Resolution of Disputes Concerning the Physical Mechanism and DC/AC Stress/Recovery Modeling of Negative Bias Temperature Instability (NBTI) in p-MOSFETs

Narendra Parihar, Uma Sharma, Subhadeep Mukhopadhyay, Nilesh Goel, Ankush Chaudhary, Rakesh Rao, 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) is due to interface trap generation (ΔN_{IT}) and trapping of holes in gate insulator traps (ΔN_{HT}). However, the isolation methods and the relative dominance of ΔN_{IT} and ΔN_{HT}, time constants of ΔN_{IT} and ΔN_{HT} for stress, recovery and associated temperature (T) activation, and whether ΔN_{IT} recovers or remains permanent after stress, are widely debated. The resolution of such disputes is necessary to develop a reliable NBTI model. This work uses carefully designed measurements and simulations to resolve the aforementioned disputes. The contribution of ΔN_{IT} and ΔN_{HT} on overall threshold voltage shift (ΔV_T) is determined. Kinetics of ΔN_{IT} and ΔN_{HT} during stress and recovery, T activation and associated time constants are verified in both large and small area devices. Existing theoretical models for ΔN_{IT} and ΔN_{HT} are benchmarked and validated against DC and AC experiments. Capability of the existing models for predicting end-of-life ΔV_T is demonstrated.**

***Index Terms*—NBTI, trap generation, hole trapping, Reaction-Diffusion (RD) model, Gate Sided Hydrogen Release (GSHR) model, extended Non-radiative Multi-Phonon (e-NMP) model.**

I. INTRODUCTION

Negative Bias Temperature Instability (NBTI), resulting in positive charge buildup in the gate insulator, is a well-known reliability issue for Silicon Oxynitride (SiON) [1] and High-K Metal Gate (HKMG) [2]-[4] p-MOSFETs. It is necessary to understand NBTI physical mechanism and develop a suitable model, a subject of intense debate [5], to reliably predict the experimental data during accelerated stress and estimate end-of-life (EOL) degradation under use condition for technology qualification.

In the past, NBTI threshold voltage shift (ΔV_T) has been attributed to only interface trap generation (ΔN_{IT}) [6], [7], to only hole trapping in pre-existing gate insulator traps (ΔN_{HT}) [8], [9], to a strongly coupled ΔN_{IT} and ΔN_{HT} mechanism [10] and to mutually uncoupled ΔN_{IT} and ΔN_{HT} processes [1], [5], [11]-[14]. The ΔN_{IT} only models cannot predict gate insulator process dependence [1], [3] and ultra-fast measured data [5]. The ΔN_{HT} only models are not consistent with experimental evidence of interface trap generation, directly estimated using Charge Pumping (CP) [1], [6], [13]-[16], Direct-Current I-V (DCIV) [5], [14], [17]-[21], subthreshold slope [22] and Low Voltage Stress Induced Leakage Current (LV-SILC) [7], [17] techniques. The coupled ΔN_{IT} and ΔN_{HT} mechanism (the two stage model) is not consistent with long-time measured stress data and cannot explain gate insulator process dependence, as discussed in [5], [23]. It is now well accepted that ΔV_T is due to mutually uncoupled contributions from ΔN_{IT} and ΔN_{HT}.

However, several aspects such as the extraction procedure of ΔN_{IT} and ΔN_{HT}, their relative dominance during stress and at EOL under use condition, time constants of ΔN_{IT} and ΔN_{HT} during stress and recovery and the associated temperature (T) activation, whether ΔN_{IT} recovers after the removal of stress or remains permanent, are debated, as listed in Table-I. As a result, the NBTI mechanism and model are also debated.

Table-I. NBTI is due to generation of interface traps (TG: ΔN_{IT}) and hole trapping in pre-existing traps (HT: ΔN_{HT}). Several aspects concerning the underlying NBTI physical mechanism are debated (see text).

Topic of debate	Model Set-A	Model Set-B
TG & HT extraction procedure	ΔV_T (from UF I-V) and ΔV_{IT} (from DCIV) measurements	P and R components of ΔV_T (from UF I-V) recovery
Dominance at long, relevant stress time	Trap generation (TG)	Hole trapping (HT)
Trap generation mechanism	Reaction-Diffusion (RD)	Gate Sided Hydrogen Release (GSHR)
Recovery of generated traps	Recoverable	Almost permanent
Trapping-detrapping time constants	Small	Large
T activation of trapping magnitude	Small	Large

As mentioned above, DCIV measurements were performed to directly estimate ΔN_{IT} contribution to overall ΔV_T; the later was measured by ultra-fast (UF, ~ µs delay) method [5], [19]-[21], [24], [25]. However, since DCIV is a slow technique (~ 10s delay) and scans only a fraction of the energy bandgap, corrections for delay and bandgap should be performed on as-measured DCIV data, so that ΔN_{IT} and ΔV_T can be compared to extract ΔN_{HT}. This framework (Set-A of Table-I) suggests ΔN_{IT} dominated ΔV_T, and can explain gate insulator process dependence across different device geometries [1], [5], [24]-[26]. It suggests lower time constants and lower T activation of ΔN_{HT} as compared to ΔN_{IT}. The recovery of ΔN_{IT} is also directly demonstrated using CP [6], [16], DCIV [18]-[20] and subthreshold slope [22] techniques. Note that ΔN_{IT} dominated ΔV_T is fully consistent with reports across multiple industries

978-1-5090-6641-4/17/$31.00 ©2017 IEEE

and technology nodes [27]. Therefore, the ΔN_{IT} only models [6], [7] can be modified to predict NBTI experiments.

Recently, a comprehensive modeling framework consisting of uncorrelated contributions from ΔN_{IT} and ΔN_{HT}, and bulk trap generation (ΔN_{OT}, for harsher stress conditions), has been proposed [24]-[26]. It calculates generation and passivation of ΔN_{IT} using the double interface H/H_2 Reaction-Diffusion (RD) model [5], and charge occupancy of generated traps and their contribution (ΔV_{IT}) by Transient Trap Occupancy Model (TTOM) [24], [25]. Density (ΔN_{HT}) and contribution (ΔV_{HT}) due to hole trapping and detrapping are calculated by using empirical expressions. Generated bulk trap density (ΔN_{OT}) is empirically calculated, and occupancy of, and contribution (ΔV_{OT}) from these traps are calculated using TTOM.

The model framework was validated against UF measured data from gate-first HKMG devices for diverse experimental conditions [24]-[26]. It can predict ΔV_T time evolution during and after DC and AC stress for different stress (V_{GSTR}) and recovery (V_{GREC}) bias, T, frequency (*f*) and pulse duty cycle (PDC), and mixed DC-AC stress under completely arbitrary time segments and for varying voltage, frequency and activity conditions. It can explain gate insulator process dependence, predict measured data from large area and mean of multiple small area devices, and can also estimate EOL degradation. Further details about Set-A (Table-I) approach and the model framework can be found in a recent book [28] and also online presentations [29]. To the best of our knowledge, we are not aware of any other model framework hitherto proposed, that can predict such wide variety of experimental results.

Although multiple measurement methods provide evidence of ΔN_{IT} generation and recovery as mentioned above, a single report [13] suggested no recovery of ΔN_{IT} after stress. In [13], CP method was used to measure ΔN_{IT} before, during and after stress. SiON devices were used and subjected to stress at high V_{GSTR} and T. The lack of ΔN_{IT} recovery is presumably due to measurement issues associated with long CP sweeps, and this aspect is explained in [30]. Moreover, the use of high V_{GSTR} and T results in relatively larger ΔN_{OT} contribution, and ΔN_{OT} shows negligible recovery [14]. Note that CP measurements simultaneously capture ΔN_{IT} and ΔN_{OT} (TDDB like generated bulk traps). Therefore, large ΔN_{OT} contribution would result in negligible recovery that is wrongly ascribed fully to ΔN_{IT}, when CP method is used under such strong stress conditions.

Nevertheless, presumed non-recoverable ΔN_{IT} was used to isolate ΔN_{IT} and ΔN_{HT} subcomponents. UF ΔV_T recovery was decomposed into recoverable (R) and permanent (P) parts, and attributed respectively to contributions (ΔV_{HT} and ΔV_{IT}) due to ΔN_{HT} and ΔN_{IT} (Set-B of Table-I) [31], [32]. In such a framework, ΔN_{HT} fully dominates ΔV_T, and therefore, ΔN_{HT} requires large time constants during stress and recovery and also requires large T activation. This is in direct conflict with the earlier approach (Set-A) and needs to be resolved.

Moreover, it was suggested that RD model is inconsistent with UF measured ΔV_T recovery [8], [33], [34]. As described in [24]-[26], RD model calculates trap density (ΔN_{IT}), but it is the occupancy of these traps (ΔV_{IT}), calculated using TTOM enabled RD model that actually contributes to overall ΔV_T for recovery. In addition, effects due to ΔV_{HT} (and possibly also ΔV_{OT}) should also be considered. Therefore, comparison of only RD model to ΔV_T recovery is basically irrelevant.

