Bias temperature instability study on switching defects in SiON and high-K gate dielectrics

Tung, Zhi Yan

2019

Tung, Z. Y. (2019). Bias temperature instability study on switching defects in SiON and high-K gate dielectrics. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/88969>

<https://doi.org/10.32657/10220/47652>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**BIAS TEMPERATURE INSTABILITY STUDY
ON SWITCHING DEFECTS
IN SION AND HIGH-K GATE DIELECTRICS**

TUNG ZHI YAN

SCHOOL OF ELECTRICAL & ELECTRONIC ENGINEERING

2019

**BIAS TEMPERATURE INSTABILITY STUDY
ON SWITCHING DEFECTS
IN SION AND HIGH-K GATE DIELECTRICS**

TUNG ZHI YAN

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2019

STATEMENT OF ORIGINALITY

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarized materials, and has not been submitted for a higher degree to any other University or Institution.

30-JAN-2019

Date

TUNG Zhi Yan

SUPERVISOR DECLARATION STATEMENT

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

30-JAN-2019

Date

Assoc Prof. ANG Diing Shenp

AUTHORSHIP ATTRIBUTION STATEMENT

This thesis contains material from five papers published in the following peer-reviewed journals and conference proceedings where I was the first and/or corresponding author.

Chapter 4, Section 4.2 is published as A. A. Boo, Z. Y. Tung, and D. S. Ang, “New Observations on the Correlation Between Hole-Trapping Transformation and SILC Generation Under NBTI Stressing,” *IEEE Electron Device Letters*, vol. 37, no. 4. pp. 369–372, April 2016, doi: 10.1109/LED.2016.2531752.

The contributions of the co-authors are as follows:

- Assoc. Prof. D. S. Ang provided the initial project direction, gave suggestions in the interpretation of the measurement data, and edited the manuscript.
- Dr. A. A. Boo collected the measured threshold voltage shift, analyzed the data collected and drafted the manuscript.
- I collected the measured gate current shift (stress-induced leakage current, SILC), analyzed the data collected and assisted in drafting the manuscript.

Chapter 4, Section 4.3 is published as Z. Y. Tung and D. S. Ang, “Transient to temporarily permanent and permanent hole trapping transformation in the small area SiON *p*-MOSFET subjected to negative-bias temperature stress,” in *Proceedings IEEE 21st International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, Jun./Jul. 2014, pp. 254–257, doi: 10.1109/IPFA.2014.6898197.

The contributions of the co-authors are as follows:

- Assoc. Prof. D. S. Ang suggested the initial project direction, provided guidance in the interpretation of the measurement data, as well as edited the manuscript.
- I prepared the experiment setup and collected all the data measured in this experiment. Also, I analyzed the data collected as well as drafted the manuscript.

Chapter 5 is published as Z. Y. Tung, and D. S. Ang, “Alteration of oxide-trap switching activity at operating condition by voltage-accelerated stressing,” in *Proceedings IEEE International Reliability Physics Symposium (IRPS)*, Apr. 2016, pp. XT.01.1–XT.01.3, doi: 10.1109/IRPS.2016.7574643, and Z. Y. Tung and D. S. Ang, “Impact of voltage-accelerated stress on hole trapping at operating condition,” *IEEE Electron Device Letters*, vol. 37, no. 5, pp. 644–647, May 2016, doi: 10.1109/LED.2016.2543239.

The contributions of the co-authors are as follows:

- Assoc. Prof. D. S. Ang provided the initial project direction, supervised the data analysis, and edited the manuscript.
- I conducted the experiment and collected all the measurement data in this project. Also, I interpreted the data collected as well as drafted the manuscript.

Chapter 6 is published as Z. Y. Tung, D. S. Ang, and Y. Gao, “On electron-trap transformation and its unexpected frequency dependence under dynamic positive-bias temperature stressing,” in *Proceedings IEEE International Reliability Physics Symposium (IRPS)*, Jun. 2014, pp. XT.10.1–XT.10.6, doi: 10.1109/IRPS.2014.6861187.

The contributions of the co-authors are as follows:

- Assoc. Prof. D. S. Ang suggested the initial project direction, gave advice in the data interpretation, and edited the manuscript.
- I set up the experiment setup and collected all the data measured in this experiment under the guidance of Dr. Y. Gao. Also, I analyzed the data measured and drafted the manuscript.

30-JAN-2019

Date

TUNG Zhi Yan

ACKNOWLEDGEMENT

First, I would like to express my utmost gratitude to my project supervisor, Dr. Ang Diing Shenp (Associate Professor) from Nanyang Technological University (NTU), for his inspiring ideas, technical guidance, stimulating discussions, positive motivation and continuous support throughout the journey of my Ph.D. study. His scientific knowledge along with his valuable experience in the device physics and electrical measurements as well as strong passion in research have inspired me to enhance my knowledge and research ability through intensive reading, logical thinking, and scientific analysis. Furthermore, his effective writing and presentation skills have enabled me to present my ideas and findings clearly through professional writing and speaking. Being a dedicated supervisor, he was willing to provide constructive advice and spend time together in solving the challenges that were faced. I am forever grateful to have such a precious opportunity to work under Dr. Ang's supervision which allowed me to discover intriguing experimental results that contributed towards understanding the underlying mechanism of BTI.

Next, I wish to express my gratefulness to Singapore Economic Development Board and GLOBALFOUNDRIES Singapore for jointly granting me a Ph.D. scholarship. In addition, I would like to express my appreciation to Mr. Leong Kam Chew (formerly from GLOBALFOUNDRIES Singapore) for his support in my scholarship application. I am thankful that I had the chance to attend classes on the fundamental knowledge of wafer fabrication in the semiconductor industry conducted by Ms. Rosaline Tan (from GLOBALFOUNDRIES Singapore), Mr. Leong Kam Chew, and Mr. Damien Leong (formerly from GLOBALFOUNDRIES Singapore) during the first year of my Ph.D. study.

I would like to express my appreciation to the lab assistants in NTU, Mr. Chow Kam Wah, Mr. Chan Siang Cheng and Ms. Seet Lye Ping, as well as application engineer from

Tektronix, Mr. Lee Yong Ping, for their technical support in the prober station and equipment.

Further, I would like to extend my gratitude to other group members (Dr. Boo Ann Ann, Dr. Gao Yuan, Dr. Yew Kwang Sing, Dr. Zhou Yu, Dr. Zhang Haizhong, Dr. Tham Wai Hoe and Mr. Toh Rui Tze) for their technical discussion and knowledge sharing. Many thanks to all parties who had supported me throughout the thrilling adventure of my Ph.D. study.

Last but not least, I dedicate this dissertation to my dearest parents and my beloved brother. Their selfless love and care have been constantly supporting and inspiring me in moving forward in my life.

TABLE OF CONTENT

STATEMENT OF ORIGINALITY	i
SUPERVISOR DECLARATION STATEMENT	ii
AUTHORSHIP ATTRIBUTION STATEMENT	iii
ACKNOWLEDGEMENT	vi
TABLE OF CONTENT.....	viii
ABSTRACT	xi
LIST OF FIGURES	xiv
LIST OF TABLES	xxvi
LIST OF ABBREVIATIONS	xxvii
LIST OF SYMBOLS	xxix
CHAPTER 1: INTRODUCTION	1
1.1 Background	1
1.2 Motivation.....	7
1.3 Objectives	9
1.4 Scope	10
CHAPTER 2: LITERATURE REVIEW	12
2.1 Introduction to Bias Temperature Instability	12
2.2 Interface Traps	17
2.3 Switching Oxide Traps	18
2.4 The Reaction-Diffusion (R-D) Model and its Discrepancies	19
2.5 Hydrogen Species in Bulk Oxide Trap Generation and its Discrepancies	23
2.6 <i>R</i> and <i>P</i> Components under NBTI	26
2.7 <i>R</i> and <i>P</i> Components under PBTI	31

2.8	Recent Studies on Single Charge Trapping on Individual Defects in Gate Dielectric	33
2.9	Chapter Summary.....	38
CHAPTER 3: BTI CHARACTERIZATION TECHNIQUES.....		39
3.1	Introduction.....	39
3.2	Motivation for Fast Measurement Methods	39
3.3	Fast Measurement Techniques.....	41
3.4	Measurement Techniques used in Experiments.....	43
3.4.1	Ultra-Fast Switching Measurement Technique.....	43
3.4.2	Keithley's Ultra-Fast BTI Package	48
3.4.3	Stress-Induced Leakage Current (SILC) measurement	55
3.5	Chapter Summary.....	56
CHAPTER 4: SWITCHING OXIDE TRAPS RECOVERY BEHAVIOR UNDER DYNAMIC NEGATIVE-BIAS TEMPERATURE INSTABILITY		57
4.1	Introduction.....	57
4.2	Correlation between Hole-Trapping Transformation and SILC Generation under NBTI Stressing	60
4.2.1	Experimental Details	61
4.2.2	Results and Discussion	62
4.3	Transient to Temporary Permanent and Permanent Hole Trapping Transformation in the Small-Area <i>p</i> -MOSFET Subjected to Negative-Bias Temperature Stress.....	68
4.3.1	Experimental Details	68
4.3.2	Results and Discussion	70
4.4	Correlation between Drain and Gate Currents of Small-Area SiON Gate <i>p</i> -MOSFETs Subjected to NBTI Stressing	83

4.4.1	Experimental Details	84
4.4.2	Results and Discussion	85
4.5	Chapter Summary.....	95
CHAPTER 5: IMPACT OF VOLTAGE-ACCELERATED STRESS (VAS) ON SWICTHING OXIDE TRAPS AT OPERATING CONDITION		97
5.1	Introduction.....	97
5.2	Experimental Details	99
5.3	Results and Discussion	101
5.4	Chapter Summary.....	124
CHAPTER 6: FREQUENCY DEPENDENCE OF ELECTRON-TRAP TRANSFORMATION UNDER DYNAMIC POSITIVE-BIAS TEMPERATURE INSTABILITY.....		125
6.1	Introduction.....	125
6.2	Experimental Details	126
6.3	Results and Discussion	130
6.4	Chapter Summary.....	146
CHAPTER 7: CONCLUSION and recommendation.....		147
7.1	Conclusion	147
7.2	Recommendations for Future Work.....	152
LIST OF PUBLICATIONS		154
REFERENCES.....		155

ABSTRACT

Device reliability remains an extremely challenging issue for the state-of-the-art complementary metal-oxide-semiconductor field-effect transistor (CMOS, MOSFET). One of the most challenging reliability issues in CMOS devices is the bias-temperature instability (BTI). Previous studies on SiON and high- k gate dielectrics (HfO₂ and HfSiON) for large area devices reveal that the transient oxide-charge trapping/detrapping by switching oxide traps (SOTs) due to dynamic bias-temperature stress (BTS), which is usually treated as a parasitic effect, may play a role on the gate oxide's long term reliability. The "fast" measurement used in the studies reveals the inconsistency in the prediction of the classical explanation for the negative-bias temperature instability (NBTI) using a reaction-diffusion (R-D) model. The studies show that the transformation of SOTs to a more permanent form, which is also known as recoverable-to-permanent (*R-to-P*) transformation, occurs when the exerted thermal-cum-electrical stress surpasses a particular threshold while the cyclical behavior of switching oxide-charge trapping and detrapping is observed when the employed thermal-cum-electrical stress is below this threshold. These observations indicate that the generation of permanent defects is possible due to the change that may occur in the oxide structure caused by the applied thermal-cum-electrical stress. The concurrent increasing of the gate current with the NBTI-induced hole trapping transformation implies that the outset of the bulk oxide trap generation is caused by the hole-trapping conversion. However, electron trapping transformation is not always accompanied by an increase in the gate current.

The *R-to-P* conversion of the SOTs in large-area CMOS devices was revealed in previous studies. The continuing scaling-down of the CMOS ICs leads to considerable interest in the study of small-area devices. The study of the drain current recovery trends on the small-area SiON/polysilicon gate *p*-MOSFETs subjected to recurring NBTI

stress/relaxation cycling presents direct experimental evidence of *R-to-P* hole-trapping transformation, which was inferred from past research on large-area devices. The emission times of hole traps are normally presumed to be time-invariant, but the results from this study suggest otherwise. Analysis from the experimental evidence demonstrates that the emission times of hole traps are not time-invariant and can increase due to the evolution of the defect sites into more structurally stable configurations. In addition, a new type of switching hole traps is observed, which exhibits intermittent charging during stress and an occasional increase in emission time by ~5 orders of magnitude.

A past study reported an observation that generation of stress-induced leakage current (SILC) linearly tracks the decrease in NBTI recovery. This indicates that the bulk oxide trap generation and the NBTI-induced hole-trapping conversion are correlated and stem from the same physical mechanism. The study of the correlation between *R-to-P* transformation of hole trapping and SILC generation in big-area devices reveals that both sequels eminently showcase similar power-law time exponents, temperature dependence, and activation energies. The revelation of a kink at the temperature of 125°C that indicates a reduction in the activation energy from the low-to-high temperature region in both *R-to-P* conversion and SILC generation signifies strong connection and a common degradation mechanism between these two effects. Also, investigating the time-dependence of drain as well as gate current shift, ΔI_d and ΔI_g respectively, in small-area devices subjected to NBTI stress reveals the positive and negative correlations along with the uncorrelated behavior between these two effects. In addition to the observation of a typical single-step I_d recovery accompanied by a change in I_g , fluctuation similar to random telegraph noise (RTN) in I_d was also detected to be accompanied by RTN-like fluctuation in I_g . Of particular interest is the relatively significant ΔI_g of ~40% accompanying the ΔI_d , which evidently indicate the involvement of bulk oxide defects. A possible mechanism that may activate and deactivate a bulk oxide trap as a trap-assisted tunneling (TAT) center is illustrated based on *ab initio*

simulations of the oxygen-vacancy defects. Two scenarios that may enable a bulk oxide trap to function as a TAT center in the positively charged state are discussed.

The impact of voltage-accelerated stress (VAS) on hole trapping probed in the operating environment was examined. The charge-capturing activity of time-zero oxide traps measured under operating condition can be altered by VAS. After the application of VAS, the time-zero oxide traps can become either more active or less active in hole capturing in the operating environment, suggesting that the trap atomic structure may have been modified by the VAS. Also, after undergoing a relaxation or an idle state for a period of time, the charge-capture activity of the oxide traps may revert to their pre-VAS state. More intriguingly, the drain-current recovery step amplitude of a single oxide charge emitted by an SOT can be altered by VAS. The observed stress-induced change of trap-switching behavior and the alteration of the drain-current degradation under operating conditions should be taken into account in the reliability evaluation of small-area devices. This is essential given the remarkably increasing impact of oxide charge trapping on device channel conduction as the size of the device reduces.

In addition to the NBTI studies, a PBTI study was performed as well. An intriguing trend was observed in the examination of the frequency dependence of dynamic positive-bias temperature instability (DPBTI) on HfO_2/TiN gate n -MOSFETs. The investigation reveals that the shallow-to-deep electron traps transformation exhibits an unanticipated positive dependence on the gate frequency as the PBTI stress progresses, i.e. the fraction of the shallow-to-deep electron trap conversion increases with frequency, causing more permanent electron traps at higher gate frequency in the investigated frequency range of 1 mHz to 1 MHz.

LIST OF FIGURES

Figure		Page
1-1	Schematic diagrams showing side and top view of a MOSFET structure. The dimensions of a MOSFET, such as gate oxide thickness, T_{ox} , channel length, L , and channel width, W , are scaled down to achieve high density and high performance. The channel length, L , is the determinant parameter and descriptor of a MOSFET technology node.	2
1-2	Simulated characteristic of gate leakage current, I_g , as a function of gate voltage, V_g , for various SiO_2 thicknesses [15]. The gate leakage is too high for most of the MOSFET practical operating condition at $V_g = 1$ V for SiO_2 with a thickness of less than 1.5 nm which consists of 3 to 4 atomic layers of SiO_2 .	4
2-1	The transition of the MOSFET lifespan limitation mechanism as a function of gate dielectric thickness. BTI is the major limiting factor as compared to HCI for device lifespan when the gate dielectric is less than 3.5 nm [46].	13
2-2	The I_d and g_m versus V_g curves of a 2 nm SiON p -MOSFET in the linear regime at the fresh condition (solid line) and after 10 ks NBTI stress (dashed line) at 125°C [30].	15
2-3	Various types of defects (charges) at different locations in a thermally grown SiO_2 on a silicon substrate [11].	16
2-4	Schematic illustration of the interfacial states on (a) silicon (111) and (b) silicon (100) surfaces [32].	18
2-5	Schematic illustration showing hole trapping under NBTI stress (left) and hole emission during relaxation at flatband voltage (right). The holes from the conducting channel during NBTI stressing may tunnel into deep localized states (a) below or (b) above the substrate Fermi level. Holes captured may tunnel back into the substrate during recovery stage [81].	19
2-6	Schematic diagram of the R-D model representing Si/ SiO_2 interface, showing: (a) the $\equiv\text{Si-H}$ defect precursor under fresh state, (b) the electrically-activated $\equiv\text{Si-H}$ defect precursor under the NBTI stressing generating the interface trap where the released hydrogen could react with SiO_2 , forming an oxide trapped charge and a hydroxyl group, and (c) the hydroxyl group diffusing away from the interface to the oxide, which is a reaction- limiting factor [92]	20
2-7	(a) Schematic energy band diagram illustrating the generation of deep level hole traps (energy states above the n -well conduction-	28

band edge E_C) caused by NBTI stressing. Shaded regions represent the energy distribution of the stress-induced hole traps. A possible hole-trap generation mechanism involves thermal- and field-assisted dissociation of the oxide defect precursors and subsequent loss of electronic charge to the conduction band of the n -well. (b) Shallower hole traps (i.e., those below E_C) are instantaneously discharged by electron tunneling from the n -well upon the termination of stress; deep-level hole traps, however, do not relax. (c) A positive gate bias lowers the hole-trap states, causing more relaxation, but some fraction may still remain since their energy states are pinned by the Si/SiO₂ conduction-band offset [122].

2-8	The two-stage model for a switching oxide trap coupled with the creation of a dangling bond at the interface. When the E' center is positively charged (in state 2), the hydrogen passivating a silicon dangling bond at the interface can move to the E' center, thereby effectively locking in the positive charge (state 4). The charge state of the created dangling bond depends on the position of the Fermi-level [131].	29
2-9	A simplified transition state diagram of a defect model describing the transition of a defect from precursor state 1' to positively charged state 2 through metastable state 2' [133].	30
2-10	A schematic energy band diagram of the HfO ₂ /TiN gate stack under (a) 0 V and (b) -1 V gate bias. An electron trapped at an energetically deep oxide trap could not be readily emitted due to the energy barrier ϕ between the trap level and the Si conduction band edge under $V_g = 0$ V relaxation. A negative gate bias 'raises' the trap level reducing the barrier ϕ which allows the trapped electron to be now easily emitted into the Si conduction band. [56]	32
2-11	(a) Conduction channel with a non-uniform potential profile induced by RDD. The carriers conduct through the percolation path during the "ON" state, as indicated by the white lines. (b) Blocking of a percolation path by an oxide defect locating above the constriction point upon hole-trapping due to NBTI effect, resulting in significant drain current degradation [137]	34
2-12	Top: Two examples of threshold voltage recovery traces of a small-area MOSFET ensuing NBTI stress. Each step refers to a hole-detraping event of a charged oxide trap. Bottom: A spectral map consisting of clusters of points wherein each cluster represent an individual defect (bottom). The clusters in the spectral map which are formed based on the step heights and emission times extracted from the recovery traces are the fingerprints of each defect [47].	35
2-13	The four states of defect modeled based on E' center properties, illustrating the existence of metastable state [47], [224].	36
3-1	NBTI degradation as a function of stress time for various delays between stress interruption and measurement under the same characterization method. The device lifespan varies for different	40

measurement delays (data after Ref. [50]).

3-2	A schematic timing diagram illustrating the OTF measurement methodology. I_{dlin} shifts and the associated g_m variations can be tracked (data after ref. [30]).	41
3-3	(a) A schematic diagram of the UFS setup for NBTI based on the PIV setup. The device under test (DUT) is stressed at a static negative voltage V_g^{dc} . Measurement at a specific interim is performed by injecting a narrow positive pulse of amplitude V_g^{p} , which momentarily brings the overall gate voltage down to the measurement level. The corresponding change in the drain current is converted into voltage (via a 50 Ω internal resistance) and measured by a high-speed digital OSC. A correction scheme accounts for the change in drain bias during measurement to yield the desired drain current at the measurement drain voltage. (b) Drain and gate voltage waveforms during measurement. Additional pulses of amplitude $V_g^{\text{p}} \pm \Delta$, where $\Delta = 50$ mV) may be added for transconductance measurement [162].	44
3-4	Waveforms of the gate voltage pulse V_g^{p} (top) injected by the PG and the corresponding I_d^{p} (bottom) captured by the OSC. V_g^{p} has a 100 ns pulse width and a 2.6 V amplitude with a 10 ns rising/falling time. The I_d^{p} tracked has a 20 ns delay that is caused by the wiring capacitance. The measured I_d^{p} is not affected by this 20 ns delay since the measurement is performed at the quasi-steady state between 60–80 th nanoseconds, while the stable waveforms are acquired after 50 ns [162].	46
3-5	(a) Drain current, $ I_d $ measured using both DC and UFS approaches by pulsing/sweeping the gate voltage from 0 to -1 V. The drain voltage is kept at -0.1 V. (a) Linear region (b) Sub-threshold region (data after Ref. [165])	47
3-6	The measure window for spot measurement in a pulse. The start-measuring and stop-measuring points are defined at 75% and 90% of each step width [164].	51
3-7	Schematic diagram of four-terminal local sense connections using the SMA to dual SSMC adapter cable (Y-cable) with Model 4225-RPM connected to DUT during experiment. Gate and drain terminals are applied with desired bias and waveform through Model 4225-RPM whereas source and bulk terminals are grounded. [164]	52
3-8	Schematic illustration for local sensing connection. The Force triax connector has center conductor connected to DUT high while the outer shield is connected to DUT low. The Sense connector is not used. [164], [166]	54
3-9	Test sequence diagram of various types of testing function waveforms that can be performed using the Ultra-Fast BTI package.	55

3-10	A schematic of the TAT mechanism aiding the carrier tunneling through the gate oxide contributing to the SILC.	56
4-1	Recovery of $ \Delta V_t^r $ per cycle (normalized to the $ \Delta V_t^r $ of the first cycle) as a function of the number of SR cycles, exhibiting constant recovery independent of the SR cycles. The R-D model simulation shows a gradual decrease in $ \Delta V_t^r $ with an increase in the number of SR cycles, which does not align with the experimental data (data after Ref. [106]).	57
4-2	A comparison of (a) $ \Delta V_t $ recovery and (b) gate current I_g in the DNBTI test only and in DNBTI test with an intermediate CVS as a function of number of SR cycles. The $ \Delta V_t $ recovery is deduced from the difference between $ \Delta V_t $ extracted at the end of recovery phase and the $ \Delta V_t $ extracted at the end of the prior stress phase, as depicted in the inset of (a) (data after ref. [52]).	58
4-3	(a) Increase in the SILC closely “tracks” the decrease in $ \Delta V_t $ recovery (b) Outstanding linear correlation between SILC (extracted at two gate voltages) and reduction in $ \Delta V_t $ recovery (data after ref. [52]).	59
4-4	(a) The shift in threshold voltage, $ \Delta V_t $ of a HfSiON gate p -MOSFET subjected to stress/relaxation cycling. The $ \Delta V_t $ recovery in the respective cycles are denoted as R_j , where j is the cycle number. (b) Stress induced leakage current (SILC) versus gate voltage (V_g) curve after 40 stress/ relaxation cycles; (i) HfSiON and (ii) SiON gate p -MOSFETs. The gate voltage was swept from -1.5 V to 1.5 V and then back to -1.5 V.	62
4-5	(a) $ \Delta R_{40,1} $ as a function of $(kT)^{-1}$, where k is the Boltzmann constant and T is temperature in Kelvin, for the HfSiON gate p -MOSFET. $\Delta R_{40,1}$ is defined as $R_{40} - R_1$, where R_{40} and R_1 represent the overall $ \Delta V_t $ recovery in the 40 th and 1 st cycle, respectively. (b) SILC generation (defined as $\Delta I_g/I_{g0}$ at $V_g = -1$ V) as a function of $(kT)^{-1}$. ΔI_g is the difference between the I_g (at $V_g = -1$ V) of the 40 th and 1 st cycle. The temperatures indicated are in degree Celsius ($^{\circ}\text{C}$)	63
4-6	(a) $ \Delta R_{40,1} $ and (b) SILC generation ($\Delta I_g/I_{g0}$) as a function of $(kT)^{-1}$ for the SiON gate p -MOSFET. The values above the plotted lines indicate the temperatures in degree Celsius.	64
4-7	Both SILC generation ($\Delta I_g/I_{g0}$) and $ \Delta R_{40,1} $ exhibit very similar dependence on stress time. Data may be fitted using a power-law relation, with closely matched exponents (as indicated).	65
4-8	Correlation plot of the power-law time exponent m for $\Delta R_{40,1}$ versus the power-law time exponent n for SILC generation. The numbers in the plots denote temperature in degree Celsius.	66

4-9	Construction of a spectral map (bottom) by the extraction of single detrapping events from three typical representative drain current degradation, $ \Delta I_d $, recovery traces (top) ensuing NBTI stress. The emission time, τ_e , of a defect is obtained by tracking the discrete step height, h , from the single detrapping events, i.e. each defect is identified by a cluster of (τ_e, h) points as shown in the spectral map (bottom).	69
4-10	End-of-stress (EOS) and end-of-recovery (EOR) $ \Delta I_d $ values, $ \Delta I_d ^{\text{eos}}$ and $ \Delta I_d ^{\text{eor}}$, respectively, of each DNBTI cycle. The total amount of $ \Delta I_d $ degradation due to 10 s stress of each cycle is represented by $ \Delta I_d ^{\text{eos}}$ while the remnant of $ \Delta I_d $ degradation after the relaxation of recoverable SOTs in 10 ks relaxation is denoted by $ \Delta I_d ^{\text{eor}}$.	71
4-11	Number of step heights and amount of $ \Delta I_d $ recovery, R , as DNBTI progresses. The number of step heights in each $ \Delta I_d $ recovery trace refers to the amount of SOTs which recover in that particular cycle during relaxation. The trend of the number of steps heights is in correspondence to the trend of the amount of R .	72
4-12	$ \Delta I_d $ recovery traces of cycles 1 to 5. (a) $ \Delta I_d $ recovery trace of cycle 1 shows that $ \Delta I_d > 0$ after 10 ks relaxation, implying that the first 10 s NBTI stress has already generated some relatively permanent defect(s). (b) Defect #5 did not recover in cycle 2, as compared to cycle 1 in (a), resulting in a greater $ \Delta I_d $ degradation at the end of the 10 ks relaxation. (c) Defect #2 and #4 were not charged during the 10 s stress of cycle 3, as can be seen from the much lower $ \Delta I_d $ degradation at the start of the relaxation phase (compared to that of cycle 1 or 2). The final degradation remains the same, indicating that defect #5 still remained charged after the 10 ks relaxation phase of cycle 3. (d) Defect #2 and #4 were being charged during NBTI stress in cycle 4 as the initial $ \Delta I_d $ degradation is similar to those in (a) and (b). Defect #4 recovered but defect #2 still remained charged owing to the final degradation after the 10 ks relaxation is higher, caused by the step height of defect #2 step height as compared to that in (b) and (c). Defect #5 still remained charged. (e) Defect #5 finally discharged the trapped hole in cycle 5, since it had been charged in cycle 2. However, defect #2 still remained charged. (f) A comparison between cycle 1 and cycle 5 shows that the $ \Delta I_d $ degradation after the 10 ks relaxation in cycle 5 is larger than that in cycle 1. The recovery step height of defect #2 is not seen and the $ \Delta I_d $ degradation after 10 ks relaxation in subsequent cycles (starting from cycle 4) is obviously higher by an amount equaling to the recovery step height of defect #2. This may be attributed to the conversion of defect #2 from a transient hole trap into a more permanent hole trap.	73
4-13	ΔI_d recovery traces of cycles 5 to 9. (a) The recovery trace of cycle 5 shows that 4 defects were discharged, but defect #2 (permanent hole trap) was not. (b) Defect #5 was inactive or not charged during the 10 s stress in cycle 6. (c) Defect #A is observed with a new step height in cycle 7. (d) Hole captured by defect #A during the 10 s	75

stress in cycle 8 was not emitted after 10 ks relaxation. (e) Part of the step height of defect #A, which is denoted by defect #A'', was discharged during the 10 s stress in cycle 9, whereas part of the step height of defect #A, which is denoted by defect #A', remained charged at the end of the 10 ks relaxation in cycle 9. (f) A comparison between cycle 5 and cycle 9 shows that the $|\Delta I_d|$ degradation after the 10 ks relaxation in cycle 9 is larger than that in cycle 5 by the amount of the step height equivalent to defect #5, which corresponds to defect #A'.

- 4-14 Evolutions of the 6 SOTs discovered based on the recovery traces in each cycle as a function of the number of DNBTI cycles. Each dynamic NBTI (DNBTI) cycle consists of a 10 s NBTI stress followed by a 10 ks relaxation. A defect is deemed to be active when it is charged and discharged in the same DNBTI cycle. It is considered to be inactive when it is not charged during the stress phase after being detrapped in the previous cycle. For example, defect #5 was charged during the stress phase in cycle 2 but was not discharged until cycle 5. Defect #2 and defect #5 transformed into a more permanent form after cycle 4 and cycle 9 respectively, when they could not discharge within the experiment window. Defect #6 is not activated until cycle 16. 80
- 4-15 A schematic illustration of the metastable state formation that produces a temporary increase in the trap emission time. Upon charge capture, the initial structural change (ΔQ_1) is minimum, thus allowing the trap to 'switch' effortlessly from the intermediate state back to the precursor state (short trap relaxation time). Through phonon interaction, further structural change (ΔQ_2) may put the defect in a metastable state. The energy barrier for the defect to revert to the intermediate state is now increased, resulting in a longer trap relaxation time. Further structural change may put the defect into a permanent state with a relaxation time that surpasses a reasonable experimental timeframe. 81
- 4-16 (a) Schematic energy band diagrams illustrating (i) hole tunneling from channel to the polysilicon gate and valence band electron tunneling from the gate to the *n*-well; (ii) two possible ways by which NBTI stress induced oxide defects contribute to an increase of gate current. Also depicted is the change in oxide band bending caused by hole trapping. The solid (open) circle denotes electron (hole). (b) Hole tunneling current I_{sd} (sensed at the tied source/drain S/D), valence band electron tunneling current I_w (sensed at the *n*-well) and gate current I_g as a function of gate voltage (V_g) obtain from charge separation measurement on a small-area SiON gate *p*-MOSFET. 87
- 4-17 Time-dependent relaxation of drain (I_d) and gate (I_g) current shifts of a small area SiON gate *p*-MOSFET following NBTI stress, demonstrating examples of the four combinations of ΔI_d and ΔI_g changes: (1) I_d recovery without change of I_g (shaded oval region); (2) I_g change without I_d recovery (shaded rectangular region); (3) I_d 88

recovery with I_g increase (dashed); (4) I_d and I_g recovery (solid)..

- 4-18 Post-NBTI relaxation characteristics of I_d and I_g of small-area p -MOSFETs. (a) Negatively correlated RTN-like fluctuations; (b) and (c) Significant I_g changes ($> 10\%$ with respect to I_{g0}) along with that of I_d . 90
- 4-19 Shifts in the charge transition level (CTL) of an oxygen vacancy (V_{O1}) defect upon changes in hole occupancy. The amount of shift is seen to vary with the position of the V_O defect in the simulated supercell (e.g. negligible shifts are obtained at another V_{O2} defect site), implying the role of local oxide chemistry. The supercell under study was attained using molecular dynamics based on the melt-and-quench scheme. Full structural relaxation was carried out at 0 K after each change in hole occupancy. 91
- 4-20 Positively correlated RTN-like fluctuations of ΔI_d and ΔI_g observed during the post-NBTI relaxation of two small-area SiON gate p -MOSFETs. (a) Small ΔI_d and ΔI_g fluctuations of $\sim 2\%$ of the respective pre-stress values; (b) “Gigantic” ΔI_d and ΔI_g fluctuations of $\sim 90\%$ of I_{d0} and $\sim 40\%$ of I_{g0} , respectively. 92
- 5-1 A schematic illustration of the electrical test sequence adopted to examine the effect of voltage-accelerated stressing (VAS) on time-zero switching oxide traps (SOTs) measured in the monitoring step under operating condition at $V_g = -1$ V 98
- 5-2 Four typical $|\Delta I_d|$ recovery traces during 100 s relaxation interval at $V_g = -0.6$ V ensuing 10 s biasing interval at operation condition of $V_g = -1$ V. Each abrupt step-like drain current, I_d , recovery corresponds to an emission event of a trapped-hole in oxide defect. $|\Delta I_d|^{\text{cor}}$ labeled represents the remnant of $|\Delta I_d|$ at the end of 100 s relaxation interval. 99
- 5-3 An instance of I_d evolution monitored under operating condition at $V_g = -1$ V using the same DUT as in Figure 5-2. The dotted lines are eye guides, attempting to delineate what appears to be an I_d degradation step due to a hole captured by an oxide trap at $\sim 10^{-4}$ s amidst the “noisy” measurement data. 100
- 5-4 TDDS spectral maps of two test devices D (left) and M (right) extracted from the corresponding $|\Delta I_d|$ recovery traces during monitoring steps before (top) and after (bottom) VAS at $V_g = -2$ V, showing reduction in defect switching activity through the clear decrease in the intensity of cluster D1 (dashed line) and the disappearance of cluster M1 (dotted line) after the VAS step.. 102
- 5-5 Hole-emission events (symbolized by the data points) during monitoring steps before and after VAS at $V_g = -2$ V illustrating (a) reduction in activity of time-zero defect D1 and (b) disappearance of time-zero defect M1 under operating condition, supporting the observation of TDDS spectral maps in Figure 5-4. Capture activity 103

of the defects is expressed in percent. Cycle numbering continues from the previous monitoring step. The line depicts the evolution of I_d degradation at the end of each relaxation interval, $|\Delta I_d|^{\text{eor}}$, during monitoring steps.

- 5-6 TDDS spectral maps of two test devices A (left) and M (right) extracted from the corresponding $|\Delta I_d|$ recovery traces during monitoring steps before (top) and after (bottom) VAS at $V_g = -2$ V, demonstrating increasing in defect switching activity through the apparent formation of a “new” cluster A5 (dashed rectangle on the left) after the VAS step relative to only one emission event (denoted by ▲) during the pre-VAS monitoring step as well as the appearance of cluster M2 (dotted rectangle on the right) after the VAS step despite no sign of activity observed prior to the VAS step. 104
- 5-7 Hole-emission events (denoted by the data points) during pre- and post-VAS monitoring steps demonstrating (a) increasing in activity of time-zero defect A5 and (b) appearance of time-zero defect M1 under operating condition, corroborating the observation of spectral maps in Figure 5-6. Cycle numbering continues from the previous monitoring step. The line outlines the evolution of I_d degradation at the end of each relaxation interval, $|\Delta I_d|^{\text{eor}}$, during monitoring steps. Capture activity of the defects is expressed in percent. 106
- 5-8 Examples of three oxide defects exhibiting different impacts on the switching or capture activity imposed by a series of VAS steps involving successively larger voltages from -2 to -3 V: (a) reduction in the capture activity shown by trap G1 after VAS at -3 V; (b) similar capture activity exhibited by trap G2 without being altered by VAS; and (c) increase in capture activity displayed by trap H2 after VAS at -3 V. Capture activity of the defects is expressed in percentage. 107
- 5-9 TDDS spectral maps of a test device J extracted from the corresponding $|\Delta I_d|$ recovery traces during monitoring steps (a) before the VAS step as well as after the VAS steps at (b) -2 V, (c) -2.5 V and (d) -3 V. The intensity of cluster J1 (dashed line) fades progressively through the successive higher VAS voltages from -2 to -3 V. Also, a reduction in the density of cluster J2 (dotted line) is due to a brief disappearance of trap J2’s activity owing to the temporary increase in emission time after the -3 V VAS step. 108
- 5-10 (a) A gradual decrease ensued by a complete disappearance in activity of trap J1 after the application of successive higher VAS voltages from -2 to -3 V. (b) A brief disappearance in activity of trap J2 (used to emit in less than 100 s in every cycle) owing to the short-term rising in emission time after applying VAS at -3 V; activity restored after the emission of the captured hole at cycle 330. Capture activity of the defects is expressed in percentage.. 109
- 5-11 TDDS spectral maps of test devices D (left) and E (right) extracted from the corresponding $|\Delta I_d|$ recovery traces during 111

monitoring steps before VAS step (top), after VAS at -2 V (middle) and after 3 months of storage (bottom), demonstrating that the activity of defects can be restored to their state prior to VAS after sufficient rest period, as seen for trap D1 and E3.

