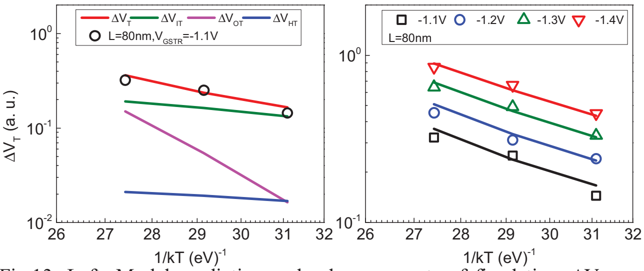

Fig.13. Left: Model prediction and subcomponents of fixed time ΔV_T as a function of T for a fixed V_{GSTR}. Right: Overall model prediction of fixed time ΔV_T as a function of T for different V_{GSTR}. Channel length of 80nm is used.

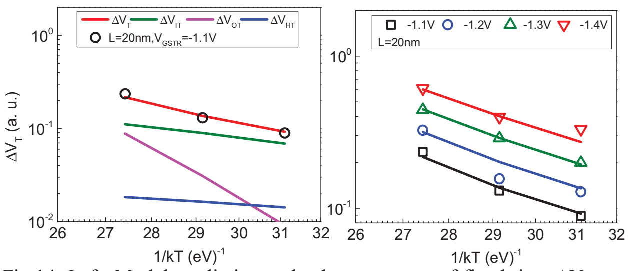

Fig.14. Left: Model prediction and subcomponents of fixed time ΔV_T as a function of T for a fixed V_{GSTR}. Right: Overall model prediction of fixed time ΔV_T as a function of T for different V_{GSTR}. Channel length of 20nm is used.

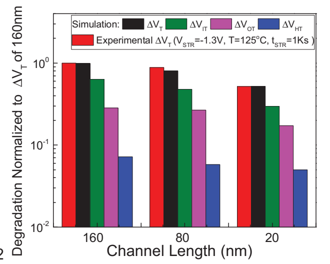

Fig.15. Comparison of model subcomponents at fixed time for different channel length (for fixed V_{STR}/T).

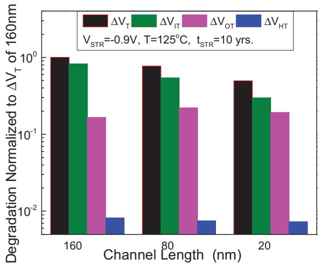

Fig.16. 10 yrs. model projection with subcomponents for different channel length. For shorter channel, ΔV_{IT} decreases; whereas ΔV_{HT} and ΔV_{OT} remain similar.

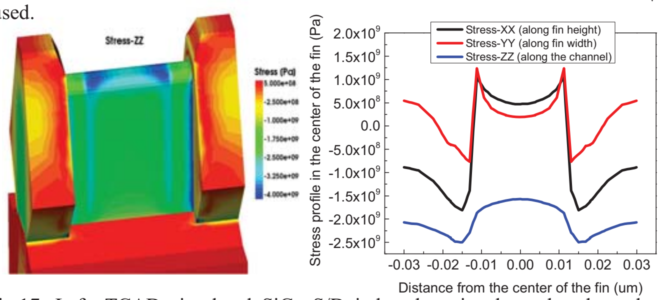

Fig.17. Left: TCAD simulated SiGe S/D induced strain along the channel length (S_{ZZ}). Right: Stress profile from source to drain in different directions of fin. Only the stress component along the channel length dominates.

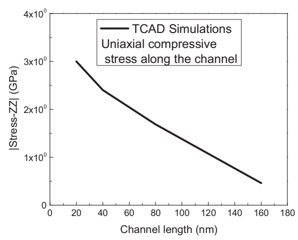

Fig.18. Channel length dependence of S_{ZZ} (along the channel) stress component.

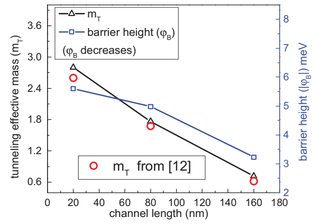

Fig.19. channel length dependence of tunneling effective mass (m_T) and barrier height (φ_B).

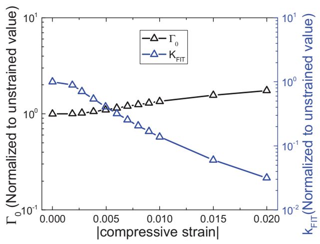

Fig.20. Mechanical Strain impact on the tunneling parameter (Γ_0) and prefacer (k_{FIT}). Data normalized to unstrained value.