Nevertheless, such presumed incompatibility has led to the development of different energy-well (energy-state) based models [8], [9], [35], [36]. In particular, the extended Non-radiative Multi-Phonon (e-NMP) model for ΔN_{HT} (ΔV_{HT}) and the Gate Sided Hydrogen Release (GSHR) model for ΔN_{IT} (ΔV_{IT}) are of interest. It is important to note that the e-NMP model was used to predict ΔV_{HT} (hence ΔV_T) recovery and is never benchmarked against stress data. Similarly, the GSHR model was only tested against a single DC stress data and not benchmarked against diverse DC and AC experiments. These aspects will also be addressed in this paper.

In this paper, carefully designed experiments are used to address and resolve the disputes mentioned in Table-I. The R versus P decomposition of ΔV_T recovery into ΔV_{HT} and ΔV_{IT}, when evaluated for different stress conditions, are shown to be incompatible respectively with e-NMP and GSHR models. The correct bandgap correction scheme has been identified for DCIV based decomposition method. Experimental proofs are provided to show smaller time constants and T activation of ΔV_{HT}. Finally, the RD and GSHR models are benchmarked against DC and AC stress data for wide range of conditions, and the inconsistencies of the GSHR model are highlighted.

All experiments done in this paper are on gate first HKMG devices. Two different devices having different gate insulator processes A and B have been used. Process B has ~2X higher pre-existing hole trap density compared to process A, which is measured by flicker noise method [37]. Measurements are done by UF measure-stress-measure method with 10µs delay to determine ΔV_T. DCIV measurements are used to determine ΔN_{IT} and its contribution (ΔV_{IT}) to overall ΔV_T. Refer to [3], [24], [25] for further details on devices and measurements.

II. Isolation of Trap Generation and Trapping

A. Permanent (P) and Recoverable (R) method:

Fig.1 shows time evolution of UF measured ΔV_T recovery for different V_{GREC} and stress time (t_{STR}). Measured data fitted with empirical universal recovery expression [38], and P and R components are determined. Use of different V_{GREC} helps to determine the actual end-of-stress (start of recovery) value. This exercise was repeated for different V_{GSTR} and T, and for both process A and B devices.

Fig.2 shows the time evolution of P and R components measured during stress at different V_{GSTR} and T for process A and B devices. Note that the magnitude of R is greater than P component. The GSHR [35], [36] and e-NMP [9] models are used to calculate the time evolution of "P" (=ΔV_{IT}) and "R" (=ΔV_{HT}) respectively for stress, and shown in Fig.2 (as lines). Downloadable simulation codes, and all details regarding the equations, implementation and parameter sensitivity analysis can be found elsewhere for GSHR [39] and e-NMP [40]. It is possible to match the time evolution of measured P and R for a particular V_{GSTR} and T and for a given process, by adjusting respectively the GSHR and e-NMP model parameters. Note that GSHR model has 11 and e-NMP model has 20 adjustable parameters, and multiple parameters can be varied to obtain the desired fitting[1]. However, both GSHR and e-NMP model become incompatible with measured stress data when V_{GSTR}, T and process are varied, when all model parameters, except the pre-existing trap density N_0, are kept constant. Note, only N_0 is varied between processes A and B to remain consistent with flicker noise measurements. Also note that it is possible to predict the stress data at a particular V_{GSTR} and T for both processes, if all model parameters are varied between process A and B. However, even then, these models would still fail to predict stress data when V_{GSTR} and T are varied (results not shown here for brevity).

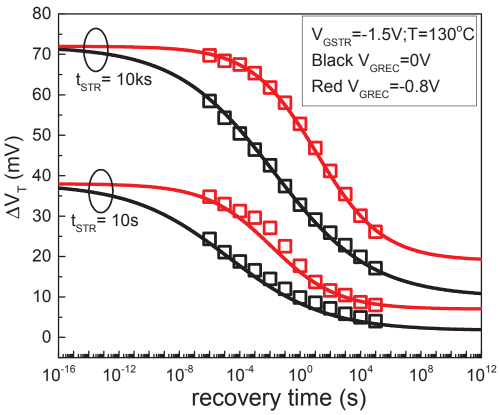

Fig.1. UF measured ΔV_T recovery after different stress time and empirical model fit to determine permanent (P) and recoverable (R) components. The P component is obtained from extrapolated long time saturated part.

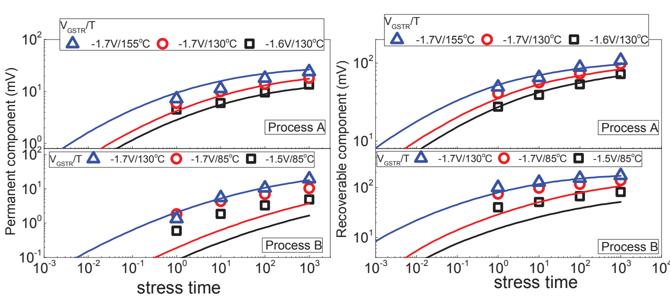

Fig.2. The time evolution of extracted permanent (P, Left) and recoverable (R, Right) components at different V_{GSTR} and T. GSHR and e-NMP models (lines) are respectively used to model P and R components. Unlike measured data, model shows saturation at higher V_{GSTR} and T.

This discrepancy is further illustrated in Fig.3. Measured long-time power law time slope (n) of P and R components is compared respectively to GSHR and e-NMP simulation, for changes in V_{GSTR}, T and process. It is evident that the models cannot predict measured data. Measured n for both P and R and for both processes show no particular trend as V_{GSTR} and T are varied. GSHR and e-NMP simulated n reduces slightly at higher V_{GSTR}, however, a drastic reduction is seen at higher T. Note that the large number of GSHR and e-NMP model parameters can be adjusted to predict experimental n for a particular V_{GSTR} and T for a particular process, but the model predictions fail for other stress conditions as shown. Except N_0, all parameters are held constant between process A and B in these simulations. It is possible to match n at a particular V_{GSTR} and T for both processes, if the model parameters for process A and B are independently adjusted. Even then, these models would be inconsistent with data as V_{GSTR} and T are varied (results are not shown here for brevity).

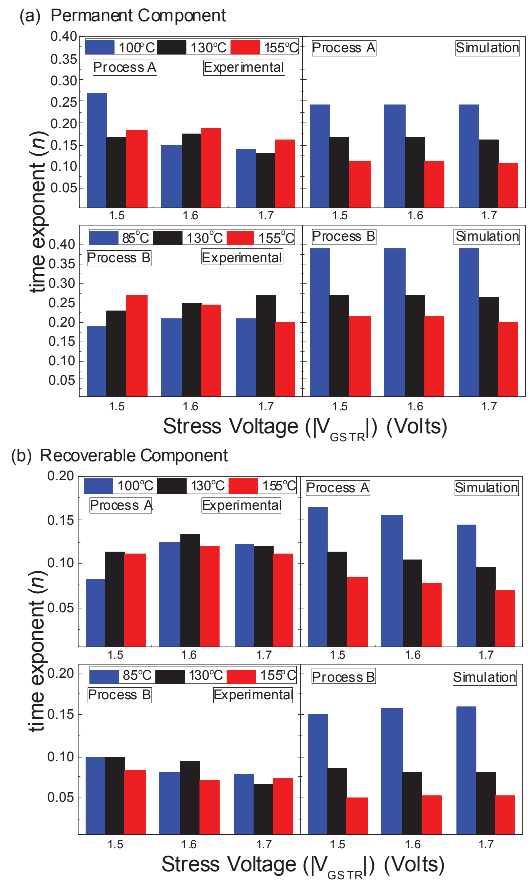

Fig.3. (Top 4 panels): Experimental (Left) and GSHR simulated (Right) time exponent of permanent (P) component for process A (Top) and process B (Bottom) for different V_{GSTR} and T. (Bottom 4 panels): Experimental (Left) and e-NMP simulated (Right) time exponent of recoverable (R) component for process A (Top) and process B (Bottom) for different V_{GSTR} and T. The simulated results are incompatible with experiments.

To even further illustrate the discrepancy, Fig.4 shows the measured T activation of P and R components at fixed stress time and the corresponding GSHR and e-NMP simulations. Both GSHR and e-NMP models cannot predict the measured T activation for P and R respectively, for various V_{GSTR} and processes as shown.

Hence, the P and R component based decomposition done for stress experiments are incompatible with physical models. This has not been realized till date, due to the overemphasis in modeling recovery without attempting to model stress [8],

[1] The complicated e-NMP model can be substituted by the much simpler NMP model having only 9 adjustable parameters [9]. However, both models produce very similar ΔV_{HT} kinetics during stress and recovery. Refer to [41] for details regarding the NMP model. Comparison of different hole trapping models will be shown elsewhere.