- 5-12 Hole-emission events (symbolized by the data points) during the monitoring steps before VAS, after VAS at -2 V and after 3-month storage, for traps (a) D1, (b) E3 and (c) B3. Traps D1 and E3 display restoration in capture activity to their state prior to VAS application after an adequate relaxation period, corroborating the observation of the TDDS spectral maps in Figure 5-11. However, the capture activity of trap B3 does not get reinstated to the level it was before the VAS step after the 3-month storage. Capture activity of the defects is expressed in percentage. The cycle numbering is continued from the previous monitoring step. The fluctuated line outlines the evolution of I_d degradation at the end of each relaxation interval, $|\Delta I_d|^{\text{cor}}$, during the monitoring steps. 113
- 5-13 TDDS spectral maps of test device F extracted from the corresponding $|\Delta I_d|$ recovery traces during the monitoring steps (a) before VAS step, after VAS at (b) -2, (c) -2.5 and (d) -2.8 V, (e) after 1-week storage, as well as (f) after reapplication of the -2.8 V VAS step. There are 3 main clusters displayed, F1, F2 and F3. The dashed line marks the value of the extracted single-trap $|\Delta I_d|$ recovery amplitude of cluster F2 during the pre-VAS 100 s relaxation interval at $V_g = -0.6$ V, serving as an eye guide to compare and illustrate the change in the amplitude of the drain-current recovery after the application of the VAS steps. 114
- 5-14 The probability density function for emission characteristics of the three defects, F1, F2 and F3 in test device F at $V_g = -0.6$ V during the pre-VAS step. The emission time of these three defects follows an exponential distribution, consistent with a first-order reaction-limited process [59]. The solid, dashed and dotted lines represent the exponential fitting equation (Eqn 5-1) and the extracted emission time constant for each trap is stated above the lines. 115
- 5-15 Probability plot of drain-current recovery amplitude of single traps F1, F2 and F3. The drain-current recovery amplitude of each individual single trap follows a lognormal distribution [210]–[215], dissimilar from the exponential distribution observed in the drain-current relaxation step amplitudes of single defects obtained from many devices [48], [143], [216]–[221]. The general fitting equation of cumulative density function (CDF) of lognormal distribution is shown in Eqn 5-2, with x in Eqn 5-2 is $|\Delta I_d|$ in this case. The μ , σ^2 and the mean ($|\Delta I_d|$ recovery amplitudes) calculated are tabulated in Table 5-1. 116
- 5-16 The evolution of the drain-current recovery amplitude, $|\Delta I_d|$, following the VAS test sequence with respect to that during the pre-VAS (time-zero) step, $|\Delta I_d|_{\text{time-zero}}$, for traps F1, F2 and F3 of test device F. Traps F1 and F3 display no apparent change in the $|\Delta I_d|$ 118

while obvious fluctuation of $|\Delta I_d|$ can be detected for trap F2. There is a missing point for the drain-current recovery amplitude of trap F2 at the first application of the -2.8 V VAS step. This is because the F2 cluster is absent from the TDDS map as shown in Figure 5-13.

- 5-17 The evolution of the emission time of three defects, F1, F2 and F3, in test device F following the VAS test sequence. The emission time is extracted from the exponential distribution fitting presented in Figure 5-14. No significant change is observed in the emission time for these three traps. There is a missing point for the emission time of trap F2 at the first application of the -2.8 V VAS step. This is because the F2 cluster is absent from the TDDS spectral map, as shown in Figure 5-13 120
- 5-18 (a) Cumulative density function for an SOT's capture characteristic at $V_g = -1$ V and (b) probability density function for the same SOT's emission characteristic at $V_g = -0.6$ V. Extremely good fitting of the data (square) to the exponential function (solid lines) can be achieved [59]. The fitting equations and the corresponding coefficients of capture and emission characteristics are stated in (a) and (b) respectively. 124
- 6-1 A sequence of electrical characterization tests implemented to a HfO_2/TiN gate n -MOSFET to examine the effect of DPBTI stress frequency on the proposed shallow-to-deep oxide trap transformation. 126
- 6-2 Four typical time-dependent $|\Delta I_d|$ relaxation curves of a fresh DUT subjected to the repeated alternate trap-filling and trap-emptying phases. The overlapping of the recovery traces implies that the fixed set of test conditions always probes a similar group of oxide traps. The fitting equation is $|\Delta I_d| = B_r [\log_{10}(1 + t_s/t_r)] + P$, where t_s is the stress time, t_r is the relaxation time, and B_r and P are constant. Throughout the experiment, the temperature was set to be 100°C while the gate voltage during relaxation was set to be zero. The device was stressed for 10^3 s prior to each relaxation trace, hence t_s is constant. The fitting coefficients B_r and P of each trace are listed in Table 6-1. 127
- 6-3 Comparison between transconductance g_m curves of the pristine state and post-DPBTI stressing. After DPBTI stress, the g_m curve shifts laterally towards more positive V_g in comparison to the pre-stress (fresh) g_m curve. However, the peak transconductance value does not degrade after DPBTI stress. This outcome indicates that the interface trap generation is insignificant during DPBTI stressing, and the positive threshold voltage shift is primarily a consequence of the electron trapping. 128
- 6-4 Energy band diagrams in the course of: (a) PBTI stressing, as well as (b) and (c) relaxation upon the removal of PBTI stressing. (a) During PBTI stressing, stress induced oxide traps with energy states (denoted by ‘-’) at or below the Si Fermi energy E_{FS} are filled with 130

electrons. (b) When the stress voltage is removed, some trap states are raised above E_{FS} and the trapped electrons are emitted spontaneously, accounting for the initial rapid PBTI recovery. (c) Electrons trapped at deeper trap states ‘see’ a relatively large emission barrier ϕ_B and, thus, require a much longer time to be emitted, accounting for the extended PBTI recovery.

- 6-5 Relaxation curve of the linear drain current shift $|\Delta I_d|$ of a DUT: 131
 (a) After the first trap-filling phase before DPBTI stress; (b) After DPBTI stressing at 1 MHz; (c) After the second trap-filling phase. R_1 and R_2 labeled in (a) and (c) represent the difference between $|\Delta I_d|$ at the end of the trap filling and trap-emptying phases (i.e the $|\Delta I_d|$ recovery amount) before and after the DPBTI stress test, respectively. A smaller R_2 compared to R_1 implies that some of the oxide traps that were active in the trap-filling/emptying cycle prior to the DPBTI stress test are no longer involved in the trap-filling/emptying cycle after the DPBTI stress test. The continuous solid lines are the fitting curves and the fitting equation is the same as the one in Figure 6-2, $|\Delta I_d| = B_r [\log_{10}(1+t_s/t_r)] + P$. The coefficients of each fitting curve are displayed in the respective figures.
- 6-6 Dependence of $|\Delta I_d|$ measured at the end-of-stress (EOS) and at the 133
 end-of-relaxation (EOR) of the last stress/relaxation cycle as a function of the gate stress frequency (50 % duty cycle). All DUTs were stressed to the same total clock time of 5×10^3 s. The difference between $|\Delta I_d|^{eos}$ and $|\Delta I_d|^{eor}$ for the 1 MHz stress could not be clearly resolved due to the finite measurement delay of ~ 1 μ s, and thus, their coincidence is an artifact. The fitting equations for the $|\Delta I_d|^{eos}$ and $|\Delta I_d|^{eor}$ data points are $|\Delta I_d|^{eos}(\mu A) = 431 - 8.5 \log_{10} f_g$ and $|\Delta I_d|^{eor}(\mu A) = 184 + 28.2 \log_{10} f_g$ respectively. $|\Delta I_d|^{eos}$ decreases with frequency in a rate of 8.5 $\mu A/dec$ while $|\Delta I_d|^{eor}$ increases with frequency in a rate of 28.2 $\mu A/dec$
- 6-7 Relaxation curves of the linear drain current shift $|\Delta I_d|$ for the trap- 135
 emptying phase before and after DPBTI stress. The larger $|\Delta I_d|$ at the start of trap-emptying phase following the DPBTI stress may be ascribed to an increase in the number of stress-induced deep-level electron trapping events. The curve with circle points (before DPBTI) is the same as the curve in Figure 6-5(a) while the curve with square points (after DPBTI) is the same as the curve in Figure 6-5(c). The solid and dashed lines are the fitting curves for the data points before and after DPBTI respectively
- 6-8 (a) $|\Delta I_d|$ evolution under the repeated application of the trap-filling 136
 and emptying phases. (b) The amount of $|\Delta I_d|$ recovery in the trap-emptying phase, R as a function of the number of trap filling/emptying cycles. Also shown is $|\Delta I_d|^{eor}$ or the remnant $|\Delta I_d|$ after each trap-emptying phase, exhibiting a progressive increase.
- 6-9 The overlapping of the $|\Delta I_d|$ relaxation curves during the trap- 137
 emptying phases before the DPBTI stress and after the recovery at -1 V subsequent to the DPBTI stress. The outcome is dissimilar to

the case illustrated in Figure 6-7, where the $|\Delta I_d|$ relaxation trace obtained during the trap-emptying phase after a 0 V recovery ensuing the DPBTI stress shifts upwards relative to the one acquired preceding DPBTI stress. This entails that the restoration of the original shallow-level oxide defects from the deep energy states is feasible via the application of a negative gate recovery voltage. The fitting equation before the DPBTI stress is $|\Delta I_d|(\mu A) = 21.5[\log_{10}(1+1000/t_r)] + 21.5$ while the fitting equation after post-DPBTI-stress recovery at -1 V is $|\Delta I_d|(\mu A) = 21.6[\log_{10}(1+1000/t_r)] + 23.0$. The constants B_r and P of these two fitting curves are comparable.

- 6-10 Increasing trend in $|\Delta R|$ with gate frequency under DPBTI stress. 139
 $|\Delta R|$ ($= |R_2 - R_1|$) represents the difference between $|\Delta I_d|$ recovery measured in the trap-emptying phase before (R_1) and after (R_2) the DPBTI stress. $|\Delta R|$, measured under DC PBTI stress, is also plotted for comparison. The DC stress time was equal to the total clock period of the dynamic stress. The different symbols denote the data from the two experimental runs (cf. Figure 6-1) carried out on the same DUT. The linear fitting equation, coefficient of determination (R-squared), and root-mean-square error (RMSE, σ) are shown in the diagram. Further studies are needed to understand the underlying significance of the coefficients in the fitting equation.
- 6-11 Gate voltage V_g as a function of time showing the comparison of the waveforms between (a) 1 MHz and (b) 1 kHz under DPBTI stress. 140
The insignificant 0.16-V overshoot (compared to the top-level V_g signal of 2 V applied) of the peak signal at 1-MHz gate frequency is not the culprit to yield higher $|\Delta R|$ obtained in Figure 6-10.
- 6-12 A simplistic schematic 2D energy-coordinate diagram illustrating an 142
initial hypothesis of the transformation of a shallow-to-deep trap state, through an intermediate state, during DPBTI stressing. The transition between the different states across different energy barriers involved is indicated by the curvy-rightwards arrow. The intermediate state is a shallow trap that can be readily formed, even under a high gate frequency, due to the small energy barrier ΔE_1 and structural change ΔQ_1 . However, its reversal to the precursor state is restricted by the larger energy barrier ($>\Delta E_1$), and thus, the defect can still remain in the intermediate state when the next stress cycle is applied. Therefore, such defects see the high gate-frequency DPBTI stress as a quasi-DC stress, which enables the defects to have ample time to overcome a greater energy barrier ΔE_2 and structural change ΔQ_2 in order to transit to a more stable deep trap state. At lower gate frequencies, however, the transition to the final deep state becomes less likely due to the longer relaxation interval, which increases the chance of the intermediate state reverting to the precursor state. It should be noted that direct transition from the precursor to the final deep state (dotted line) entails a much larger energy barrier.

LIST OF TABLES

Table		Page
4-1	Extracted values of n , A , m and B of the fitting curves on SILC generation ($\Delta I_g/I_{g0}$) and $ \Delta R_{40,1} $ for HfSiON and SiON.	67
5-1	The values of μ , σ^2 and the mean ($ \Delta I_d $ recovery amplitudes) calculated for the lognormal distribution curve shown in Figure 5-15.	117
6-1	Comparison of coefficients A , B and P of the fitting equation for 4 relaxation traces in Figure 6-2. Each coefficient for each relaxation trace is almost the same, supporting the overlapping of these 4 relaxation traces.	144

LIST OF ABBREVIATIONS

AC	alternating current
ACS	Automated Characterization Suite
AHR	anode hydrogen release
ALD	atomic layer deposition
BTI	bias temperature instability
BTS	bias-temperature stress
CDF	cumulative density function
CMOS	complementary metal-oxide semiconductor
CTL	charge transition level
CVS	constant voltage stress
DC	direct current
DIBL	drain-induced barrier lowering
DLTS	deep level transient spectroscopy
DNBTI	dynamic negative-bias temperature instability
DPBTI	dynamic positive-bias temperature instability
DUT	device under test
EOR	end of relaxation
EOS	end of stress
EOT	electrical oxide thickness
ESR	electronic spin resonance
HCI	hot carrier injection
HDL	Harry Diamond Lab, recently known as the Army Research Lab
HKMG	high- k /metal gate
IC	integrated circuit
LRME	lattice-relaxation multi-phonon emission
MOSFET	metal-oxide-semiconductor field-effect transistor
NBTI	negative-bias temperature instability
OSC	oscilloscope
PBTI	positive-bias temperature instability
PG	pulse generator
PIV	pulsed current-voltage

R-D	reaction-diffusion
RBT	remote bias tee
RMSE	root-mean-square error
RTN	random telegraph noise
RTS	random telegraph signal
SCE	short channel effect
SILC	stress-induced leakage current
SMU	source measurement unit
SoC	system on chip
SOT	switching oxide trap
SR	stress/recovery
SRAM	static random-access memory
TAT	trap-assisted tunneling
TDDDB	time-dependent dielectric breakdown
UFS	ultra-fast switching
VAS	voltage-accelerated stress
VASP	Vienna <i>ab initio</i> simulation package
VBE	valence-band electron

LIST OF SYMBOLS

I_d	drain current
I_g	gate current
V_g	gate voltage
V_t, V_{th}	threshold voltage
V_d	drain voltage
V_{DD}	supply voltage
V_{FB}	flatband voltage
g_m	transconductance
L	channel length
W	channel width
T_{ox}	gate dielectric thickness
C_{ox}	gate oxide capacitance per unit area
φ_s	work function of semiconductor material
φ_{MS}	work function difference between metal and semiconductor
φ_B	bulk fermi potential
T	temperature
k	Boltzmann constant
q	electrical charge, 1.602×10^{-19} C
n_i	semiconductor intrinsic carrier concentration
N_D	doping concentration in semiconductor material
ϵ_0	permittivity of free space
ϵ_S	permittivity of semiconductor medium
Q_f	fixed charge density
Q_{it}	interface trap density
Q_{ot}	oxide charge density
Q_B	bulk charge density
R	recoverable component
P	permanent component
V_O	oxygen-vacancy
τ_c	capture time constant
τ_e	emission time constant

CHAPTER 1: INTRODUCTION

1.1 Background

The transistor, one of the greatest inventions in human history, was invented by William Shockley, Walter Brattain and John Bardeen on 16th December, 1947 at Bell Telephone Laboratories [1] to replace electron vacuum tubes. In the late 1950s, the planar integrated circuits (ICs), which consist of transistors and other electronic components such as diodes, capacitors and resistors, were invented by Jack Kilby [2]–[4] and Robert Noyce [5]. With the birth of the transistor and ICs, along with the favorable attributes arising from the down-scaling of the transistor, the semiconductor industry underwent unprecedented, enormous technological and economical development for five decades. In the modern era, transistors, which serve as the fundamental building blocks in all the ICs, are extensively manufactured in semiconductor-based products that are used widely in almost every type of technology, including electronic gadgets, home appliances, medical instruments, automobiles, airplanes, and satellites. The most efficient transistor from a technological and economical point of view is the silicon-based metal-oxide-semiconductor field-effect transistor (MOSFET), which was invented by Martin M. (John) Atalla and Dawon Kahng at Bell Labs in 1959 [6]. The efficiency of the MOSFET allows it, together with other electronic components, to form ICs and systems on chip (SoCs), where analog, digital, mixed-signal and radio frequency functions can be performed on a single chip substrate. The MOSFET's prodigious performance advancement in ICs or a SoC over the past half a century is credited to the relentless down-scaling of MOSFET.

The continuous, immense down-scaling of the MOSFET over the decades that was propelled by the commercial demands of great performance and low cost follows the well-

known Moore's law [7]–[9], that predicted doubling the amount of MOSFETs being fabricated on an IC and an exponential reduction in the minimum feature size roughly every 18 months. Scaling down the dimension of MOSFETs (Figure 1-1) enables the integration of a larger number of MOSFETs into a single IC chip. Hence, increasing the density of the MOSFETs leads to growth in the functionality and complexity of the ICs despite the low manufacturing cost per chip. The dimensions of the MOSFET, specifically the gate length or channel length, L , are scaled down to instigate greater transistor current, $I_{d,sat}$ (Eqn. 1-1), which leads to an increase in the switching speed of the devices and shortens the circuit delay, resulting in a higher operating frequency. Furthermore, down-scaling the gate length or channel length decreases the chip's computation time as well as reduces the power consumption and dissipation.

Eqn. 1-1 shows the operating drain current at saturation region, $I_{d,sat}$:

$$I_{d,sat} = \frac{W}{L} \mu C_{ox} \frac{(V_g - V_t)^2}{2} \quad (\text{Eqn. 1-1})$$

where L is the channel length, W is the channel width, μ is the mobility of charger carriers, C_{ox} is oxide capacitance per unit area, V_t is threshold voltage and V_g is gate voltage.

Figure 1-1. Schematic diagrams showing side and top view of a MOSFET structure. The dimensions of a MOSFET, such as gate oxide thickness, T_{ox} , channel length, L , and channel width, W , are scaled down to achieve high density and high performance. The channel length, L , is the determinant parameter and descriptor of a MOSFET technology node.

However, as the density of MOSFETs and the complexity of the ICs grew, the heat produced during the operation of the MOSFETs could not be scaled down, leading to a constant increase of the temperature of the MOSFETs or ICs during operation. Consequently, the degradation of the off-current and the increase in the threshold voltage are unavoidable, causing a rise in the power density over time. Furthermore, scaling down of the supply voltage in proportion to the linear dimensions of the MOSFETs is unfavorable for safeguarding the performance of high-density ICs [10]. As a result, the electric field of the transistor increases during operation or on-state, leading to severe undesirable reliability issues.

The advancement of lithography technology, along with the excellent electrical properties of silicon dioxide, SiO_2 , as an insulator and its use as the gate oxide, have been key elements in contributing to the successful reduction of the MOSFET's size. SiO_2 , which is thermally grown on the single crystalline silicon substrate is extremely stable and of high quality, with a superior Si/ SiO_2 interface that has a low intrinsic interface trap [11] density of $\sim 10^{10} \text{ cm}^{-2} \text{ eV}^{-1}$ [12]. The relatively large $\sim 9 \text{ eV}$ bandgap in SiO_2 makes it an outstanding electrical isolator which can avoid electrical or signal interference between the controlling gate and the channel current beneath the gate oxide in the Si substrate. Further, SiO_2 is able to withstand the fabrication process with a high thermal budget owing to its stable chemical and thermal properties. Moreover, SiO_2 exhibits a high breakdown electric field of $\sim 15 \text{ MV/cm}$ for ultrathin layers and $\sim 10 \text{ MV/cm}$ for thicker layers. Also, SiO_2 has a low hole and electron trapping rate. All these notable SiO_2 properties are extremely vital for a desirable MOSFET performance.

Based on Eqn. 1-1, in order to achieve higher $I_{d,\text{sat}}$ for better MOSFET performance, other than down-scaling the channel length, L , as discussed previously, the gate oxide capacitance per unit area, C_{ox} , needs to be raised. The Eqn. 1-2 below shows the C_{ox} equation.

$$C_{ox} = \frac{\epsilon_0 k}{T_{ox}} \quad (\text{Eqn. 1-2})$$

where ϵ_0 is the absolute permittivity of free space, k is the relative permittivity dielectric constant of the gate dielectric material and T_{ox} is the thickness of the gate dielectric which in this case is considered to the SiO_2 .

To increase C_{ox} , the gate oxide thickness T_{ox} must be scaled down in proportion to the channel length and width of the MOSFETs in order to ensure the gate control over the channel with higher C_{ox} . However, issues surface when the thickness of conventional SiO_2 that forms the gate dielectric hits its physical limit of less than 2 nm at the technology nodes of 180 nm and 130 nm. The uniformity of the gate oxide thickness which only consists of a few atomic layers is eminently difficult to control during the fabrication process [13]. Even a small variation in a very thin gate oxide during the manufacturing process can lead to a large disparity in the device performance. Another problem caused by the thinning of the gate oxide is the exponential increase of the gate leakage current due to quantum tunneling in the ultra-thin gate oxide [14]–[17]. Figure 1-2 illustrates the simulated I_g - V_g characteristic

Figure 1-2. Simulated characteristic of gate leakage current, I_g , as a function of gate voltage, V_g , for various SiO_2 thicknesses [15]. The gate leakage is too high for most of the MOSFET practical operating condition at $V_g = 1$ V for SiO_2 with a thickness of less than 1.5 nm which consists of 3 to 4 atomic layers of SiO_2 .

for various SiO₂ thicknesses. The gate leakage current density exceeds 1 A/cm² at 1 V gate voltage for the 1.5 nm ultrathin SiO₂ layer due to direct tunneling of electrons [16], which leads to high power consumption and interference between the gate and channel currents. These effects are unsuitable for most of the practical applications. Also, failure of the full-bandgap formation in SiO₂ below the thickness of 7 Å [18], [19] causes high densities of charge carriers to tunnel through the gate oxide, leading to the escalation of trap generation, which ultimately provokes oxide breakdown [20]–[22]. Consequently, the difficulty in controlling the uniformity of the SiO₂ layer during fabrications and the rapid increase of the gate leakage current are the limiting factors for down-scaling the SiO₂ thickness. In addition to these two limiting factors, another reliability issue is the boron penetration into the substrate from the gate electrode through the ultrathin gate dielectric due to the same reduced thickness of the SiO₂ layer [23], [24].

Due to these limiting factors encountered in scaling down the gate oxide thickness as discussed above, increasing the gate oxide dielectric constant, k , to raise C_{ox} based on Eqn. 1-2 is an alternative, feasible method to tackle the further down-scaling of the gate oxide. There are several ways to increase k . One of the methods is to add nitrogen to the SiO₂ via nitrogen deposition on top of the thin SiO₂ layer through remote or decoupled plasma nitridation [25], [26]. Besides acting as a barrier that prevents boron penetration from the p^+ polysilicon gate, the SiON (SiON k value is 4 ~ 7) formed from deposition of nitrogen onto the SiO₂ (SiO₂ k value is 3.9) raises the dielectric constant, allowing equivalent oxide thickness (EOT) scaling without a severe gate leakage current [27]. Other materials with a higher dielectric constant were suggested to replace SiO₂/SiON in order to guarantee the physical thickness of gate dielectric which should be thick enough to minimize gate leakage current and to have a thin EOT to preserve the desired gate control over the channel [16].

High- k /metal gate (HKMG) stacks have been utilized in the technology node of 45 nm and beyond to raise k in order to allow the EOT to be further scaled towards 1 nm [28], [29].

The much larger dielectric constant k of transition metal oxides leads to higher capacitance and good control over the channel by the gate, as well as the ability to acquire a physically thicker oxide layer with a sub-1-nm EOT for gate current leakage reduction. The extensively accepted rare-earth metal oxides candidates used to form a high- k dielectric are hafnium (Hf) based oxides, such as HfO_2 , HfSiO_x and HfSiO_xN_y . Besides having a high dielectric constant to gain higher capacitance and have better control over the channel current, the use of Hf-based oxides as the gate oxide offers some other advantages as well, such as a large (5.8 eV) bandgap barrier with conduction band offset of 1.4 eV to reduce the gate leakage current, good compatibility with the existing CMOS process, and superb thermal stability with Si. Nonetheless, the reliability issue that arises from the use of high- k dielectrics as gate oxides is complex as the dielectric compositions are intricate.

The implementation of SiON and high- k dielectrics in state-of-the-art CMOS technologies leads to severe reliability challenges. Currently, the bias temperature instability (BTI) effect is one of the most detrimental reliability problems [30]–[34] that has captivated considerable attention owing to the critical impact of BTI on limiting the circuit performance since 1966 [35]. BTI has been attributed to the progressive degradation of device parameters, such as threshold voltage V_{th} , drain current I_d , and transconductance g_m of MOSFETs, with time, reducing the lifespan of ICs. Also, compared to the SiO_2 p -MOSFET, the SiON p -MOSFET is more susceptible to the detrimental effect of negative-bias temperature instability (NBTI) [31], bringing the attention back to NBTI in the early 2000s. Despite knowing that nitrogen is the primary culprit [36]–[39], the hidden mechanism of BTI degradation is still not thoroughly comprehended. The BTI reliability issue is more critical for high- k dielectrics because of the larger amount of inherent defects at the high- k /Si interface and in the high- k oxide [40]. Besides NBTI, positive-bias temperature instability (PBTI) also severely impacts the performance of the HKMG stack n -MOSFET [41]–[45].

1.2 Motivation

The thinning of the gate oxide thickness below 3.5 nm due to the down-scaling of CMOS technology render NBTI degradation a more alarming reliability issue as compared to hot carrier injection (HCI) [46]. Also, reduction of the supply voltage, V_{DD} reduces the relative significance of HCI effects, as compared with BTI. Therefore, a significant amount of attention has been paid to the study of the BTI in the semiconductor industry and academic institutes. Lately, there has been a considerable interest in studying the reliability of small-area devices, as trap-charging events of individual defects may cause significant degradation in the parametric shift in small-area devices [47], [48] as compared to large-area devices, where the degradation contribution of individual defects is averaged out. Hence, studying the BTI effect in small-area devices allows the analysis of the nature and properties of individual SOTs. This would serve as evidence to understand the fundamental roots of the degradation caused by individual traps and allow the counterchecking of experimental results previously reported for large-area devices, thus significantly contributing to the understanding of the physics behind the degradation of the gate oxide caused by the BTI effect. Previous studies have reported that there are 2 components in BTI – a recoverable (*R*) component, in which the damage can be reversed, and a relatively slow recovering or permanent (*P*) component [49], [50]. The transient-to-permanent traps and shallow-to-deep traps transformations in large-area devices were revealed in [51]–[56]. These discoveries for large-area devices triggered the studies of the same phenomena of individual defects in small-area devices. Additionally, the voltage dependence on the capture and emission times of individual defects has also been studied in [47], [57]–[59]. It was discovered that the change in capture behavior stems from the effect of the applied V_g on the capture barrier energy and the local inversion charge density. This prompted an interest in investigating the change in the capture behavior of individual defects at the same V_g and without any apparent change in the local inversion charge density after the

application of the voltage-accelerated stress (VAS). Moreover, studies using large-area devices [52], [54] revealed that the generation of SILC corresponding to the reduction in recoverable threshold voltage shift, which is caused by NBTI, indicates that the transformation of recoverable hole trapping into permanent fixed charged traps is correlated to the generation of bulk oxide traps. Hence, this triggered the need to study the correlation between these two events based on temperature dependence, activation energy, and power-law time exponent. Additionally, it motivated the examination of the correlation between hole trapping, denoted by the degradation of the drain current, and the bulk oxide trap generation, indicated by SILC, in small-area devices through the time-dependent relaxation of the drain and gate current shifts after NBTI stress. The study of BTI in small-area devices is crucial to understand the trapping characteristic of the individual defects in view of the severity of the individual defects on the performance of the small-area devices, especially static random-access memory (SRAM), which dominates the microprocessor chip and has a bit cell size that scales by ~50 % for every new technology node [60].

Furthermore, it is important and worthwhile to study the reliability of high- k oxides since SiO₂ and SiON gate oxides were substituted by high- k gate dielectric for technology nodes of 45 nm and beyond. The previous experiments that were reported mostly focused on the effect of direct-current (DC) stress on the dynamic PBTI in HKMG stacks n -MOSFETs. Most of the time the circuits or transistors are operated in alternating-current (AC) bias environment, i.e. the MOSFETs in the ICs are repeatedly toggled between “on” and “off” states. Hence, it is critical as well as intriguing to study the effect of the frequency or AC stress on HKMG stacks MOSFETs in terms of the transformation of shallow-to-deep electron-traps to further understand the physics and failure mechanisms of the high- k gate dielectric.

1.3 Objectives

The objectives of this thesis are listed below based on the motivation expressed in Section 1.2.

- (1) To investigate the individual SOTs (R) behavior and its conversion to permanent charged traps (P) in small-area device under NBTI stress.
- (2) To scrutinize the correlation between the R -to- P transformation of oxide traps and the generation of bulk oxide traps based on temperature dependence, activation energy, and power-law time exponent.
- (3) To examine the correlation between hole trapping denoted by the degradation of the drain current and the bulk oxide trap generation indicated by the SILC, in small-area devices through time-dependent relaxation of the drain and gate current shift after NBTI stress.
- (4) To study the impact of VAS on time-zero SOTs at operating condition. The switching (capture/emission) activity and the drain current degradation amount for these SOTs at operation condition are analyzed before and after VAS application.
- (5) To study the BTI effect on the nature and properties of individual SOTs in small-area devices that would serve as evidence to understand the fundamental roots of degradation caused by individual traps. It would also aid the counterchecking of the experimental results previously reported for large-area devices, significantly contributing to the understanding of the physics behind the degradation of the gate oxide due to the BTI effect.
- (6) To explore the frequency-dependence of electron traps in HKMG n -MOSFETs under PBTI stressing.

1.4 Scope

There are seven chapters in this report and the description of each chapter's contents is given below.

Chapter 1 presents the background, motivation, objectives and scope of this thesis.

Chapter 2 explores the fundamental knowledge and history of BTI. It also includes recent studies on the recoverable and permanent components of NBTI and PBTI degradation in SiON and HKMG MOSFETs as well as recent studies on small-area devices.

Chapter 3 provides the measurement methods employed in the experiments reported in this thesis such as the ultra-fast switching (UFS) method, Keithley's Ultra-Fast BTI Package and stress-induced leakage current (SILC) measurements. In addition, the history and various types of fast measurements are presented as well.

Chapter 4 investigates the behavior of switching oxide traps recovery under DNBTI. The experimental evidences for the correlation between the recoverable-to-permanent transformation of switching oxide traps and generation of stable oxide bulk traps in large-area *p*-MOSFETs based on temperature dependence, activation energy, and power-law time exponent are discussed. Further, results from the examination of switching-to-permanent transformation of hole trapping in small-area devices subjected to the DNBTI test are elaborated. Moreover, the outcome of the study of the correlation between hole trapping, indicated by degradation of drain current, and the bulk oxide trap generation, denoted by SILC, in small-area devices through the time-dependent relaxation of the drain and gate current shift after NBTI stress is discussed.

Chapter 5 focuses on the impact of voltage-accelerated stress (VAS) on time-zero switching oxide traps under operating conditions. The results obtained from different gate voltages used in the VAS steps to study the change in the switching activity of oxide traps in the operating environment are analyzed. Effects of the post-VAS relaxation on the SOTs are demonstrated as

well. Also, the observation of the change in drain current degradation caused by individual defects after the VAS step is discussed.

Chapter 6 examines the frequency-dependence of electron traps in HfO_2/TiN n -MOSFETs under PBTI stress. The possible explanation for the experimental observation is discussed.

Chapter 7 summarizes the findings reported in this thesis and recommends some possible studies for the future.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction to Bias Temperature Instability

The lifetime of MOSFETs is constrained due to its inevitable degradation during operation. Bias temperature instability (BTI) is one of the causes contributing to the critical degradation, limiting circuit performance. As the name implies, when MOSFETs are subjected to a high electric stress field at an elevated temperature, systematic degradation is observed in the device parameters, such as drain current, threshold voltage, subthreshold slope, transconductance, channel mobility, *etc.* [32]. The threshold voltage of MOSFETs is the most important device parameter and allows the gate voltage to “switch” the MOSFET on or off. Ideally, the threshold voltage of MOSFETs in a circuit is presumed to be unchanged within an extremely small designed range of values. In general, an IC is deemed to be malfunctioning when the threshold voltage shift is more than 50 mV or the variation of g_m degradation is more than 10% [61]. An IC may not function if the value of the threshold voltage moves out of the designed tolerance range. For instance, the functionality of logic circuits could fail because of the asymmetric degradation in the timing paths owing to the drain current degradation that is caused by the BTI effect. Furthermore, gate overdrive degradation, caused by an increase in the threshold voltage, leads to noise margin lowering in the standard random access memory as well as drain current reduction and frequency degradation in ring oscillators [62]. Therefore, in the semiconductor industry, the device parameter degradation induced by BTI is identified as a severe reliability issue that can cause circuits to malfunction, as there are billions of MOSFETs in an IC chip.

Research on BTI degradation was only an academic subject and was not heeded in the industry prior to the year 1999, since the hot carrier injection (HCI) was a more critical

reliability issue at the time. Kimizuka *et al.* [46] reported that scaling down of the CMOS technology renders NBTI degradation a more serious reliability issue as compared to HCI. The device lifespan would be shorter under NBTI degradation as compared to that of HCI when the gate dielectric thickness is scaled below 3.5 nm based the SIA roadmap shown in Figure 2-1 [46]. Thereafter, many studies have been performed in the academia and industry in order to study the mechanism and impact of BTI as well as determine ways to reduce the BTI effect. As the CMOS technology progressed, the channel length (gate length) reduced and became comparable to the depletion region near the drain and source terminal, causing the short channel effect (SCE) which leads to threshold voltage roll-off, drain-induced barrier lowering (DIBL), and punch-through. The buried channel was then substituted with a surface channel in MOSFETs to ease the SCE [63]. However, the oxide defects and interface traps affect the conducting channel in surface-channel device more easily as compared to the buried-channel device, causing BTI degradation to be more severe in surface-channel device. Moreover, the gate oxide thickness shrank aggressively from 95 nm in 1970's to less than 2 nm in the latest technology in order to attain low power operation and high current drive. However, in order to maintain the circuit operation margin, the supply voltage could not be lowered proportionally, resulting in a drastically increased

Figure 2-1. The transition of the MOSFET lifespan limitation mechanism as a function of gate dielectric thickness. BTI is the major limiting factor as compared to HCI for device lifespan when the gate dielectric is less than 3.5 nm [46].

electric field in the thinner gate oxide even under normal operating conditions. The operating temperature increases as well due to the existence of more devices per unit area, a direct result caused by the scaling down of MOSFETs. Both these factors, the very high electric field in the gate oxide and the elevated operating temperature significantly enhance the severity of BTI degradation [64]. Furthermore, the introduction of nitrogen into gate oxide, which was intended to minimize the gate leakage and lessen the penetration of boron from the p^+ -polysilicon gate through the gate oxide into the channel and silicon substrate, had the side effect of promoting NBTI degradation [31], [39], [65], [66].

As the name implies, bias temperature instability investigation utilizes the electric field (bias) and temperature stressing in an accelerated life test to study the BTI effects on MOSFETs' behavior. For this purpose, the stressing temperature is usually set in the range of 50 – 250 °C, while the electric field stress that is applied across the gate oxide is generally in the range of 6 – 10 MV/cm because 6 MV/cm is the operating oxide field and 10 MV/cm or more could cause hot carrier degradation, which is another reliability issue that should be excluded in the BTI study [67]. According to the applied gate bias polarity, BTI can be classified into two categories:

- (1) negative-bias temperature instability (NBTI) – when negative gate bias is applied, and
- (2) positive-bias temperature instability (PBTI) – when positive gate bias is used.

Practically, NBTI is studied on p -MOSFETs since the negative bias is required to turn on the p -MOSFETs. Similarly, PBTI is investigated on n -MOSFETs, as n -MOSFETs operate under the positive gate biasing condition. Figure 2-2 illustrates the degradation observed upon the application of 10 kiloseconds (ks) NBTI stress at the elevated temperature of 125°C on a typical p -MOSFET, where the threshold voltage increases while both the drain current and maximum transconductance decrease in comparison to the fresh (i.e. unstressed) condition [30].