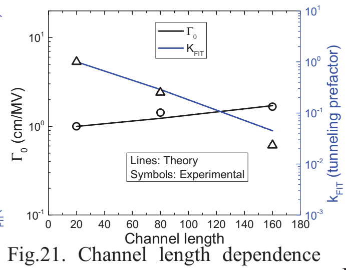

Fig.21. Channel length dependence of Γ_0 and 'k_{FIT}' obtained from both theoretical calculations and experiments. The reduction in 'k_{FIT}' is responsible for reduction of ΔV_{IT}.

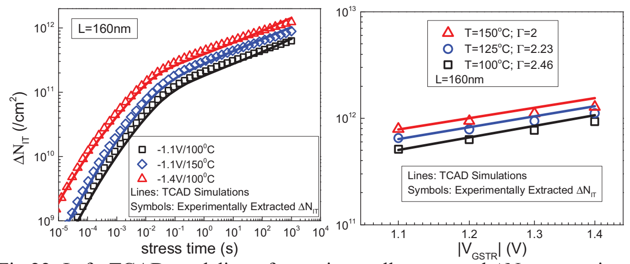

Fig.22. Left: TCAD modeling of experimentally extracted ΔN_{IT} stress time kinetics for different V_{GSTR} and T conditions. Right: TCAD prediction of fixed time experimentally extracted ΔN_{IT} as a function of V_{GSTR} for different T. Channel length of 160nm is used.

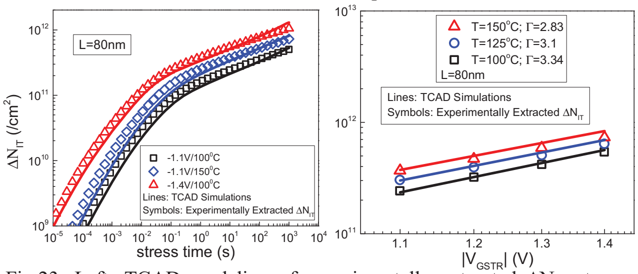

Fig.23. Left: TCAD modeling of experimentally extracted ΔN_{IT} stress time kinetics for different V_{GSTR} and T conditions. Right: TCAD prediction of fixed time experimentally extracted ΔN_{IT} as a function of V_{GSTR} for different T. Channel length of 80nm is used.

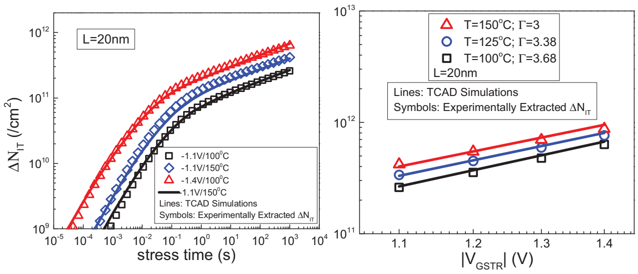

Fig.24. Left: TCAD modeling of experimentally extracted ΔN_{IT} stress time kinetics for different V_{GSTR} and T conditions. Right: TCAD prediction of fixed time experimentally extracted ΔN_{IT} as a function of V_{GSTR} for different T. Channel length of 20nm is used.

REFERENCES

[1] S. Mahapatra, N. Goel, S. Desai, S. Gupta, B. Jose, S. Mukhopadhyay, K. Joshi, A. Jain, A. E. Islam and M. A. Alam "A comparative study of different physicsbased NBTI models," *IEEE Trans. Electron Devices*, vol. 60, no. 3, pp. 901–916, Mar. 2013.

[2] J. H. Stathis, Souvik Mahapatra, Tibor Grasser, "Controversial issues in negative bias temperature instability", *Microelectronics Reliability*, Volume 81, 2018, Pages 244-251.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] A. E. Islam, H. Kufluoglu, D. Varghese, S. Mahapatra, and M. A. Alam, "Recent issues in negative-bias temperature instability: Initial degradation, field dependence of interface trap generation, hole trapping effects, and relaxation," in *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2143–2154, Sep. 2007.

[10] N. Parihar et. al.,**"**Modeling of Process (Ge, N) Dependence and Mechanical Strain Impact on NBTI in RMG HKMG SiGe FDSOI p-MOSFETs and p-FinFETs," Accepted in SISPAD 2018.

[11] Sentaurus^{TM} Process user guide, N-2017.09.

[12] https://nanohub.org/resources/bandstrlab.

[13] Sentaurus^{TM} Device user guide, N-2017.09.

[14] S. Mishra *et al*., "TCAD-Based Predictive NBTI Framework for Sub-20-nm Node Device Design Considerations," *IEEE Trans. Electron Devices*, vol. 63, no. 12, pp. 4624-4631, Dec. 2016.

[15] Guangyu Sun, "Strain effects on hole mobility of Silicon and Germanium P-type metal oxide semiconductor field effect transistors," Ph.D. Dissertation, University of Florida, 2007.