[9]. These discrepancies are not surprising, as the very basic premise of permanent ΔV_{IT} is never substantiated by several experiments. As mentioned, multiple experiments suggest the recovery of ΔN_{IT} (trap density) after stress. Furthermore, as explained in detail in [24]-[26], recovery of ΔV_{IT} (electrically active traps that contribute to ΔV_T) is not only due to ΔN_{IT} recovery but also due to fast electron capture in generated interface traps. Therefore, the P versus R decomposition and their respective assignments to ΔV_{IT} and ΔV_{HT} (set-B, Table-I) cannot be physically justified.

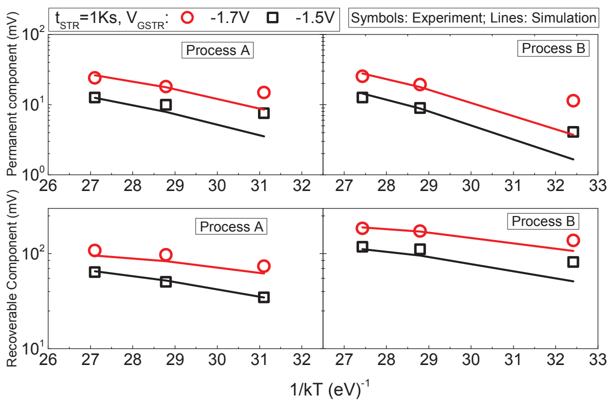

Fig.4. Temperature dependence of permanent (P, Top) and recoverable (R, Bottom) components for process A (Left) and process B (Right). The GSHR model simulations for P and e-NMP model simulations for R are shown as lines. The simulated results are incompatible with experiments.

B. DCIV based method:

Fig.5 shows the DCIV measured time evolution of ΔN_{IT} for (a) stress and (b) recovery. Stress measurements are done for two different delays, and show lower magnitude and higher n at higher delay due to recovery. Note, DCIV is a slow method and has a default delay of ~ 10s, which is sufficient to cause recovery of ΔN_{IT}. Time evolution of recovery is measured for different stress time and universal behavior is observed when plotted versus normalized time. Empirical universal recovery expression [38] is used to correct the as-measured stress data as shown in Fig.5 (a). Note that measured data using different delay show similar magnitude and n (~0.16) after correction. Fig.5 also shows as measured and corrected n for different (c) V_{GSTR} and (d) stress T. The universality of power-law time exponent n (~0.16) after such delay correction is observed for different devices and experimental conditions; refer to [19], [20], [42] for further details.

DCIV method scans traps that are physically located at and near the channel-gate insulator interface, in a limited energy zone centered around the middle of Si energy bandgap [21]. On the other hand, V_T measurements scan the entire Si energy bandgap. Hence, delay corrected ΔN_{IT} must also be corrected for bandgap, before comparing its contribution (ΔV_{IT}) against ΔV_T. Fig.6 plots the time evolution of UF measured ΔV_T and DCIV measured ΔV_{IT} after delay and bandgap corrections, and extracted ΔV_{HT} (=ΔV_T–ΔV_{IT}). The bandgap correction is performed assuming different energy-zone scenarios, for trap generation in (a) narrow zone only in the bottom half, (b)

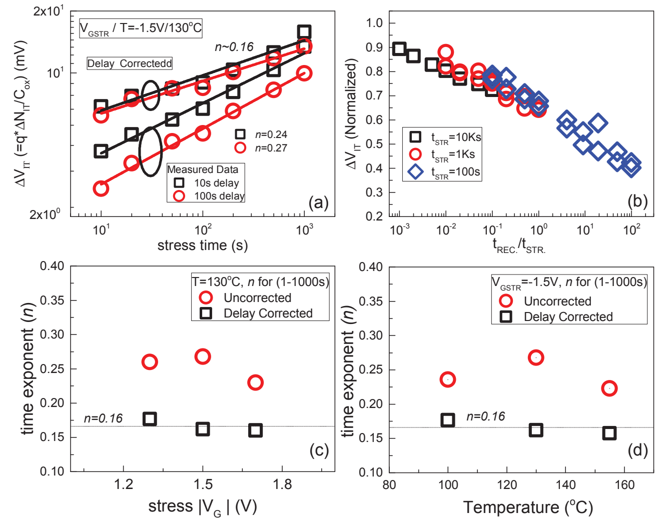

Fig.5 (a) Time evolution of ΔN_{IT} from DCIV measurements at two different delays, before and after delay correction, (b) time evolution of ΔN_{IT} recovery from DCIV measurements for different stress time. Also shown long-time power-law time exponents of measured ΔN_{IT} time evolution data during stress, before and after delay correction, as a function of (c) stress V_G and (d) stress T.

narrow zone in top and bottom halves, (c) wide zone only in the bottom half, and (d) wide zone covering both halves, of the Si bandgap. The magnitude of ΔV_{IT} and the magnitude and time exponent n of ΔV_{HT} get impacted by the correction scheme. Schemes A through C exhibit the domination of non-saturating ΔV_{HT}, but D shows the domination of ΔV_{IT} and saturated ΔV_{HT}. Note that correction scheme D was used in our earlier works [24], [25].

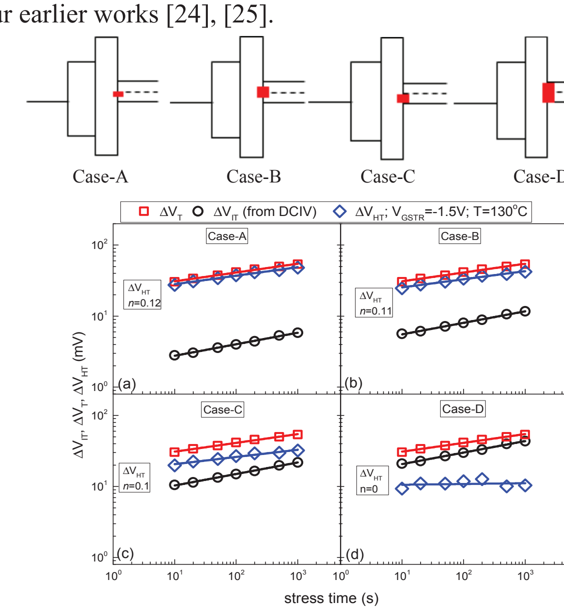

Fig.6 Time evolution of ultra-fast measured ΔV_T, DCIV measured ΔV_{IT} after delay and bandgap corrections and extracted ΔV_{HT}, for different bandgap correction cases: (a) narrow spread of donor like ΔN_{IT} in only bottom half, (b) narrow spread of donor like ΔN_{IT} in top and bottom halves, (c) broad spread of donor like ΔN_{IT} in only bottom half, and (d) broad spread of donor like ΔN_{IT} in top and bottom halves, of the Si bandgap. The relevant energy zones are shown using a patch in the energy band diagrams. The distribution of interface traps is assumed to be uniform in energy for a particular spread. The kinetics of extracted ΔV_{HT} depends on the bandgap correction scheme.

To choose the correct bandgap correction scheme, the time evolution of measured ΔV_{HT} is determined for process A and B devices at different V_{GSTR} and T, and shown in Fig.7. ΔV_{HT} extraction is done using the method shown in Fig.6, and all possible bandgap corrections discussed in cases A through D (of Fig.6) has been used. The lines in Fig.7 are from e-NMP model simulations for ΔV_{HT}. The e-NMP model parameters are calibrated to predict measured data at a particular V_{GSTR} and T, and the model is then used to predict data at different V_{GSTR} and T. Note that e-NMP model is not consistent for the corrections schemes A through C, which show non-saturation of ΔV_{HT}. The model can predict data for a given V_{GSTR} and T, but cannot predict across different V_{GSTR} and T, which is true for both processes. However, the model is consistent with the correction scheme D that shows saturated ΔV_{HT} for different V_{GSTR} and T and for different gate insulator processes.

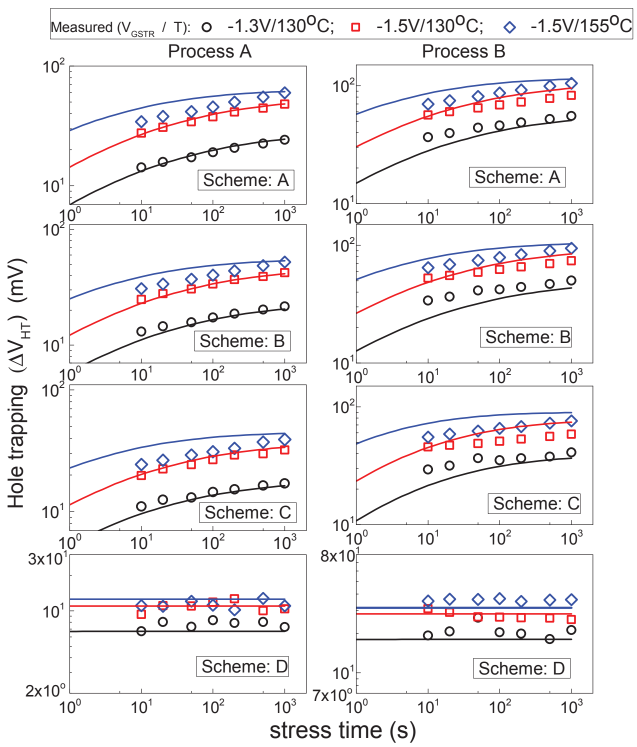

Fig.7 Time evolution of extracted ΔV_{HT} for process A and B devices, by using ultra-fast measured ΔV_T and DCIV measured and corrected ΔV_{IT} for different schemes shown in Fig.6. The extraction is done at different V_{GSTR} and T. Lines are ΔV_{HT} simulated using the e-NMP model. The e-NMP model is consistent with extracted ΔV_{HT} from scheme D.