The BTI research works on CMOS with SiO₂ and SiON gate oxides reveal that NBTI degradation in the *p*-MOSFET is much worse than PBTI degradation in the *n*-MOSFET, which was deemed to be insignificant [30]. One of the possible reasons behind this is that the threshold voltage shifts in *n*- and *p*-MOSFETs are influenced differently by the interface traps being donors and acceptors in the lower half and upper half of the bandgap respectively [68]. Hence, the investigation of NBTI degradation in *p*-MOSFETs with SiO₂ and SiON gate oxides is widely researched. Nonetheless, as the technology advances, the HKMG stack is used for sub-45 nm technology nodes. PBTI degradation in *n*-MOSFETs is equal or more critical than NBTI degradation in *p*-MOSFETs [45]. Therefore, studies on both NBTI and PBTI degradation on *p*- and *n*-MOSFETs, respectively, are equally vital.

To demonstrate the possible physical origins of the BTI phenomenon, the NBTI in a *p*-MOSFET is used in the following example, but similar approaches can be applied for PBTI on an *n*-MOSFET. The MOSFET operates by controlling the charges in the channel through the gate. The threshold voltage of a MOSFET is defined as the gate voltage that generates a potential at the Si surface for which deep inversion sets in at the interface between the oxide layer and the substrate of the transistor, effectively “inverting” the conductivity type of the

Figure 2-2. The I_d and g_m versus V_g curves of a 2 nm SiON *p*-MOSFET in the linear regime at the fresh condition (solid line) and after 10 ks NBTI stress (dashed line) at 125°C [30].

substrate [69]. The threshold voltage, V_{th} , of a p -MOSFET is expressed by Eqn. 2-1.

$$V_{th} = V_{FB} - 2\phi_B - \frac{|Q_B|}{C_{ox}} \quad (\text{Eqn. 2-1})$$

The three parameters in this equation, the flatband voltage V_{FB} , the Fermi potential ϕ_B and the bulk charge density Q_B , are given below:

$$V_{FB} = \phi_{MS} - \frac{Q_f}{C_{ox}} - \frac{Q_{it}(\phi_S) + Q_{ot}}{C_{ox}} \quad (\text{Eqn. 2-2})$$

$$\phi_B = \frac{kT}{q} \ln \frac{N_D}{n_i} \quad (\text{Eqn. 2-3})$$

$$|Q_B| = \sqrt{4q\epsilon_s\phi_B N_D} \quad (\text{Eqn. 2-4})$$

where C_{ox} is the oxide capacitance per unit area, ϕ_{MS} is the difference between the work function of the metal and semiconductor, N_D is the doping concentration in the bulk of the semiconductor, n_i is the semiconductor's intrinsic carrier concentration, ϵ_s is the permittivity of the semiconductor, Q_{ot} is the oxide charge density, Q_{it} is the interface trap density, and Q_f is the fixed charge density. The terms and positions of the variety of the charges (traps and defects) that are found in a thermally grown SiO_2 on a silicon substrate are denoted in Figure 2-3 [11].

Figure 2-3. Various types of defects (charges) at different locations in a thermally grown SiO_2 on a silicon substrate [11].

It can be observed from Equations 2-1 to 2-4 stated above that there are several parameters that contribute to the value of V_{th} in a MOSFET. The oxide thickness and substrate doping density are usually not affected by the NBTI stress [70], while the Q_f , which is commonly linked to process conditions that grow a gate dielectric layer, is deemed insignificant in the advanced MOSFET technology. Hence, Q_{it} , in which the interface trapped charge occupancy is dependent on the surface potential ϕ_s , and Q_{ot} are the only possible parameters that could engender a threshold voltage shift, ΔV_{th} , in a p -MOSFET during operations. This is conveyed in Eqn. 2-5.

$$\Delta V_{th} = - \frac{\Delta Q_{it}(t) + \Delta Q_{ot}(t)}{C_{ox}} \quad (\text{Eqn. 2-5})$$

Based on Eqn. 2-5, either or both of Q_{it} and Q_{ot} are believed to increase in order to cause a negative threshold voltage shift as the NBTI stress progresses.

2.2 Interface Traps

The existence of interface traps between the gate oxide and the silicon substrate is due to dangling bonds, interfacial lattice mismatch, introduction of foreign impurities on the silicon surface during the manufacturing process, radiation, and bond-breaking processes. The generation of interface traps at the Si/SiO₂ interface could be accelerated by NBTI stress, promoting NBTI degradation. In the ideal bulk substrate, each Si atom is bonded to four neighboring Si atoms in a tetrahedral manner. However, lattice mismatch occurs during oxidation. Most Si atoms bond to oxygen atoms while some Si bonds are left dangling at the surface as shown in Figure 2-4. Hence, the interface traps are interface defect states that are formed due to these unsatisfied bonding configurations, contributing to ΔV_{th} , as denoted in Eqn. 2-5. Passivation of the dangling bonds via annealing in the hydrogen forming gas is performed during fabrication. However, the Si-H bonds would break under nominal stress conditions, as the bonding energy of the Si-H bond is relatively weak [71], which would lead to the regeneration of the interface traps, causing ΔV_{th} . The defect dominating the

interface traps at the Si/SiO₂ interface is called P_b center. As shown in Figure 2-4, this P_b center [72]–[75] is a $[\text{Si}_3\equiv\text{Si}\cdot]$ center having unbounded central-atom orbital perpendicular to the interface and directed into a vacancy in the oxide immediately above it. Two P_b center variants, P_{b0} and P_{b1} centers, which are detected through electronic spin resonance (ESR) measurement, have been recognized to pose a significant impact on NBTI degradation [76] and the properties of these P_{b0} and P_{b1} centers are still under investigation [77], [78].

2.3 Switching Oxide Traps

In addition to the interface traps, the oxide traps, which are also known as border traps, slow states, and anomalous positive charges [79], [80], were discovered in the gate dielectric to be another prime cause of the BTI degradation that contributes to the threshold voltage shift. The oxide traps, being generally electrically neutral in pristine devices, can be charged positively or negatively by capturing holes or electrons, respectively, via tunneling during the on-state of devices. Usually, hole trapping occurs in a p -MOSFET under NBTI, while electron trapping occurs in an n -MOSFET under PBTI. The hole trapping mechanism can be used as an example to illustrate the trapping mechanism of a switching oxide trap, as schematized in Figure 2-5 [81]. During the operation or stressing of a p -MOSFET under negative gate voltage, the oxide traps in the gate dielectric capture the holes from the conducting channel through tunneling. The trapped holes would be subsequently released

Figure 2-4. Schematic illustration of the interfacial states on (a) silicon (111) and (b) silicon (100) surfaces [32].

through tunneling during the recovery stage voltage at zero or flatband voltage and will be injected later into the n -type substrate.

The “switching” oxide trap behavior portrayed in the E'_{γ} center, which is a hole caught at an oxygen vacancy, in the SiO_2 gate oxide was revealed in [82]–[87]. Further, more paramagnetic defects called K_N centers were discovered in the SiON gate dielectric and may be possible contributors to the switching oxide traps [88], [89]. This was supported through deep-level transient spectroscopy (DLTS) measurements, revealing that nitride oxides SiON have a greater amount of hole traps than the thermal oxide SiO_2 owing to the oxygen deficiency that is caused by the removal of oxygen during the nitridation process [90].

Generally, it is very challenging to differentiate between the oxide traps and interface traps spatially for the time range of practical measurement following the down-scaling of the gate oxide thickness to several angstroms [91].

2.4 The Reaction-Diffusion (R-D) Model and its Discrepancies

In 1977, Jeppson and Svensson proposed the Reaction-Diffusion (R-D) model (a

Figure 2-5. Schematic illustration showing hole trapping under NBTI stress (left) and hole emission during relaxation at flatband voltage (right). The holes from the conducting channel during NBTI stressing may tunnel into deep localized states (a) below or (b) above the substrate Fermi level. Holes captured may tunnel back into the substrate during recovery stage [81].

hydrogen transport model) to describe NBTI degradation [92], and this model has been widely cited to expound the NBTI phenomenon [71]. Being solely developed based on the interface trap, the R-D model was mainly employed to elucidate the interface trap generation mechanism at the Si/SiO₂ interface, which led to threshold voltage degradation in the state-of-the-art MOSFET. This R-D model is a two-stage model where the first stage is the generation of the interface traps due to the breaking of the Si-H bonds while the second stage is the diffusion of the hydrogen-related group that escaped in the first stage.

A schematic diagram of the R-D model is illustrated in Figure 2-6. At pristine state, Si-H trap precursors, formed by passivating the dangling bond during the hydrogen forming gas annealing, can be found at the Si/SiO₂ interface (Figure 2-6(a)). Upon the application of negative gate stress, some of the energetic inversion holes generated in the channel are seized by the weakly passivated Si-H bonds at the Si/SiO₂ interface through field assisted tunneling. This yields the Si dangling bonds (P_b centers) and hydrogen at the interface after the electrochemical reaction between the Si-H bonds and the holes that were captured from the inversion layer (Figure 2-6(b)). This interface trap generation is the first stage of the R-

Figure 2-6. Schematic diagram of the R-D model representing Si/SiO₂ interface, showing: (a) the $\equiv$ Si-H defect precursor under fresh state, (b) the electrically-activated $\equiv$ Si-H defect precursor under the NBTI stressing generating the interface trap where the released hydrogen could react with SiO₂, forming an oxide trapped charge and a hydroxyl group, and (c) the hydroxyl group diffusing away from the interface to the oxide, which is a reaction-limiting factor [92].

D model. The escaped hydrogen reacts with the SiO₂, forming a hydroxyl group, which diffuses away from the Si/SiO₂ interface into the SiO₂ bulk, and an oxide trapped charge near the interface. The latter is responsible for the threshold voltage variation ΔV_{th} and lower transconductance of the MOSFETs (Figure 2-6(c)). This diffusion process is the second stage of the R-D model. Upon removal of the gate stress in the recovery process, those nearby available hydrogen atoms instantly re-passivate the Si-H dangling bonds while those hydrogen atoms that are far apart from the interface diffuse back slowly and recombine with the dangling Si-H bonds at a slower rate.

The R-D model representation of the expressions is as follows:

where $X_{\text{interface}}$ represents the species released from the interface.

From the R-D model perspective, the NBTI degradation is manipulated by the hydrogen diffusion process, as the hydrogen remnants at the interface may repassivate the dangling silicon bonds. The R-D model, which is based on the Si-H bond breaking and hydrogen diffusion, has been extensively fine-tuned [93]–[95] to explain many experimental observations about the NBTI observed based on the “slow” stress-measurement-stress method. These observations include the following:

- (1) NBTI-induced threshold voltage degradation represented by $\Delta V_{th} \sim A \exp(-nE_D/kT) t^n$ with $n \sim 0.25$ and $E_D \sim 0.5$ eV [93], [96];
- (2) Self-annealing of a portion of interface traps upon removal of the NBTI stress [95]–[97];
- (3) Revealing the dispersive transport characteristic of the NBTI temperature dependence [98], [99];

(4) The discovery of a long-term decreasing of the NBTI time-exponent n from 0.25 to 0.13-0.16 [50], [100].

Recently, as NBTI studies advanced, it became clear that the R-D model could not explain some of the new experimental results. Study in [101] reported that depassivation of Si-H bonds by protons released in the Si substrate is consistent with the observed interface traps generation and the spontaneous Si-H bond breaking step is not energetically favourable. The recovery of the NBTI is fast and almost spontaneous upon the removal of stressing; hence, different exponents are obtained due to variations in the measurement delay. Longer measurement delays artificially yield larger power-law exponents, whereas fast measurement with shorter measurement delays generally harvest a much smaller power-law exponent of ~ 0.1 [102]. The temperature dependence of the power-law exponent reported [98] is also inconsistent with the R-D model. Additionally, the different polarity dependences of the reported NBTI recoveries [38] are incoherent with the R-D model, which states that the NBTI recovery behavior is independent of the gate voltage polarity applied. However, NBTI recovery is accelerated under a positive gate bias. Furthermore, experimental results demonstrate that the recovery has a logarithmic time dependence over a very broad range of time [103], inconsistent with the simulation results of the R-D model in which the recovery lasts for about three decades of time. Moreover, the interface traps hardly recover based on the charge-pumping measurement [30], contradicting the R-D model explanation about re-passivation of the interface traps during NBTI recovery. In addition, the NBTI recovery in each relaxation interval is constant in the dynamic NBTI (DNBTI) test, independent of the number of DNBTI stress/recovery (SR) cycles test conducted on the devices [104]–[106]. The parametric shift induced in the stress phase is nearly the same to that recovered during the relaxation phase in each DNBTI stress/relaxation cycle after the quasi-saturation of the built-up interface trap generation, i.e. the superimposition of the cyclical parametric fluctuation on a background of relatively

permanent quasi-constant interface degradation [105]–[107]. This cyclical parametric fluctuation supports the explanation established on the hole capture and emission of a group of oxide defects called switching oxide traps that may consistently switch between the neutral and positively charged state. A contemporary study based on time-dependent defect spectroscopy illustrates the persistent capture and emission behavior of individual oxide traps in small area MOSFETs [47], [108].

2.5 Hydrogen Species in Bulk Oxide Trap Generation and its Discrepancies

The defect generation in SiO_2 and at the SiO_2/Si interface is still widely deemed to be closely attributed to hydrogen in view of the numerous superb past studies such as [36], [109], even though the outcome of DNBTI studies evidently do not favor the conviction that NBTI recovery is based on the hydrogen species back diffusion. One of the mechanisms proposed for bulk oxide trap generation is anode hydrogen release (AHR) [110]. Constant gate voltage, V_g , stressing applied for time-dependent dielectric breakdown (TDDB) testing may cause the hydrogen species to be liberated from the anode interface when the “bombardment” of energetic hot carriers on the interface dissociates the Si-H bonds [111]. The core of AHR is that the bulk oxide trap generation is propelled by interface trap generation. Since interface trap generation has been elucidated to be one of the main factors for NBTI [30] and as NBTI stress is quite similar to TDDB stress (except for the smaller V_g magnitude), some bulk oxide traps are expected to be created by the NBTI stress. Some studies [112], [113] asserted the consistency with the AHR model after noticing a raise in the gate current, which is SILC, suggesting bulk oxide trap creation upon the application of NBTI stress. It is believed that Si–H bonds at the Si interface and holes in the inversion layer participate in an electrochemical reaction that causes the hydrogen species release [92], [114]. Also, the hydrogen species release may be caused by the Si–H bonds at the Si interface reacting with the protons released in the Si substrate [101].

If bulk oxide trap generation occurs during the application of NBTI stress, the parametric variation degree under the pulsed gate condition should increase with stress time. This is due to the larger amount of bulk oxide defects leading to more captured holes during NBTI stress and correspondingly more trapped holes would be released during relaxation. Nonetheless, this standpoint does not match with the results from the DNBTI investigation [115]. A relative permanent non-negligible amount of V_t shift was observed after 30 stress/relaxation cycles (Fig. 3 in Ref. [115]) and if this entire V_t shift were to be ascribed to interface trap creation caused by Si-H bond dissociation, a significant amount of hydrogen species would have been liberated in such a thin oxide layer. Additionally, their interaction with the oxide network would have resulted in bulk oxide trap creation. Intriguingly, there was no increase in the gate current (Fig. 4 in Ref [115]). Further, the V_t shift and recovery after each stress and relaxation interval, respectively, were equal and constant throughout the DNBTI test. This indicates that there is no alteration in the hole trapping level at the oxide defects during the DNBTI test. The absence of SILC supports the constant V_{th} recovery, implying that no additional bulk oxide defects were generated. This was further endorsed by the absence of a clear variation in drain-current hysteresis, which was uncovered by pulsed I-V measurement (Figs. 1 and 5 in Ref. [115]).

Whilst the evidence from [115] does not favor the role of hydrogen in oxide trap generation, other investigations show that hydrogen species participates in NBTI degradation. For example, a SiON *p*-MOSFET with post-metallization annealing in hydrogen exhibits greater interface trap generation under NBTI stress as compared to that of annealing in deuterium [112]. This result coincides with the giant isotope effect suggested in elucidating the decrease of interface trap creation in *n*-MOSFETs with deuterium annealing during the channel hot-electron stress application [116], [117]. Deuterium atoms travel away from the dangling Si bonds at a slower rate, since the deuterium isotope is heavier than the hydrogen isotope. Thus, deuterium atoms have a

greater likelihood of reforming the bonds with the dangling Si bonds, leading to lower net interface trap generation. Moreover, reduction in the Si p -type substrate depletion capacitance ensuing NBTI stress may be due to the deactivation of the boron dopant in the channel caused by the released hydrogen through the boron-hydrogen complex formation, indicating hydrogen transport across the MOS structure [70], [118], [119].

The discrepancy in the outcome mentioned above between study [115] and other studies [112], [70], [118], [119] may be due to the minimization in the degree of interaction between hydrogen and the oxide network in the ultra-thin gate oxide, as the hydrogen is quite mobile in SiO₂. Hence, while part of the interface degradation may be related to hydrogen, it does not contribute to a notable positive charge trapping impact in the ultra-thin gate oxide. Moreover, the discrepancy mentioned may be due to the decrease in the significance of hydrogen owing to replacement of Si-H bonds with stronger Si-F and Si-N bonds through the incorporation of fluorine via BF₂ drain/source implantation and nitrogen respectively. No pronounced P_b defect center creation was detected by ESR measurement in the SiON gate oxide subjected to NBTI stress, unlike the occurrence in SiO₂. Instead, a k_N defect center, i.e. hole-trapping site, was detected via ESR measurement, implying that hole trapping governs the SiON gate p -MOSFET degradation rather than interface trap generation. Other studies [38], [120]–[122] also contributed supportive evidences demonstrating the minor contribution of stress-induced interface trap generation on the total V_{th} shift through the high-frequency charge pumping measurement.

In short, while several investigations have proposed the possible role of hydrogen in NBTI [112], [118], its implication seems to be relatively less significant in the state-of-the-art, ultra-thin SiON and high- k gate oxide, in which the degradation is dominated by charge trapping [30], [36], [38]–[45], [47], [104]–[108], [115], [120]–[124], [80].

2.6 *R* and *P* Components under NBTI

As mentioned in the previous sections, NBTI degradation is attributed to two kinds of defects – interface traps and switching oxide traps. The switching oxide traps in the gate dielectric may capture holes from the conducting channel during the operation of a *p*-MOSFET subjected to negative gate voltage.

Studies reported that NBTI degradation can be categorized into two components: a recoverable (*R*) component in which the damage can be reversed and a relatively slow recovering or permanent (*P*) component [49], [50]. The permanent component refers to the defect that fails to reinstate to its initial state within the experiment time frame. For instance, an oxide trap in a *p*-MOSFET gate dielectric capturing a hole during NBTI stress is deemed to be a permanent component when it fails to emit the hole upon the removal of stress during a given relaxation period during an experiment. A huge part of the NBTI degradation could be reversed upon stress termination, especially since most of the damage recovers instantaneously when this occurs [49], [50]. As the technology node advances, causing the scaling down of gate oxide thickness, it becomes challenging to distinguish between the bulk oxide defects and interface traps. Thus, the terms *R* and *P* components substitute the bulk oxide and interface defects.

Generally, interface traps P_b centers are believed to be the major contribution of the *P* component [125] while the *R* components are switching oxide traps, including oxygen vacancy E' centers ($O_3 \equiv Si^+ \cdot Si \equiv O_3$) [87], [126], and nitrogen-related K_N centers ($N_3 \equiv Si^+ \cdot Si \equiv N_3$) [89]. Some studies reported that the *R* and *P* components are uncorrelated, and thus, each component is attributed to its own individual mechanism. H. Reisinger *et al.* [127] reported that during the application of NBTI stress, degradation caused by bulk hole trapping saturates within 10 ms and the degradation after this is interface trap generation based on diffusion-limited R-D model. Further, V. Huard's study [128] proposed that NBTI degradation is a combination of two independent *P* and *R* components. The *P* component,

which is constituted of an equal proportion of positive fixed charges and interface traps, is explained by the reaction-limited hydrogen transfer that is favored by the hole trapping on the oxygen bridge in the dielectric close to interface. In contrast, the R component, which is made of hole trapping/detrapping processes for the pre-existing defects attributed to oxygen vacancies and nitrogen-related centers in the SiON dielectric, is described by the field-assisted lattice-relaxation multi-phonon emission (LRME) model.

On the other hand, some researchers believe that there is a correlation between the R and P components that contribute to NBTI degradation in p -MOSFETs. Teo and Ang *et al.* [104]–[106], [129] revealed that there are fast and slow recoverable components in the SiON gate dielectric p -MOSFETs during the 40-cycle dynamic NBTI (DNBTI) recovery phase. The DNBTI degradation is attributed to the “switching” hole-trapping mechanism being superimposed on a relatively permanent interface-state generations and deep-level hole-traps. The recoverable threshold voltage shift, $|\Delta V_{th}|$, attributed to switching hole traps, is constant regardless of the number of DNBTI stress/recovery (SR) cycles. This outcome [104]–[106] and the R-D model do not match and this has been discussed in Section 2.4. Moreover, Boo *et al.* [52] discovered the phenomenon of the transformation of switching hole traps (R component) into permanent hole traps (P component) as well as the correlation between the converted permanent hole traps and the generation of bulk oxide traps in SiON gate dielectric p -MOSFETs. A decrease in the recoverable $|\Delta V_{th}|$, indicating R -to- P transformation, is accompanied by an increase in SILC, which indicates bulk oxide trap generation. Hence, it is proposed that these two events are correlated and stem from the same physical mechanism. Furthermore, Boo *et al.* [51] also presented that the increase in temperature yields a similar observation of the existence of a direct correlation between the generation of SILC and the reduction of NBTI recovery in the DNBTI test in SiON gate dielectric p -MOSFETs. The conversion of switching oxide traps into permanent hole traps is promoted by the higher temperature. This is because the defect sites and phonons interact

with greater intensity at a higher temperature, which boosts the structural change tendency at the trap sites. These two studies [51], [52] have addressed the discrepancy in the post-NBTI stress SILC generation discussed in Section 2.5. Further experimental evidence for the correlation between switching-to-permanent hole trapping conversion and bulk oxide trap (SILC) generation based on temperature, activation energies, and power-law time exponents will be presented in Section 4.2.

The extended emission time of trapped holes in a direct tunneling SiON gate oxide and the gate-polarity dependence of NBTI recovery [121], [122] implies that energy levels of

Figure 2-7. (a) Schematic energy band diagram illustrating the generation of deep level hole traps (energy states above the n -well conduction-band edge E_C) caused by NBTI stressing. Shaded regions represent the energy distribution of the stress-induced hole traps. A possible hole-trap generation mechanism involves thermal- and field-assisted dissociation of the oxide defect precursors and subsequent loss of electronic charge to the conduction band of the n -well. (b) Shallower hole traps (i.e., those below E_C) are instantaneously discharged by electron tunneling from the n -well upon the termination of stress; deep-level hole traps, however, do not relax. (c) A positive gate bias lowers the hole-trap states, causing more relaxation, but some fraction may still remain since their energy states are pinned by the Si/SiO₂ conduction-band offset [122].

some of the hole traps are strategically located at a position where electrons from Si encounter a significant barrier (Figure 2-7). These deep-level hole traps may be attributed to a conversion process similar to that reported in [52].

A progressive reduction of NBTI recovery is also noticed in high- k dielectric/metal gate (HKMG) stacks (HfO_2/TiN and HfSiON/TiN) p -MOSFET under DNBTI test [54], [55]. Similar to the SiON p -MOSFET, the recoverable-to-permanent hole-trapping transformation and the correlation between this transformation and SILC generation are also observed in the HKMG stacks during NBTI stressing [54]. Nonetheless, the extent of hole-trap conversion differs with the gate dielectric type [55]. Furthermore, the “threshold” temperature and oxide field beyond which the beginning of reduction in NBTI recovery are different for different gate dielectrics, which may be attributed to the different ionicity of

Figure 2-8. The two-stage model for a switching oxide trap coupled with the creation of a dangling bond at the interface. When the E' center is positively charged (in state 2), the hydrogen passivating a silicon dangling bond at the interface can move to the E' center, thereby effectively locking in the positive charge (state 4). The charge state of the created dangling bond depends on the position of the Fermi-level [131].

the gate dielectrics.

In addition, T. Grassler *et al.* [130] suggested that hole trapping and defect generation act in a coupled fashion based on the data showing the initial log-like degradation, followed by a power-law-like degradation. Additionally, the stress degradation at the examined temperatures and voltages can be perfectly scaled to a universal form, making the contribution of the two independent mechanisms for hole trapping and defect generation highly unlikely. Hence, a two-stage model (extending the Harry Diamond Lab (HDL), which is more recently known as Army Research Lab, switching trap model [126] with an E'/P_b -H complex) based on the established properties of E' centers was proposed for NBTI as shown in Figure 2-8 [131]. At stage one (HDL switching trap model), holes are trapped in near-interfacial oxygen vacancies through the multiphonon-field-assisted tunneling during the application of NBTI stress. At the second stage, the increased hole concentration considerably enhances the creation of poorly recoverable defects such as P_b -centers in SiO_2 layers and K_N -centers in oxynitrides. Also, the hydrogen exchange and defect reversibility were also discovered in high- k oxide stacks ($\text{Si/SiO}_2/\text{HfO}_2$) [132]. A defect complex, which comprises a suboxide Hf-Si bond and an interfacial dangling bond (P_b center), facilitates the hopping of H atoms across the Si-SiO₂ interface and promotes the H shuttling between the two minimum-energy wells of a single defect complex under the influence of switched-bias and temperature stress. This mechanism contributes to creation of interface traps and leads to oxide-trapped charge. Moreover, T. Grassler *et al.* [133] suggested that there is a metastable state 2' (Figure 2-9) that acts as a low-pass filter between the neutral precursor

Figure 2-9. A simplified transition state diagram of a defect model describing the transition of a defect from precursor state 1' to positively charged state 2 through metastable state 2' [133].

state 1 and positively charged state 2 to explain the results of the NBTI frequency dependence.

2.7 *R* and *P* Components under PBTI

In recent years, considerable attention has been paid to the PBTI effect on HKMG stack due to its critical impact on *n*-MOSFET reliability and performance [41]–[45], [134]. However, in contrast to NBTI, interface trap generation is insignificant for PBTI. The PBTI degradation is virtually attributed to electron trapping in high-*k* oxide defects [43]. There are numerous similarities between electron trapping under PBTI and hole trapping under NBTI:

- (1) Rapid increase of stress induced ΔV_{th} at the initial stage, followed by an asymptotic increase indicates the presence of both slow and fast electron trapping [44];
- (2) High-order time span of post-stress ΔV_{th} relaxation implies the existence of both deep (slow detrapping) and shallow (fast detrapping) electron traps [42];
- (3) Gradual decrease in ΔV_{th} recovery under cyclical alternating stress/relaxation phases when the employed stress surpasses a certain threshold suggest the occurrence transient-to-permanent (shallow-to-deep) electron-trap transformation [56].

Similar to the deep hole traps reported in NBTI studies [38], [61], [62], [120], the presence of deep electron traps is uncovered by modifying the polarity of gate voltage V_g halfway through post-stress relaxation. Gao *et al.* [56] revealed that after the significant slowing down of recovery rate indicated by a “plateau” manifested at a gate bias of $V_g = 0$ V, a change in V_g to -1 V triggers another sudden ΔV_{th} reduction or recovery, implying that a portion of trapped electrons is located at relatively deep states. The electrons in deep traps requires a long time to emit under $V_g = 0$ V, owing to the barrier between the Si conduction band edge E_c and the trap energy level, as shown in Figure 2-10. “Raising” the trap level towards E_c by applying $V_g = -1$ V reduces the emission time from these deep electron traps,

leading to an abrupt decrease in ΔV_{th} . The remnant of ΔV_{th} is ~ 0 , indicating that the influence of the interface trap generation is negligible when compared to electron trapping in HfO_2 [43]. The negligible degradation of the SiO_x interfacial layer in the HKMG stack under PBTI stress supports the observation of the insignificant PBTI effect in a SiO_2 gate n -MOSFET. The deep electron traps are much less common in SiO_2 compared to that in high- k dielectrics unless very long stress times and very high stress fields are applied on the SiO_2 . In addition, Gao *et al.* [56] reported that the electron detrapping is relatively constant and independent of the dynamic PBTI stress/relax cycle at the low oxide field of 5.5 MV/cm suggesting that similar groups of electron traps are always undergoing charging and discharging. Nonetheless, a gradual reduction in the recoverable ΔV_{th} observed at the higher oxide fields of 7 MV/cm implies that a portion of the electron trap states (R component) are transformed into deeper levels (P component), enlarging the barrier for electron emission from the trap states into the Si conduction band (Figure 2-10) and thereby increasing the emission time of the trapped electrons. Unlike NBTI, shallow-to-deep electron-trap transformation under PBTI is not accompanied by a correlated SILC generation, suggesting

Figure 2-10. A schematic energy band diagram of the HfO_2/TiN gate stack under (a) 0 V and (b) -1 V gate bias. An electron trapped at an energetically deep oxide trap could not be readily emitted due to the energy barrier ϕ between the trap level and the Si conduction band edge under $V_g = 0$ V relaxation. A negative gate bias ‘raises’ the trap level reducing the barrier ϕ which allows the trapped electron to be now easily emitted into the Si conduction band. [56]

that distinct defects are active under PBTI [56]. These experimental results exhibiting shallow-to-deep electron-trapping transformation under PBTI stressing were explained by Gu *et al.* using a vacancy-interstitial (V_O - O_i) defect pair model [135]. In this model, a V_O - O_i defect pair may evolve into a more stable state, with O_i migrating from its original location to a different lattice vacancy position, marked by a substantial reduction in the system energy at the end of the O_i migrating process and corresponding to an electron-trap state shifting from a shallow to a deeper energy level. In addition to explaining the shallow-to-deep electron-trap state conversion, this model addresses the long-term PBTI recovery involving the reverse transformation of an electron-trap state from a deep state to a shallower level with the simultaneous trapped-electron emission.

2.8 Recent Studies on Single Charge Trapping on Individual Defects in Gate Dielectric

BTI is generally studied on large-area devices, where many defects react concurrently. Hence, the individual defects characteristics corresponding to the BTI effect are averaged out, leading to difficulties in identifying and examining the properties of individual defects. Thanks to the continuous down-scaling of transistors to the nano-scale level and the reduction of the thickness as well as area of gate dielectric above the channel region, the examination of the individual defects that contribute to the observable macroscopic behavior is feasible. These individual defects were first reported in [136]. There are only a handful of defects or SOTs present in the gate oxide layer above the channel region in small-area MOSFETs [47]. Unlike large-area devices, each small-area MOSFET behaves differently due to random dopant distribution (RDD). This RDD causes a non-uniform potential profile in the channel, leading to carrier conduction through the percolation paths. In a small-area MOSFET, a single trapping event of a defect located above the constriction point may block the percolation path of a carrier conduction (Figure 2-11), resulting in a

significant shift in the threshold voltage and severe drain current degradation [48], [137]. These single-trapping events of each individual defects are reflected as steps in the threshold voltage or drain current recovery traces which look like a “staircase” [108], [138], as shown in Figure 2-12 [47]. The clusters in a spectral map (Figure 2-12) which are formed based on the step heights and emission times extracted from the recovery traces are the fingerprints of each of the defects. This defect-identification method is called time dependent defect spectroscopy (TDDS) [47]. The threshold voltage shift or drain current degradation (as indicated by the step height in the recovery traces) that is induced by each defect varies depending on the interaction of the defect charge with the underlying random discrete dopants in the channel [139]. Further, the behaviors of the individual defects are analyzed through the capture and emission mean time constants that indicate the probability of the occurrence of the stochastic trapping/detrapping events of individual defects [47], [59], [108], [140]. This charging/discharging of individual oxide defects in MOSFETs is also reflected as random telegraph noise (RTN) or random telegraph signal (RTS) which have been studied since the 1980’s [141]. It was discovered that the RTN and NBTI phenomena are attributed to the same defects and mechanisms [108], [142], [143], [144]

Figure 2-11. (a) Conduction channel with a non-uniform potential profile induced by RDD. The carriers conduct through the percolation path during the “ON” state, as indicated by the white lines. (b) Blocking of a percolation path by an oxide defect locating above the constriction point upon hole-trapping due to NBTI effect, resulting in significant drain current degradation [137].

with RTN being the result of the channel/gate dielectric system being in the state of quasi-equilibrium while the NBTI recovery corresponded to the perturbed system returning to this quasi-equilibrium. Grasser *et al.* [108] reported that the HDL switching trap model [126] (stage one in Figure 2-8) can explain the RTN and the NBTI recoverable component, which are caused by charge trapping in switching oxide defects due to the NBTI stress activating the traps with a larger time constant and the RTN only activating the defects with shorter time constant.

Several characteristics of defect charging/discharging events were revealed in [47]:

(1) Both capture and emission time constants are widely distributed, without diffusion participation (not supporting R-D model), and temperature dependence being consistent with nonradiative multiphonon theory. No temperature-independent elastic tunneling

Figure 2-12. Top: Two examples of threshold voltage recovery traces of a small-area MOSFET ensuing NBTI stress. Each step refers to a hole-detrapping event of a charged oxide trap. Bottom: A spectral map consisting of clusters of points wherein each cluster represent an individual defect (bottom). The clusters in the spectral map which are formed based on the step heights and emission times extracted from the recovery traces are the fingerprints of each defect [47].

process was found [145], [146].

(2) The capture time constant has a strong dependency on the stress bias. However, the extent of the bias dependency of the emission time constant at low voltage is affected by the defect configuration.

(3) Pre-existing oxide traps dictate the subsequent charge trapping.

(4) The presence of metastable states (Figure 2-13) is supported by defect disappearance, transient RTN, frequency dependence [133], as well as uncorrelated capture and emission time constants.

(5) RTN and BTI recoverable components are attributed to the same defects.

Further, a single trace consisting of two subsequent emission events with nearly identical statistics was observed. A defect pair with a similar configuration as well as properties was proposed amongst the three possible scenarios discussed including doubly charged defect, multi-step tunneling and defect pairs [130].

Furthermore, the DNBTI test on small-area devices (Section 4.2) has revealed that a defect can transform from a recoverable component, R , to a more permanent fixed positively-charged state, P , and switching hole traps may exhibit intermittent charging

Figure 2-13. The four states of defect modeled based on E' center properties, illustrating the existence of metastable state [47], [224].

during stress as well as an occasional increase in emission time. These observations were supported by an NBTI study [147] that indicated that the defects can be volatile, i.e. they could be deactivated and re-activated at the same site with the same properties. The defects may become temporarily inactive in both neutral and positive states. Additionally, a portion of these defects may be reactivated, for instance, by a bias pulse toward accumulation (electron injection). Also, the capture time constants of oxide traps can be altered spontaneously or by a bake step [147]. Moreover, Grassler *et al.* reported that their examination of the permanent component, P , revealed the following:

- (1) P is relatively smaller than the recoverable component R ,
- (2) P is primarily attributed to the oxide traps instead of interface traps (P_b centers),
- (3) P has a similar bias dependence behavior as that of R ,
- (4) P is unlikely to dictate device lifespan [148],
- (5) P reaches a saturation value dependent on the gate bias,
- (6) P can be reversed slowly at a constant gate bias, and
- (7) Creation of additional defects at high temperature, which are primarily uncharged around V_{th} , contributes to P at a higher V_g [149].

In addition, post-stress degradation build-up following the NBTI stress was discovered at a longer stress time and higher temperatures with a low stress voltage [150]. Further, the study using the CV measurements reported in [151] suggested that depending on the transistor type and stress bias environment, the created slowly-recovering switching traps are attributed to a combination of oxide and interface traps.

Recently, several models have been proposed to explain the NBTI phenomena and the experimental data observed such as hydrogen bridges and hydroxyl E' centers [152]. However, after discovering the volatility of oxide defects including the R -to- P transformation as well as deactivation and re-activation of the defects at the same site with the same properties [147], the four states of hydroxyl E' centers were suggested over the

hydrogen bridges. This hypothesis was supported by another study [153] based on the defects behavior in amorphous silicon dioxide. Further, Grasser *et al.* proposed the gate-sided hydrogen release model to explain the P component of NBTI [148], [149], the post-stress degradation build-up that follows NBTI [150], and the NBTI/PBTI induced slowly-recovering near-interfacial states generation [151].

It is crucial to conduct a BTI study on small-area devices in order to understand the trapping characteristic of individual defects from the perspective of the severity of the individual defects on the performance of the small-area devices, especially in SRAMs, which dominate the microprocessor chip area and has a bit cell size that scales by ~50 % with each new technology node [60].

2.9 Chapter Summary

This chapter explores the fundamental knowledge, history and recent studies conducted on BTI. The two causes that contribute to BTI degradations, interface traps generation and oxide traps, have been presented. The previous, famous R-D model along with the contributions of hydrogen species to bulk oxide traps generation and its discrepancy have been discussed as well. The amount of recovery under dynamic BTI stressing is constant, independent of the number of stress/recovery cycles. This suggests that charge trapping, which is attributed to the oxygen vacancy or E' center, dominates the degradation. This is inconsistent with the R-D model. Additionally, the discoveries of R and P components as well as R -to- P oxide-trap transformation in large-area devices have been elaborated upon. Furthermore, recent studies on single-trapping of individual defects in small-area devices have been summarized. The study of small-area MOSFETs is critical, as BTI threatens the performance of small-area devices, especially in SRAMs.