To illustrate further, Fig.8 compares the long-time power-law time exponent *n*, extracted from measured and simulated time evolution of ΔV_{HT}, corresponding to bandgap correction cases A through C. Results are shown for both process A and B devices and for different V_{GSTR} and T. Note, simulated *n* is grossly incompatible with measured data for different devices and experimental conditions, when case A through C bandgap correction schemes are used. Measured *n* shows no trend with V_{GSTR} and T, however, e-NMP simulated *n* reduces slightly at higher V_{GSTR}, while a drastic reduction is seen for higher T. Hence, non-saturated ΔV_{HT} with large trapping time constant (Set-B, Table-I) is not consistent with e-NMP calculations. Saturated ΔV_{HT} at long stress time and hence smaller trapping time constant (Set-A, Table-I) is consistent with the e-NMP model, and with earlier reports [1], [5], [11]-[14], [21], [24]-[27], and this is also discussed in detail in [28], [29].

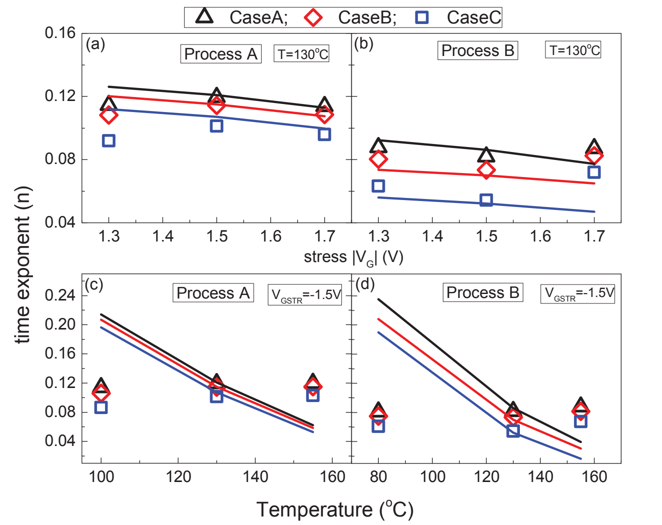

Fig.8 Long-time power-law time exponents of extracted ΔV_{HT} time evolution data during stress, for (a, c) process A and (b, d) process B devices, versus (a, b) stress V_G and (c, d) stress T. Data form bandgap correction schemes A through C (see Fig.6) are shown. Lines are e-NMP model simulations. The model calculations are not consistent with experimental data.

Fig.9 shows the T activation of measured ΔV_{HT} at different V_{GSTR} for process A and B, obtained from bandgap correction schemes A through D. Schemes A through C show higher T activation energy (E_A) of measured ΔV_{HT} for process A when compared to process B, while scheme D shows similar E_A for both processes. Moreover, obtained E_A is higher for schemes A through C compared to scheme D. The corresponding e-NMP model simulations are also shown. Simulated results are

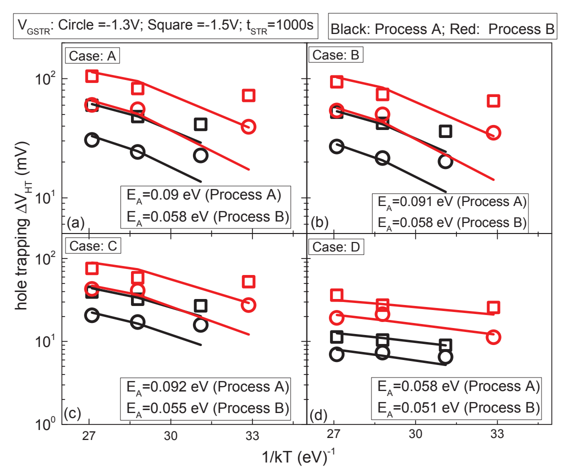

Fig.9 Temperature dependence of extracted ΔV_{HT} at different V_{GSTR} but fixed stress time, for process A and B devices, for all different extraction schemes A through D (see Fig.6). Lines are e-NMP model simulations. Measured E_A values are shown. The e-NMP simulations are consistent with data obtained only using scheme D.

grossly incompatible with the measured data for all schemes except scheme D. Note, the e-NMP model show higher E_A for process B compared to A for correction schemes A through C, contrary to measured data. Therefore it is important to note that high E_A for ΔV_{HT} (Set-B, Table-I) is not consistent with e-NMP model. However, e-NMP model can predict measured E_A (lower value) for both processes for scheme D.

Low E_A of ΔV_{HT} (Set-A, Table-I) obtained from scheme D is consistent with e-NMP model, and with earlier reports [1], [5], [12]-[14], [24]-[26], and is also discussed in [28], [29]. Moreover, scheme D isolation results in ΔV_{IT} dominated ΔV_T and higher E_A for ΔV_{IT} during stress; refer to [5], [24]-[26] for additional discussions.

C. *Comparison of different methods:*

Fig.10 (a) shows the time evolution of UF measured ΔV_T and prediction by an empirical model [43]. The model uses power-law time dependence of ΔV_{IT} (=A_1. t^n) with exponent n~0.16 (consistent with only delay corrected DCIV data) and a saturated ΔV_{HT} for long stress time, which is consistent with e-NMP simulations and the comprehensive framework [24], [25]. The ΔV_{IT} and ΔV_{HT} subcomponents are also shown in Fig.10 (a). The delay and band gap corrected DCIV measured data (scheme D, Fig.6) can be expressed as $\Delta V_{IT} = A_2$. t^n with n~0.16. Fig.10 also illustrates the comparison of A_1 (from UF ΔV_T decomposition) and A_2 (from delay and energy bandgap corrected DCIV data) for experiments done on process A and B devices, for different (b) V_{GSTR} and (c) stress T. Note that A_1 is very similar to A_2 for different V_{GSTR} and T and for both processes. This verifies the bandgap correction scheme D and suggests that generated interface traps have broad distribution in the Si energy band gap, which is consistent with [17].

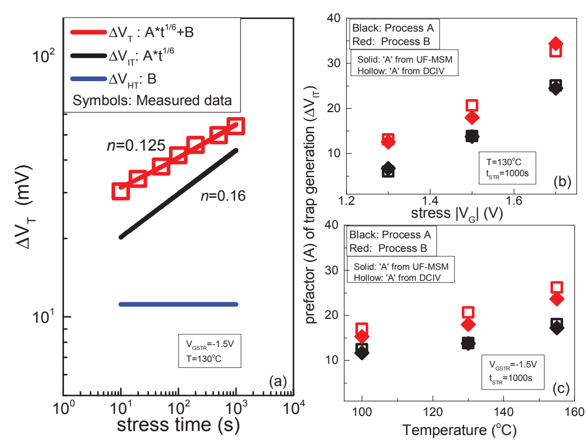

Fig.10 (a) Time evolution of ultra-fast measured ΔV_T and empirical model fitting using ΔV_{IT} (=A_1. t^n) and ΔV_{HT} (=B). Also shown comparison of A_1 from this calculation and A_2 obtained by fitting ΔV_{IT} time evolution (=A_2. t^n) from measured DCIV data after delay and bandgap correction using scheme D (see Fig.6), for process A and B devices, for different (b) stress V_G and (c) stress T. DCIV correction scheme D is consistent with empirical model.

This clearly justifies the delay and bandgap correction (scheme D) used in previous reports [5], [24], [25]. Therefore, the DCIV based isolation technique is verified, and this results in ΔV_{IT} dominated ΔV_T, saturated ΔV_{HT} and low T activation E_A for ΔV_{HT} (Set-A, Table-I). Moreover, it is of no surprise that the related model framework [24]-[26] can explain diverse set of experiments as mentioned in the introduction.

Finally, as interface traps recover after stress, as shown in Fig.5 (b) and multiple other reports, the usual assignment of P and R components respectively to ΔV_{IT} and ΔV_{HT} cannot be justified. To illustrate, Fig.11 shows the comparison of ΔV_{IT} extracted from P and R decomposition and using DCIV data, after corrections for delay and bandgap (scheme-D, Fig.6), for variations in (a) stress time, (b) V_{GSTR} and (c) stress T. It is important to note that P assignment underestimates magnitude of ΔV_{IT} and overestimates its n and E_A compared to corrected DCIV data. Therefore, the GSHR model developed based on the presumably P component also faces significant challenges to predict experimental trap generation data, and this will be addressed in detail later in this paper.