CHAPTER 3: BTI CHARACTERIZATION TECHNIQUES

3.1 Introduction

In this chapter, various BTI measurement techniques are discussed. Also, the BTI characterization methods utilized in the experiments that are presented in Chapter 4, 5 and 6 are explained, such as the I_d - V_g measurement that extracts the time-zero threshold voltage and current, the Ultra-Fast Switching (UFS) method that examines the threshold voltage shift under BTI stressing and recovery, the usage of the Ultra-Fast BTI module from the Keithley Automated Characterization Suite (ACS) software that supports the sub-microsecond pulse characterization, and the gate current measurement that tracks the stress-induced leakage current (SILC).

3.2 Motivation for Fast Measurement Methods

At the early stage, static BTI stressing is generally applied to study the BTI effect, in which a constant and continuous BTI stressing is applied to the MOSFET gate terminal with the drain and source grounded at elevated temperature. However, transistors do not always operate under constant gate bias environment in IC. Instead, the MOSFETs inside an operating IC are repeatedly switched between the “on” and “off” states at frequency range of megahertz and beyond. Hence, studying the dynamic BTI of the degradation and performance of the MOSFETs is crucial. In the dynamic BTI study, a significant part of the BTI degradation can be recovered upon the removal of stress [49], [50]. The recovery rate of the BTI is instantaneous and faster upon the termination of stress at first, which contradicts the R-D model that states that the BTI recovery is relatively slow at the initial stage but progressively advances at a much faster rate. This spontaneous and notable BTI

recovery leads to complications in BTI characterization. The BTI degradation observed under direct current (D.C.) measurements is largely underestimated. The device lifespan estimation fluctuates with various time delays between the stress interruption and measurement [50], [154], as shown in Figure 3-1.

BTI experimental data is markedly affected by the fast recovery upon the termination of stress [49], [50]. The results from the NBTI test are sensitive to the measurement delay [50], [154] and the positive gate bias applied [155]. Therefore, a precise measurement, capturing the BTI stress and recovery, the threshold voltage shift ΔV_{th} and the linear drain current shift, is needed to build a complete picture of the underlying NBTI degradation mechanisms and their physical origins. The time delay between the stress termination and measurement should be minimized, as it cannot be totally eliminated because of the limitation of the equipment. Therefore, several fast measurement approaches have been innovated to shorten the measurement delay in order to obtain the BTI results with the greatest possible accurate. [50]

Before the development of fast BTI measurement methods, the conventional DC BTI measurement, which is the “stress – measure – stress” (SMS) method, is commonly used to

Figure 3-1. NBTI degradation as a function of stress time for various delays between stress interruption and measurement under the same characterization method. The device lifespan varies for different measurement delays (data after Ref. [50]).

monitor the degradation caused by BTI stressing by extracting the threshold voltage shift ΔV_{th} . In this SMS method, firstly, a fresh device is measured to acquire the pristine state of the linear I_d - V_g curve by sweeping the gate voltage from 0 V to the operating voltage and applying a small voltage at the drain terminal. After that, the gate terminal of the device is subjected to a stress voltage. During this stress application the stress voltage is periodically interrupted following a series of selected time intervals that is defined by the users to perform the linear I_d - V_g measurements in order to trace the degradation caused by the stress voltage. The measurement delay incurred after each periodic interruption to perform the measurements during the stress application causes a large amount of ΔV_{th} recovery before the actual measurement takes place. The measurement delay lies within the range of tens of seconds to a few seconds. Hence, this measurement delay leads to the underestimation of BTI degradations and a misleading analysis on BTI effect in the device's lifespan.

3.3 Fast Measurement Techniques

Hitherto, many fast measurement approaches have been established to minimize the measurement delay during the BTI recovery upon stress termination in order to accurately measure the fast recovery of BTI [156]. The measurement delay is the interim off-stress time between the BTI stress removal and the start of the device parameter measurements. A

Figure 3-2. A schematic timing diagram illustrating the OTF measurement methodology. I_{dlin} shifts and the associated g_m variations can be tracked (data after ref. [30]).

renowned example is the on-the-fly (OTF) method [157], in which the BTI measurements are taken by monitoring I_d degradation, ΔI_d , and real-time transconductance, g_m , at a gate stress voltage, $V_{g,\text{stress}}$, without terminating the electrical stress to minimize the potential recovery effect. This approach maintains a quasi-constant gate voltage during both stress and measurement periods by adding small pulses around the $V_{g,\text{stress}}$ during the measurement, as shown in Figure 3-2. Through this method, in addition to the monitoring of the linear drain current, $I_{d\text{lin}}$, at $V_{g,\text{stress}}$, $I_{d\text{lin}}$ at $V_{g,\text{stress}} \pm \Delta V_{\text{sense}}$ can be traced, where ΔV_{sense} is the voltage change in the pulsing during the measurement [30]. Additionally, the real-time g_m variation can be evaluated (based on Eqn. 3-1) during the stress application interval, and hence, threshold voltage shift, ΔV_{th} , can be assessed from ΔI_d using g_m .

$$g_m(V_g, t) = \left(\frac{\partial I_{d\text{lin}}}{\partial V_g} \right)_{t, V_{\text{th}}} \Rightarrow g_m(V_g, t) \cong \frac{\Delta I_{d\text{lin}}}{\Delta V_{g\text{sense}}} \quad (\text{Eqn. 3-1})$$

This technique allows the threshold voltage shift to be determined at the stress level necessary to avoid the fast recovery caused by the stress interruption in the conventional SMS method. Nonetheless, OTF is not able to accurately measure the “time-zero” threshold voltage as the reference value, since the device starts to degrade rapidly upon stressing [102]. Hence, OTF returns the higher power-law exponent in the initial stage of stress, and the ΔV_{th} examined at stress voltage is different from the ΔV_{th} obtained in the subthreshold region [102]. Furthermore, the additional degradation that is induced during the measurement taken under stress conditions is not pertinent to normal operation, leading to the overestimation of the BTI impact on the MOSFETs [158]. Additionally, the data resulted from the measurements performed with this gate voltage stress regime are sensitive to systematic errors [159].

Other measurement methods, such as Fast $I_d V_g$ and Ultra-Fast V_t , embrace the technique in which the BTI measurements are captured during the application of an extremely short interruption. This brings down the stress voltage level to a favorable measuring voltage

level and, thus, reduces the chance of inducing unwanted degradations during stress conditions. Ultra-Fast I_dV_g , invented by Kerber *et al* [160] and later improved by Shen *et al* [161], uses pulse generator to inject a single pulse of 2 μ s, rising from 0 to 1 V, at gate terminal and applies a small voltage of 0.1 V at drain terminal to measure the I_d - V_g characteristic in ~ 1 μ s during the rising or falling edge of the pulse. The drain current measured is converted to voltage by using a low noise pre-amplifier and is tracked by the oscilloscope together with the gate voltage. Therefore, ΔV_{th} can be extracted through the generated I_d - V_g curve from the oscilloscope output. The Ultra-Fast V_t method extracts V_{th} from gate voltage using a negative feedback loop with a measurement time that is less than 1 μ s [103].

3.4 Measurement Techniques used in Experiments

3.4.1 Ultra-Fast Switching Measurement Technique

Du and Hu *et al.* developed the Ultra-Fast Switching (UFS) method [162], [163] by combining the customary pulsed I - V setup with a simple linear drain current correction method. It has the ability to minimize the measurement delay down to sub-100 ns and reduce gate pulsing down to 40 ns by using a Keithley 4200 Pulse IV setup. The I_d - V_g curve (transfer curve) is constructed based on the measured drain current within a ~ 50 ns delay at a defined gate voltage. Hence, the whole or part of the transfer curve can be measured with the minimal cumulative recovery effect through this pulsed current-voltage (PIV) measurement method.

The UFS measurement technique leverages on the well-developed commercial PIV setup, as shown in Figure 3-3(a) [162], in which the self-heating of the transistor is minimized during the current-voltage (I-V) measurement, guaranteeing signal integrity at an extremely short pulse-width. Based on the Keithley 4200 Semiconductor Characterization System, the UFS measurement setup consists of a pulse generator (4200-

SCP2), an oscilloscope (4200-SCP2), two source measurement units (4200-SMU), two remote bias tees (4200-RBT) and a 3-port power divider.

The pulse generator (PG) and oscilloscope (OSC) are employed to obtain the rapid transient voltage source and contemporaneous current capturing. The PG injects a single voltage pulse, and the OSC reads the current change due to the injection of said pulse. The two remote bias tees (RBT) are used to couple and decouple the DC and the periodic-pulse signals. Each RBT is connected to the gate and drain terminal of the device under test (DUT), respectively. The RBT at the DUT's gate terminal couples the DC voltage (output

Figure 3-3. (a) A schematic diagram of the UFS setup for NBTI based on the PIV setup. The device under test (DUT) is stressed at a static negative voltage V_g^{dc} . Measurement at a specific interim is performed by injecting a narrow positive pulse of amplitude V_g^p , which momentarily brings the overall gate voltage down to the measurement level. The corresponding change in the drain current is converted into voltage (via a 50 Ω internal resistance) and measured by a high-speed digital OSC. A correction scheme accounts for the change in drain bias during measurement to yield the desired drain current at the measurement drain voltage. (b) Drain and gate voltage waveforms during measurement. Additional pulses of amplitude $V_g^p \pm \Delta$, where $\Delta = 50$ mV) may be added for transconductance measurement [162].

from SMU) and the single voltage pulse (output from the pulse generator) to provide a signal output of the single voltage pulse that is superimposed on the DC bias. The corresponding drain current change induced by the single voltage pulse injection during the BTI measurement is tracked using the pulse current, which is decoupled from the DC current originating from the drain current output by the RBT that is connected at the drain terminal. The pulse current is read by the OSC. The 3-port power divider, which utilizes a $16.67 \text{ } \Omega$ resistor in each port to split the electrical power equally among the three connectors, is connected to the DUT's gate terminal to provide an impedance matched signal path [164].

The UFS technique [162], [163] can be applied for both NBTI and PBTI. The NBTI is used as an example to elaborate the UFS technique's operating principle. During NBTI stressing, a static negative gate voltage, V_g^{dc} , is applied to a p -MOSFET through the gate RBT SMU DC input, which results in a corresponding constant DC drain current, I_d^{dc} , that is captured by the drain RBT SMU DC output. During a selected time interval, a single 100 ns positive voltage pulse, V_g^{p} , is fed from the PG that is connected to the gate RBT pulse input. This narrow pulse V_g^{p} superimposes on the V_g^{dc} to bring down the overall gate voltage at the gate RBT output ($V_{g,\text{meas}} = V_g^{\text{dc}} + V_g^{\text{p}}$), lowering the level of the I_d measurement. Hence, I_d ($I_{d,\text{meas}} = I_d^{\text{dc}} + I_d^{\text{p}}$) reduces instantaneously upon gate pulsing, but the I_d^{dc} which is traced by the drain RBT SMU DC output, remains unchanged. The change in the I_d , which is triggered by the V_g^{p} and reflected as I_d^{p} , is captured by the high-speed digital OSC that is connected to the drain RBT AC terminal. A voltage V_d^{p} is registered by automatically converting I_d^{p} using the $50 \text{ } \Omega$ internal resistor in the OSC ($V_d^{\text{p}} = I_d^{\text{p}} \times 50 \text{ } \Omega$). This current-to-voltage conversion using resistance overcomes the amplifier-based method bandwidth constraint [134], [139], enabling the reduction of the rapid pulse transition time to only a few nanoseconds. A series of multiple gate pulses are needed to form a whole I_d - V_g curve, since only one drain measurement can be performed for each gate voltage pulse

that has been pulsed to a desired measurement voltage. The corresponding waveforms of V_g and V_d are illustrated in Figure 3-3(b). The fresh I_d - V_g curve is measured before the application of stress to serve as a reference for the subsequent post-stress stages by using $V_g^{dc} = -1$ V, $V_d^{dc} = -0.1$ V, and multiple gate pulses to build a whole I_d - V_g sweep. $V_g^{dc} = -1$ V is used instead of $V_g^{dc} = 0$ V to avoid the measurement errors caused by waveform overshoot and inconsistency between SMU and OSC. This can be done as long as the V_g^{dc} used to measure the fresh I_d - V_g curve does not generate significant NBTI degradation. During the relaxation phase, V_g^{dc} is set to 0 or a positive value depending on the specific experiment being performed, and a measurement sequence that is similar to the one

Figure 3-4. Waveforms of the gate voltage pulse V_g^p (top) injected by the PG and the corresponding I_d^p (bottom) captured by the OSC. V_g^p has a 100 ns pulse width and a 2.6 V amplitude with a 10 ns rising/falling time. The I_d^p tracked has a 20 ns delay that is caused by the wiring capacitance. The measured I_d^p is not affected by this 20 ns delay since the measurement is performed at the quasi-steady state between 60–80th nanoseconds, while the stable waveforms are acquired after 50 ns [162].

elaborated above is implemented.

The UFS measurement method is capable of tracking I_d within 100 ns of applying the gate pulse [162]. The pulse waveforms of the V_g^p injected by the PG and the corresponding I_d^p that was captured by OSC during measurement are presented in Figure 3-4. The gate voltage pulse V_g^p rises quickly from 0 to 2.6 V with a 10 ns rising/falling time and a 100 ns pulse width. The corresponding I_d^p measured has a 20 ns delay, caused by the wiring capacitance. Nonetheless, the measured I_d^p is not affected by this 20 ns delay, as the measurement is implemented at the quasi-steady state, which is the “flat” portion of the V_g^p and I_d^p waveforms, delivering the mean values of V_g^p and I_d^p between the 60th and 80th nanosecond respectively. Therefore, the immanent delay between these two signals is not crucial. This desired attribute aids in mitigating the limitation of the ramped-voltage technique [134] in which the gate voltage and drain current waveform synchronization is pivotal. Although the UFS method is able to attain a 40 ns pulse width and stable waveform around 50 ns, the best selection of the pulse width is recorded to be between 70 and 100 ns considering the voltage overshoot and damping effect.

To verify the UFS measurement method, DC-measured and UFS-measured I_d - V_g curves were compared. An excellent match was observed between the DC-measured and UFS-

Figure 3-5. (a) Drain current, $|I_d|$ measured using both DC and UFS approaches by pulsing/sweeping the gate voltage from 0 to -1 V. The drain voltage is kept at -0.1 V. (a) Linear region (b) Sub-threshold region (data after Ref. [165]).

measured I_d - V_g curves in the linear region and down to $\sim 5 \mu\text{A}$ in the subthreshold region, as displayed in Figure 3-5 (data after Ref.[165]), indicating that the UFS technique is reliable for transfer curve measurement. The RBT connected to the drain terminal in the UFS method degrades the resolution limit to $\sim 5 \mu\text{A}$, whereas the SMU has a high resolution down to the range of $\sim \text{pA}$. The resolution limit caused by the RBT is not a limitation, as the sub-threshold current in advanced technology is anticipated to rise as a result of the down-scaling of the channel length and gate dielectric thickness. Hence, the UFS characterization method can be used in characterizing BTI of state-of-the-art leading-edge technological nodes.

The UFS method is able to effectively minimize the recovery effect during the measurement by using the very short 100 ns gate pulse. In addition, the insignificant cumulative recovery at the end of the aggregate measurement period grants a vital advantage for the accurate measurement of an I_d - V_g curve over an extended period of up to $\sim 5 \text{ s}$, but with the recovery consistently kept constant (approximately) at a minimum level which is determined by a very short delay of less than 100 ns [163]. This negligible cumulative recovery is achieved by an extremely short (1%) duty cycle of pulse that enables the reverting of the possible recovery during each 100 ns pulse to its level prior to the stress interruption and before the incurring of the next V_g pulse for the I_d measurement. The very short 1% duty cycle is attained by alternating the 100 ns pulse and the $10 \mu\text{s}$ V_g stress level during the I_d measurement. Using this approach, each data point that constructs the I_d - V_g curve is recorded under a similar degradation state, eliminating the possible I_d - V_g curve alteration that may occur from the DC voltage sweep technique.

3.4.2 Keithley's Ultra-Fast BTI Package

Although in-house BTI characterizations such as UFS discussed in section 3.4.1 are able to provide reliable measurements, there are several technical challenges that require

addressing. Firstly, the $\log(\text{time})$ scale is favorable in the BTI test, but the standard pulse generators are made to create waveforms on a fixed recurring interim. Further, the oscilloscope, transimpedance amplifier, and current probes have their own defined performance specifications that may not be optimized to operate in collaboration. The integration of these independent components to deliver optimum performance across a broad dynamic range in order to achieve accurate and precise current measurement at fast speeds is challenging. Additionally, splitters and bias-tees are usually used to build in-house characterization systems, restricting the performance of the test setup. A bias-tee may restrict the bandwidth from 100 ns to 10 μs , which is suitable for high speed measurements, but the intermediate measurement range of 10 μs to DC as well as the pre-stress and post-stress DC measurements could not be performed as part of the stress-measure setup. Moreover, as data streaming is not supported in traditional oscilloscopes, the results can only be transferred out at the end of the experiments. A significant amount of data needs to be transferred out for post-processing on a computer. In addition, since the outcome of the experiments cannot be analyzed till the data is transferred out upon their completion, termination of the experiment due to catastrophic failures or parametric shifts in real-time is impossible. The Keithley's Ultra-Fast BTI Package is designed to overcome these challenges [164], [166].

The Keithley's Ultra-Fast BTI Package [164], [166] uses AC and DC reliability stress-measure techniques. Ensuring that the source instrumentation would not be the restricting factor when performing low-level measurements and offering current sensitivity down to tens of picoamps, this Ultra-Fast BTI Package supports sub-microsecond pulse characterization of the drain current at a reduced drain voltage, minimizing the drain-to-source fields that could otherwise skew the test results. The low current measurement (<10 nA resolution and noise) can be performed with a fast pace (at least one reading within 1 μs) and visibility (1 μs off-stress window includes source transition of rise, fall and settling).

The Ultra-Fast BTI Package [164], [166] integrates Keithley's advanced DC I-V and ultra-fast I-V measurement capabilities with an automatic test executive software to deliver the most sophisticated BTI measurement platform on leading-edge silicon CMOS technology. This Ultra-Fast BTI Package is built on the Keithley Model 4200-SCS parameter analyzer with a Model 4225-PMU Ultra-Fast I-V Module, two Model 4225-RPM Remote Amplifier/Switches, Automated Characterization Suite (ACS) software, Ultra-Fast BTI Test Project Module, and effective connections.

The Model 4225-PMU Ultra-Fast I-V Module [164], [166] is the hardware core of the ultra-fast I-V measurement capability and is essential for BTI characterization in microseconds. This module, which incorporates a sophisticated two-channel waveform generator with high-speed voltage and current measurement capabilities, a deep measurement buffer, and a real-time test execution engine, supports four measurement types, including spot, step sweep, smooth sweep and sample. A test sequence can be formed by combining multiple measurement types. The outcomes consisting of the mathematical means of multiple samples in the measurements, which are calculated automatically by the system's real-time test execution engine to reduce the data volume, are cascaded back to the high-level test module for near-real-time analysis and test termination.

The Model 4225-RPM Remote Amplifier/Switch [164], [166] is developed to maximize the current measurement sensitivity of the Model 4225-PMU Ultra-Fast I-V Module. The remote pre-amplifier (RPM) together with the pulse measurement unit (PMU) are needed for the ultra-fast BTI characterization, as BTI measurements must be as fast as possible in order to reduce the recovery effects. RPM helps in positioning the fast-speed measurement circuits and pulse source near the DUT, minimizing the cable length and corresponding cable parasitic effects, including cable capacitance, source overshoot and load-line effects, to render the round-trip propagation delay shorter than the rise or fall time of the defined pulse.

The test software used with the ultra-fast BTI module is called Automated Characterization Suite (ACS) [164], [166]. Stress environments, stress timing, and a variety of measurement types, such as spot, sweep, and sample, are easily defined with this software. Together with the ACS software, the high-precision, low-level measurements of BTI stress degradation and recovery effects as well as pre- and post-stress measurements can be characterized by the ultra-fast BTI module. Characterization of BTI degradation and recovery behavior utilizing DC or AC stress and longer stress-measure sequences with single pulse charge trapping (SPCT) measurements can be performed using the ACS.

The key advantages of using Keithley's Ultra-Fast BTI Package with 4225-PMU, 4225-RPM, and ACS software are stated below [164], [166]:

(1) *Speed and sensitivity*

Since BTI is a highly dynamic phenomenon that requires fast and sensitive measurements for accurate characterization, this ultra-fast BTI measurement system provides an optimum integration of measurement sensitivity and speed for ultra-fast BTI characterization as well as accurate ultra-fast outcomes without the utilization of interconnects and RF structures.

(2) *Reducing undesired drain-to-source field*

Minimizing the strength of the drain-to-source electric field is crucial during BTI

Figure 3-6. The measure window for spot measurement in a pulse. The start-measuring and stop-measuring points are defined at 75% and 90% of each step width [164].

experiments in order to avoid unwelcome charge displacement and hot carrier injection effect. Since the drain current is proportional to the drain-to-source field, the more sensitive the drain current measurement is, the lower the required drain voltage will be. This Keithley's Ultra-Fast BTI Package is able to perform low current measurements (<10 nA

Figure 3-7. Schematic diagram of four-terminal local sense connections using the SMA to dual SSMC adapter cable (Y-cable) with Model 4225-RPM connected to DUT during experiment. Gate and drain terminals are applied with desired bias and waveform through Model 4225-RPM whereas source and bulk terminals are grounded. [164]

resolution and noise) which enables the use of lower drain voltages in order to yield more precise outcomes.

(3) *Minimizing recovery effect and relaxation time*

The single-point I_d spot measurements and I_d - V_g step sweeps with ten measurement points can be accomplished within 1 and 10 μ s respectively.

(4) *Intuitive test sequence development*

A test sequence of up to 20 chained-up measurements can be created without coding or script-writing by selecting one or more measurement types, such as spot, step sweep, smooth sweep, or sample, and the appropriate values for the measurement parameters are to be entered in the intuitive interface.

The 4 measurement types in the Ultra-Fast BTI Package [164], [166] are described as follows:

(1) *Spot*

The spot measurement returns a single mean value from the average of the samples performed on the measure window during gate and drain pulsing. Figure 3-6 shows the measure window for the spot mean measurements. The start-measuring and stop-measuring points are defined at 75% and 90% for both the amplitude and the base level. The number of samples measured on the amplitude and the base level is dependent on the size of the portions to be sampled and the sampling rate. Generally, the sampling rate is considered to be 200×10^6 samples per second.

(2) *Step Sweep*

The step sweep is similar to a conventional DC SMU sweep, in which each step in the sweep consists of a settling period and an integration (or averaging) period. The measure window for each step follows the measure window of the spot, where the start-measuring and stop-measuring points are defined at 75% and 90% of each step width.

(3) *Smooth Sweep*

The signal in a smooth sweep is constantly sampled throughout the sweep without any settling times.

(4) *Sample*

The signal in a sample measurement is continuously sampled at a constant gate and drain voltage conditions.

Figure 3-7 [164] schematically illustrates the four-terminal local sensing connections using the SMA to dual SSMC adapter cable (Y-cable) with the Model 4225-RPM connected to the DUT during the experiments. Figure 3-8 [164], [166] depicts the local sensing connection in greater detail. Each of the Force and Sense triax connectors has a central conductor and an outer shield. The central conductor of the Force connector is connected to the DUT high, while the outer shield is connected to DUT low. The Sense connector is not utilized. The gate and drain terminals, which are connected to the Force connector's central conductor, are applied with the desired bias and waveform using the RPM controlled by the ACS software, whereas the source and bulk terminals connected to the Force connector's outer shield are grounded. Various types of testing waveforms can be performed using the Ultra-Fast BTI Package depending on the parameters defined by the users in the ACS software, as illustrated in Figure 3-7. [164]. In the experiments conducted in this report, the BTI stress and relaxation voltages were applied at the gate terminal, while

Figure 3-8. Schematic illustration for local sensing connection. The Force triax connector has center conductor connected to DUT high while the outer shield is connected to DUT low. The Sense connector is not used. [164], [166]

the BTI recovery behavior was characterized by monitoring the drain current trend throughout the relaxation period. The SUSS 200 mm Prober Station was used as the probe station. The experimental details of each individual experiment run will be discussed in detail in the following chapters.

3.4.3 Stress-Induced Leakage Current (SILC) measurement

Oxide trap generation can also be monitored through stress-induced leakage current (SILC) measurement. SILC is the increase in the gate current, I_g , relative to the fresh condition caused by the generated stress-induced oxide defect in the progressive degraded gate oxide during BTI stressing. To measure the SILC, first the I_g of a pristine device has to be measured before the BTI stress application. Next, during the SILC measurement, a gate voltage above the threshold voltage and much lower than the stressing voltage is applied in order to shun oxide trap generation during SILC measurement, while the source, drain and bulk terminals are grounded. The temperature during the SILC measurement must be the same as the BTI stressing temperature throughout the experiment in order to obtain a fair I_g comparison between pre- and post-BTI stress.

The SILC phenomenon is commonly attributed to the trap-assisted tunneling (TAT) mechanism [167]. The probability of carrier tunneling between the gate and bulk across the gate oxide depends on the energy barrier in the gate oxide, which has to be overcome. The

Figure 3-9. Test sequence diagram of various types of testing function waveforms that can be performed using the Ultra-Fast BTI package [164].

carrier tunneling probability increases with the presence of oxide defects that split the energy barrier into several levels, aiding the carriers in tunneling successively through the thinner energy barriers in the gate oxide (Figure 3-10). An increase in I_g due to SILC after the application of BTI stress generally indicates the occurrence of stress-induced oxide defect generation during BTI stress, as an increase in the defects generated in the oxide raises the likelihood of carrier tunneling via the TAT mechanism.

3.5 Chapter Summary

The imperfection of the traditional slow DC measurement (stress-measure-stress) and the impact of the BTI recovery effect, which is caused by the measurement delay, on the BTI degradation characterization have been highlighted in this chapter. Various fast measurement methods, which were developed to overcome the BTI measurement delay, have been presented. The characterization methods used in the experiments, such as the UFS measurement system, Keithley's Ultra-Fast BTI package, and the SILC measurement approach, have been elaborated upon. The details of each of the experiments conducted will be discussed in the following chapters.

Figure 3-10. A schematic of the TAT mechanism aiding the carrier tunneling through the gate oxide contributing to the SILC.

CHAPTER 4: SWITCHING OXIDE TRAPS RECOVERY BEHAVIOR UNDER DYNAMIC NEGATIVE-BIAS TEMPERATURE INSTABILITY

4.1 Introduction

As discussed in Section 2.4, the R-D model was widely utilized to describe NBTI degradation events [92], [97]. However, this R-D model could not explain the studies reported in [30], [38], [43], [98], [102]–[97], which revealed that NBTI degradation is attributed to hole trapping caused by the oxide trap along with the interface trap generation.

Both fast and slow recoverable components were observed by Teo and Ang *et al.* [104]–[106], [129] in SiON gate dielectric *p*-MOSFETs during the 40-cycle DNBTI recovery phase. The DNBTI degradation is attributed to a symmetric “switching” hole-trapping mechanism superimposing on a rather permanent interface-state and deep-level hole-traps

Figure 4-1. Recovery of $|\Delta V_t^F|$ per cycle (normalized to the $|\Delta V_t^F|$ of the first cycle) as a function of the number of SR cycles, exhibiting constant recovery independent of the SR cycles. The R-D model simulation shows a gradual decrease in $|\Delta V_t^F|$ with an increase in the number of SR cycles, which does not align with the experimental data (data after Ref. [106]).

generations. The recoverable threshold voltage shift, $|\Delta V_t|$, attributed to switching hole traps, is constant regardless of the number of DNBTI SR cycles, as shown in Figure 4-1 [106]. Moreover, Boo *et al.* [52] discovered the conversion of switching hole traps (R component) into permanent trap holes (P component) and the correlation between the converted permanent trap holes and the generation of bulk oxide traps in SiON gate dielectric p -MOSFETs. Compared to the constant recoverable $|\Delta V_t|$ under the dynamic NBTI (DNBTI) test only [105]–[107], [115], the amount of recoverable $|\Delta V_t|$ decreases when a constant voltage stress (CVS) is applied in the interval of DNBTI, as shown in Figure 4-2(a) [52]. An intermediate CVS test incorporated into the DNBTI test transforms some of the switching hole trapping into a more permanent charged trap, implying that considerable oxide-trap conversion only happens after a certain threshold of the oxide field has been exceeded. Intriguingly, this study [52] also uncovered that the gradual decrease in $|\Delta V_t|$ recovery under DNBTI with an intermediate CVS test is accompanied by an increase in the SILC, as shown in Figure 4-2(b). There is no SILC observed in the DNBTI only test. An excellent linear correlation was reported between the gradual decrease in the recoverable

Figure 4-2. A comparison of (a) $|\Delta V_t|$ recovery and (b) gate current I_g in the DNBTI test only and in DNBTI test with an intermediate CVS as a function of number of SR cycles. The $|\Delta V_t|$ recovery is deduced from the difference between $|\Delta V_t|$ extracted at the end of recovery phase and the $|\Delta V_t|$ extracted at the end of the prior stress phase, as depicted in the inset of (a) (data after ref. [52]).

$|\Delta V_t|$ and the increase in SILC, as illustrated in Figure 4-3 [52]. Based on the observation of the significant increase in the SILC corresponding with the substantial decrease in the recoverable $|\Delta V_t|$, Boo *et al.* [52] suggests that the generation of stable oxide bulk traps is correlated to the transformation of transient hole trapping into more permanently trapped holes and that they stem from the same physical mechanism. A similar observation was reported by Gao *et al.* using *p*-MOSFETs with HKMG HfO_2/TiN and HfSiON/TiN [54], as discussed in Section 2.6. Further experimental evidence for the correlation between these two events based on temperature dependence, activation energy, and the power-law time exponent will be presented in Section 4.2.

The *R*-to-*P* transformation of SOTs studied earlier [51]–[56] and the correlation between the *R*-to-*P* transformation of SOTs and generation of bulk oxide traps from the investigations [51], [52] discussed previously were using large-area devices. Owing to the down-scaling of the transistor to nano-scale dimensions, the gate dielectric area has been getting smaller as well. Only a handful of defects or SOTs are present in the gate oxide layer above the channel region that may cause a critical stochastic impact on the performance of operation of MOSFETs [47], [139]. Hence, an investigation of the switching-to-permanent transformation of the hole-trapping in small-area devices subjected to the DNBTI test was performed. Interesting results were obtained from this study, which

Figure 4-3. (a) Increase in the SILC closely “tracks” the decrease in $|\Delta V_t|$ recovery (b) Outstanding linear correlation between SILC (extracted at two gate voltages) and reduction in $|\Delta V_t|$ recovery (data after ref. [52]).

will be presented in Section 4.3. These results indicate that the emission times of hole traps are not time-invariant (as normally presumed) but can increase due to the evolution of the defect sites into more structurally stable forms. The experimental evidence presented in Section 4.2 and 4.3 triggered the urge to examine the correlation between hole trapping, denoted by the degradation of the drain current, and the bulk oxide trap generation, indicated by the SILC, in small-area devices through the time-dependent relaxation of the drain and gate current shifts after NBTI stress. The results of this examination will be discussed in Section 4.4.

4.2 Correlation between Hole-Trapping Transformation and SILC Generation under NBTI Stressing

Recently, studies conducted on small-area *p*-MOSFETs [47] corroborated earlier findings, which stated that the cyclical trapping and detrapping of holes by oxide and interface traps are mainly responsible for the parametric fluctuations observed under DNBTI stressing [106], [168]. It was also found that after a period of stressing, a particular trap can be transformed into a semi-permanent charged state, implying that the applied electrical stressing may have changed the nature of the trap thus allowing the hole to remain trapped for a longer period. This finding will be discussed in Section 4.3 and is supported by the study reported in [147]. This observation aligns with the earlier result that showed a progressive decrease in NBTI recovery under repeated stressing and relaxation, indicating that some of the trapped holes which were able to emit within the given relaxation interval in the initial stage were unable to do so at a later stage [51], [54].

In addition, a generation of SILC that linearly tracks the decrease in NBTI recovery has been observed [51], [52], [54]. The correlated changes have prompted a hypothesis that both effects share a common origin, i.e. the same traps responsible for the cyclical hole capture and emission behavior also account for the generated SILC after they have

transformed into a more permanently charged state. In this study, new experimental evidence on the correlated changes of the two effects is presented in terms of temperature dependence, activation energy, and the power-law time exponent.

4.2.1 Experimental Details

The DUTs used were TiN/HfSiON (HKMG) and polysilicon/SiON gate stacks *p*-MOSFETs. The drawn channel width/length for DUTs with high-*k* gate stack and SiON gate dielectric are 2 μm /40 nm and 10 μm /60 nm respectively. The TiN/HfSiON gate stack is made up of a 2.5 nm HfSiON layer, grown via atomic-layer deposition on a 0.9-nm thermal SiO_x interfacial layer in a gate-first integration flow. The Hf composition is ~70 atomic percent and the equivalent oxide thickness is 1.6 nm. The SiON is 1.8 nm thick and was achieved via decoupled plasma nitridation of an in situ steam-generated SiO₂ layer.

Alternating NBTI stress and relaxation intervals, each lasting 1×10^3 s, were applied for 40 times (arbitrarily determined) to a DUT at a given temperature. The gate voltages applied during NBTI stress and relaxation were -2 and 0 V respectively. Temperatures varying from 100-200°C were examined and a pristine DUT was used at each temperature. Stress induced degradation was monitored in terms of the threshold voltage shift $|\Delta V_t|$, extracted from linear transfer curves measured at specific timings using the ultra-fast switching method [162]. Figure 4-4(a) illustrates the typical variation in $|\Delta V_t|$ over several alternating NBTI stress/relaxation cycles. The $|\Delta V_t|$ recovery amount during the relaxation interval, denoted as R , was determined using the difference between $|\Delta V_t|$ measured at the end of the stress and relaxation intervals as shown. R provides a measure of the extent of trapped-hole emission during the relaxation interval.

In a separate set of experiments, a DUT was first subjected to the same NBTI stress gate voltage at -2 V for 10^3 s. This was followed by gate current (I_g) measurement, using a quasi-static bi-directional gate-voltage sweep from -1.5 to $+1.5$ V and back, for the purpose

of monitoring SILC generation. The above procedure was then repeated on the same DUT 40 times. Figure 4-4(b) illustrates typical SILC generation in the HfSiON and SiON DUTs after numerous stress intervals. It should be noted that no apparent reduction in the SILC was yielded by the excursion to +1.5 V and back, implying that a majority of the oxide traps accountable for the SILC are relatively stable.

4.2.2 Results and Discussion

Figure 4-5(a) shows the $|\Delta R_{40,1}|$ versus $(kT)^{-1}$ plot for the HfSiON gate p -MOSFETs; k is Boltzmann constant, T is temperature in Kelvin and $\Delta R_{40,1} (= R_{40} - R_1)$ is the difference in

Figure 4-4. (a) The shift in threshold voltage, $|\Delta V_t|$ of a HfSiON gate p -MOSFET subjected to stress/relaxation cycling. The $|\Delta V_t|$ recovery in the respective cycles are denoted as R_j , where j is the cycle number. (b) Stress induced leakage current (SILC) versus gate voltage (V_g) curve after 40 stress/ relaxation cycles; (i) HfSiON and (ii) SiON gate p -MOSFETs. The gate voltage was swept from -1.5 V to 1.5 V and then back to -1.5 V.

$|\Delta V_t|$ recovery between the 40th and the 1st cycles. Each data point denotes the average value for several test devices. A non-Arrhenius behavior over the range of temperatures studied, from 100-175°C, can be observed. Two piecewise linear fittings of the data points, demarcated by $T \sim 398$ K or 125°C, may be attempted. The best fit yields activation energies E_a of 0.76 eV and 0.09 eV for $T < 398$ K and $T > 398$ K, respectively. Similar results are acquired for $\Delta R_{j,1}$ extracted at other j th cycles ($30 \leq j < 40$). The lower boundary of j is restricted by the measurement resolution for R , which is ~ 1 mV.