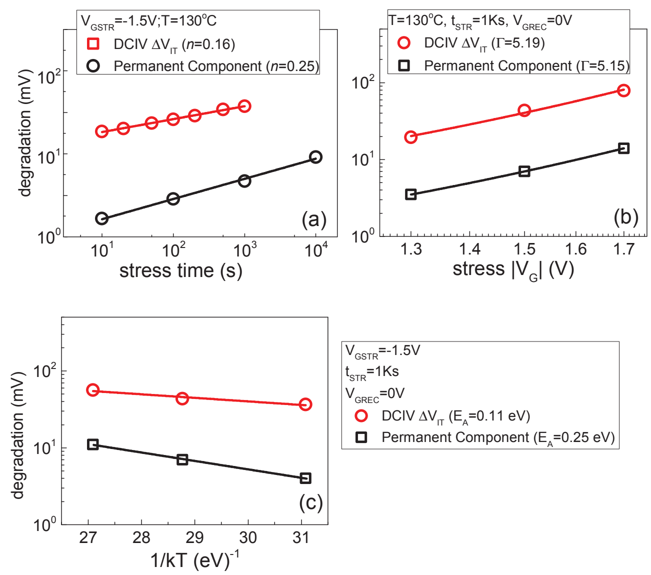

Fig.11 Comparison of permanent (P) component (=ΔV_{IT}) to DCIV measured and corrected ΔV_{IT} versus (a) stress time, (b) stress V_G and (c) stress T. The estimation using P component grossly differs from DCIV based estimate.

III. Features of Hole Trapping-Detrapping

A. *Impact on stress:*

Fig.12 (a) shows measured ΔV_T at fixed stress time versus T for an extended T range. Experimental results clearly show non-Arrhenius behavior in the full T range, but measured data can be fitted using two separate Arrhenius dependencies, with higher and lower E_A respectively at higher and lower range of T. Higher E_A at higher T is due to the ΔV_{IT} domination of ΔV_T, as ΔV_{IT} has higher E_A than ΔV_{HT}. Note that E_A for the higher T part is similar to DCIV data as ΔV_{HT} contribution is relatively small. However, the contribution from ΔV_{IT} starts to get suppressed as T is reduced below 0ºC, and at lower T, ΔV_{HT} dominates ΔV_T, causing reduction in E_A. Also note, E_A for the lower T part matches well with E_A shown in Fig.9 (d). Process B devices are used in these experiments, having relatively higher ΔV_{HT} contribution than process A devices, to highlight the predominant role of ΔV_{HT} at lower T.

Fig.12 (b) shows the time evolution of measured ΔV_T over an extended T range. The power-law time exponent *n* reduces gradually with the reduction in T below 0ºC. This is also fully consistent with gradual dominance of ΔV_{HT} contribution as T is reduced, and with the saturation of ΔV_{HT} at longer stress time. Therefore, the lower time constant for hole trapping and the associated low T activation E_A are independently verified.

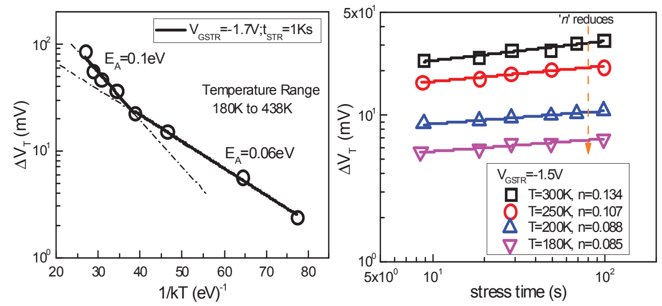

Fig.12 (a) Temperature dependence of measured ΔV_T at fixed stress time for extended T range. Lines are Arrhenius model fit to data, separately done at higher and lower T ranges. (b) Time evolution of measured ΔV_T for different T, over extended T range. Reduction in E_A and *n* at lower T are consistent with the hole trapping features.

Fig.13 (a) plots the T activation of UF measured ΔV_T while Fig.13 (b) plots the UF measured power-law time exponent *n* versus V_{GSTR} for process A and B devices. Process B shows lower E_A and lower *n* compared to process A devices. Note that process B has ~2X higher pre-existing hole trap density and hence relatively higher ΔV_{HT} contribution than process A devices [24]-[26]. Since ΔV_{HT} saturates at longer stress time and has lower E_A, higher contribution from ΔV_{HT} reduces the *n* and E_A of overall ΔV_T for process B devices as shown. It is important to remark that for both process A and B devices, the overall ΔV_T is still dominated by the ΔV_{IT} subcomponent.

Therefore, it is unequivocally established that longer time ΔV_T is dominated by ΔV_{IT} while ΔV_{HT} saturates, ΔV_{IT} shows higher E_A than ΔV_{HT}, and DCIV based isolation scheme D is consistent with e-NMP model and can explain NBTI process dependence. As mentioned before, these features (see Table-I, Set-A) have been used to develop the comprehensive model framework [24]-[26].

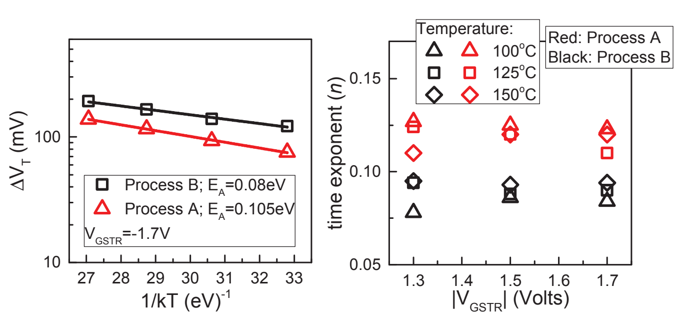

Fig.13. (a) T activation of UF measured ΔV_T and (b) stress V_G dependence of UF measured long-time power-law time exponent *n* for process A and B devices. Lower E_A and *n* for process B are consistent with the hole trapping features.

B. Impact on recovery:

Fig.14 shows UF measured time evolution of ΔV_T recovery for various V_{GSTR}, V_{GREC}, stress T and t_{STR}, as well as model prediction using the comprehensive framework [24]-[26]. As discussed before, TTOM enabled RD model is used for ΔV_{IT}. The e-NMP model is used to predict ΔV_{HT} (shown in bottom panels), and this is unlike earlier reports [24]-[26] that used empirical equations. The relative contributions from ΔV_{IT} and ΔV_{HT} at the beginning of recovery are determined from the decomposition of stress results. Note that the e-NMP model shows small time constant for the hole detrapping process at lower V_{GREC}, while a fraction of holes do not detrap at larger V_{GREC}. This is consistent with the empirical calculations used in [24]-[26]. Note, a large non-recoverable ΔV_{HT} is obtained at higher $|V_{GREC}|$ as a large fraction of traps remain above the Fermi level in this recovery condition. It is also important to remark that the larger number of e-NMP model parameters can be adjusted to achieve large time constants for recovery. However, when the parameters are optimized to predict both the stress and recovery experiments, the e-NMP model cannot support large time constants. Refer to [40] for details.

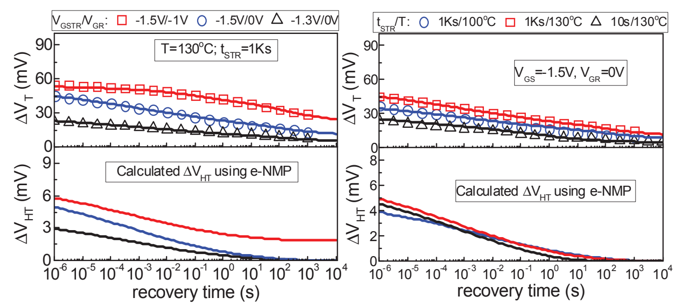

Fig.14. Time evolution of UF measured ΔV_T recovery for different stress and recovery bias (Left) and stress time and temperature (Right) and prediction by RD + TTOM + e-NMP model. RD + TTOM calculate ΔV_{IT} recovery, e-NMP calculates ΔV_{HT} recovery. Symbols: measured data, lines: model. The e-NMP model is consistent with smaller time constant for hole detrapping.

Fig.15 (a) shows ΔV_T recovery time evolution in multiple small area devices from UF measurements. Step like recovery is observed, whenever a hole gets detrapped from pre-existing traps, an electron gets captured in generated interface traps, or an interface trap gets passivated. The mean can be predicted by the composite model framework as shown in the bottom panel, with TTOM enabled RD model for ΔV_{IT} and e-NMP for ΔV_{HT} subcomponents. Once again, e-NMP simulations do not exhibit large time constants, and this contradicts previous assertions (Set-B, Table-I).