As the number of stress/relaxation cycles increases, R has been observed to either remain constant or decrease gradually [51], [52], [54], [105], [106]. A constant R indicates that it is always the similar group of traps that is capturing and emitting holes during the stress and ensuing relaxation interval respectively [105], [106]. This inference is supported by measurements on small-area p -MOSFETs, which revealed the highly consistent, cyclical hole-trapping/detrapping behavior of individual traps [47]. The decrease in R indicates that some of the trapped holes that were emitted in the 1st relaxation interval remained trapped after the same relaxation period in a later cycle [51], [52], [54]. It has been proposed that

Figure 4-5. (a) $|\Delta R_{40,1}|$ as a function of $(kT)^{-1}$, where k is the Boltzmann constant and T is temperature in Kelvin, for the HfSiON gate p -MOSFET. $\Delta R_{40,1}$ is defined as $R_{40} - R_1$, where R_{40} and R_1 represent the overall $|\Delta V_t|$ recovery in the 40th and 1st cycle, respectively. (b) SILC generation (defined as $\Delta I_g / I_{g0}$ at $V_g = -1$ V) as a function of $(kT)^{-1}$. ΔI_g is the difference between the I_g (at $V_g = -1$ V) of the 40th and 1st cycle. The temperatures indicated are in degree Celsius ($^{\circ}\text{C}$).

the repetitive application of an oxide stress field, coupled with phonon-induced effects, may cause permanent structural relaxation of the oxygen vacancy defects (major source of hole traps) and convert them into stable hole-trapping sites [51], [52], [54]. In recent studies of small-area *p*-MOSFETs [169], [170], the direct observation of a loss in the ability of a trap to emit the trapped hole after a certain number of stress/relaxation cycles lends strong support to the proposed explanation [51], [52], [54].

The general explanation for the observed non-Arrhenius behavior of $\Delta R_{40,1}$ in Figure 4-5(a) is provided as follows. The phonon-induced effects are relatively weak at low temperatures, and therefore, a certain energy barrier would have to be overcome in order to change the structure of a given trap. An increase in temperature leads to a decrease in the energy barrier due to enhanced phonon-induced effects, which help in the structural change of the trap. The structural change of the trap at high temperature (low E_a value) may be related to hydrogen-related process. The proton released in the Si substrate [101] and hydrogen released from the Si-H bond [59], [101], [131], [148] may be the possible causes in aiding the structural relaxation of trap to become a more permanent charged trap. However, the true reason behind the existence of the critical temperature of $\sim 125^\circ\text{C}$, beyond which a decrease in E_a occurs, remains a question which needs further investigation.

Intriguingly, the non-Arrhenius behavior of $\Delta R_{40,1}$ is also reflected in the results of SILC

Figure 4-6. (a) $|\Delta R_{40,1}|$ and (b) SILC generation ($\Delta I_g/I_{g0}$) as a function of $(kT)^{-1}$ for the SiON gate *p*-MOSFET. The values above the plotted lines indicate the temperatures in degree Celsius.

generation (Figure 4-5(b)). Similar to $\Delta R_{40,1}$, a change in E_a occurring at the temperature of 125°C should be noted. A similar critical temperature signaling a change in the mechanism driving SILC generation in the HfSiO/SiO_x stack was earlier reported by Sahhaf *et al.* [171]. Piecewise linear fitting of the SILC versus $(kT)^{-1}$ plot yields an E_a of 0.12 eV for $T > 398$ K and an E_a of 0.71 eV for $T < 398$ K, which is in excellent alignment with the values acquired from the $\Delta R_{40,1}$ versus $(kT)^{-1}$ plot. The close correspondence between the SILC and $\Delta R_{40,1}$ data in terms of temperature dependence and activation energy furnishes another piece of pivotal evidence (on top of the direct proportionality observed earlier [51], [52]), which states that the two effects originate from a common degradation mechanism. A non-Arrhenius behavior of SILC generation under electrical stressing has been previously reported [171]–[173] but the underlying mechanisms are still unclear. Its correlation with $\Delta R_{40,1}$ revealed in this study indicates that the transient-to-permanent hole-trapping conversion could be a possible mechanism.

Measurements from the SiON gate *p*-MOSFETs produce similar results – a correlated non-Arrhenius behavior of both $\Delta R_{40,1}$ and SILC (Figure 4-6). However, some subtle differences should be highlighted as compared to the results of the HfSiON *p*-MOSFET: (1) A slower roll-off in the SILC as compared to ΔR is observed in the SiON device for temperatures below 100 °C, which is evident from the smaller E_a of 0.76 eV for the SILC as

Figure 4-7. Both SILC generation ($\Delta I_g/I_{g0}$) and $|\Delta R_{40,1}|$ exhibit very similar dependence on stress time. Data may be fitted using a power-law relation, with closely matched exponents (as indicated).

compared to the E_a of 0.91 eV for $\Delta R_{40,1}$. No apparent difference, however, can be noted in the HfSiON device. A non-negligible SILC contribution from the stress-induced interface traps, due to the much thinner SiON dielectric [115], may have caused a more gradual SILC roll-off. Further studies are required to validate this hypothesis. (2) The extracted activation energies are smaller for the HfSiON device. This may be due to the more ionic nature of HfSiON, and thus, the effect of the applied stress field on the structural change in this dielectric is stronger than that in the SiON dielectric [54].

Another indication that $\Delta R_{40,1}$ and SILC may share a common origin lies in the exceptional agreement between the time exponents of the two effects. Figure 4-7(a) demonstrates an example of the evolution of SILC generation in the HfSiON device on a set of log-log axes. The temporal data may be fitted using a power law, with the time exponent n being ~ 0.3 . Interestingly, the evolution of $|\Delta R_{40,1}|$ also exhibits a power-law time dependence, with a time exponent m that closely matches that of SILC generation. The correlation between n and m is consistently observed at other temperatures, as depicted in Figure 4-8(a). A similar observation applies to the SiON device as shown in Figure 4-7(b) and Figure 4-8(b). The power law fitting equation for SILC and $|\Delta R_{40,1}|$ are

$$\text{SILC} = A(t_s)^n \quad (\text{Eqn. 4-1})$$

Figure 4-8. Correlation plot of the power-law time exponent m for $\Delta R_{40,1}$ versus the power-law time exponent n for SILC generation. The numbers in the plots denote temperature in degree Celsius.

$$|\Delta R_{40,1}| = B(t_s)^m \quad (\text{Eqn. 4-2})$$

where m and n are power-law time exponent, while A and B are constants. The A , B , m and n values extracted from Figure 4-7 are tabulated in Table 4-1. The power-law time exponents values extracted in this experiment are similar to that of the classical exponent, α , reported for SILC [174]–[177]. The α reported in [174]–[177] is in the range from ~0.22 to 0.47. The various values of α observed was probably caused by the gate-processing technology [177].

The power-law time exponent of a parametric shift is usually regarded as the reflection of a particular degradation mechanism (e.g. an exponent of 0.16 has been linked with a molecular-hydrogen-driven degradation process whereas an exponent of 0.5 is associated with an atomic-hydrogen-driven process [178]). However, such associations are still to be confirmed, given the lack of direct experimental evidence on the degradation mechanisms and the recent experimental evidence that does not support the hydrogen transport model [47], [105], [106], [170]. For instance, an exponent of 0.16 may also be explained using a model involving hole trapping at oxygen vacancy defects [179]. Nonetheless, the excellent agreement between the exponents n and m further supports the hypothesis that both $\Delta R_{40,1}$ and SILC are driven by the same degradation mechanism; although, its exact nature would remain elusive until further details about the atomic defects are available.

Parameter	SILC ($\Delta I_g/I_{g0}$)		$ \Delta R_{40,1} $	
	n	A	m	B
HfSiON	0.27	0.74	0.29	$0.05 \times 10 \text{ mV}$
SiON	0.47	1.15×10^{-3}	0.44	$2.66 \times 10^{-4} \text{ V}$

Table 4-1. Extracted values of n , A , m and B of the fitting curves on SILC generation ($\Delta I_g/I_{g0}$) and $|\Delta R_{40,1}|$ for HfSiON and SiON.

4.3 Transient to Temporary Permanent and Permanent Hole Trapping Transformation in the Small-Area *p*-MOSFET Subjected to Negative-Bias Temperature Stress

Although the microscopic mechanisms of NBTI remain controversial, it is widely recognized that hole trapping is one of the causes behind this phenomenon and accountable for the degradation in the drain current as well as the increase in the threshold voltage shift during NBTI stress. The transient nature of hole trapping, which presents one of the major challenges in modeling NBTI, has been shown to be primarily associated with the capture and emission of holes by switching oxide traps (SOTs) under pulsed gate stressing [106]. Recent studies on large area devices [51]–[56] suggest that a portion of the SOTs may be gradually transformed into a more permanent form (P) as a result of the progressive decrease in the recoverable component (R) of NBTI as dynamic stressing progresses. Hitherto, there is no direct evidence of the proposed SOT transformation that has been presented. Lately, there has been a considerable interest in small-area devices which avail themselves to a detailed study of the charge capture/emission events at individual oxide defects from discrete steps in the time dependent NBTI recovery trace [47], [48], [59], [180], [181]. In this study, direct evidence for the proposed SOT transformation is presented. After several stressing and relaxing cycles, the emissions times of some of the SOTs are clearly prolonged, indicating that they have been changed into a more permanent form.

4.3.1 Experimental Details

The DUTs used were *p*-MOSFETs, with a p^+ polysilicon gate and 1.8 nm SiON gate dielectric prepared by decoupled plasma nitridation. The drawn channel length and width are 60 nm and 120 nm respectively. A DUT was subjected to alternating 10 s NBTI stress and 10 ks relaxation phases each for 30 times at 100°C. The devices were subjected to a constant gate voltage $V_g = -1.8$ V (oxide field ~ 8.5 MV/cm) during the stress phase,

followed by $V_g = 0$ V during relaxation phase, with the drain bias V_d constantly set at -0.1 V throughout the experiment. A stress phase and the ensuing relaxation phase are defined as one DNBTI cycle. During relaxation, the linear drain current I_d recovery was measured by pulsing gate voltage V_g from 0 to -0.7 V at the drain-current resolution of 10 μ A, completing the measurement with a delay of less than 1 μ s. I_d degradations are proportional to the threshold voltage shifts, ΔV_{th} , and therefore to defect trapping. In small-area devices, only a handful of SOTs are present. The SOTs can be individually identified through the different discrete single-trap drain-current recovery amplitudes (step heights) and the corresponding emission times in the I_d recovery traces using the spectral map obtained from time-dependent defect spectroscopy [47].

Figure 4-9. Construction of a spectral map (bottom) by the extraction of single detrapping events from three typical representative drain current degradation, $|\Delta I_d|$, recovery traces (top) ensuing NBTI stress. The emission time, τ_e , of a defect is obtained by tracking the discrete step height, h , from the single detrapping events, i.e. each defect is identified by a cluster of (τ_e, h) points as shown in the spectral map (bottom).

4.3.2 Results and Discussion

The drain current degradation, $|\Delta I_d|$, was plotted with respect to the pre-stress value during each 10 ks recovery. Three typical representative of $|\Delta I_d|$ recovery traces are shown in Figure 4-9(top). In small-area (ultra-scaled) devices, only a handful of defects are present. Each of them can contribute to a significant amount of degradation in the drain current during stress. Hence, step-like $|\Delta I_d|$ recovery traces were observed during relaxation. Each recovery step in the relaxation traces carries a discrete I_d step height, h , of the hole-detrapping event and a emission time, τ_e , corresponding to the detrapping event. These step heights, h and the corresponding emission times, τ_e are used to construct a spectral map, as shown in Figure 4-9(bottom), which is used to identify and differentiate the SOTs. The clusters of (τ_e, h) points located on the spectral map are believed to be the “fingerprints” of individual SOTs [47].

$|\Delta I_d|$ values at the end of stress (EOS), $|\Delta I_d|^{\text{eos}}$, and the end of recovery (EOR), $|\Delta I_d|^{\text{eor}}$, of every DNBTI cycle are plotted in Figure 4-10. $|\Delta I_d|^{\text{eos}}$ shows the total amount of $|\Delta I_d|$ degradation caused by each 10 s stress while $|\Delta I_d|^{\text{eor}}$ shows the remaining amount of $|\Delta I_d|$ degradation after each 10 ks relaxation phase. This 10 ks relaxation stage allows the recoverable SOTs to revert upon the removal of the negative gate bias stress. The $|\Delta I_d|^{\text{eos}}$ values can be observed to be noisier than the $|\Delta I_d|^{\text{eor}}$ values. This is due to the different number of traps being charged up in each stress phase of the DNBTI cycle. Some traps that are active during some of the DNBTI cycles may be inactive during other DNBTI cycles. This will be discussed subsequently in more detail. Figure 4-11 illustrates that the trend of the number of SOTs (number of step heights) that recover during the relaxation phase corresponds to the trend of the amount of drain current that gets restored at the end of each relaxation phase.

Two upward steps labeled (i) and (ii) are detected in Figure 4-10, in which (i) and (ii) represent the degradation amount of $|\Delta I_d|$ after relaxation phase of cycle 1 and 2,

respectively. A non-zero $|\Delta I_d|$ of $\sim 0.2 \mu\text{A}$ is evident at the end of the first relaxation phase (Figure 4-10(i)). This may be attributed to either:

(1) permanent interface traps (e.g. dangling Si orbitals that formed by dissociated Si-H bonds) that charge up spontaneously every time the I_d measurement was performed at $V_g = -0.7 \text{ V}$; or

(2) oxide defect(s) that are already being transformed into relatively permanent forms by the prior 10 s stress phase, leading to the trapped hole(s) not being emitted during the 10 ks relaxation interval.

Scenario (2) is supported by the observation that the $|\Delta I_d|$ never returned to ~ 0 in the remaining 29 recovery traces in Figure 4-10, implying that some permanent defects were formed during the very first stress phase. Based on the observation of the upward step labeled (ii) in Figure 4-10, it is proposed that there is another SOT that was converted into a more permanent form, resulting from the 10 s stress phase in cycle 2. This is because the $|\Delta I_d|$ at the end of the 10 ks relaxation (EOR) of cycle 2 is higher than that of cycle 1 and the

Figure 4-10. End-of-stress (EOS) and end-of-recovery (EOR) $|\Delta I_d|$ values, $|\Delta I_d|^{\text{eos}}$ and $|\Delta I_d|^{\text{eor}}$, respectively, of each DNBTI cycle. The total amount of $|\Delta I_d|$ degradation due to 10 s stress of each cycle is represented by $|\Delta I_d|^{\text{eos}}$ while the remnant of $|\Delta I_d|$ degradation after the relaxation of recoverable SOTs in 10 ks relaxation is denoted by $|\Delta I_d|^{\text{eor}}$.

$|\Delta I_d|^{\text{eos}}$ values for both cycles are comparable as well as the $|\Delta I_d|^{\text{eor}}$ values from subsequent cycles do not fall below the $|\Delta I_d|^{\text{eor}}$ value of cycle 2. Further, there is a high chance of interpreting the upward step and downward step of EOR labeled (iii) and (iv), respectively in Figure 4-10 as an SOT in which the hole was captured during the 10 s stress at cycle 4 but was not released until during the 10 ks relaxation at cycle 5. However, these two interpretations were found to be inaccurate after a thorough investigation of the individual $|\Delta I_d|$ recovery traces. Hence, it can be only known that there is a transformation of some of the SOTs to a more permanent form based on the EOS and EOR trend alone, but it is not possible to identify the exact defects or SOTs that transform from transient to permanent form. This will be discussed later in this section.

The first five cycles of $|\Delta I_d|$ recovery traces are plotted in Figure 4-12. Each labeled defect was identified through drain-current recovery amplitudes or step heights and the emission times of a single trap in each of the recovery traces. $|\Delta I_d|$ recovery trace of cycle 1 in Figure 4-12(a) supports the observation of upward step (i) in Figure 4-10 where there is a

Figure 4-11. Number of step heights and amount of $|\Delta I_d|$ recovery, R , as DNBTI progresses. The number of step heights in each $|\Delta I_d|$ recovery trace refers to the amount of SOTs which recover in that particular cycle during relaxation. The trend of the number of steps heights is in correspondence to the trend of the amount of R .

Figure 4-12. $|\Delta I_d|$ recovery traces of cycles 1 to 5. (a) $|\Delta I_d|$ recovery trace of cycle 1 shows that $|\Delta I_d| > 0$ after 10 ks relaxation, implying that the first 10 s NBTI stress has already generated some relatively permanent defect(s). (b) Defect #5 did not recover in cycle 2, as compared to cycle 1 in (a), resulting in a greater $|\Delta I_d|$ degradation at the end of the 10 ks relaxation. (c) Defect #2 and #4 were not charged during the 10 s stress of cycle 3, as can be seen from the much lower $|\Delta I_d|$ degradation at the start of the relaxation phase (compared to that of cycle 1 or 2). The final degradation remains the same, indicating that defect #5 still remained charged after the 10 ks relaxation phase of cycle 3. (d) Defect #2 and #4 were being charged during NBTI stress in cycle 4 as the initial $|\Delta I_d|$ degradation is similar to those in (a) and (b). Defect #4 recovered but defect #2 still remained charged owing to the final degradation after the 10 ks relaxation is higher, caused by the step height of defect #2 step height as compared to that in (b) and (c). Defect #5 still remained charged. (e) Defect #5 finally discharged the trapped hole in cycle 5, since it had been charged in cycle 2. However, defect #2 still remained charged. (f) A comparison between cycle 1 and cycle 5 shows that the $|\Delta I_d|$ degradation after the 10 ks relaxation in cycle 5 is larger than that in cycle 1. The recovery step height of defect #2 is not seen and the $|\Delta I_d|$ degradation after 10 ks relaxation in subsequent cycles (starting from cycle 4) is obviously higher by an amount equaling to the recovery step height of defect #2. This may be attributed to the conversion of defect #2 from a transient hole trap into a more permanent hole trap.

permanent degradation of $\sim 0.2 \mu\text{A}$ at the end of the 10 ks relaxation and the explanation has been proposed earlier. Defect #5 is absent from the $|\Delta I_d|$ recovery trace of cycle 2 (Figure 4-12(b)). After a detailed investigation, it can be seen that defect #5 was trapped during the 10 s stress phase of cycle 2. The difference in the EOR between cycle 1 and 2 is of the step height of defect #5, while the $|\Delta I_d|^{\text{eos}}$ values are almost the same for both cycles. Hence, this upward step (ii) in Figure 4-10 is caused by the charged defect #5 that could not discharge the trapped hole in the 10 ks relaxation window of the cycle 2. The $|\Delta I_d|^{\text{eos}}$ value of cycle 3 is lower than that of cycle 1 and 2 as shown in Figure 4-10 because defects #2 and #4 were not charged during 10 s stress of cycle 3, as illustrated in Figure 4-12(c). The hole was still trapped in defect #5 at the end of the 10 ks relaxation of cycle 3. In cycle 4 (Figure 4-12(d)), holes were captured by defect #2 and #4 during the 10 s stress phase, as the EOS of cycle 4 is comparable to that of cycle 1 and 2 (Figure 4-10). Defect #4 was discharged but defect #2 remained charged at the end of the 10 ks relaxation of cycle 4. This is shown by the upward step (iii) in Figure 4-10 which is caused by the amount of the step height due to defect #2. This means that defects #2 and #5 remained charged at the end of the 10 ks relaxation of cycle 4. In cycle 5 (Figure 4-12(e)), there are four step heights, compared to only three step heights in cycle 4. This means that one of the defects #2 or #5, which remained charged in previous cycle, was discharged during the 10 ks relaxation in cycle 5. This is supported by the downward step (iv) in Figure 4-10. It is deduced that this trap is defect #5 instead of defect #2, as the height of downward step detected in Figure 4-10 labeled (iv) or Figure 4-12(e) is similar to the step height tagged as defect #5 ($\sim 170 \text{ nA}$) instead of step height marked with defect #2 ($\sim 210 \text{ nA}$) in cycle 1 (Figure 4-12(a)). Therefore, the hole trapped in defect #5 was emitted, while defect #2 remained charged after the 10 ks relaxation in cycle 5.

Figure 4-13 illustrates $|\Delta I_d|$ recovery traces of cycle 5 to 9. As mentioned earlier, defect #5 was charged during stress phase in cycle 1 but it was not discharged until the relaxation

phase of cycle 5. However, defect #5 was inactive, i.e. it did not capture a hole during the 10 s stress phase in cycle 6. This is supported by the $|\Delta I_d|^{\text{eos}}$ value in cycle 6, which is lower than the $|\Delta I_d|^{\text{eos}}$ value in cycle 5 by the step height of defect #5, while the $|\Delta I_d|^{\text{eor}}$ values for both cycle 5 and 6 remain at a similar level, as shown in Figure 4-10 or Figure 4-13(a) and (b). In cycle 7 (Figure 4-13(c)), the $|\Delta I_d|^{\text{eos}}$ value is higher than that of cycle 5. It is observed that there is a new large step (~ 300 nA) in the $|\Delta I_d|$ recovery traces of cycle 7, which is not observed in the previous cycles. Thus, it is labeled as defect #A. Defect #2 remained charged, since its recovery step height is nowhere to be observed in cycles 5, 6 and 7. In

Figure 4-13. $|\Delta I_d|$ recovery traces of cycles 5 to 9. (a) The recovery trace of cycle 5 shows that 4 defects were discharged, but defect #2 (permanent hole trap) was not. (b) Defect #5 was inactive or not charged during the 10 s stress in cycle 6. (c) Defect #A is observed with a new step height in cycle 7. (d) Hole captured by defect #A during the 10 s stress in cycle 8 was not emitted after 10 ks relaxation. (e) Part of the step height of defect #A, which is denoted by defect #A", was discharged during the 10 s stress in cycle 9, whereas part of the step height of defect #A, which is denoted by defect #A', remained charged at the end of the 10 ks relaxation in cycle 9. (f) A comparison between cycle 5 and cycle 9 shows that the $|\Delta I_d|$ degradation after the 10 ks relaxation in cycle 9 is larger than that in cycle 5 by the amount of the step height equivalent to defect #5, which corresponds to defect #A'.

cycle 8, this new defect #A was charged in the stress phase, but it was not discharged during the relaxation phase. This is illustrated by the $|\Delta I_d|^{\text{cor}}$ value of cycle 8, which is higher than that of cycle 7 by the amount of the step height of defect #A, as shown by the upward step (v) in Figure 4-10 or Figure 4-13(d). At the end of cycle-8 relaxation, it can be seen that the other defects #1, #3 and #4 were discharged but the defect #A and #2 were charged. In cycle 9, the $|\Delta I_d|^{\text{eos}}$ and $|\Delta I_d|^{\text{cor}}$ values are lower than those of cycle 8 (Figure 4-10 or Figure 4-13(d) and (e)), while the number of step heights in both $|\Delta I_d|$ recovery traces of cycles 8 and 9 are the same (Figure 4-11 or Figure 4-13(d) and (e)), which is 3 discrete step heights corresponding to defects #1, #3, and #4. This means that defects #1, #3 and #4 were charged during the stress phase in cycle 9 along with the previously charged defect #2 and #A. With the present state of the individual step heights corresponding to defects #1, #3 and #4 in cycle 9, the only possibility is that a hole was emitted from either defect #2, defect #A or both during the 10 s stress phase in cycle 9. It is unusual for holes to be detrapped during the stress phase, but it may occur owing to a structural change or the hole overcoming the barrier and being released due to the electric field. Neither the step height of defect #2 (~210 nA) nor the step height of defect #A (~300 nA) correspond to the difference between the $|\Delta I_d|^{\text{eos}}$ values of cycles 8 and 9 (~ 130 nA). A subsequent comparison between cycles 7 and 9 revealed that the difference in $|\Delta I_d|^{\text{cor}}$ values of these two cycles corresponds to the step height of defect #5. It is worth noting that there is no trace or step height of defect #5 that appears in subsequent cycles after cycle 9. Thus, the question is what defect #A is and what actually happens within cycles 7 to 9 that is unusual and complicated. There are three possible scenarios:

S1: Defect #A was a new trap generated or had only started to get activated during the 10 s stress in cycle 7. Hence, this new trap was discharged during the relaxation of cycle 7. In cycle 8, it was then charged again when stress was applied but was not discharged during the relaxation in cycle 8. However, in cycle 9, it was discharged during the 10 s stress phase

while a hole was captured by defect #5. Since defect #5 has a step height that is lower than that of defect #A, the $|\Delta I_d|^{\text{eos}}$ of cycle 9 is lower than that of cycle 8. Transforming into a more permanent component, defect #5 remained charged after the relaxation phase in cycle 9, as the discharging step of defect #5 was not observed in the subsequent cycles. It is possible that this permanent component is a new activated defect that converts into a permanent component upon activation in the placement of defect #5, but the probability of this is low. The probability of this new defect having the same step height as defect #5 as well as the appearance of this defect causing the disappearance of defect #5 (since there is no trace of defect #5 in the subsequent cycles) is almost impossible. Hence this trap has to be defect #5.

S2: Defect #A could be of the defect pair [182] of defect #5. The defect pair was charged during the stress and discharged during the relaxation in cycle 7. It is seen as a one-step emission event, as the emission time is in the order of 10^{-1} s, while the reported defect pair can be seen as two separated emission events, as the emission time reported is within 10^{-2} s [182]. The defect pair altered the structure of defect #5, causing the emission time of the defect pair of defect #5 (initially of the order of $>10^3$ s) to be shorter (of the order of 10^{-1} s). This defect pair was charged again during stress but was not discharged during relaxation in cycle 8. However, during the stress phase in cycle 9, the structure around defect #5 underwent relaxation formation, leading to the disappearance of the defect pair; thus, detrapping a hole from this defect pair of defect #5 that disappeared during the stress phase, leaving behind the singly charged defect #5 at the end of the stress phase of cycle 9. This hole trapped in defect #5 did not emit during the relaxation in cycle 9, leading to the higher $|\Delta I_d|^{\text{cor}}$ by the amount of the step height of defect #5 in cycle 9 as compared to the $|\Delta I_d|^{\text{cor}}$ of cycle 7.

S3: Defect #A may be doubly charged [134], [183]–[186], capturing two holes during the 10 s stress, and the holes were detrapped during the relaxation phase in cycle 7 (Figure

4-13(c)). Since the emission time is in the order of 10^{-1} s, it is suspected that a hole was detrapped from a doubly charged defect to singly charged one in an interval on the order of 10^{-1} s, while the second hole was subsequently emitted after than 10^{-1} s. Due to the limited time resolution of the measurement, the observed result is seen to be a huge step height while it actually consists of two step heights corresponding to two hole-capturing events. In cycle 8, these two holes were not detrapped during the relaxation phase upon being captured in the stress phase, as indicated by the difference in $|\Delta I_d|^{\text{eor}}$ between cycle 7 and cycle 8 (Figure 4-13(c) and (d)). In cycle 9, as compared to the recovery trace in cycle 7, the $|\Delta I_d|^{\text{eos}}$ is lower but the $|\Delta I_d|^{\text{eor}}$ is higher. The differences in $|\Delta I_d|^{\text{eos}}$ and $|\Delta I_d|^{\text{eor}}$ between cycles 7 and 9 are denoted by defect #A'' and defect #A' respectively. The total of the step heights of defects #A' and A'' is comparable to the step height of defect #A. This could imply that the doubly charged defect #A, which did not recover in cycle 8, released a hole during the stress in cycle 9, represented by defect #A'' step height, becoming a singly charged defect at the EOS of cycle 9. Further, this singly charged defect was not detrapped at the EOR of cycle 9, as denoted by defect #A'. Interestingly, the step height of defect #A' is comparable to the step height of defect #5. Hence, it can be concluded that defect #A' is defect #5, since defect #5, which has converted into a more permanent form, is not observed in the subsequent cycles and cannot disappear without any reason.

Regardless of the three possible scenarios proposed, defect #5 is not detected in the subsequent cycles after cycle 9. There occurs a permanent degradation by the amplitude of the step height contributed by defect #5, indicating that it becomes a permanent hole trap, as shown by (vii) in Figure 4-10 and Figure 4-13(f). Also, defect #2, which was initially an SOT, transformed into a more permanent form after the stress phase in cycle 4. Figure 4-12 (f) compares the $|\Delta I_d|$ degradation at the end of the 10 ks relaxation of cycle 1 and cycle 5 in order to illustrate that defect #2 remains charged after the 10 ks relaxation of cycle 5. Defect #2 was absent, along with the accompanying permanent $|\Delta I_d|$ degradation caused by

the step height of defect #2 in the subsequent cycles. The emission time of defect #2 initially was of the order of 10^{-2} s in cycles 1 and 2, but it was not observed after being charged during the 10 s stress phase in cycle 4. The emission time of defect #2 was increased by more than $\sim 10^7$ orders of magnitude, i.e. from 10^{-2} s to more than 10^5 s. Hence, defect #2, initially being an SOT, was transformed into a permanent hole trap after the 10 s stress in cycle 4. From the detailed investigation of $|\Delta I_d|$ recovery traces (Figure 4-12 and Figure 4-13) in addition to the examination of the EOS and EOR trends (Figure 4-10), an accurate interpretation of the transient nature of the defects by the DNBTI test is obtained. An inaccurate analysis may occur if the investigation is solely based on the EOS and EOR trends (Figure 4-10). Nonetheless, these EOS and EOR trends still display the present state of the permanent degradation of the $|\Delta I_d|$ values caused by the transformation of SOTs into permanent hole traps; however, it may not indicate which SOTs convert into more permanent forms.

After the analysis and investigation of all 30 cycles of $|\Delta I_d|$ recovery traces of this DUT, there are 6 defects (considering scenarios S2 and S3 discussed above, and excluding S1 that states defect #A is a new activated trap) discovered in this DUT during the experiment window. The evolutions of these 6 defects are illustrated in Figure 4-14. It should be noted that a detailed examination of the $|\Delta I_d|$ recovery traces from all tested devices reveals the following four categories of SOTs, and the DUT with the 6 defects mentioned is used to exemplify the classification of the SOTs that were uncovered.

Category 1: Cyclical SOTs

Defects #1 and #3 are typical instances of “cyclical” SOTs [47], [51]–[56], [59], [106], [180], as shown in Figure 4-14. The presence of Defects #1 and #3 in every $|\Delta I_d|$ recovery trace implies that they were charged in every stress phase and their capture time constants are much shorter than the applied stress period. The consistent emission behaviors suggest that the structural changes of the defects, which occur upon hole-trapping, are minimal and

fully reversible (within the chosen relaxation interval). These regular charging and discharging events of the SOTs observed in small-area devices further support the inconsistency of the R-D model discussed in Section 2.4.

Category 2: *R-to-P* (Recoverable-to-permanent) SOTs

Defects #2, #5 and #6 are examples of SOTs whose emission times increased during the course of the experiment. They initially exhibited the behavior of an intermitted SOT (SOTs in Category 4 which will be discussed later) and then were converted to a permanent form when the DNBTI progressed. The explanation of the conversion of the defects #2 and #5 were elaborated previously, and defect #6 will be discussed in the next category.

Category 3: Newly activated SOTs in later cycle

Defect #6 is the newly activated oxide trap [30], [187] that became active in cycle 16.

Figure 4-14. Evolutions of the 6 SOTs discovered based on the recovery traces in each cycle as a function of the number of DNBTI cycles. Each dynamic NBTI (DNBTI) cycle consists of a 10 s NBTI stress followed by a 10 ks relaxation. A defect is deemed to be active when it is charged and discharged in the same DNBTI cycle. It is considered to be inactive when it is not charged during the stress phase after being detrapped in the previous cycle. For example, defect #5 was charged during the stress phase in cycle 2 but was not discharged until cycle 5. Defect #2 and defect #5 transformed into a more permanent form after cycle 4 and cycle 9 respectively, when they could not discharge within the experiment window. Defect #6 is not activated until cycle 16.

Label (viii) in Figure 4-10 refers to the charged defect #6 that was charged during the 21st stress phase but was not discharged during the ensuing relaxation phase. The hole was then emitted from defect #6 in the 22nd relaxation phase, which is denoted by Figure 4-14 and (ix) in Figure 4-10. After that, a hole was trapped in defect #6 during 26th stress phase and remained charged until the last cycle (30th cycle) of the experiment as indicated by Figure 4-14 and (x) in Figure 4-10. This newly generated SOT #6 is deemed to be transformed into a more permanent form in the experiment window after cycle 26.

Category 4: “intermittent” SOTs

The behavior of defect #4 is rather complex, comprising of a periodic absence from the I_d recovery trace (i.e. not charged during the prior stress phase) and, more importantly, a

Figure 4-15. A schematic illustration of the metastable state formation that produces a temporary increase in the trap emission time. Upon charge capture, the initial structural change (ΔQ_1) is minimum, thus allowing the trap to ‘switch’ effortlessly from the intermediate state back to the precursor state (short trap relaxation time). Through phonon interaction, further structural change (ΔQ_2) may put the defect in a metastable state. The energy barrier for the defect to revert to the intermediate state is now increased, resulting in a longer trap relaxation time. Further structural change may put the defect into a permanent state with a relaxation time that surpasses a reasonable experimental timeframe.

temporary but significant increase in its emission time (from ~ 0.01 -1s to greater than 10^4 s). Defect #4 was active during the 1st and 2nd stress phases but not during the 3rd. As compared to the 1st and 2nd relaxation phases, the reduced $|\Delta I_d|$ at the beginning and a similar $|\Delta I_d|$ at the end of the 3rd relaxation phase confirm that the SOT was not charged during the 3rd stress phase (Figure 4-10(a)-(c)). After that, it could be repeatedly charged and discharged in the subsequent cycles except for the 15th and 23rd stress phases, where it was not charged. A temporary increase in emission time occurred in cycle 24, in which it was charged during the stress phase but did not get discharged until the 26th relaxation phase (i.e. an increase in emission time by ~ 20 ks). After that, it exhibited a cyclical behavior again till the last cycle of the experiment (cycle 30). A part of the behavior of this SOT is similar to the so-called disappearing traps that do not introduce any permanent degradation, as reported in [47], [59]. Such inconsistent charging during stress may be attributed to the capture time constant being comparable to the stress period. However, the temporary yet significant increase of emission time is a new behavior that has not been reported previously. This defect #4 is coined an ‘intermittent’ SOT due to the ‘togglng’ it exhibits between quasi-permanent and cyclical behavior. This behavior supports the existence of metastable trap states [133], in addition to two switching states (the precursor and intermediate states), as depicted in the schematic two-dimensional energy-coordinate diagram in Figure 4-15. Only a relatively small energy barrier is needed to be overcome by the intermediate state to revert to the precursor state. This allows the SOT to switch easily from the intermediate state back to the precursor state, leading to a fast relaxation (arrow 1). However, the defect in the metastable state, induced by further structural change, requires a longer recovery time due to the large restoration energy barrier that exists between the metastable state and intermediate state (arrow 2). Moreover, a greater extent of structural change, which depends on the phonon interaction and local atomic configuration, may bring the defect to a permanent state where

the emission time exceeds the reasonable experimental timeframe (Defect #2, #5 and #6 discussed previously).

The observation of this transformation of SOTs into permanent traps in this experiment demonstrates that the emission times of these SOTs are initially shorter than the relaxation time, but they transform into permanent traps whose emission times are longer than the recovery time of the experiment window. Our observation on this transformation is supported by the data, which shows that each of the permanent degradation step height at the EOR of $|\Delta I_d|$ recovery traces is accompanied by an absence or disappearance of an SOT with a similar step height in the $|\Delta I_d|$ recovery traces. This observation does not correspond with the additional unknown defects, which have the emission times longer than the relaxation time of the experiment timeframe, reported in [59], but is supported by the study in [147].

4.4 Correlation between Drain and Gate Currents of Small-Area SiON Gate *p*-MOSFETs Subjected to NBTI Stressing

Random telegraphic noise (RTN) in the MOSFET drain current (I_d) was traditionally accredited to charge exchange, via tunneling, between oxide traps and the inversion channel [188], [189]. Recent observations on the correlated gate-current (I_g) RTN in ultra-thin gate devices were considered as an electrostatic effect of trap occupancy fluctuations on the oxide field by some studies [190]–[192]. To account for the relatively large RTN time constants and activation energy as compared to those related to a direct-tunneling process [191]–[196], a non-radiative multi-phonon (NMP) trap-assisted tunneling (TAT) model was proposed [192], [197]. The NMP model attributes I_g RTN to oxide traps toggling between a TAT-inactive and TAT-active state. Such transitions are usually accompanied by trap occupancy changes (thus affecting I_d as well) and are determined primarily by the structural adjustments that occur at a trap site due to electron-phonon interaction. Therefore, the time

constants and activation energy involved should be distinguished from those of tunneling-based capture/emission processes [191], [193].

All the past studies on RTN have focused mainly on the n -MOSFET, showing both a positive and negative correlation between I_d and I_g . In order to explain the correlated decrease of I_d and increase of I_g , the trap was assumed to be situated very close to the Si channel interface, in a manner that it could remain constantly charged (leading to the I_d decrease), while in a TAT-active state (leading to I_g increase) [192], [197]. However, as will be shown later, the near-Si-interface traps only account for minor fluctuations ($\sim 1\%$) in I_g .