Small hole detrapping time constant is verified using Time Dependent Defect Spectroscopy (TDDS) measurement plots shown in Fig.15 (b). In TDDS study, a small area device is repetitively stressed and the recovery traces are measured by UF method [44]. Step like recovery is observed, which is due to hole detrapping from pre-existing traps, electron capture in generated interface traps, or passivation of interface traps. In Fig.15 (b), stress conditions are suitably adjusted (lower T, short stress time) and process B samples are used to ensure ΔV_{HT} domination of ΔV_T at the end of stress. Therefore, the recovery steps are primarily due to hole detrapping, showing low time constants. This verifies the consistency of e-NMP model simulation for hole detrapping during recovery. Once again, this contradicts previous assumptions of long recovery due to hole detrapping (Set-B, Table-I). Lower time constant for hole detrapping is consistent with previous results from different gate insulator processes and device dimensions, see [24]-[26] for a detailed discussion.

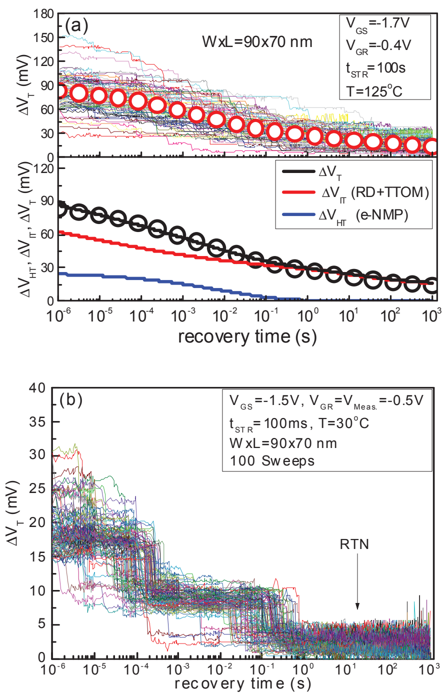

Fig.15. (a) Time evolution of UF measured ΔV_T recovery in multiple small area devices, both individual (lines) and mean (symbols) traces are shown (top panel). Mean recovery (symbols) predicted by RD + TTOM + e-NMP model (lines, bottom panel). (b) Multiple TDDS recovery traces in small area device. Results confirm smaller time constant for hole detrapping process.

IV. Trap Generation Kinetics

Previous sections have established that ΔV_{HT} saturates at long time while ΔV_{IT} evolves in time and dominates ΔV_T. In this section, the predictive capabilities of the TTOM enabled RD [24], [25] and GSHR [35], [36] models for interface trap generation will be verified against DC and AC stress data.

A. *Prediction of DCIV measurements:*

Fig.16 shows the time evolution of DCIV measured and corrected (for delay and bandgap) ΔV_{IT} at different V_{GSTR} and T for DC stress. The RD and GSHR model predictions are shown as lines. Model parameters are adjusted to predict the measured data at lowest V_{GSTR} and T, and then kept constant to predict higher stress conditions. Note that the RD model has only 3 adjustable parameters and can successfully predict measured data, refer to [25] for details[2]. However, the GSHR model shows saturation at higher V_{GSTR} and T and therefore is not compatible with the measured data. Note, it is possible to readjust GSHR model parameters to predict measured data at any particular V_{GSTR} and T. However, the model would fail to predict data as V_{GSTR} and T are varied.

[2] All RD model simulations shown in this section are performed using same parameter set. Refer to [25] for details.

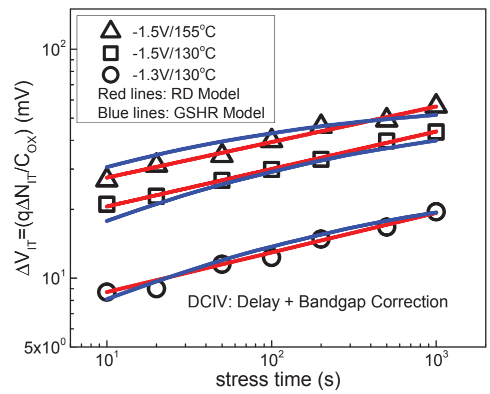

Fig.16. Time evolution of ΔV_{IT} (from delay and bandgap corrected DCIV), along with RD and GSHR model prediction for different stress V_G and T. RD model can while GSHR model cannot predict the measured data.

Fig.17 compares DCIV measured (and corrected) power-law time exponent *n* (at long stress time) to that from RD and GSHR model simulations at different V_{GSTR} and stress T. DC stress is used. The measured *n* is independent of V_{GSTR} and T (*n*~0.16), which is predicted by the RD model. Note, a slight reduction in measured *n* at higher V_{GSTR} and T is due to the field reduction effect (as increased ΔV_T reduces the oxide field at longer stress time). However, GSHR model prediction of *n* is drastically different from measurements. Simulated *n* reduces slightly at higher V_{GSTR}, but the reduction is drastic at higher T. Note that field reduction is not invoked in either RD or GSHR simulations; therefore, reduction of *n* especially at higher T is an inherent feature of the GSHR model. The large number of GSHR model parameters can be easily adjusted to predict *n* at any particular V_{GSTR} and T. However, it is very challenging for the GSHR model to predict *n* across different V_{GSTR} and T (using a consistent set of parameters) as shown.

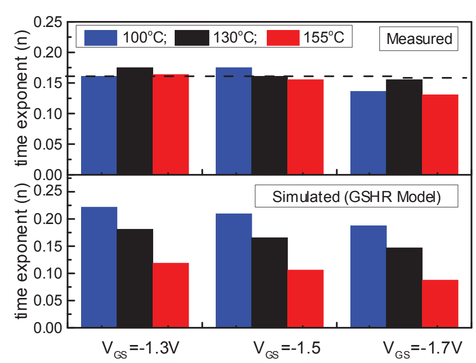

Fig.17. Measured (DCIV, both delay and bandgap corrected, top panel) and GSHR model simulated (bottom panel) power-law time exponent (*n*) for different stress V_G and temperature. The RD model value is shown as dotted line in the top panel. RD model can while GSHR model cannot predict the measured data.

Fig.18 plots T activation of DCIV measured and corrected (delay and bandgap) ΔV_{IT} for different V_{GSTR} for DC stress.

The RD and GSHR model predictions are shown as lines. Measured data exhibit similar E_A across different V_{GSTR}, and can be predicted by the RD model. However, GSHR model cannot predict measured data, and the simulated E_A reduces at higher V_{GSTR}. Once again, this is due to the strong saturation shown by GSHR simulation at longer time and higher stress T, which is an inherent feature of this model. Note that the large number of GSHR model parameters can be adjusted to predict E_A at a particular V_{GSTR}. However, the model cannot predict E_A, with a fixed parameter set, across all V_{GSTR}.

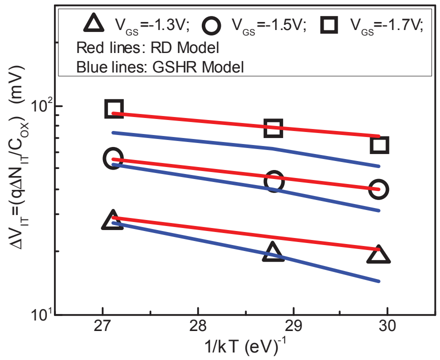

Fig.18. T dependence of measured ΔV_{IT} (from DCIV after delay and bandgap corrections) at different stress V_G. The RD and GSHR model calculations are also shown. Symbols: measured data, lines: model calculation. RD model can while GSHR model cannot predict the measured data.

B. Prediction of DC and AC stress:

Fig.19 shows the time evolution of DCIV measured ΔV_{IT} (corrected for delay and band-gap) and UF measured ΔV_T for DC and AC stress. For AC stress, measurements are done at the end of pulse off phase (Mode-B) [24]. ΔV_{IT} is calculated using the RD and GSHR models for both DC and AC stress. ΔV_T for DC stress is calculated using the RD plus e-NMP and GSHR plus e-NMP models. ΔV_T for AC stress is calculated using the TTOM enabled RD plus e-NMP and the GSHR plus e-NMP models. The model parameters are adjusted to predict DC stress at a single V_{GSTR} and T, and then used to predict the AC stress. Measured data can be predicted using the RD model framework, while the GSHR framework is inconsistent with experimental results. Note that the inherent saturation of GSHR model is stronger for AC stress. The capability of the

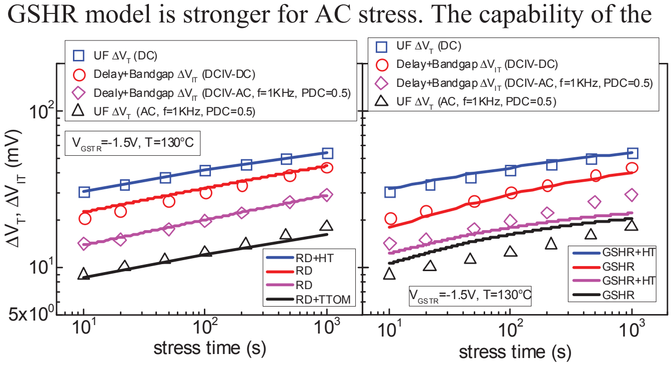

Fig.19. Time evolution of UF measured ΔV_T and DCIV measured ΔV_{IT} (delay and bandgap corrected) for DC and AC stress, and prediction by RD and RD + TTOM (Left) and GSHR (Right) model. Symbols: measured data, lines: model calculation. RD model can while GSHR model cannot predict the measured data.