On the other hand, studies on the p -MOSFET are relatively few, and not all cases of the I_d and I_g RTN correlation have been observed. As noted in [192], the case wherein I_d and I_g both increase (decrease) due to trapped-hole emission (capture) is yet to be reported. This experiment presents time-dependent I_d and I_g relaxation characteristics of small-area oxynitride (SiON) gate p -MOSFETs subjected to negative-bias temperature stressing. All combinations of I_d and I_g fluctuations, including the correlated increase (or decrease) of I_d and I_g , are revealed. Of interest are cases involving significant changes of I_g ($\sim 40\%$), which to-date have not been reported [192]. The model based on electrostatic effect [193], [194] or near-Si-interface traps [192], [197] could not explain these significant I_g changes, which imply the role of bulk-oxide traps. Possible scenarios where a positively charged bulk-oxide trap may function as a TAT-active center are proposed to explain the correlated decrease of I_d and increase of I_g .

4.4.1 Experimental Details

The devices under testing (DUTs) used small-area p -MOSFETs, having p^+ polysilicon/SiON gate stack with a drawn channel length and width of 60 nm and 120 nm, respectively. The SiON thickness is 1.8 nm. The results presented are for DUTs stressed at a gate voltage of -1.8 V for 10 s at 100°C , followed by a 10 ks relaxation at $V_g = -0.5$ V

during which both I_d and I_g were continuously monitored [196], with a small bias of -0.1 V at the drain, using the Keithley 4200 SCS parameter analyzer. The stress-cum-relaxation cycle was repeated at least 40 times on each DUT.

4.4.2 Results and Discussion

A. *Valence-Band Electron Tunneling*

Unlike the I_g of a thin-gate n -MOSFET which comprises of only the direct electron-tunneling component, the I_g of a thin-gate p -MOSFET is made up of two current components, namely direct hole-tunneling from the channel towards the gate and valence-band electron (VBE) tunneling from the p^+ polysilicon gate towards the n -well (Figure 4-16(a)). It is therefore essential to clarify the relative contributions of these two currents towards I_g . To this end, charge separation measurement [188] was performed, and a typical result is shown in Figure 4-16(b). The hole-tunneling current is probed by the source/drain current (I_{sd}) and the VBE tunneling current is probed by the n -well current (I_w) [188]. As can be observed, the hole-tunneling current dominates I_g across almost the entire V_g measurement range. At $V_g = -0.5$ V (the gate bias where I_d and I_g were monitored during relaxation), the nominal I_{g0} or I_{sd0} is of $\sim 2 \times 10^{-11}$ A and I_{w0} is well below the measurement limit ($\sim 10^{-14}$ A) — the subscript ‘0’ denotes the pre-stress value. Hence, any SILC component in I_g after NBTI stress can only be ascribed to an increase of the hole-tunneling current, deemed to occur in two possible manners (Figure 4-16(a)(ii)): (1) trap-assisted tunneling (TAT) via bulk oxide traps; (2) tunneling via interface traps. To verify this inference, I_w was also monitored during post-NBTI relaxation and was observed to remain below the measurement limit for the stress conditions used in this work.

B. *I_g -Contribution of an Interface Trap*

The contribution of an interface trap towards I_g is expected to be minor, as an interface trap does not lead to a “percolation path” within the gate oxide (i.e. it does not significantly

alter the tunneling barrier). An estimation of the I_g -contribution of a TAT-active interface trap may be made based on its relation to the direct tunneling time constant τ_c , $\tau_e \propto \exp(T_{ox}/T_0)$ [198], where T_{ox} is the tunnel oxide thickness and T_0 is a constant depending on the oxide quality and applied voltage. The above estimation can be corroborated by an earlier study on the NBTI of large-area p -MOSFETs with a similar SiON thickness [115]. As shown in [115], the SILC (defined as $\Delta I_g/I_{g0}$) arising from stress induced interface traps of density $\sim 10^{12} \text{ cm}^{-2}$ is ~ 0.7 or 70 %. For this interface trap density, there would be an average of 70 traps in a $60 \times 120 \text{ nm}^2$ small-area p -MOSFET, indicating that the SILC contribution per trap is $\sim 1\%$. It should be noted that this estimation is made based on the peak $\Delta I_g/I_{g0}$ value (@ $V_g \sim -0.2 \text{ V}$) and thus would be lower at more negative V_g values.

C. *Uncorrelated I_d and I_g Fluctuations*

In order to facilitate discussion, NBTI stress induced I_d and I_g shifts are defined as follows: $\Delta I_d(t) = I_{d0} - I_d(t)$; $\Delta I_g(t) = I_g(t) - I_{g0}$, where $I_d(t)$ and $I_g(t)$ are the currents at a given relaxation time t . Positive ΔI_d and ΔI_g after NBTI stress refer to reduced drive current (due to hole trapping) and increased gate leakage (due to stress induced activation of TAT centers) respectively. During relaxation, the downward transition of ΔI_d or ΔI_g at a particular relaxation time signifies recovery caused by the emission of a trapped hole or deactivation of a TAT center and vice versa. Figure 4-17 illustrates all four combinations of ΔI_d and ΔI_g changes that were observed during post-NBTI relaxation.

The shaded oval regions in Figure 4-17(a), (b) and (d) illustrate examples of where the recovery of I_d does not correspond to any change in I_g (i.e. they are uncorrelated). In some cases (e.g. Figure 4-17(b)), the recovery in I_d is temporary, as it is followed by a transition back to the original level, presenting a RTN-like behavior. The lack of any change in I_g implies that the annihilated trapped-hole site is not an active TAT center. Past studies [121], [122] have shown the generation of trapped holes with rather deep energy states after NBTI stress (Figure 4-16(a)(ii)). These deep-level hole traps do not contribute to the TAT of holes

but can degrade I_d . The annihilation of a deep-level trapped hole would therefore result in I_d recovery without a corresponding change in I_g .

Figure 4-16. (a) Schematic energy band diagrams illustrating (i) hole tunneling from channel to the polysilicon gate and valence band electron tunneling from the gate to the n -well; (ii) two possible ways by which NBTI stress induced oxide defects contribute to an increase of gate current. Also depicted is the change in oxide band bending caused by hole trapping. The solid (open) circle denotes electron (hole). (b) Hole tunneling current I_{sd} (sensed at the tied source/drain S/D), valence band electron tunneling current I_w (sensed at the n -well) and gate current I_g as a function of gate voltage (V_g) obtain from charge separation measurement on a small-area SiON gate p -MOSFET.

An increase or decrease of I_g without any I_d recovery can also be seen, as denoted by the shaded rectangular regions in Figure 4-17(a) and (b). Such behavior implies the corresponding activation or deactivation of a TAT center without a change in its charge state, leading to I_d being unaffected. The defect concerned may be neutral or a trapped-hole

Figure 4-17. Time-dependent relaxation of drain (I_d) and gate (I_g) current shifts of a small area SiON gate p -MOSFET following NBTI stress, demonstrating examples of the four combinations of ΔI_d and ΔI_g changes: (1) I_d recovery without change of I_g (shaded oval region); (2) I_g change without I_d recovery (shaded rectangular region); (3) I_d recovery with I_g increase (dashed); (4) I_d and I_g recovery (solid).

site whose structure is continuously modified by phonons in such a manner that its energy level may either be aligned or misaligned to the carrier tunneling path (Figure 4-16(a)(ii)) without the emission of the trapped hole. It should be noted that all uncorrelated changes in I_g observed in this work are relatively minor ($\sim 2\%$ compared to I_{g0}), suggesting that the defects involved are near-interface or interface traps.

D. *Negatively Correlated I_d and I_g Fluctuations*

The dashed-line regions in Figure 4-17(b) and (c) represent negative correlation, where the recovery of I_d triggers an *increase* of I_g . Negatively correlated RTN-like fluctuations are also observed in the post-NBTI ΔI_d and ΔI_g relaxation characteristics (Figure 4-18(a)). While such behavior has been observed in the I_d and I_g RTN of the n -MOSFET, it has not yet been reported for the p -MOSFET [192]. A possible explanation lies in the electrostatic effect of hole trapping [190], which reduces the tunneling of holes (and, hence, I_g) due to an increase in the oxide barrier (Figure 4-16(a)(ii)). In some experiments, the reduction of I_g below the pre-stress level was observed after NBTI stress (e.g. Figure 4-17(a)), indicating that hole trapping could indeed overcompensate the increase arising from the stress induced TAT centers. The emission of trapped holes lessens the electrostatic effect and thus increases I_g . However, it should be noted that this explanation is limited to small changes in I_g (of a couple of percent). From a comprehensive simulation study [190], it was found that the electrostatic effect of oxide charging/discharging only accounted for a maximum of 1 % of the change in I_g .

In numerous cases, a much more significant I_g change ($>10\%$ of I_{g0}), which could not be accounted for by only the electrostatic effect, can be seen accompanying the change in I_d (e.g. Figure 4-18(b) and (c)). The significant I_g change must also necessarily imply that it is the bulk oxide defect, and not the interface trap, that is rendered an active (inactive) TAT center (thus creating (eliminating) a partial percolation path in the oxide) after hole emission (capture).

A possible explanation may be found in the shift in the energy level of a bulk oxide defect due to the structural alteration that accompanies the change in charge occupancy. This proposition is supported by an *ab initio* simulation study of the oxygen-vacancy (V_O) defect (a major source of hole trapping centers in the gate oxide) using Vienna *ab initio* simulation package (VASP) described in the literature [199]. The plane wave pseudo-potential method within the density functional theory was adopted. As shown in Figure 4-19, changes in the hole occupancy can lead to a non-negligible shift in the charge transition level (CTL) (akin to the energy level of the defect in the oxide bandgap) [200]; hole emission causes the CTL to shift downwards and vice-versa. For instance, in the positively

Figure 4-18. Post-NBTI relaxation characteristics of I_d and I_g of small-area p-MOSFETs. (a) Negatively correlated RTN-like fluctuations; (b) and (c) Significant I_g changes ($> 10\%$ with respect to I_{g0}) along with that of I_d .

charged state, the CTL of a bulk oxide defect may be situated above the carrier tunneling path (cf. Figure 4-16(a)(ii)), and hence, the defect is an inactive TAT center. After the emission of the trapped hole, the downward shift in the CTL may cause it to coincide with the tunneling path, rendering it an active TAT center. It should be noted that for another V_O defect site in the same supercell, only minor shifts occur in the CTL, implying the role of local oxide chemistry. Thus, changes in the hole occupancy at this site would not affect I_g .

E. Positively Correlated I_d and I_g Fluctuations

Of particular interest in this work is also the positive correlation depicted in Figure 4-17(d) (solid line), where the I_d recovery triggers a corresponding recovery (*decrease*) of I_g . Similar to the negative correlation case, such correspondence is not limited to the downward ΔI_d and ΔI_d steps shown in Figure 4-17(d). Positively correlated RTN-like ΔI_d and ΔI_g fluctuations are also observed (Figure 4-20). A salient point here is also the wide ranging ΔI_g , fluctuating from <5% in Figure 4-17(d) and Figure 4-20(a) to ~40% in Figure 4-20(b), similar to the negatively correlated counterparts shown in Figure 4-17 and Figure

Figure 4-19. Shifts in the charge transition level (CTL) of an oxygen vacancy (V_{O1}) defect upon changes in hole occupancy. The amount of shift is seen to vary with the position of the V_O defect in the simulated supercell (e.g. negligible shifts are obtained at another V_{O2} defect site), implying the role of local oxide chemistry. The supercell under study was attained using molecular dynamics based on the melt-and-quench scheme. Full structural relaxation was carried out at 0 K after each change in hole occupancy.

4-18. In Figure 4-20(b), extremely significant fluctuations of ΔI_d of several microamperes (amounting to $\sim 90\%$ of $I_{d0} \sim 3 \mu\text{A}$) are also observed, in contrast to the changes of tens of nanoamperes seen in previous figures. This implies that the defect concerned may be situated above a main channel constriction path arising from the random dopant distribution effect [139], [143].

Similar observations have been made from RTN measurement of the p -MOSFET, where an increase in I_d (due to trapped-charge emission) is accompanied by a decrease of I_g and vice versa [193]. It was argued that the electrostatic effect arising from trapped-hole emission may provide an explanation for both the negatively and positively correlated ΔI_d and ΔI_g changes [193]. Although hole trapping in the gate oxide retards the tunneling of holes, it can enhance VBE tunneling from the gate (Figure 4-16(a)(ii)). Due to the opposite impact of hole trapping on these two tunneling mechanisms, the emission of a trapped hole

Figure 4-20. Positively correlated RTN-like fluctuations of ΔI_d and ΔI_g observed during the post-NBTI relaxation of two small-area SiON gate p -MOSFETs. (a) Small ΔI_d and ΔI_g fluctuations of $\sim 2\%$ of the respective pre-stress values; (b) “Gigantic” ΔI_d and ΔI_g fluctuations of $\sim 90\%$ of I_{d0} and $\sim 40\%$ of I_{g0} , respectively.

may, therefore, cause either an increase or a decrease in I_g . However, considering that our measurement of I_g was carried out in the regime where $I_w \ll I_{sd}$ (Figure 4-16 (b)); I_w remained below the measurement limit after NBTI stress and the electrostatic effect had a limited impact on I_w , it may be concluded that the electrostatic effect-enhanced I_w is unable to account for the positively correlated ΔI_g fluctuations observed.

Therefore, the positively correlated ΔI_d and ΔI_g changes necessarily imply that the active TAT center concerned must be constantly positively charged in order to degrade I_d while contributing to the I_g increase. Upon emission of the positive trapped charge, the TAT center is deactivated, thus restoring I_d and decreasing I_g . The same inference also applies to the positively correlated ΔI_d and ΔI_g changes observed on the n -MOSFET [193], [194], [197] for which the electrostatic explanation is clearly invalid. To fulfill the above requirement, Bury *et al.* [197] assumed in their nonradiative multi-phonon (NMP) model that the defect concerned was located very close to the oxide/Si interface and in such a manner that it would appear to be charged constantly (thus degrading I_d) while acting as an active TAT center contributing to the I_g increase. This restriction on the location of the defect would, however, limit the model to minor fluctuations in ΔI_g (<5%) (e.g. Figure 4-20(a)).

To account for more significant changes in ΔI_g (e.g. Figure 4-20(b)), the involvement of bulk oxide defects must be considered. For bulk oxide defects, however, the charge capture and emission time constants are generally comparable. This implies that the net charging of the defect site would be zero, which could not explain the correlated changes in ΔI_d . Two scenarios have been proposed in this study that may help fulfill the requirement of the bulk oxide defect being positively charged while at the same time acting as a TAT center:

(1) *Conversion of pre-existing negatively charged defect sites to neutral ones by NBTI stress* – Process-related impurities, such as B and F, can give rise to pre-existing negatively charged defect sites in the gate [201], [202]. Transition to a neutral state due to charge

compensation by hole trapping during NBTI stress degrades I_d . The subsequently neutral defect sites could then also function as TAT centers, increasing I_g .

(2) *Defects that can toggle between a single (+) and double (+2) positively charged state* – The V_O defect may momentarily attain a +2 state by readily losing the dangling-bond electron on one of the Si–Si atoms to the inversion channel when the electron energy level is aligned to the top of the Si valence band. The additional trapped-hole is then emitted towards the gate valence band. A recent study [192] stated that the trapped-hole emission towards the gate valence band entails weak electron-phonon coupling and could occur with nearly no energy barrier. Hence, trapped-hole emission towards the gate may then readily occur (+2 $\rightarrow$ +) with a subsequent re-transition back to the +2 state via electronic charge exchange with the empty states near the top of the Si valence band. In an ultra-thin gate oxide, such electronic charge exchanges between the bulk oxide defect and the inversion channel and gate may occur at time on scales that are much shorter than those required for structural change to occur at the defect site through phonon interaction. This disparity may then create a steady-state trap-assisted conduction process at a single positively charged state that momentarily degrades I_d and increases I_g before phonon-induced structural alteration deactivates the TAT center (cf. Figure 4-19). The existence of +2 defect sites may be corroborated by the results from the time-dependent defect spectroscopy study [182], which revealed two trapped-hole emission events falling in the same cluster in the defect spectroscopy map.

In addition to the explanations on the possible scenarios happening in the positively and negatively correlated RTN-like ΔI_d and ΔI_g in small-area MOSFETs, a similar sequence of capture, relaxation, and emission events has been envisioned to occur in 1/f noise observed in large MOSFETs [203], [204]. Essentially, 1/f noise is a summation of several RTN fluctuators that are simultaneously active. Hence, similar mechanisms and oxygen-vacancy-related defects (oxygen vacancy centers) which are responsible for the 1/f noise in large

MOSFETs are also the culprits for the RTN observed in the small-area MOSFETs with a thin layer of gate dielectric.

4.5 Chapter Summary

New experimental evidence for a correlated evolution of switching-to-permanent hole trapping transformation (as measured by the decrease in NBTI recovery after a given number of stress/relaxation cycles) and SILC generation under NBTI stressing on large-area *p*-MOSFETs has been presented in Section 4.2. Both effects are shown to exhibit a very closely matched temperature dependence (over the range of 100 to 200°C), activation energies and power-law time exponents, further highlighting the strongly correlated behaviors of the two phenomena. Also, a “corner” temperature (~125°C), which marks an increase in the activation energy from the high to low temperature regime, was revealed in both the trapped-hole transformation and SILC generation. These findings corroborate an earlier hypothesis [51], [52], [54] that both effects originate from a common degradation mechanism.

In addition to the evolution of switching-to-permanent hole trapping transformation observed in large-area devices, a direct examination of the recoverable-to-permanent conversion of individual SOTs in small-area SiON *p*-MOSFETs was presented in Section 4.3. The progressive increase in the emission times of SOTs (defect #2, #5 and #6) provides experimental evidence for an initially transient hole trap becoming more permanently charged as the NBTI stress is repeatedly applied. The results confirm our earlier inference of transient-to-permanent hole-trapping transformation, derived from the decreasing recovery per cycle of a large-area device subjected to NBTI stress/relaxation cycling [51], [52], [54], [55], [106]. While the nature of such defects remains elusive, the study clearly demonstrates that the trap parameters, such as the emission time constant, are not time-invariant but evolve continuously during stress. The evolution is believed to be linked to

phonon-facilitated local defect structural changes that occur upon charging/discharging. In addition, a new type of switching hole traps (defect #4), exhibiting intermittent charging during stress and an occasional increase in emission time by ~5 orders of magnitude (from ~0.01–1s to greater than 10^4 s), was observed. Such inconsistent charging during stress may be attributed to the capture time constant being comparable to the stress period. However, the temporary yet significant increase in emission time is a new behavior and has not been reported previously, supporting the proposed existence of metastable trap states [133].

Moreover, the study of post-NBTI time-dependent relaxation of I_d and I_g in small-area SiON p -MOSFETs that was discussed in Section 4.4 led to the discovery of uncorrelated and correlated (both positive and negative) changes for both currents, matching the various combinations of I_d and I_g RTN observed on the n -MOSFET. Enhancement of VBE tunneling that may lead to an increase of I_g after NBTI stress and, hence, a positively correlated I_d and I_g recovery (upon trapped-hole emission), was ruled out using the charge separation measurement. While phonon-induced activation/deactivation of the interface traps as TAT centers may account for the relatively minor correlated I_g changes, it was argued that the involvement of bulk oxide traps was necessary in order to address the much more significant I_g changes (up to ~40%) that were observed. The possible activation/deactivation of a bulk oxide defect as a TAT center upon hole emission/capture that may explain the negatively correlated ΔI_g changes, has been illustrated via an *ab initio* simulation study on the V_O defect. Two scenarios that could lead to a steady-state hole conduction through a single positively charged bulk oxide trap, which accounts for the positively correlated ΔI_g changes, were also proposed.

CHAPTER 5: IMPACT OF VOLTAGE-ACCELERATED STRESS (VAS) ON SWITCHING OXIDE TRAPS AT OPERATING CONDITION

5.1 Introduction

Voltage-accelerated stressing (VAS) is one of the most commonly used methods for assessing the reliability of electronic devices, including the degradation and breakdown of the MOSFET gate oxide. All this while, the attention has been focused on the effect of the newly created oxide traps on parametric shifts and it has been assumed that the applied stress does not affect the pre-existing or time-zero oxide traps (i.e. already present under the operating condition). This assumption is generally valid for large-area MOSFETs, given:

- (1) The usually low density of these traps ($\sim 10^{10} \text{ cm}^{-2}$);
- (2) The relatively insignificant perturbation a trap-occupancy change causes on the overall I_d .

However, the same assumption may not hold true in the case of small-area MOSFETs. It has been shown that the random trapping/detrapping of inversion charges by a few oxide traps in a small-area MOSFET can have a significant impact on the I_d , especially if a trap is located near the channel percolation path arising from the random fluctuations of dopants in the channel region [47], [48], [57], [58], [60], [143], [181]. There have been numerous studies on the statistical analysis of such random charge trapping/detrapping events, through the momentary application of a gate stress voltage and monitoring of the post-stress relaxation characteristics [47], [152], [205].

Although already inferred earlier via the analysis of the post-stress relaxation of large-area *p*-MOSFETs subjected to dynamic negative-bias temperature instability (NBTI) stressing [52], [54], [55], [129], it was only recently confirmed that NBTI stressing can increase the hole-emission times of certain oxide traps (as discussed in Section 4.3 and also reported in [147]), i.e. an oxide trap that can regularly emit the trapped hole within the relaxation interval would subsequently fail to do so, indicating a change from a temporary to a permanent charged state. The results imply that the nature of an oxide trap can evolve under NBTI stressing, possibly due to a change in its atomic structure. While evidence of the change in a trap's emission characteristics has been reported, the effect of VAS on the capture characteristics has not been discovered. In this work, it is stated that the likelihood of an oxide trap to capture a hole under operating condition can be altered after a VAS step. In particular, it is found that an oxide trap can be made to capture a hole either more or less frequently after a prior high voltage stress, indicating that the capture probability in addition to the emission probability of the trap may be altered by the applied stress. This study should be distinguished from past studies (e.g. [47], [57]–[59]) that examined the voltage dependence of a trap's capture characteristics. In these studies, the change in the capture

Figure 5-1. A schematic illustration of the electrical test sequence adopted to examine the effect of voltage-accelerated stressing (VAS) on time-zero switching oxide traps (SOTs) measured in the monitoring step under operating condition at $V_g = -1$ V.

behavior stems from the effect of the applied V_g on the capture barrier energy and the local inversion charge density. In contrast, the change in capture behavior revealed here was observed at the same V_g and without any apparent change in the local inversion charge density after VAS.

5.2 Experimental Details

The DUTs used were small-area p -MOSFETs with 60 nm and 120 nm drawn length and width respectively. The gate stack consists of p^+ polysilicon gate and 1.8 nm SiON gate dielectric developed by decoupled plasma nitridation.

An electrical test sequence illustrated in Figure 5-1 was implemented on a total of 38 DUTs at 100°C to investigate the impact of VAS on the capture activity of the time-zero switching oxide traps (SOTs) measured in the monitoring step under operating condition of $V_g = -1$ V. In step (1), a pristine DUT was first subjected to 100 monitoring cycles, each comprising a 10 s biasing interval under operating condition at $V_g = -1$ V, followed by a 100 s relaxation interval at $V_g = -0.6$ V, during which the linear drain current I_d was continuously measured at a constant drain voltage of $V_d = -0.1$ V and the $|\Delta I_d|$ recovery curves were obtained as shown in Figure 5-2. An abrupt downward step in the $|\Delta I_d|$ recovery

Figure 5-2. Four typical $|\Delta I_d|$ recovery traces during 100 s relaxation interval at $V_g = -0.6$ V ensuing 10 s biasing interval at operation condition of $V_g = -1$ V. Each abrupt step-like drain current, I_d , recovery corresponds to an emission event of a trapped-hole in oxide defect. $|\Delta I_d|^{\text{eor}}$ labeled represents the remnant of $|\Delta I_d|$ at the end of 100 s relaxation interval.

curve indicates that a $|\Delta I_d|$ recovery step originates from the emission of a trapped hole. This $|\Delta I_d|$ recovery step height together with the emission time serve as a means for identifying the SOT that captured a hole during the $V_g = -1$ V interval. The individual oxide traps are identified through the recovery steps in the $|\Delta I_d|$ recovery curves instead of recognizing the individual oxide defects via measuring I_d degradation directly under operating condition at $V_g = -1$ V. This method was employed because an “on-the-fly” approach for studying the capture characteristics of a trap by continuously monitoring the I_d evolution at the applied stress voltage is theoretically sound but difficult to implement experimentally. Figure 5-3 shows an example of the I_d evolution measured directly under operating condition at $V_g = -1$ V. A common difficulty faced is posed by the strong influence of the measurement noise due to a change in the measurement range, making the step degradation in the I_d arising from a hole-capturing event barely visible and difficult to identify. Hence, a more common approach is to observe the time-dependent emission behaviors at a lower voltage (i.e. a relaxation voltage of -0.6 V in this experiment) and deduce the hole-capturing characteristics from there.

After 100 monitoring cycles, VAS at $V_g = -2$ V, which was equivalent to an oxide field of ~ 9 MV/cm, was then applied on the DUT for 100 s, followed by 100 s relaxation at -0.6

Figure 5-3. An instance of I_d evolution monitored under operating condition at $V_g = -1$ V using the same DUT as in Figure 5-2. The dotted lines are eye guides, attempting to delineate what appears to be an I_d degradation step due to a hole captured by an oxide trap at $\sim 10^{-4}$ s amidst the “noisy” measurement data.

V, which is regarded as step (2). After the post-VAS relaxation at -0.6 V had reached steady state, another 100 monitoring cycles were reapplied under the same conditions as stated in step (1), which is referred to as step (3), in order to determine the impact of VAS on the switching activity of the time-zero oxide traps identified in the pre-VAS monitoring step. Steps (2) and (3) were then repeated on the same DUT with a successively higher VAS of -2.5 V and -3 V for each repetition.

Next, the DUTs that had undergone the application of VAS previously were stored at room temperature. After storage for 3 months, the DUTs were again subjected to 100 monitoring cycles at 100°C in order to explore the effect of the long relaxation period on the switching activities of defects, which were previously altered by VAS.

5.3 Results and Discussion

In this section, all the observations from the experiment will be elaborated, followed by a discussion about the possible mechanism behind the observations. By comparing the pre- and post-VAS occupancy or capture activity of a trap under operating condition at $V_g = -1$ V within 10 s during the 100 monitoring cycles, the impact of VAS on the hole-capturing characteristic of the trap can be examined. The capture activity of an oxide trap at $V_g = -1$ V within 10 s during the 100 monitoring cycles is defined as the number of times a trapped-hole emission event (downward recovery step in $|\Delta I_d|$ recovery curves) related to the oxide defect during the 100 s relaxation interval is detected with respect to the 100 monitoring cycles applied. The observation of a trapped-hole event linked to a particular oxide defect during the 100 s relaxation interval implies that the defect was able to capture a hole during the prior -1 V biasing interval.

The $|\Delta I_d|$ recovery step height together with the emission time for each event of a trapped-hole discharged by an oxide defect in the 100 recovery curves measured during the monitoring step were plotted as a “spectral map” enabling the identification of active oxide

defects under operating condition at $V_g = -1$ V [180]. Examples of spectral maps illustrating time-zero active oxide defects under operating condition at $V_g = -1$ V are depicted in Figure 5-4(a)(i). Among the 38 DUTs tested, 17 DUTs shows that a VAS step at $V_g = -2$ V imposes different effects on the capture activity of the time-zero SOTs at $V_g = -1$ V. Apart from the expected appearance of new oxide defects, two intriguing results are consistently obtained after the VAS step by comparing the pre- and post-VAS spectral map; an oxide trap capture activity or switching activity can be either (1) decreased or (2) increased after the VAS step. In the first case, 15 out of the 17 DUTs possess SOTs that exhibit behavior of being less likely to capture a hole after the VAS step; while in the second case, 11 out of the 17 DUTs having SOTs which evolve to be more likely to capture a hole after the VAS step. In order to clearly elucidate the individual defects in various devices, each defect is labeled with an alphabet followed by a number. For examples, trap A3 present trap #3 in device A while trap B2 denotes trap #2 in device B.

Instances described by case (1) showing a decrease in the switching or capture activity of SOTs after an application of VAS at -2 V are depicted in Figure 5-4 and Figure 5-5 for

Figure 5-4. TDDS spectral maps of two test devices D (left) and M (right) extracted from the corresponding $|\Delta I_d|$ recovery traces during monitoring steps before (top) and after (bottom) VAS at $V_g = -2$ V, showing reduction in defect switching activity through the clear decrease in the intensity of cluster D1 (dashed line) and the disappearance of cluster M1 (dotted line) after the VAS step.

trap D1 and M1. The TDDS spectral map of device D in the pristine state under operating condition at $V_g = -1$ V (Figure 5-4(a)(i)) shows 2 distinct clusters, D1 and D2, corresponding to two active time-zero SOTs. The intensity of each cluster in the spectral maps reflects the activity of the corresponding SOT. The I_d shift at the end of each relaxation interval, $|\Delta I_d|^{\text{eor}}$, always goes to ~ 0 in all the pre-VAS 100 monitoring cycles (Figure 5-5), proving that there is no activation of other defects apart from these 2 SOTs. Nevertheless, the TDDS spectral map is remarkably different after the VAS step illustrated in Figure 5-4(a)(ii). The intensity of cluster D1 (dashed line in Figure 5-4(a)) is clearly reduced due to a decrease in the number of hole-emission events during the post-VAS monitoring step (Figure 5-5(a)). This suggests that trap D1 becomes less active under the operating condition after the VAS step. In Figure 5-5, each data point denotes a hole-emission event monitored at -0.6 V in a particular monitoring cycle, which means that the

Figure 5-5. Hole-emission events (symbolized by the data points) during monitoring steps before and after VAS at $V_g = -2$ V illustrating (a) reduction in activity of time-zero defect D1 and (b) disappearance of time-zero defect M1 under operating condition, supporting the observation of TDDS spectral maps in Figure 5-4. Capture activity of the defects is expressed in percent. Cycle numbering continues from the previous monitoring step. The line depicts the evolution of I_d degradation at the end of each relaxation interval, $|\Delta I_d|^{\text{eor}}$, during monitoring steps.

defect was activated and a hole was captured during the prior biasing at -1 V in that monitoring cycle. Evidently, fewer data points (hole-emission events) are observed at the post-VAS monitoring step (open symbols in Figure 5-5(a)) for trap D1. The evolution of I_d degradation at the end of each relaxation interval, $|\Delta I_d|^{\text{eor}}$, during the monitoring steps is also plotted in Figure 5-5(a) as a function of the monitoring-cycle number. An increase in $|\Delta I_d|^{\text{eor}}$ is expected after VAS step compared to the pre-VAS $|\Delta I_d|^{\text{eor}}$ due to the generation of new traps induced by stress. Apart from that, there is no significant change observed in $|\Delta I_d|^{\text{eor}}$ throughout the post-VAS 100 monitoring cycles. It can be clearly observed that there is no evident corresponding reduction in $|\Delta I_d|^{\text{eor}}$ for each of the three hole-emission events of trap D1 in the post-VAS monitoring step. This indicates that the missing of data points or decreasing of hole-emission events of trap D1 after the VAS step is caused by the reduction in the likelihood of the defect being activated to capture a hole under the operating condition and not by the activated defect failing to emit a trapped hole in every relaxation

Figure 5-6. TDDS spectral maps of two test devices A (left) and M (right) extracted from the corresponding $|\Delta I_d|$ recovery traces during monitoring steps before (top) and after (bottom) VAS at $V_g = -2$ V, demonstrating increasing in defect switching activity through the apparent formation of a “new” cluster A5 (dashed rectangle on the left) after the VAS step relative to only one emission event (denoted by $\blacktriangle$) during the pre-VAS monitoring step as well as the appearance of cluster M2 (dotted rectangle on the right) after the VAS step despite no sign of activity observed prior to the VAS step.

interval during the monitoring step. In addition to the reduction in the capture activity of the oxide defects due to VAS, it is interesting to discover the complete disappearance of some of the clusters, such as trap M1 in Figure 5-4(b), from the spectral maps and no emission event being observed after the VAS step. Trap M1 is an extreme instance of reduction in the switching activity of the defect, in which no “activity” or emission event of the defect concerned is detected after the VAS step, even though the defect was apparently active before the VAS step, as illustrated in Figure 5-5(b). Since the increase in $|\Delta I_d|^{\text{cor}}$ that was induced by stress after the VAS step is greater than the trap M1 $|\Delta I_d|$ step height, it is inconclusive whether trap M1, which was a switching oxide trap prior to VAS, has been transformed by VAS into a permanent form [52], [129], [131], [206]. Another possible reason for no emission event being observed in trap M1 after VAS could be the complete failure to activate trap M1 to capture a hole under the operating condition after the VAS step.

On the other hand, an opposite result is observed in case (2), represented by trap A5 and trap M2 in Figure 5-6 and Figure 5-7, demonstrating an increase in switching or capture activity of SOTs after the VAS step. The TDDS spectral map of device A in the pristine state under operating condition at -1 V (Figure 5-6(a)(i)) exhibits a relatively inactive time-zero oxide trap, symbolized by the solid triangle (▲), in addition to the 4 individual clusters, A1–A4, that refer to 4 active time-zero SOTs. Before the application of the VAS step, an oxide defect is rather inactive, activated only once to capture and emit a hole in the 100 monitoring cycles (Figure 5-7(a)). However, after the VAS step, a “new” cluster A5 visibly emerges in the same region as that of the inactive defect that was only activated once prior to the VAS step, as depicted in Figure 5-6(a)(ii). The hollow triangle (△) symbols denoting emission events in Figure 5-7(a) clearly show the significant activity of a defect with an almost identical $|\Delta I_d|$ step height as that of the relatively inactive defect prior to the VAS step. Based on the highly similar $|\Delta I_d|$ step height and emission time, the appearance of the

cluster A5 indicates an increase in the switching or capture activity of the defect under the operating condition after the VAS step. In addition to the increase in the capture activity of the defects induced by VAS, it is interesting to note that some of the clusters, such as cluster M2 in Figure 5-6(b), appear after the VAS step. Trap M2 is an extreme example of an increase occurring in the switching or capture activity of the defect. The oxide trap concerned was inactive and unable to capture a hole at -1 V prior to the VAS step since there is no emission event observed during the pre-VAS 100 monitoring cycles. However, after the VAS step, it becomes active and able to be activated by -1 V intervals to capture a hole in the post-VAS 100 monitoring cycles, and thus, emission events are detected. This is illustrated in Figure 5-7(b). It is important to point out that $|\Delta I_d|^{\text{eor}}$ always returns to ~ 0 in all the pre-VAS 100 monitoring cycles (Figure 5-7), implying that the absence of the emission event in the relaxation intervals during the pre-VAS 100 monitoring cycles is not a consequence of the defect failing to emit the trapped hole but rather because the defect is

Figure 5-7. Hole-emission events (denoted by the data points) during pre- and post-VAS monitoring steps demonstrating (a) increasing in activity of time-zero defect A5 and (b) appearance of time-zero defect M1 under operating condition, corroborating the observation of spectral maps in Figure 5-6. Cycle numbering continues from the previous monitoring step. The line outlines the evolution of I_d degradation at the end of each relaxation interval, $|\Delta I_d|^{\text{eor}}$, during monitoring steps. Capture activity of the defects is expressed in percent.

less prone to be activated to capture a hole under the operating condition before the VAS step.

Apart from the reduced and increased capture activity of switching defects after VAS application discussed earlier, the switching activities of some defects remain unaltered by the VAS step. No apparent change in the number of emission events signifies that the capture activity remains similar after VAS step, as shown by trap G1, G2 and H2 illustrated in Figure 5-8. Hence, it is compelling to explore the effect imposed by the higher VAS voltage, such as -2.5 and -3 V, on the defects probed under operating condition at -1 V, especially for those defects that are unaffected by the VAS at -2 V.