RD framework is verified by prediction of measured data at different V_{GSTR} and T for both DC and AC stress, and at different PDC and *f* for AC stress, without adjusting any model parameters. These results have already been discussed in earlier reports [24], [25], [42].

Fig.20 plots the PDC and *f* dependence of measured ΔV_{IT} and ΔV_T at fixed stress time for Mode-B AC stress. ΔV_{IT} and ΔV_T are respectively obtained from time evolution of delay and bandgap corrected DCIV and UF measured data. RD and TTOM enabled RD models (along with e-NMP) respectively predict measured ΔV_{IT} and ΔV_T for different duty cycle and frequency. The framework can predict the "S" shaped PDC dependence, the absence of kink (for ΔV_{IT}) and its presence (for ΔV_T) near 100% PDC, and the frequency independence of ΔV_{IT} and ΔV_T. In contrast, the GSHR model (along with e-NMP) cannot predict AC duty cycle and frequency data. It is important to remark that identical model parameters have been used for DC and AC stress for both frameworks. Hence, Fig.16 through Fig.20 clearly show the predictive capability of RD based framework, while the GSHR based framework faces significant challenges to predict the experimental data. This limitation of GSHR model has hitherto not reported, as it was never used to predict stress data under wide range of DC and AC stress conditions.

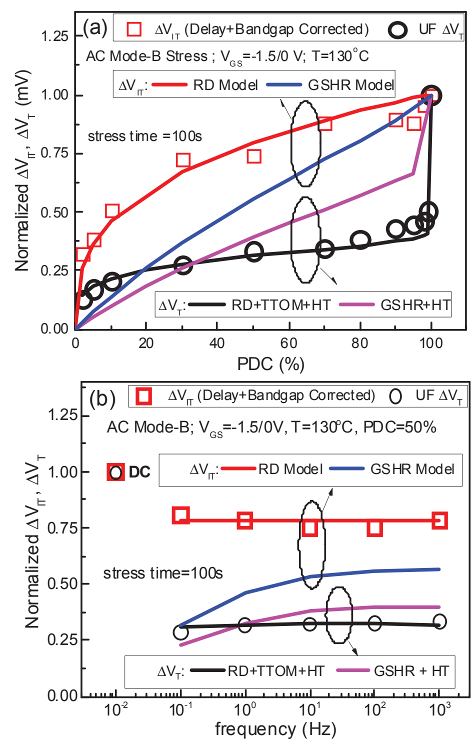

Fig.20 (a) UF measured ΔV_T and DCIV measured ΔV_{IT} (delay and bandgap corrected) versus duty cycle and (b) frequency, and prediction by RD and RD+TTOM and GSHR model frameworks. Symbols: measured data, lines: model calculation. RD model can while GSHR model cannot predict the measured data.

C. Prediction of end-of-life degradation:

Fig.21 plots time evolution of UF measured ΔV_T at high V_{GSTR} for short time and at low V_{GSTR} for long time for DC stress. The RD (for ΔV_{IT}) plus e-NMP (for ΔV_{HT}) and GSHR (for ΔV_{IT}) plus e-NMP (for ΔV_{HT}) frameworks are used to predict the short time ΔV_T at different V_{GSTR} and the model parameters are obtained. Note that the RD model framework can predict data at various V_{GSTR} and T and for both DC and AC stress, as shown in detail in [24]-[26]. However as shown above, the GSHR framework cannot predict data at different T and also for AC stress. Hence for simplicity, fixed T DC stress data are used for parameter calibration. The models are then used to predict long time DC stress data at low V_{GSTR} but identical T. The RD based framework can predict but GSHR framework underestimates long-time ΔV_T data as shown. As long-time ΔV_T is dominated by ΔV_{IT}, the RD model can, while the GSHR model (due to inherent saturation) cannot, be successfully used for estimation of EOL NBTI degradation.

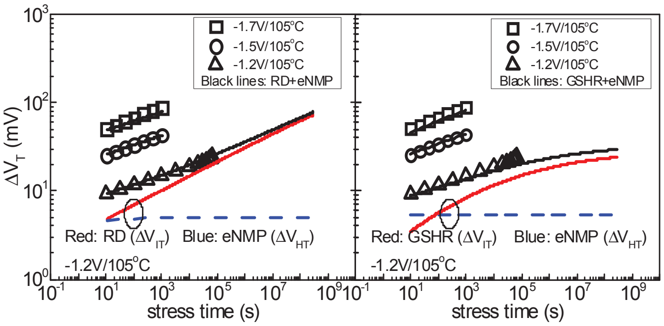

Fig.21. Time evolution of UF measured ΔV_T (symbols) at different stress V_G and for different duration of DC stress. The RD + e-NMP (Left) and GSHR + e-NMP (Right) framework based prediction (lines) are shown. RD model can while GSHR model cannot predict long-time measured data.

V. Conclusion

Existing NBTI mechanisms and models are benchmarked against diverse DC and AC experimental data in HKMG p-MOSFETs having various gate insulator processes. The often -used P versus R assignment technique, respectively for ΔV_{IT} and ΔV_{HT}, has been found to be inconsistent with GSHR and e-NMP model simulations. Hence, resulting ΔV_{HT} domination of ΔV_T, large time constant and large E_A for ΔV_{HT} for stress are not physically justifiable. However, obtained ΔV_{HT} from correct DCIV based isolation is much smaller than ΔV_{IT}, and the former shows saturation at long stress time and low E_A, consistent with e-NMP simulations. These features can also predict the gate insulator process dependence of NBTI. The kinetics of ΔV_{IT} dominates that of ΔV_T at longer time for both DC and AC stress. The TTOM enabled RD model along with e-NMP model can respectively predict the kinetics of ΔV_{IT} and ΔV_{HT} during DC and AC stress, as well as for recovery after DC stress, for both large and small area devices. The time constant associated with hole detrapping for recovery is shown to be small, and this is verified by TDDS experiments in small area devices. In contrast, the GSHR model cannot predict the ΔV_{IT} kinetics during DC and AC stress. The RD model based framework can, but the GSHR based framework cannot, predict long-time degradation for EOL calculation.

Acknowledgement

M. A. Alam and A. E. Islam (Purdue), S. Ramey (Intel), A. Kerber (Globalfoundries), J. Stathis (IBM), G. Groeseneken and J. Franco (IMEC), A. Oates (TSMC), V. Reddy (TI), V. Huard (ST) for stimulating discussions.

References

[1] S. Mahapatra, K. Ahmed, D. Varghese, A. E. Islam, G. Gupta, L. Madhav, D. Saha, and M. A. Alam, “On the physical mechanism of NBTI in silicon oxynitride p-MOSFETs: Can differences in insulator processing conditions resolve the interface trap generation versus hole trapping controversy?” in *Proc. Int. Rel. Phys. Symp.*, 2007, pp. 1–9.

[2] E. Cartier, A. Kerber, T. Ando, M. M. Frank, K. Choi, S. Krishnan, B. Linder, K. Zhao, F. Monsieur, J. Stathis, and V. Narayanan, “Fundamental aspects of HfO2-based high-k metal gate stack reliability and implications on *tinv*-scaling,” in *Proc. IEDM*, 2011, pp. 18.4.1–18.4.4.

[3] K. Joshi, S. Mukhopadhyay, V. Chaudhary, N. Nanaware, B. Rajamohnan,T. Sato, M. Bevan, A. Wei, A. Noori, B. McDougal, 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,” in *Proc. Int. Rel. Phys. Symp.*, 2013, pp. 4C.2.1.

[4] 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,” in *Proc. Int. Rel. Phys. Symp.*, 2013, pp. 4C.5.1.

[5] 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 Physics-Based NBTI Models,” IEEE Trans. Electron Devices, 60, 901 (2013).

[6] D. Varghese, D. Saha, S. Mahapatra, K. Ahmed, F. Nouri, and M. A. Alam, “On the dispersive versus Arrhenius temperature activation of NBTI time evolution in plasma nitrided gate oxides: Measurements, theory, and implications,” in *IEDM Tech. Dig.*, 2005, pp. 684–687.

[7] A. T. Krishnan, C. Chancellor, S. Chakravarthi, P. E. Nicollian, V. Reddy, A. Varghese, R. B. Khamankar, S. Krishnan, and L. Levitov, “Material dependence of hydrogen diffusion: Implications for NBTI degradation,” in *IEDM Tech. Dig.*, 2005, pp. 688–691.