The successively higher VAS voltages of -2.5 V and -3 V impose different effects on the switching or capture activity of the defects under operating condition at -1 V. Figure 5-8 depicts the examples of the capture activity behaviors of oxide traps G1 and G2 in device G

Figure 5-8. Examples of three oxide defects exhibiting different impacts on the switching or capture activity imposed by a series of VAS steps involving successively larger voltages from -2 to -3 V: (a) reduction in the capture activity shown by trap G1 after VAS at -3 V; (b) similar capture activity exhibited by trap G2 without being altered by VAS; and (c) increase in capture activity displayed by trap H2 after VAS at -3 V. Capture activity of the defects is expressed in percentage.

as well as trap H2 in device H. Trap G1 shown in Figure 5-8(a) is highly active under operating condition prior to the VAS step, with hole-emission events detected in almost every monitoring cycle. This trap remains as active as before and its likelihood to capture a hole under the operating condition is not altered by VAS at -2 and -2.5 V. However, only after the application of VAS at -3 V, the capture activity of trap G1 under the operating condition decreases drastically, where only 3 hole-emission events are detected out of 100 monitoring cycles. This outcome should be compared with that in Figure 5-5(a), where a decrease in capture activity is able to be detected in trap D1 after a VAS step at -2 V. Likewise, trap H2 denoted in Figure 5-8(c) is relatively less active under the operating condition prior to the VAS step. The activeness of this trap stays nearly the same under operating condition after the VAS at -2 and -2.5 V. However, after VAS at -3 V, the capture activity of trap H2 under the operating condition increases immensely, wherein the emission events are tracked in almost every monitoring cycle. This indicates that a much higher VAS, such as -3 V, is needed to make certain oxide traps more capture-active at the

Figure 5-9. TDDS spectral maps of a test device J extracted from the corresponding $|\Delta I_d|$ recovery traces during monitoring steps (a) before the VAS step as well as after the VAS steps at (b) -2 V, (c) -2.5 V and (d) -3 V. The intensity of cluster J1 (dashed line) fades progressively through the successive higher VAS voltages from -2 to -3 V. Also, a reduction in the density of cluster J2 (dotted line) is due to a brief disappearance of trap J2's activity owing to the temporary increase in emission time after the -3 V VAS step.

operating condition, contrasting to certain defects that can become more capture-active after being triggered by a relatively lower VAS at -2 V as shown by trap A5 in Figure 5-7(a). On the other hand, some of the oxide traps remain comparatively intact under the operating condition without being affected by VAS. This scenario is typified by a very active trap G2 (Figure 5-8(b)) in device G in which the capture activity under the operating condition is relatively unaffected even after VAS at -3 V. Nonetheless, such a change in the emission characteristic does not always accompany the change in capture characteristic (Figures 5-4, 5.5, 5-6, 5-7), indicating that the capture and emission characteristics of a defect are not correlated in general.

Furthermore, it is observed that the capture activity of an oxide trap under operating

Figure 5-10. (a) A gradual decrease ensued by a complete disappearance in activity of trap J1 after the application of successive higher VAS voltages from -2 to -3 V. (b) A brief disappearance in activity of trap J2 (used to emit in less than 100 s in every cycle) owing to the short-term rising in emission time after applying VAS at -3 V; activity restored after the emission of the captured hole at cycle 330. Capture activity of the defects is expressed in percentage.

condition decreases gradually as indicated by the progressive fading of the cluster intensity after the VAS steps at successively higher voltages, as instantiated by trap J1 in device J shown in Figure 5-9 and Figure 5-10(a). Intriguingly, no hole-emission event can be spotted during the 100 monitoring cycles after VAS at -3 V. It has been discussed previously that the total disappearance of the hole-emission event is inconclusive due to two possible reasons: (1) the trap could ultimately fail to capture a hole under the operating condition after the VAS step and thus, no hole-emission event is detected; or (2) the emission time of the defect increases after the VAS step because the switching defect transforms into a permanent fixed charge upon capturing a hole, failing to emit the captured hole, and thus, no hole-emission is observed, as exemplified by trap J2 in the same device illustrated in Figure 5-9 and Figure 5-10(b). The hole-emission events of trap J2 are clearly probed in every cycle of the 100 monitoring cycles before VAS as well as after VAS of -2 and -2.5 V, indicating that the concerned trap is a highly active switching oxide trap, constantly capturing a hole under the operating condition at -1 V and subsequently emitting the trapped-hole during the relaxation interval at -0.6 V in every monitoring cycle before the VAS step at -3 V. This is supported by the consistent dense intensity of cluster J2 in Figures 5-9(a)-(c). After the VAS step at -3 V, the intensity of cluster J2 decreases as shown in Figure 5-9(d). At first glance, it is thought that the intensity reduction of cluster J2 is promoted by the lowering in the capture activity of trap J2 after the VAS, similar to the scenario delivered by trap D1 (Figure 5-4(a) and Figure 5-5(a)) discussed previously. However, after paying scrupulous attention to the inspection of the J2 emission events during the 100 monitoring cycles after VAS at -3 V (Figure 5-10(b)), it was discovered that the intensity reduction of cluster J2 is due to a brief loss of trap J2 activity caused by the temporary increase in the emission time after the VAS at -3 V. No trace of the hole-emission event of this trap can be observed from monitoring cycles 301 to 329; but at cycle 330, the hole-emission event starts to reoccur and is detected again regularly in nearly every

cycle after monitoring cycle 330. The hole-emission event of this trap at monitoring cycle 330 is accompanied by an abrupt downward recovery step of $|\Delta I_d|^{\text{cor}}$ whose value of recovered $|\Delta I_d|^{\text{cor}}$ is equivalent to the step height of the trap. After the hole-emission event at monitoring cycle 330, the switching or capture activity of the trap reverts to the previous level of activeness in switching or capturing and emitting a hole consistently in every monitoring cycle. This is the notable evidence substantiating that instead of trap J2 losing the likelihood of capturing a hole under the operating condition during the monitoring cycles, the defect fails to discharge the trapped hole owing to an increase in emission time engendered by the application of a high VAS at -3 V. This causes a brief absence of emission events from monitoring cycles 301 till 329 after the VAS step [169], [170]. Nonetheless, no emission events can be observed for trap J1 when the monitoring step was re-executed after three weeks of storage and after a subsequent biasing at $V_g = +2$ V for 100 s, which saw the $|\Delta I_d|^{\text{cor}}$ value being reduced to 61% of the $|\Delta I_d|^{\text{cor}}$ value after the -3 V VAS step.

Figure 5-11. TDDS spectral maps of test devices D (left) and E (right) extracted from the corresponding $|\Delta I_d|$ recovery traces during monitoring steps before VAS step (top), after VAS at -2 V (middle) and after 3 months of storage (bottom), demonstrating that the activity of defects can be restored to their state prior to VAS after sufficient rest period, as seen for trap D1 and E3.

After being stored for an adequate relaxation period at room temperature, interestingly, some of the defects demonstrate that the capture activity can be restored to their state prior to VAS application. Figure 5-11 and Figure 5-12 illustrate examples of this type for trap D1 and E3. As discussed previously, the density of cluster D1 in device D is obviously reduced after VAS at -2 V (dashed line in Figure 5-11(a)(ii)) compared to that prior to the VAS step (dashed line in Figure 5-11(a)(i)), due to a decrease in the number of hole-emission events of trap D1 during monitoring step after the VAS step, implying that trap D1 becomes less capture-active under the operating condition after the VAS step. The monitoring step was reapplied on this device after 3 months of storage. Surprisingly, the number of hole-emission events was restored to a value similar to that prior to the VAS step, as illustrated in Figure 5-12(a). This suggests that the capture activity of trap D1 is reinstated to its state prior to the VAS step after the 3-month storage ensuing the VAS step. On the other hand, defect D2 in the same device exhibits no significant change in the capture activity throughout the experiment. Further, similar to trap A5, an increase in the intensity of cluster E3 in device E is apparent after VAS at -2 V (dotted red line in Figure 5-11(b)(ii)) compared to that before the VAS step (dotted red line in Figure 5-11(b)(i)). This is because the number of hole-emission events of trap E3 rises during the monitoring step after the VAS step, implying that trap E3 has become more active in capturing a hole under the operating condition after the VAS step. After 3 months of storage, the monitoring step was reemployed on this device. Remarkably, the number of hole-emission events reduced, and was reinstated to a level similar to what it was before the VAS step, as depicted in Figure 5-12(b), indicating the restoration of trap E3's capture activity to its state prior to the VAS step after the 3-month storage. On the contrary, no apparent change is detected for the traps E1 and E2 in the same device throughout the experiment.

On the other hand, the capture activity of some defects cannot be reinstated after the storage. An example of this is given by trap B3 depicted in Figure 5-12(c). After VAS at -2

V, a reduction in hole-emission events of trap B3 is observed during the monitoring step, implying that the capture activity under operating condition after the VAS has decreased. After storage for 3 months, the number of hole-emission events detected during the monitoring step remains low, being dissimilar to the number obtained during the pre-VAS monitoring step. This is due to the capture activity of trap B3 not being restored to its state prior to the VAS step after the 3-month storage.

In addition to the alteration of the switching or hole-capture activity of time-zero defects at operating condition by VAS, it is compellingly observed that the single-trap $|\Delta I_d|$

Figure 5-12. Hole-emission events (symbolized by the data points) during the monitoring steps before VAS, after VAS at -2 V and after 3-month storage, for traps (a) D1, (b) E3 and (c) B3. Traps D1 and E3 display restoration in capture activity to their state prior to VAS application after an adequate relaxation period, corroborating the observation of the TDDS spectral maps in Figure 5-11. However, the capture activity of trap B3 does not get reinstated to the level it was before the VAS step after the 3-month storage. Capture activity of the defects is expressed in percentage. The cycle numbering is continued from the previous monitoring step. The fluctuated line outlines the evolution of I_d degradation at the end of each relaxation interval, $|\Delta I_d|^{\text{eor}}$, during the monitoring steps.

recovery amplitude during operating condition can be changed by VAS. This observation should be distinguished from the past studies in [48], [207], [208] which examined the voltage dependence of the step heights induced by charge trapping into defect states with the change in local inversion charge density. On the contrary, the fluctuation in the $|\Delta I_d|$ recovery amplitude detected here was monitored at the same V_g without any apparent change in the local inversion charge density after VAS.

The alteration in the single-trap $|\Delta I_d|$ recovery amplitude during a 100 s emission interval at $V_g = -0.6$ V after the VAS steps was observed in the test device F. Figure 5-13 illustrates the TDDS spectral maps of test device F, displaying three major clusters, F1, F2 and F3. The dashed line is drawn at the value of the extracted single-trap $|\Delta I_d|$ recovery amplitude of cluster F2 during the pre-VAS 100 s relaxation interval at $V_g = -0.6$ V (derived through lognormal fitting (Figure 5-15) that will be explained later), which coincides with

Figure 5-13. TDDS spectral maps of test device F extracted from the corresponding $|\Delta I_d|$ recovery traces during the monitoring steps (a) before VAS step, after VAS at (b) -2, (c) -2.5 and (d) -2.8 V, (e) after 1-week storage, as well as (f) after reapplication of the -2.8 V VAS step. There are 3 main clusters displayed, F1, F2 and F3. The dashed line marks the value of the extracted single-trap $|\Delta I_d|$ recovery amplitude of cluster F2 during the pre-VAS 100 s relaxation interval at $V_g = -0.6$ V, serving as an eye guide to compare and illustrate the change in the amplitude of the drain-current recovery after the application of the VAS steps.

the kernel of the cluster F2 (Figure 5-13(a)). This line acts as a guide with respect to which one can compare the change in the amplitude of the $|\Delta I_d|$ recovery after a series of applications of successively higher voltages in VAS steps and storage (Figures 5-13(b)-(f)). The emission times of traps F1, F2, and F3 before VAS step, as depicted in Figure 5-14, follow an exponential distribution, consistent with a first-order reaction-limited process [59]. Hence, the emission time of each cluster or defect can be extracted by using the exponential distribution fitting shown in Figure 5-14. The fitting equation is

$$\text{Probability density} = \left(\frac{t_e}{\tau_e}\right) \left[\exp\left(-\frac{t_e}{\tau_e}\right)\right] \quad (\text{Eqn. 5-1})$$

where t_e is the relaxation time upon stress termination while τ_e is the emission time constant of an SOT during relaxation. ($V_g = -0.6$ V in this case). The emission time constant of an SOT is a stochastic event and it is one of the characteristics of an SOT. Different SOTs have different unique τ_e . The emission time constant of an SOT is the average time used by

Figure 5-14. The probability density function for emission characteristics of the three defects, F1, F2 and F3 in test device F at $V_g = -0.6$ V during the pre-VAS step. The emission time of these three defects follows an exponential distribution, consistent with a first-order reaction-limited process [59]. The solid, dashed and dotted lines represent the exponential fitting equation (Eqn 5-1) and the extracted emission time constant for each trap is stated above the lines.

an SOT to release a captured charge at a given gate voltage. The chance for the emission of a captured charge by an SOT is high around this emission time constant.

The analysis of cluster F2 is as follows. As mentioned earlier, the dashed line in Figure 5-13 is drawn at the value of the extracted single-trap $|\Delta I_d|$ recovery amplitude of cluster F2 during the pre-VAS 100 s relaxation interval at $V_g = -0.6$ V (derived through lognormal fitting (Figure 5-15) that will be explained later), which coincides with the kernel of the cluster F2, (Figure 5-13(a)). After the application of VAS at -2 V (Figure 5-13(b)), a cluster (cluster X) obtained during the monitoring step is positioned below cluster F2 obtained during the pre-VAS monitoring step. We can see that the kernel of this cluster X obtained after the -2 V VAS step is below the dashed line. The emission times and the $|\Delta I_d|$ recovery amplitudes of cluster F2 obtained during the pre-VAS monitoring step and cluster X obtained during the monitoring step after the -2 V VAS step are compared. The emission

Figure 5-15. Probability plot of drain-current recovery amplitude of single traps F1, F2 and F3. The drain-current recovery amplitude of each individual single trap follows a lognormal distribution [210]–[215], dissimilar from the exponential distribution observed in the drain-current relaxation step amplitudes of single defects obtained from many devices [48], [143], [216]–[221]. The general fitting equation of cumulative density function (CDF) of lognormal distribution is shown in Eqn 5-2, with x in Eqn 5-2 is $|\Delta I_d|$ in this case. The μ , σ^2 and the mean ($|\Delta I_d|$ recovery amplitudes) calculated are tabulated in Table 5-1.

times of these two clusters extracted (first two points from the left of trap F2 in Figure 5-17) are found to be almost the same. However, the $|\Delta I_d|$ recovery amplitude of cluster X obtained during the monitoring step after the -2 V VAS step is lower than that of cluster F2 obtained during the pre-VAS monitoring step (first two points from the left of trap F2 in Figure 5-16). Furthermore, after the VAS step at -2.5 V (Figure 5-13(c)), a cluster (cluster Y) obtained is positioned above cluster X obtained after the -2 V VAS step, but the kernel of cluster Y is still below the dashed line. Since the emission times obtained for clusters F2, X, and Y are almost the same (Figure 5-17) but the $|\Delta I_d|$ recovery amplitudes are different, it is deduced that these clusters originate from the same trap F2, exhibiting different $|\Delta I_d|$ recovery amplitudes due to the application of VAS. It is quite unlikely that clusters F2, X, and Y are caused by different traps, since the possibility of three different traps with the same emission time occurring in the same device is extremely low, and the emission times as well as the $|\Delta I_d|$ recovery amplitudes of defects are stochastic events [47], [209]. Moreover, it is quite unlikely that cluster F2 disappears and gets replaced by new clusters with similar emission time but lower $|\Delta I_d|$ recovery amplitudes after VAS at -2 and -2.5 V. Thus, clusters F2, X, and Y most likely arise from the same defect, which is trap F2.

Now, it is being inferred that the single-trap $|\Delta I_d|$ recovery amplitude of trap F2 changes, since the clusters F2 (Figure 5-13(b) and (c)) are below the dashed line after the application of the VAS at -2 and -2.5 V. Intriguingly, cluster F2 is nowhere to be detected (Figure 5-13(d)) after the VAS step at -2.8 V, similar to the observation found for the behavior of the traps M1 and J1 discussed previously. It has been elaborated previously that the total

Parameter	Trap F1	Trap F2	Trap F3
μ	-15.392	-15.914	-15.636
σ^2	0.070	0.121	0.108
Mean, (μA)	0.214	0.130	0.171

Table 5-1. The values of μ , σ^2 and the mean ($|\Delta I_d|$ recovery amplitudes) calculated for the lognormal distribution curve shown in Figure 5-15.

disappearance of the hole-emission event is inconclusive due to two reasons:

(1) the trap could ultimately fail to capture a hole under the operating condition after the VAS step and, thus, no hole-emission event is detected; or

(2) the emission time of the oxide trap increases after the VAS step because the switching oxide trap transforms into a fixed charge trap upon capturing a hole, failing to emit the hole captured, and thus, no hole-emission is observed.

After the 1-week storage, cluster F2 reappears above the dashed line (Figure 5-13(e)), indicating that the $|\Delta I_d|$ recovery amplitude is higher than that in the pre-VAS state. After that, the VAS step with -2.8 V was reemployed, but this time cluster F2 does not disappear and is below the dashed line, indicating a change in the $|\Delta I_d|$ recovery amplitude. In short, the cluster F2 shifts above and below the dashed line, portraying that the amplitude of the $|\Delta I_d|$ recovery of cluster F2 are not the same after successively higher VAS steps and storage.

In our experiment, the single-trap $|\Delta I_d|$ recovery amplitude distribution follows a lognormal distribution [210]–[215], unlike the exponential distribution discovered in the

Figure 5-16. The evolution of the drain-current recovery amplitude, $|\Delta I_d|$, following the VAS test sequence with respect to that during the pre-VAS (time-zero) step, $|\Delta I_d|_{\text{time-zero}}$, for traps F1, F2 and F3 of test device F. Traps F1 and F3 display no apparent change in the $|\Delta I_d|$ while obvious fluctuation of $|\Delta I_d|$ can be detected for trap F2. There is a missing point for the drain-current recovery amplitude of trap F2 at the first application of the -2.8 V VAS step. This is because the F2 cluster is absent from the TDDS map as shown in Figure 5-13.

drain-current relaxation step amplitude of single defects obtained from many devices [48], [143], [216]–[221]. The $|\Delta I_d|$ recovery amplitude distributions of traps F1, F2 and F3 fit well with lognormal distribution, as shown in Figure 5-15. The general equation for a cumulative density function (CDF) of lognormal distribution [210]–[215] is

$$\text{CDF} = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left[\frac{\ln x - \mu}{\sqrt{2}\sigma} \right] \quad (\text{Eqn. 5-2})$$

where μ and σ are the location and scale parameters, respectively, for the normally distributed logarithm $\ln(X)$ while X is the random variable which is $|\Delta I_d|$ recovery amplitude in the analysis of Figure 5-15. The actual mean and variance for a lognormal distribution of random variable X are

$$\text{Mean} = \exp \left[\mu + \frac{\sigma^2}{2} \right] \quad (\text{Eqn. 5-3})$$

$$\text{Variance} = [\exp(\sigma^2) - 1][\exp(2\mu + \sigma^2)] \quad (\text{Eqn. 5-4})$$

The mean of the $|\Delta I_d|$ recovery amplitude of each trap, calculated based on μ and σ^2 extracted from the lognormal distribution and Eqn 5-3, is used to track the drain-current recovery amplitude alteration after each VAS step. Table 5-1 shows the values of μ , σ^2 and calculated mean of the $|\Delta I_d|$ recovery amplitude. The slope of the fitting is determined by the σ^2 value. The slope which has higher gradient has smaller σ^2 value. Hence, curve with higher gradient of slope indicates that the spread of the $|\Delta I_d|$ recovery amplitudes from the mean value is smaller.

The evolution of the mean of the $|\Delta I_d|$ recovery amplitudes of traps F1, F2 and F3 during the monitoring cycles ensuing the application of each VAS step with respect to the time-zero mean of the $|\Delta I_d|$ recovery amplitude, $|\Delta I_d|_{\text{time-zero}}$, is presented in Figure 5-16. No significant change was detected in the $|\Delta I_d|$ recovery amplitude for traps F1 and F3. However, the $|\Delta I_d|$ recovery amplitude of trap F2 fluctuates significantly throughout the series of VAS step applications. A missing point in the $|\Delta I_d|$ recovery amplitude of trap F2 is observed at the first application of the -2.8 V VAS step due to the absence of the cluster

F2 in the TDDS map illustrated in Figure 5-13. The possible reasons behind the $|\Delta I_d|$ recovery amplitude fluctuation will be discussed later.

To further support the argument that the three detected traps F1, F2 and F3 are not of other traps, the time emission of each of the clusters concerned is extracted using the exponential distribution as displayed in Figure 5-14. Figure 5-17 demonstrates the trend of the extracted emission times for clusters F1, F2 and F3 following the VAS test series. The consistency in emission times of these three clusters throughout the VAS test sequence, together with the $|\Delta I_d|$ recovery amplitudes discussed previously, implies that the three clusters in the spectral maps (Figure 5-13) must originate from the same individual defects, traps F1, F2 and F3 respectively. These three traps show no visible changes in their emission times throughout the VAS step sequence. Similarly, a missing point for the emission time of trap F2 at the first application of the -2.8 V VAS step is due to the absence of the F2 cluster in the TDDS map, as shown in Figure 5-13.

Because of the lack of direct evidence for the atomic nature of the oxide traps, the mechanisms behind the observed changes in the trap capture activity after VAS are not

Figure 5-17. The evolution of the emission time of three defects, F1, F2 and F3, in test device F following the VAS test sequence. The emission time is extracted from the exponential distribution fitting presented in Figure 5-14. No significant change is observed in the emission time for these three traps. There is a missing point for the emission time of trap F2 at the first application of the -2.8 V VAS step. This is because the F2 cluster is absent from the TDDS spectral map, as shown in Figure 5-13.

known at this stage. Considering the increasing oxide field (~ 9 MV/cm at $V_g = -2$ V) due to thickness scaling, a general explanation may be found in the possible distortion of the oxide network and the associated changes in the local atomic configuration of the trap sites that may occur upon the application of the stress, thus changing the characteristics of the traps. A more specific hypothesis on the deactivation of an oxygen vacancy defect, by hydrogen released from the dissociation of Si-H bonds at the interface has been proposed [59], [131], [152]. It should be noted that while the capture activity of a trap is affected after VAS, no shift is observed in its emission-time distribution (Figure 5-4(a) and Figure 5-6(a)). More study is needed to clarify whether the capture and emission characteristics have to necessarily be correlated.

Moreover, previous simulations [222] have demonstrated that any variation in the local inversion charge density that affects a trap's capture characteristic would also be reflected by a change in the I_d recovery step that occurs upon the emission of the trapped charge. However, it can be observed that the average I_d recovery steps associated with SOTs displaying visible change in switching or capture activity after the VAS step remain almost identical to that measured prior to the VAS step (Figures 5-5(a), 5-7 (a), 5-8 and 5-10). This indicates that the possible role of a variation in the local inversion charge density on the capture characteristics may be ruled out. Nonetheless, fluctuations of the I_d recovery step of an SOT without alteration in switching or capture activity (trap F2 in this experiment) are detected in the operating condition after the application of the VAS steps, as shown in Figures 5-13 and 5-16. Other traps, which are near the trap F2 or share the same percolation path with trap F2, may be activated and charged during the VAS, inducing changes and reorganization of the percolation path and hence, increasing or reducing the I_d degradation contribution of trap F2 during the operating condition [222].

It is known that a complete capture time-constant extraction of all VAS-affected traps would strengthen the observation in changing of the capture characteristic of time-zero

SOTs. However, the capture time-constants are not derived due to following reason. A 100-cycle monitoring step, which yields a single data point on the cumulative distribution function plot, utilizes more than 3 hours to complete, since each monitoring cycle consists of a 10 s capture interval at $V_g = -1$ V and a 100 s emission interval at $V_g = -0.6$ V. To achieve a decent cumulative distribution plot (from which the capture time constant is extracted through curve fitting) for a test device, one typically has to lengthen the capture interval to 500 s or longer. Doing so increases the experimental time significantly, as one would need on average ~2–3 days to measure a complete trap occupancy curve. As the experiment involved many devices to establish the statistical significance of the observations, not every test device was fully characterized for the SOT's capture time constants before the VAS step so as to keep the total experimental time at a reasonable level. For some traps, which were characterized before the VAS step, the capture and emission processes are found to follow the exponential distribution shown in Figure 5-18, in agreement with the study reported in [59].

The fitting equation of a cumulative density function for an SOT's capture characteristic at $V_g = -1$ V in Figure 5-18(a) is

$$\text{Occupancy} = \left(\frac{\tau_e^H}{\tau_e^H + \tau_c^H} \right) \left[1 - \exp\left(\frac{-t_s}{\tau_c^H} \right) \right] \quad (\text{Eqn. 5-5})$$

where t_s is the stressing time or capture time interval while τ_c^H and τ_e^H are the capture and emission time constants during stress ($V_g = -1$ V in this case) respectively. The τ_c^H and τ_e^H extracted from the data in Figure 5-18(a) are 3.16 s and 9.04 s respectively. τ_c^H being shorter than τ_e^H is expected since $V_g = -1$ V is a stress voltage to the SOT analyzed in Figure 5-18(a). This SOT has a high chance to capture a charge at $V_g = -1$ V if the stressing time is longer than the τ_c^H (in this case 3.16 s) since τ_c^H is shorter than τ_e^H . The fitting equation of a probability density function for the same SOT's emission characteristic during relaxation ($V_g = -0.6$ V in this case) in Figure 5-18(b) is

$$\text{Probability density} = \left(\frac{t_e}{\tau_e^L} \right) \left[\exp \left(-\frac{t_e}{\tau_e^L} \right) \right] \quad (\text{Eqn. 5-6})$$

where t_e is the relaxation time upon stress termination while τ_e^L is the emission time constant at $V_g = -0.6$ V (in this case). The τ_e^L extracted from the data in Figure 5-18(b) is 1.58 s. τ_e^L being shorter than τ_e^H is expected since an SOT tends to emit a captured charge at a lower voltage and an SOT has a higher chance to emit the trapped charge around the relaxation time of τ_e^L (1.58 s in this case). Since the capture and emission characteristics of an SOT are stochastic events, every SOT has its unique τ_c^H , τ_e^H , and τ_e^L at different voltages. For an SOT to have a high chance to capture a charge during stress, τ_c^H must be shorter τ_e^H . τ_e^L is shorter than τ_e^H since emission of a captured charge is likely to happen at lower voltage. The relative comparison amongst τ_c^H , τ_e^H , and τ_e^L extracted matches with the literature reported in [59].

Moreover, it should be noted that a given VAS has different impacts on the traps. As illustrated in Figures 5-5, 5-7, 5-8 and 5-10, some traps can be affected after a relatively low-voltage stress, whereas others are affected only after stresses involving higher voltages. Hence, the experiment was repeated several times on a given test device (at different stress voltages) and on as many test devices as possible to determine the statistical importance of the observations. In view of the above constraints, certain test conditions would have to remain fixed during the entire course of the experiments. It is understood that a complete time-constant extraction of all VAS-affected traps would strengthen but not change the conclusion. To make the results presented more quantitative, the capture activity of a trap is defined as the percentage of emission events to the total number of cycles applied (if a trap displays an emission event during relaxation, it had captured a hole during the biasing interval), and the value is given in Figures 5-5, 5-7, 5-8 and 5-10. This percentage pertains to a single data point on the CDF plot, and a change in it would necessarily reflect a change of the capture characteristic.

5.4 Chapter Summary

It is found that VAS can either increase or decrease the likelihood of a time-zero oxide trap (i.e. already active under operating condition) to capture a hole. Out of the 75 oxide traps probed in 38 small-area *p*-MOSFETs, 19 traps exhibit decreased capture activity while 11 traps demonstrate increased capture activity after the -2 V VAS step. Moreover, it is observed that a given VAS step has different effects on the oxide traps, i.e. some traps can be affected by a relatively low stress voltage while others are affected only by higher stress voltages. Additionally, a sufficiently long relaxation period after a VAS step may restore the capture activity of a time-zero oxide trap to the pre-VAS state. Furthermore, the drain current recovery step height during the monitoring cycle can be altered by VAS. Given that VAS is commonly used on the gate oxide and that an oxide trap can have a substantial effect on the performance of small-area MOSFETs, further consideration should be given to the possible impact the findings may have on the reliability testing of small-area devices.

Figure 5-18. (a) Cumulative density function for an SOT's capture characteristic at $V_g = -1$ V and (b) probability density function for the same SOT's emission characteristic at $V_g = -0.6$ V. Extremely good fitting of the data (square) to the exponential function (solid lines) can be achieved [59]. The fitting equations and the corresponding coefficients of capture and emission characteristics are stated in (a) and (b) respectively.

CHAPTER 6: FREQUENCY DEPENDENCE OF ELECTRON-TRAP TRANSFORMATION UNDER DYNAMIC POSITIVE-BIAS TEMPERATURE INSTABILITY

6.1 Introduction

In addition to the NBTI studies reported in Chapter 4 and 5, the positive-bias temperature instability (PBTI) degradation was examined as well. It has been discussed in Chapter 2 that the HKMG stack suffers from the PBTI effect having adverse impacts on the *n*-MOSFET performance [45], [134]. This detrimental PBTI effect posts a major concern on mature HKMG stack technology [41], [43]. Similar to NBTI, PBTI induced degradation recovers significantly upon electrical stress termination [44]. However, unlike NBTI, which entails both hole trapping and interface trap creation, PBTI emerges primarily from electron trapping in the high-*k* oxide [41]–[45], [134]. Hence, PBTI recovery during the relaxation stage is predominantly a result of the emission of electron trapped at the oxide defects during the prior stressing phase.

The extent of PBTI recovery progressively decreased as the HfO₂/TiN gate *n*-MOSFET was subjected to repeated stress/relaxation cycles (DPBTI test) [56], similar to the NBTI stressing of the *p*-MOSFET [51], [54]. This outcome implies that a part of the electron trapping is not temporary, and it is gradually transformed into a more permanent trapping during PBTI stressing. It is believed that when given sufficient stressing time, oxide traps may evolve structurally into more stable forms when the associated energy barriers are overcome by the applied electrical *cum* thermal stresses. These changes in turn lead to a shift of the energy states to deeper levels, thus ‘locking-in’ the trapped electrons. Since the

build-up of relatively permanent electron trapping ultimately limits device and circuit lifetimes, an examination of the above effect in terms of the gate frequency is pertinent, as transistors and circuits always operate under an AC bias environment. In the course of this investigation, an unexpected result was obtained. The evolution of electron trapping into a more permanent form is found to exhibit a positive dependence on the gate frequency. Subjecting a HfO_2/TiN gate n -MOSFET to dynamic PBTI stressing at a higher gate frequency is found to result in a larger part of the electron trapping being converted into a more permanent form than when stressing at a lower gate frequency.

6.2 Experimental Details

The DUTs used in this experiment were n -MOSFETs with TiN gates. The high- k dielectric stack consisted of a 3 nm hafnium (IV) oxide (HfO_2) formed via atomic layer deposition (ALD) on a 0.9 nm SiO_x interfacial layer. The equivalent oxide thickness is 1.4 nm determined from capacitance-voltage measurement. The channel length and width are 0.225 μm and 10 μm respectively.

The experimental methodology which explains the frequency dependence of the electron-trapping transformation occurring under positive-bias temperature stressing is

Figure 6-1. A sequence of electrical characterization tests implemented to a HfO_2/TiN gate n -MOSFET to examine the effect of DPBTI stress frequency on the proposed shallow-to-deep oxide trap transformation.

schematically outlined in Figure 6-1. First, a DUT was biased at a relatively low, constant gate voltage (V_g) of 1.4 V for 10^3 s, during which some of the oxide traps in the HfO_2 were charged negatively upon capturing the electrons that were injected from the inversion channel. This portion of the experiment is marked as the trap-filling phase. At the end of the 10^3 s interval, the linear drain current I_d was measured through the gate signal pulsing from 1.4 V to 1 V. Next, the DUT underwent a relaxation state at $V_g = 0$ V for 10^4 s. During this relaxation phase, trapped electrons from the previous trap-filling phase were emitted, and thus, this part of the experiment is referred to as the trap-emptying phase. The electron

Figure 6-2. Four typical time-dependent $|\Delta I_d|$ relaxation curves of a fresh DUT subjected to the repeated alternate trap-filling and trap-emptying phases. The overlapping of the recovery traces implies that the fixed set of test conditions always probes a similar group of oxide traps. The fitting equation is $|\Delta I_d| = B_r[\log_{10}(1+t_s/t_r)] + P$, where t_s is the stress time, t_r is the relaxation time, and B_r and P are constant. Throughout the experiment, the temperature was set to be 100°C while the gate voltage during relaxation was set to be zero. The device was stressed for 10^3 s prior to each relaxation trace, hence t_s is constant. The fitting coefficients B_r and P of each trace are listed in Table 6-1.

detrapping progress was tracked using the linear I_d measured at 1 V at selected intervals. $|\Delta I_d|$ is defined as the magnitude of the difference between the linear I_d measured at a particular time with respect to the fresh linear I_d monitored prior to any stressing. Subjecting the DUT repeatedly to alternate trap-filling and trap-emptying phases yielded nearly identical (i.e. overlapping) time-dependent $|\Delta I_d|$ shift curves, as shown in Figure 6-2, similar to the case of the DNBTI test [105], [107]. This indicates that it is always a similar group of oxide traps that is probed by the fixed set of test conditions. The moderate V_g applied in the trap-filling phase ensured a constant amount of PBTI recovery (Fig. 2 in [56]), and Figure 6-8(b) in this chapter), i.e. there was no significant amount of recoverable-to-permanent (*R-to-P*) electron-trapping conversion during this test.

After the trap-filling and trap-emptying phases, the DUT was subjected to DPBTI stress test, at a higher stress V_g , under a given gate frequency for 5×10^3 s. The top and bottom level of the gate pulse were 2 V (stress voltage) and 0 V (relaxation voltage), respectively,

Figure 6-3. Comparison between transconductance g_m curves of the pristine state and post-DPBTI stressing. After DPBTI stress, the g_m curve shifts laterally towards more positive V_g in comparison to the pre-stress (fresh) g_m curve. However, the peak transconductance value does not degrade after DPBTI stress. This outcome indicates that the interface trap generation is insignificant during DPBTI stressing, and the positive threshold voltage shift is primarily a consequence of the electron trapping.

while the duty cycle was 50%. During the DPBTI stress test, no I_d measurement was made. After the DPBTI stress test, a 10^4 s relaxation at a fixed $V_g = 0$ V was applied to the DUT to allow the PBTI recovery to reach quasi-saturation in order to minimize the effect of the $|\Delta I_d|$ recovery on subsequent tests (see later discussion). The trap-filling and emptying phases were then reapplied on the DUT to examine the impact of the DPBTI test on the electron trapping/detrapping characteristics of the similar group of oxide traps probed prior to the dynamic PBTI test. Subsequently, a 10^3 s relaxation at a negative V_g of -1 V was applied on the DUT, which was immediately ensued by the reapplication of the trap-filling and emptying phases. The purpose of the negative V_g relaxation phase will be explained in the following section. The same DUT was then subjected repeatedly to the entire test series several times but at different gate frequencies each time during the DPBTI stressing phase. The gate frequencies used in this experiment ranged from 1 mHz to 1 MHz. In general, high gate frequency refers to the frequency which is above 10 GHz. However, to facilitate the discussion and explanation in this experiment, the high frequency refers to the frequencies increase towards 1 MHz while the low frequency refers to the frequencies decrease towards 1 mHz.

The DUT degradation was monitored in terms of the linear drain current shift $|\Delta I_d|$, in which the linear, I_d , was measured during the 10^4 s trap-emptying phase by pulsing V_g from 0 V to 1 V with a measurement delay of ~ 0.7 μ s. This fast measurement was achieved through the BTI module using ACS software in a Keithley 4200 SCS parameter analyzer, used in the same system setup as was discussed in Section 3.4.2. The drain voltage was biased at 0.1 V, and the temperature was set to 100°C for the entire experiment. An equivalent shift in the threshold voltage, $|\Delta V_{th}|$, was not converted from $|\Delta I_d|$, as this conversion would entail additional I_d measurements required for the determination of the transconductance g_m . This would, in turn, result in more recovery and lead to the device degradation being underestimated.

6.3 Results and Discussion

There are two possible factors degrading the device performance due to DPBTI stressing, namely stress induced interface trap generation and electron trapping. Although the I_d - V_g curve was not measured during the period of 10^4 -s trap-emptying phase as mentioned in the previous section, I_d - V_g curves were obtained for the pristine state prior to the trap-filling phase and after the DPBTI stress to construct the transconductance g_m curves in order to examine these two possible factors of the device degradation. The typical g_m curves of the pristine state and post-DPBTI stress are illustrated in Figure 6-3. It can be observed that the maximum g_m does not degrade after DPBTI stressing as compared to the maximum g_m of the pristine state. This indicates that there is no significant interface trap generation induced by stress (within the experimental window studied) [43]. On the other hand, the g_m curve after DPBTI stressing shifts laterally towards a more positive V_g

Figure 6-4. Energy band diagrams in the course of: (a) PBTI stressing, as well as (b) and (c) relaxation upon the removal of PBTI stressing. (a) During PBTI stressing, stress induced oxide traps with energy states (denoted by ‘-’) at or below the Si Fermi energy E_{FS} are filled with electrons. (b) When the stress voltage is removed, some trap states are raised above E_{FS} and the trapped electrons are emitted spontaneously, accounting for the initial rapid PBTI recovery. (c) Electrons trapped at deeper trap states ‘see’ a relatively large emission barrier ϕ_B and, thus, require a much longer time to be emitted, accounting for the extended PBTI recovery.

implying that I_d degradation directly probes electron trapping in the dielectric stack.