[8] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P. Wagner, F. Schanovsky, J. Franco, M. T. Luque and M. Nelhiebel, "The Paradigm Shift in Understanding the Bias Temperature Instability: From Reaction–Diffusion to Switching Oxide Traps," in *IEEE Trans. Electron Devices*, vol. 58, no. 11, pp. 3652-3666 (2011).

[9] T. Grasser, “Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities,” *Microelectron. Reliab.*, vol. 52, no. 1, pp. 39–70, (2012).

[10] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel, “A two stage model for negative bias temperature instability,” in *Proc. Int. Rel. Phys. Symp.*, 2009, pp. 33–44.

[11] J. H. Lee, W. H. Wu, A. E. Islam, M. A. Alam, and A. S. Oates, “Separation method of hole trapping and interface trap generation and their roles in NBTI reaction–diffusion model,” in *Proc. Int. Rel. Phys. Symp.*, 2008, pp. 745–746.

[12] S. Mahapatra, V. D. Maheta, A. E. Islam, and M. A. Alam, “Isolation of NBTI stress generated interface trap and hole trapping components in PNO p-MOSFETs,” in *IEEE Trans. ectron Devices*, vol. 56, no. 2, pp. 236–242, Feb. 2009.

[13] V. Huard, “Two independent components modeling for negative bias temperature instability,” in *Proc. Int. Rel. Phys. Symp.*, 2010, pp. 33–42.

[14] S. Mahapatra, A. Islam, S. Deora, V. Maheta, K. Joshi, A. Jain, and M. Alam, “A critical re-evaluation of the usefulness of R–D framework in predicting NBTI stress and recovery,” in *Proc. Int. Rel. Phys. Symp.*, 2011, pp. 6A.3.1–6A.3.10.

[15] S. Mahapatra, P. Bharath Kumar, and M. A. Alam, “Investigation and Modeling of Interface and Bulk Trap Generation During Negative Bias Temperature Instability of pMOSFETs,” IEEE Trans. Electron Devices, vol. 51, pp. 1371 (2004).

[16] Y. Mitani, "Influence of nitrogen in ultra-thin SiON on negative bias temperature instability under AC stress," in *Proc. IEDM, 2004*, pp. 117-120.

[17] J. H. Stathis, G. LaRosa, and A. Chou, “Broad energy distribution of NBTI-induced interface states in p-MOSFETs with ultra-thin nitride oxide,” in *Proc. Int. Rel. Phys. Symp.*, 2004, pp. 1–7.

[18] J. P. Campbell, P. M. Lenahan, A. T. Krishnan, and S. Krishnan, “NBTI: An atomic-scale defect perspective,” in *Proc. Int. Rel. Phys. Symp.*, 2006, pp. 442–447.

[19] S. Mukhopadhyay, K. Joshi, V. Chaudhury, N. Goel, S. De, R. K. Pandey, K. V. R. M. Murali and S. Mahapatra, “Trap Generation in IL and HK layers during BTI / TDDB stress in scaled HKMG N and P MOSFETs”, in *Proc. Int. Rel. Phys. Symp.*, 2014, pp.GD.3.1-GD.3.11.

[20] S. Mukhopadhyay and S. Mahapatra, "An Experimental Perspective of Trap Generation Under BTI Stress," in *IEEE Trans. Electron Devices*, vol. 62, no. 7, pp. 2092-2097, July 2015.

[21] S. Mukhopadhyay, N. Goel, S. Mahapatra, "A Comparative Study of NBTI and PBTI Using Different Experimental Techniques," in *IEEE Trans. Electron Devices*, vol.63, no. 10, pp. 4038-4045, (2016).

[22] S. Tsujikawa, T. Mine, K. Watanabe, Y. Shimamoto, R. Tsuchiya, K. Ohnishi, T. Onai, J. Yugami, and S. Kimura, “Negative bias temperature instability of pMOSFETs with ultra-thin SiON gate dielectrics,” in *Proc. Int. Rel. Phys. Symp.*, 2003, pp. 183–188.

[23] S. Gupta, B. Jose, K. Joshi, A. Jain, M. A. Alam, and S. Mahapatra, “A comprehensive and critical re-assessment of 2-stage energy level NBTI model”, in *Proc. Int. Rel. Phys. Symp.*, 2012, pp. XT.3.1-XT.3.6.

[24] N. Goel, T. Naphade and S. Mahapatra, "Combined trap generation and transient trap occupancy model for time evolution of NBTI during DC multi-cycle and AC stress," in *Proc. Int. Rel. Phys. Symp.*, 2015, pp. 4A.3.1-4A.3.7.

[25] N. Parihar, N. Goel, A. Chaudhary and S. Mahapatra, "A Modeling Framework for NBTI Degradation Under Dynamic Voltage and Frequency Scaling," in *IEEE Trans. Electron Devices*, vol. 63, no. 3, pp. 946-953 (2016).

[26] 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,” in *IEEE Trans. Electron Devices,* vol. 64, no. 1, pp. 256-263, (2017).

[27] S. Mahapatra, V. Huard, A. Kerber, V. Reddy, S. Kalpat and A. Haggag, “Universality of NBTI -From devices to circuits and products”, in *Proc. Int. Rel. Phys. Symp.*, 2014, pp.3B.1.1-3B.1.8.

[28] S. Mahapatra, “*Fundamentals of Bias Temperature Instability in MOS Transistors*, 1st ed. New Delhi, India: Springer, 2015.

[29] http://nanohub.org/courses/NBTI

[30] M. F. Li et al., "Understand NBTI Mechanism by Developing Novel Measurement Techniques," in *IEEE Trans. on Device and Materials Reliability*, vol. 8, no. 1, pp. 62-71, (2008).

[31] B. Kaczer, T. Grasser, Ph. J. Roussel, J. Martin-Martinez, R. O’Connor, B. J. O’Sullivan, G. Groeseneken, "Ubiquitous relaxation in BTI stressing—New evaluation and insights," in *Proc. Int. Rel. Phys. Symp.*, 2008, pp. 20-27.

[32] B. Kaczer, T. Grasser, J. Martin-Martinez, E. Simoen, M. Aoulaiche, P. J. Roussel, and G. Groeseneken, “NBTI from the perspective of defect states with widely distributed time scales,” in *Proc. Int. Rel. Phys. Symp.*, 2009, pp. 55–60.

[33] H. Reisinger, O. Blank, W. Heinrigs, A. Muhlhoff, W. Gustin, and C. Schlunder, “Analysis of NBTI degradation- and recovery-behavior based on ultra-fast VT-measurements,” in *Proc. Int. Rel. Phys. Symp.*, 2006, pp. 448–453.

[34] 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, pp. 3586–3593, Nov. 2014.

[35] T. Grasser, M. Waltl, Y. Wimmer, W. Goes, R. Kosik, G. Rzepa, H. Reisinger, G. Pobegen, A. El-Sayed, A. Shluger, and B. Kaczer, "Gate-sided hydrogen release as the origin of "permanent" NBTI degradation: From single defects to lifetimes," in *Proc. IEDM,* 2015, pp. 20.1.1-20.1.4.

[36] T. Grasser *et al*., "The “permanent” component of NBTI revisited: Saturation, degradation-reversal, and annealing," in *Proc. Int. Rel. Phys. Symp.*, 2006, pp. 5A-2-1-5A-2-8.

[37] G. Kapila, N. Goyal, V. D. Maheta, C. Olsen, K. Ahmed, and S. Mahapatra, “A comprehensive study of flicker noise in plasma nitrided SiON p-MOSFETs: Process dependence of pre-existing and NBTI stress generated trap distribution profiles,” in *Proc. IEDM*, 2008, pp. 1–4.

[38] T. Grasser, W. Gos, V. Sverdlov and B. Kaczer, "The Universality of NBTI Relaxation and its Implications for Modeling and Characterization," in *Proc. Int. Rel. Phys. Symp.*, 2007, pp. 268-280.

[39] https://nanohub.org/resources/24943.

[40] https://nanohub.org/resources/24942.

[41] G. Rzepa et **al., “**Efficient Physical Defect Model Applied to PBTI in High-k Stacks**”** in *Proc. Int. Rel. Phys. Symp.*, 2017.

[42] N. Goel, K. Joshi, S. Mukhopadhyay, N. Nanaware, and S. Mahapatra, “A comprehensive modeling framework for gate stack process dependence of DC and AC NBTI in SiON and HKMG p-MOSFETs,” Microelectron. Reliab., vol.54, pp. 491, (2014).

[43] K. Joshi, S. Mukhopadhyay, N. Goel, N. Nanaware and S. Mahapatra, “A detailed study of gate insulator process dependence of NBTI using a compact model”, IEEE Trans. Electron Devices, p.408, v.61, (2014).

[44] 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," in *Proc. Int. Rel. Phys. Symp.*, 2010, pp. 16-25.