From Figure 6-2, it is observed that the relaxation traces exhibit an apparent recovery lasting for 10 decades, i.e. from 10^{-6} s to 10^4 s, suggesting that the associated oxide defects responsible for this process involve a broad energy levels distribution, as illustrated in Figure 6-4. Figure 6-4(a) depicts that electrons are captured by the stress induced oxide defects that have energy states at or below the Si Fermi energy level E_{FS} in the course of PBTI stressing. Electrons captured at shallow-level oxide defects having energy states near or above the Si conduction band edge E_C , as portrayed in Figure 6-4(b), are readily emitted upon the termination of PBTI stressing, accounting for the short term fast $|\Delta I_d|$ recovery. The emission barrier ϕ_B illustrated in Figure 6-4(c) represents the difference between the E_C

Figure 6-5. Relaxation curve of the linear drain current shift $|\Delta I_d|$ of a DUT: (a) After the first trap-filling phase before DPBTI stress; (b) After DPBTI stressing at 1 MHz; (c) After the second trap-filling phase. R_1 and R_2 labeled in (a) and (c) represent the difference between $|\Delta I_d|$ at the end of the trap filling and trap-emptying phases (i.e the $|\Delta I_d|$ recovery amount) before and after the DPBTI stress test, respectively. A smaller R_2 compared to R_1 implies that some of the oxide traps that were active in the trap-filling/emptying cycle prior to the DPBTI stress test are no longer involved in the trap-filling/emptying cycle after the DPBTI stress test. The continuous solid lines are the fitting curves and the fitting equation is the same as the one in Figure 6-2, $|\Delta I_d| = B_r [\log_{10}(1+t_s/t_r)] + P$. The coefficients of each fitting curve are displayed in the respective figures.

and the defect energy level. By perceiving either a very small positive ϕ_B or a negative ϕ_B after the stress V_g is set to zero, the electrons trapped at shallow-level oxide defects could be quickly emitted ensuing the termination of PBTI stress. On the other hand, a longer emission time is needed for those electron trapped at deep-level oxide defects, as they see a much larger positive ϕ_B . The emission of these electrons accounts for $|\Delta I_d|$ recovery which extends to several thousand seconds or more. This extended $|\Delta I_d|$ recovery progress is also observed in the time-dependent $|\Delta I_d|$ recovery traces for the trap-emptying phase prior to the application of the DPBTI stress test, for the relaxation interval after the DPBTI stress test, and the following trap-emptying phase, as exemplified in Figure 6-5.

The parameter labeled R_1 in Figure 6-5(a) depicts the difference between the $|\Delta I_d|$ at the start and end points of the $|\Delta I_d|$ recovery curve obtained from the trap-emptying phase prior to the DPBTI stress test. It should be noted that $|\Delta I_d|$ at the ending point of the trap-filling phase is equivalent to $|\Delta I_d|$ at the starting point of trap-emptying phase. Hence, R_1 assesses the amount of recoverable trapped electrons that are emitted from the oxide traps distributed over a certain energy range ΔE as shown in Figure 6-4(b). At the end of recovery, the non-zero $|\Delta I_d|^{\text{cor}}$ ($|\Delta I_d|$ at the end of relaxation) is considered to account for the electrons trapped at deeper levels which are not emitted during the experimental window time frame because their emission times are much longer than the relaxation period in this experiment.

The $|\Delta I_d|$ recovery curve that follows immediately after DPBTI stressing (at 1 MHz) is shown in Figure 6-5(b). The evidently larger $|\Delta I_d|$ at the start of this relaxation phase (or end of the DPBTI stressing phase) as compared to that of the prior trap-emptying phase may be attributed to the greater number of oxide traps being activated or filled by the larger stress V_g of 2 V. This is also made apparent by the greater $|\Delta I_d|$ recovery within the 10^4 s interval. Furthermore, the $|\Delta I_d|$ recovery (~ 285 μA) during the relaxation that follows DPBTI stressing is larger than the $|\Delta I_d|$ recovery ($R_1 \sim 232$ μA) during the trap-emptying phase before the DPBTI stress test. This is ascribed to the larger number of trapped electrons

discharging from the greater amount of activated oxide defects through the 2-V stress. The higher stress V_g had also resulted in an increased amount of deep-level electron trapping, as indicated by the larger remnant $|\Delta I_d|$ at the end of the relaxation interval, as shown in Figure 6-5(b). The reason for including a 10^4 s relaxation interval after the DPBTI stress test is stated as follows. Figure 6-6 shows the dependence of $|\Delta I_d|^{\text{eos}}$ and $|\Delta I_d|^{\text{eor}}$, measured at the end of stress (EOS) and end of recovery (EOR) interval respectively, on the last stress/relax cycle of DPBTI stress as a function of the applied gate pulse frequency (50% duty cycle). The same 5×10^3 s clock period was maintained for all the gate frequencies applied on all DUTs. The overlapping of $|\Delta I_d|^{\text{eos}}$ and $|\Delta I_d|^{\text{eor}}$ points for the 1 MHz stress is an artifact and could not be distinctly resolved owing to the finite measurement delay of ~ 1 μ s. The steady decrease in $|\Delta I_d|^{\text{eos}}$ is expected, since only oxide defects with increasingly shorter capture

Figure 6-6. Dependence of $|\Delta I_d|$ measured at the end-of-stress (EOS) and at the end-of-relaxation (EOR) of the last stress/relaxation cycle as a function of the gate stress frequency (50 % duty cycle). All DUTs were stressed to the same total clock time of 5×10^3 s. The difference between $|\Delta I_d|^{\text{eos}}$ and $|\Delta I_d|^{\text{eor}}$ for the 1 MHz stress could not be clearly resolved due to the finite measurement delay of ~ 1 μ s, and thus, their coincidence is an artifact. The fitting equations for the $|\Delta I_d|^{\text{eos}}$ and $|\Delta I_d|^{\text{eor}}$ data points are $|\Delta I_d|^{\text{eos}}(\mu\text{A})=431-8.5\log_{10}f_g$ and $|\Delta I_d|^{\text{eor}}(\mu\text{A})=184+28.2\log_{10}f_g$ respectively. $|\Delta I_d|^{\text{eos}}$ decreases with frequency in a rate of 8.5 μ A/dec while $|\Delta I_d|^{\text{eor}}$ increases with frequency in a rate of 28.2 μ A/dec.

time constant can be charged as the gate frequency is increased. In other words, there are fewer charged oxide traps leading to a smaller $|\Delta I_d|^{\text{eos}}$, as each of the stress intervals is shorter at a higher gate frequency. However, despite the decrease in the number of trapped electrons in oxide defects, it is intriguing to observe the rising trend of $|\Delta I_d|^{\text{eor}}$ as the gate frequency increases. This behavior may be attributed to the highly asymmetrical charging and discharging characteristics of oxide traps, as this is also observed for hole trapping under NBTI stressing [131]. As shown in Fig. 4 in [131], a very short NBTI stress pulse can give rise to a long relaxation process whose duration extends for many orders of magnitude. Due to this effect, the rate and extent of the $|\Delta I_d|$ recovery after the last relaxation interval varies with the gate frequency. The continuous $|\Delta I_d|$ recovery arising from the DPBTI stress would then influence the $|\Delta I_d|$ recovery determined in the ensuing trap-filling/emptying phases, which are aimed at probing the impact of the DPBTI stress on a given group of oxide traps. In view of this, a sufficiently long relaxation interval was added after the DPBTI stress in order to ensure that the post-DPBTI-stress $|\Delta I_d|$ recovery had approached *quasi-saturation* before the trap-filling/emptying phases were reapplied.

After the trap-filling phase was reapplied, the resulting $|\Delta I_d|$ relaxation curve is as presented in Figure 6-5(c), with R_2 labeled indicating the difference between the $|\Delta I_d|$ at the start and end points of the $|\Delta I_d|$ recovery trace. By comparing the $|\Delta I_d|^{\text{eor}}$ between Figure 6-5(b) and (c), it results that $|\Delta I_d|^{\text{eor}}$ of the trap-emptying phase (Figure 6-5(c)) and $|\Delta I_d|^{\text{eor}}$ at the end of the relaxation interval following the DPBTI stress are almost indistinguishable, indicating an insignificant additional $|\Delta I_d|$ recovery during the trap-filling/emptying cycle after the DPBTI test. This is because the $|\Delta I_d|$ recovery ensuing the DPBTI stress had approached the quasi-saturation state prior to the reapplication of the trap-filling/emptying cycle.

Figure 6-7 compares of the $|\Delta I_d|$ recovery curves obtained from the trap-emptying phases during pre- and post-DPBTI stress. The number of trapped electron discharging

from the recoverable oxide traps during the trap-emptying phase after DPBTI stress, $R_2 \sim 212 \mu\text{A}$, is smaller than the one before the DPBTI stress, $R_1 \sim 232 \mu\text{A}$, as shown in Figure 6-7. Moreover, it can be observed from Figure 6-7 that the $|\Delta I_d|$ recovery rate in the trap-emptying phase after DPBTI test is evidently reduced compared to that before the DPBTI test. These observations in Figure 6-7 lead to two possible interpretations:

(1) the larger remnant $|\Delta I_d|^{\text{eor}}$ at the end of recovery period ensuing the DPBTI stress shown in Figure 6-5(b) suggests that there were more electrons trapped in deeper energy levels in the oxide layer, causing a reduction in the oxide field that might have subsequently decreased the electron trapping events in the following trap-filling phase. Therefore, the PBTI recovery during the subsequent trap-emptying phase would be correspondingly reduced.

(2) Oxide traps at shallow energy levels shifted to deeper trap states during the DPBTI

Figure 6-7. Relaxation curves of the linear drain current shift $|\Delta I_d|$ for the trap-emptying phase before and after DPBTI stress. The larger $|\Delta I_d|$ at the start of trap-emptying phase following the DPBTI stress may be ascribed to an increase in the number of stress-induced deep-level electron trapping events. The curve with circle points (before DPBTI) is the same as the curve in Figure 6-5(a) while the curve with square points (after DPBTI) is the same as the curve in Figure 6-5(c). The solid and dashed lines are the fitting curves for the data points before and after DPBTI respectively.

stress executed at a higher stress V_g . Thus, electrons trapped in the deeper trap states, which would need longer time to emit, were not discharged during the post-DPBTI-stress relaxation interval, attributing to the higher remnant $|\Delta I_d|^{\text{eor}}$, as shown in Figure 6-5(b). Consequently, these oxide defects were not involved in the following trap-filling/emptying cycle. Gao *et al.* [56] reported a progressive reduction in the PBTI recovery amount per cycle when an *n*-MOSFET is repeatedly stressed and relaxed. They also reported that stressing at higher stress V_g exhibits a more pronounced extent of reduction. The capturing of electrons by oxide traps is believed to prompt a structural change in the oxide defects to more stable forms, which in turn provokes a conversion of shallow trap energy levels to deeper trap states.

The two explanations stated previously are further detailed here as follows. To prove explanation (1), alternating trap-filling/emptying phases were subjected to a similar pristine DUT for 30 cycles to monitor the $|\Delta I_d|$, as illustrated in Figure 6-8(a). The progressive increase in the remnant $|\Delta I_d|$ at the end of each trap-emptying phase, $|\Delta I_d|^{\text{eor}}$, presented in Figure 6-8(b) indicates that the deep-level electron trapping increases gradually with the trap-filling/emptying cycle numbers. Nevertheless, a constant amount of $|\Delta I_d|$ recovery or

Figure 6-8. (a) $|\Delta I_d|$ evolution under the repeated application of the trap-filling and emptying phases. (b) The amount of $|\Delta I_d|$ recovery in the trap-emptying phase, R as a function of the number of trap filling/emptying cycles. Also shown is $|\Delta I_d|^{\text{eor}}$ or the remnant $|\Delta I_d|$ after each trap-emptying phase, exhibiting a progressive increase.

recoverable oxide traps, labeled as R in Figure 6-8(b), throughout the 30-cycle trap-emptying phases suggests that the number of electron re-emitted in each of these 30-cycle trap-emptying phases is similar and that the applied test conditions always probe the similar group of oxide traps. This outcome demonstrates that the electron capturing events during the trap-filling phases were not influenced by the gradual increase in the deep-level electron trapping events, demonstrating that explanation (1) is incorrect.

The results obtained in Figure 6-8, however, substantiate explanation (2), which previously stated that the lowered $|\Delta I_d|$ recovery rate and reduced R_2 during the post-DPBTI-stress trap-emptying phase, shown in Figure 6-5(c) and Figure 6-7, are due to a part

Figure 6-9. The overlapping of the $|\Delta I_d|$ relaxation curves during the trap-emptying phases before the DPBTI stress and after the recovery at -1 V subsequent to the DPBTI stress. The outcome is dissimilar to the case illustrated in Figure 6-7, where the $|\Delta I_d|$ relaxation trace obtained during the trap-emptying phase after a 0 V recovery ensuing the DPBTI stress shifts upwards relative to the one acquired preceding DPBTI stress. This entails that the restoration of the original shallow-level oxide defects from the deep energy states is feasible via the application of a negative gate recovery voltage. The fitting equation before the DPBTI stress is $|\Delta I_d|(\mu A)=21.5[\log_{10}(1+1000/t_r)]+21.5$ while the fitting equation after post-DPBTI-stress recovery at -1 V is $|\Delta I_d|(\mu A)=21.6[\log_{10}(1+1000/t_r)]+23.0$. The constants B_r and P of these two fitting curves are comparable.

of the oxide defects located at the shallow energy levels that were converted to deeper trap states during the DPBTI stressing. Therefore, these oxide defects are no longer active in the reapplied trap-filling/emptying cycle subsequent to the DPBTI stress. Figure 6-9 depicts the overlapping of trap-emptying-phase $|\Delta I_d|$ relaxation curves before DPBTI stress and after recovery at -1 V, which further attests explanation (2). The DUT was subjected to a 10^3 s relaxation at $V_g = -1$ V after the post-DPBTI-stress trap-emptying phase. After this negative-bias recovery test was employed, the DUT underwent the trap-filling/emptying phases once again. It is remarkable to observe the overlapping of the trap-emptying-phase $|\Delta I_d|$ recovery trace after negative V_g relaxation with the pre-DPBTI-stress trap-emptying-phase $|\Delta I_d|$ recovery curve. This is in stark contrast to the process depicted in Figure 6-7 that shows an upward shift of the $|\Delta I_d|$ relaxation trace during the trap-emptying phase after a recovery at 0 V succeeding the DPBTI stress, as compared to the one obtained prior to DPBTI stress. This fascinating outcome implies that the previously ‘lost’ PBTI recoverable oxide traps during the post-DBPTI-stress trap-emptying phase owing to the high- V_g DPBTI stress is henceforth restored after the negative V_g recovery. During the negative relaxation biasing, the negative V_g “elevates” the energy levels of deep trap states (some of which originated from shallower trap states), allowing the emission of electrons that were captured in these deep oxide traps. After this electron emission, The emptied deep-level oxide traps are either annihilated or shifted back to shallower energy states. The oxide traps that are restored back to shallower energy levels could henceforth behave as switching oxide traps, capturing and emitting electrons during the following trap-filling and trap-emptying steps, respectively. Moreover, the overlapping and virtually identical $|\Delta I_d|$ recovery traces obtained from the trap-emptying phases before DPBTI stress and after the post-DPBTI-stress negative gate biasing relaxation, illustrated in Figure 6-9, indicate the absence of additional oxide defects generation by the larger V_g adopted during the operation of DPBTI stressing. This deduction is justified by the study in gate current measurement under PBTI

stressing (Fig. 4 in [56]), which validates that no SILC was generated under the investigated stress environments.

The impact of various gate frequencies under DPBTI stress on the difference between the $|\Delta I_d|$ recoverable component, $|\Delta R|$, monitored during the pre-DPBTI-stress (R_1) and post-DPBTI-stress (R_2) trap-emptying phase, i.e. $|\Delta R| = |R_2 - R_1|$, is illustrated in Figure 6-10. As mentioned earlier, R_1 and R_2 examine the number of recoverable trapped electrons that are emitted from the oxide traps within a given energy window ΔE (Figure 6-4(b)) probed by the trap-emptying phases before and after DPBTI stress. The difference between R_1 and R_2 , i.e. $|\Delta R|$, can be deemed as the volume of transformation of the oxide traps, which are probed within a given energy window ΔE by the trap-emptying phases (Figure 6-4(b)), from

Figure 6-10. Increasing trend in $|\Delta R|$ with gate frequency under DPBTI stress. $|\Delta R|$ ($= |R_2 - R_1|$) represents the difference between $|\Delta I_d|$ recovery measured in the trap-emptying phase before (R_1) and after (R_2) the DPBTI stress. $|\Delta R|$, measured under DC PBTI stress, is also plotted for comparison. The DC stress time was equal to the total clock period of the dynamic stress. The different symbols denote the data from the two experimental runs (cf. Figure 6-1) carried out on the same DUT. The linear fitting equation, coefficient of determination (R-squared), and root-mean-square error (RMSE, σ) are shown in the diagram. Further studies are needed to understand the underlying significance of the coefficients in the fitting equation.

shallow energy levels to deeper energy states by DPBTI stress. It is intriguing to note that $|\Delta R|$ increases with the gate frequency, as shown in Figure 6-10. This result signifies that at a higher frequency, the shallow-to-deep energy-state conversion of oxide traps intensifies. It should be pointed out that the same trend, which is $|\Delta R|$ increasing with DPBTI-stress gate frequency, is observed in other similar DUTs that have the same attributes. Despite this, the exact $|\Delta R|$ values vary for different devices.

In addition, it is important to check whether the waveforms of the DPBTI stress applied in the experiment, especially a high gate frequency such as 1 MHz, would produce any artifacts in the results obtained. The reason is that for a high gate frequency, such as 1 MHz, the waveform has a very short period of 1 μ s, which allows the gate voltages at the top level (2 V) and bottom level (0 V) of the signal to stay only for 0.5 μ s each. The waveform of 1 MHz may be distorted severely if the used analyzer cannot generate the high-frequency signals with fast rising and falling times (with respect to the duration of the gate voltage pulse) as well as stable peak signals. Figure 6-11 demonstrates the comparison of the waveforms between 1 MHz (Figure 6-11(a)) and 1 kHz (Figure 6-11(b)) generated by the Keithley Analyzer using the ACS software with rising and falling times of 0.1 μ s, respectively. From Figure 6-11(a), the peak signal of the 1 MHz gate frequency was 2.16 V

Figure 6-11. Gate voltage V_g as a function of time showing the comparison of the waveforms between (a) 1 MHz and (b) 1 kHz under DPBTI stress. The insignificant 0.16-V overshoot (compared to the top-level V_g signal of 2 V applied) of the peak signal at 1-MHz gate frequency is not the culprit to yield higher $|\Delta R|$ obtained in Figure 6-10.

at the end of the rising of the signal from bottom level, which overshoot by 0.16 V, as compared to the top-level gate voltage V_g of 2 V applied under DPBTI stress. After the overshoot, the signal quickly dropped back to the 2 V level before falling back to 0 V at the end of the stress period of 0.5 μ s under the 1 MHz gate frequency. The insignificant overshoot of the peak signal by 0.16 V is not the cause of the higher $|\Delta R|$, which may be regarded as an artifact or unreal outcome. Therefore, the higher $|\Delta R|$ observed at 1 MHz gate frequency under the DPBTI stress (Figure 6-10) is a credible result that is unaffected by the insignificant overshoot at the peak of the top-level signal. Further evidence supporting this fact is stated as follows. Figure 6-11(b) illustrates the gate voltage signal generated under 1 kHz DPBTI stress, exhibiting a stable waveform without any overshoot. The other gate frequencies used, such as 100 kHz, 10 kHz, 100 Hz, 10 Hz, 1 Hz, 0.1 Hz, and 1 mHz, exhibit stable waveforms (not shown), similar to the case of 1 kHz, as shown in Figure 6-11(b). The $|\Delta R|$ observed at 1 kHz is larger than the $|\Delta R|$ monitored at lower frequencies, even though the waveforms of 1 kHz and of the frequencies lower than 1 kHz are equally stable without any overshoot. As a consequence, a higher $|\Delta R|$ at a higher frequency or the rising trend of $|\Delta R|$ with an increase in gate frequency is real and not an artifact created by the overshooting of the high-frequency signal.

The mechanism behind this increasing trend of $|\Delta R|$ with gate frequency of DPBTI stress is explained as follows. Grasser *et al.* suggested an intermediate trapping state “resided in-between” the precursor state and a more permanent trapped-hole state by studying the hole trapping under DPBTI stress [133]. The existence of the intermediate trapping state is considered necessary to explain the reason behind the decrease in hole-trapping events at a high gate frequency (1 MHz). In the initial stage of looking for a credible explanation for the outcome observed in Figure 6-10, a similar model was found, that employed the simplistic two-dimensional energy-coordinate diagram, as illustrated schematically in Figure 6-12. It was hypothesized that the lowering of the energy barrier for

electron trapping by the intermediate trapping state, which involved a lower extent of structural change, allowed the oxide defect to trap an electron rapidly even during the structural change, allowed the oxide defect to trap an electron rapidly even during the

Figure 6-12. A simplistic schematic 2D energy-coordinate diagram illustrating an initial hypothesis of the transformation of a shallow-to-deep trap state, through an intermediate state, during DPBTI stressing. The transition between the different states across different energy barriers involved is indicated by the curvy-rightwards arrow. The intermediate state is a shallow trap that can be readily formed, even under a high gate frequency, due to the small energy barrier ΔE_1 and structural change ΔQ_1 . However, its reversal to the precursor state is restricted by the larger energy barrier ($>\Delta E_1$), and thus, the defect can still remain in the intermediate state when the next stress cycle is applied. Therefore, such defects see the high gate-frequency DPBTI stress as a quasi-DC stress, which enables the defects to have ample time to overcome a greater energy barrier ΔE_2 and structural change ΔQ_2 in order to transit to a more stable deep trap state. At lower gate frequencies, however, the transition to the final deep state becomes less likely due to the longer relaxation interval, which increases the chance of the intermediate state reverting to the precursor state. It should be noted that direct transition from the precursor to the final deep state (dotted line) entails a much larger energy barrier.

application of a high gate frequency. This proposition seems to be coherent with the trend portrayed in Figure 6-10, indicating that an extremely short stress pulse could activate the related oxide traps. Although the electrons could be captured rapidly by the defect sites under an extremely short stress pulse, the oxide defects could not readily emit the trapped electrons upon the termination of stress, due to the higher energy barrier required for electron detrapping. Therefore, the electrons remain trapped in the oxide defects during the relaxation interval at $V_g = 0$ V, even under the pulsed gate stressing. Raising the gate pulsing frequency, which reduces the top-level V_g stress interval and bottom-level of relaxation interval, would increase the likelihood of this scenario. Hence, with the same total clock period in this experiment, DPBTI stressing at a higher gate pulsing frequency of 1 MHz, wherein each cycle exhibits a shorter top-level 2 V stress interval of 0.5 μ s and bottom-level of 0 V relaxation interval of 0.5 μ s, may in effect be identical to a DC stress for the oxide traps, expediting the eventual transformation of the shallow trap states to more stable and deeper energy levels. On the other hand, DPBTI stressing at a lower gate frequency of 1 mHz with the same total clock period used in this experiment, featuring a corresponding longer 0 V relaxation interval of 500 s in each DPBTI stressing cycle, enhances the chances of the oxide defects emitting trapped electrons and reversing to their precursor state, therefore, decreasing the possibility of the conversion of shallow trap states to deeper energy levels.

In order to verify the proposed hypothesis by using the same electrical test series outlined in Figure 6-1, the application of the DPBTI stress was supplanted by a DC stress with the stressing time of 5×10^3 s, same as the total clock period used in DPBTI frequency stress. As explained previously, it is believed that some oxide traps would regard DPBTI stressing as quasi-DC stress. The experimental result from the DC stress is plotted in Figure 6-10. Remarkably, the $|\Delta R|$ from the DC stress does not fall in the same region as the $|\Delta R|$ from the high-frequency DPBTI stress test, and instead, is noticeably smaller than the $|\Delta R|$

from high gate frequencies and comparable to those $|\Delta R|$ obtained from the ultralow gate frequency under DPBTI stress. Therefore, the hypothesis proposed (Figure 6-12), however, cannot be corroborated by the outcome of this DC stress test. In light of this discovery from the experimental result, it appears that a different mechanism may be responsible for the rising trend of $|\Delta R|$ with an increasing gate frequency. Further investigation is required to explicate the nature of the underlying mechanism responsible for this outcome. Nevertheless, this study has unveiled the electron trapping conversion of oxide defects from a shallow trap state to a deeper energy level and its positive dependency of gate frequency under DPBTI stressing, entailing that defect behaviors could actually vary with the applied electrical stress conditions. If this phenomenon is ubiquitous, it may have a significant implication on PBTI modeling.

The relaxation traces in Figures 6-2, 6-5, 6-6, 6-7 were tried to be fitted with equation reported in [131], [223] and the equation is

$$|\Delta I_d| = B_r [\log_{10}(1 + t_s/t_r)] + P \quad (\text{Eqn 6-1})$$

where B_r is the recovery rate which is temperature and electric field dependent, P is the roughly permanent contribution depending on the stress time, t_s is the stress time and t_r is the relaxation time. Throughout the experiment, the temperature was set to be 100°C while the gate voltage during relaxation was set to be zero. The device was stressed for 10^3 s prior to each relaxation trace. The fitting coefficients B_r and P of each trace in Figure 6-2 are

Coefficient	Trace 1	Trace 2	Trace 3	Trace 4
B_r (μA)	32.0	32.2	32.4	32.5
P (μA)	16.2	16.3	17.1	14.0

Table 6-1. Comparison of coefficients A , B and P of the fitting equation for 4 relaxation traces in Figure 6-2. Each coefficient for each relaxation trace is almost the same, supporting the overlapping of these 4 relaxation traces.

listed in Table 6-1. Table 6-1 shows that each coefficient is almost the same in these four relaxation traces. This further supports the fact that the overlapping of these 4 relaxation traces indicates that a similar group of oxide traps is always probed by the fixed set of test conditions.

From Figure 6-5, the coefficient B_r for all three relaxation traces are different, while the terms P , which is the contribution of the permanent degradation, of the curves in Figures 6-5(b) and 6-5(c) are almost the same and are different from that of the curve in Figure 6-5(a). As explained previously, DPBTI frequency stress caused the transformation of recoverable traps to permanent traps. Since the relaxation curve in Figure 6-5(b) was tracked after the 5-kilosecond DPBTI frequency stress which has caused more deep-level electrons trap in the gate dielectric, hence the B_r and P are different from the one before DPBTI frequency stress shown in Figure 6-5(a). Also, coefficients of the curve in Figure 6-5(c) are different from those in Figure 6-5(a) since the trap-emptying phase relaxation curve in Figure 6-5(c) was affected by the effect from the prior DPBTI frequency stress.

The relaxation trace before DPBTI stress in Figure 6-7 is the same as the curve in Figure 6-5(a). As discussed previously, the overlapping of the trap-emptying-phase $|\Delta I_d|$ recovery trace after negative V_g relaxation with the pre-DPBTI-stress trap-emptying-phase $|\Delta I_d|$ recovery curve in Figure 6-7 indicates that the previously ‘lost’ PBTI recoverable oxide traps during the post-DBPTI-stress trap-emptying phase owing to the high- V_g DPBTI stress is henceforth restored after the negative V_g recovery. These two relaxation traces in Figure 6-7 are fitted with Eqn 6-1 and the constants B_r and P in these two fitting curves are matched. This further supports the explanation elaborated previously.

It is to note that the constants B_r and P shown in Table 6-1 (c.f. Figure 6-2) are different from those exhibited by the relaxation traces in Figure 6-2(a) and Figure 6-7. This is because the data presented in Figure 6-2 is of different DUT from the data portrayed in Figure 6-2(a) and Figure 6-7. However, the DUTs used have the same attributes. Hence, the

coefficient B_r is device dependent as well in addition to the temperature and electric-field dependent [131], [223].

Although Eqn 6-1 can fit the relaxation curves in Figures 6-2, 6-5, 6-7 and 6-9 with R -squared of more than 0.8, a better fitting model may be needed to improve the fitting of the relaxation curves. Hence, further studies are needed to explore the improvement in the fitting equation and to elucidate the physical mechanisms behind the coefficients of the improved fitting model. Nonetheless, the main contribution of this experiment is the discovery of the positive gate frequency dependence of shallow-to-deep electron-trapping transformation under DPBTI.

6.4 Chapter Summary

A detailed examination of the gate frequency dependence of a recently revealed shallow-to-deep transformation of electron-trapping states was conducted in this work. Results show an unexpected positive dependence on the gate frequency, i.e. the extent of the shallow-to-deep conversion of oxide traps is found to increase, resulting in more permanent electron trapping occurring at a higher gate frequency. The outcome from DC stress performed for the same stress duration as the total clock period used in DPBTI frequency stress demonstrates a much lower extent of shallow-to-deep transformation, similar to that seen for DPBTI stressing at ultralow gate frequencies. This observation does not support the initial hypothesis based upon intermediate trapping states. In view of this observation, the initial hypothesis would have to be abandoned, and further work is required to shed light on the underlying mechanism.

CHAPTER 7: CONCLUSION AND RECOMMENDATION

7.1 Conclusion

Reduction of the gate oxide thickness below 3.5 nm due to the down scaling of CMOS technology renders BTI degradation as one of the detrimental reliability problems that severely impact the performance of state-of-the-art CMOS devices, especially the nano-scaled small-area devices such as SRAM. In such small-area devices with a thin oxide thickness, only a handful of defects or SOTs are present in the gate oxide layer above the channel region. Owing to the random dopant distribution in the gate oxide layer, a defect located on the crucial percolation channel path may cause a detrimental degradation of the device performance upon capturing a charge from the channel during operation. Hence, studying the BTI effect in small-area devices is vital to understand the nature and properties of individual SOTs, which would serve as evidences to understand the fundamental roots of the degradation caused by individual traps and to countercheck the experimental results of large-area devices reported previously. Thus, this would significantly contribute to the understanding of the physics behind the degradation of the gate oxide caused by the BTI effect.

Fast measurements are needed to monitor the BTI effect due to its fast recovery. The UFS and Keithley's Ultra-Fast BTI Package utilized in the experiments are able to capture the BTI recovery effect with a very short time delay, providing reliable measurement results.

Examination of the correlation between *R-to-P* transformation of hole-trapping and SILC generation in the SiON and HfSiON gate dielectric under NBTI stressing demonstrates that both phenomena exhibit very closely matched power-law time exponents,

activation energies, and most importantly, a non-Arrhenius temperature dependence (over the range of 100–200°C) with a “corner” temperature ($\sim 125^\circ\text{C}$), which marks an increase in the activation energy from the high to low temperature regime. All these findings further highlight the strongly correlated behaviors of these two phenomena and both effects originate from a common degradation mechanism.

In the view of the evolution of the switching-to-permanent hole-trapping transformation observed in large-area devices, a direct examination of the recoverable behavior of the individual SOTs was performed in small-area SiON *p*-MOSFETs subjected to NBTI stress. There are four main categories of SOTs that can be observed from this experiment.

The first group comprises “cyclical” SOTs that are consistently charged/discharged upon capturing/emitting holes during alternating stress/relaxation phases. This indicates that the capture time constants of these defects are much shorter than the applied stress period, and the defect structural changes that occur upon hole-trapping are minimal as well as fully reversible (within the chosen relaxation interval), allowing the trapped hole to be emitted. These regular charging and discharging events of the SOTs in small-area devices further highlight the inconsistency of the R-D model.

The second group is “*R-to-P*” SOTs. The progressive increase in the emission times of these SOTs provides experimental evidence for an initially transient hole trap becoming more permanently charged as the NBTI stress is repeatedly applied. The results confirm the inference regarding the transient-to-permanent hole-trapping transformation that was derived from the decreasing recovery per cycle of large-area devices subjected to NBTI stress/relaxation cycling. While the nature of such defects remains elusive, the study clearly shows that trap parameters, such as the emission time constant, are not time-invariant but evolve continuously during stress. This evolution is believed to be linked to the repetitive application of an oxide stress field, coupled with phonon-facilitated local defect structural changes, which may in turn cause permanent structural relaxation of oxygen vacancy

defects (major source of hole traps) and convert them into stable hole-trapping sites. The direct observation of a trap losing its ability to emit the trapped hole after a certain number of stress/relaxation cycles in this experiment strongly support the proposed explanation.

The third group includes the “newly activated” SOTs that are activated in subsequent cycles, i.e. they are not time-zero SOTs. This implies that these defects can be activated or freshly created by the NBTI stress.

The last group of SOTs is named “intermittent SOTs. These SOTs exhibit intermittent charging during stress and an occasional increase in emission time (by ~5 orders of magnitude – from ~0.01–1s to greater than 10^4 s in this study). Such inconsistent charging during stress may be attributed to the capture time constant being comparable to the stress period. However, the temporary yet significant increase of the observed emission time supports the existence of metastable trap states, which lead to long relaxation times, in addition to two switching states (the precursor and intermediate states), which lead to the quick relaxation.

After the analysis of the outcomes from the two experiments mentioned above, it is intriguing to explore the post-NBTI time-dependent relaxation of I_d and I_g and the correlation between hole-trapping and bulk oxide traps in small-area SiON *p*-MOSFETs. The experiment uncovered uncorrelated and correlated (both positive and negative) changes between both currents (with a total of four combinations). An I_d recovery without a corresponding change in I_g (i.e. uncorrelated) indicated that the annihilated deep-level trapped-hole site is not an active TAT center. Further, an increase or decrease in I_g without a corresponding change in I_d (i.e. uncorrelated) implied a corresponding activation or deactivation of a TAT center without a change in its charge state, indicating that I_d was not altered. Moreover, there were two cases of negatively correlated I_d and I_g fluctuations, such as a recovery of I_d triggering either (1) a small increase in I_g , or (2) a relatively larger increase in I_g (> 10% with respect to fresh I_g). Case (1) may be explained by the

electrostatic effect of hole trapping, where the emission of trapped holes reduces the electrostatic effect and, thus, increases I_g . Case (2) may be attributed to possible activation/deactivation of a bulk oxide defect as a TAT center upon hole emission/capture and this explanation was supported via an *ab initio* simulation study of the V_O defect. Most intriguingly, there were also two cases of positively correlated I_d and I_g recovery, namely (i) a small recovery of I_d triggering a small decrease in I_g ($< 5\%$ with respect to fresh I_g) and (ii) a huge degradation/recovery of I_d ($\sim 90\%$ with respect to fresh I_d) accompanying a large increase/decrease in I_g ($\sim 40\%$ with respect to fresh I_g). While case (i) may be accounted for by the phonon induced activation/deactivation of interface traps as TAT centers, it is argued that the involvement of bulk oxide traps is necessary to address the much more significant I_g changes in case (ii). Two scenarios that could lead to a steady-state hole conduction through a single positively charged bulk oxide trap, which accounts for the positively correlated I_d and I_g recovery, are also proposed. It is to be noted that an increase of I_g after NBTI stress due to enhancement of VBE tunneling and the subsequent positively correlated I_d and I_g recovery (upon trapped-hole emission) is ruled out using charge separation measurement.

An examination regarding the impact of voltage-accelerated stress (VAS) on time-zero SOTs monitored at operating condition revealed that the VAS can either increase or decrease the likelihood of a time-zero oxide trap capturing a hole at operating condition. There were 75 defects detected in the 38 small-area *p*-MOSFETs used. 11 traps displayed increased capture activity while 19 traps demonstrated reduced capture activity after the VAS step at -2 V. Furthermore, the oxide traps are affected differently by a given VAS step. The capture activity of some traps may be altered by a relatively low stress voltage while higher stress voltages are needed to change the capture activity of other defects. Additionally, the capture activity of defects may be reduced progressively by applying successively higher stress voltages. Moreover, the capture activity of a time-zero oxide trap

can be restored to the pre-VAS state if given a substantial recovery period after the VAS step. These findings may be explained by a possible stress-induced oxide network distortion and the corresponding alterations to the local atomic configuration of the defect sites, which may occur during stress application, thus modifying the properties of the defects. A more specific inference of an oxygen vacancy defect deactivated by the hydrogen released from dissociated Si-H bonds at the interface has been suggested. In addition, VAS can modify the drain current recovery step height without changing the capture activity of defects during the monitoring cycle. This may be due to other traps, which are near or share the same percolation path with the defects concerned, being activated and charged during the VAS, inducing changes and reorganization of the percolation path, and hence, increasing or reducing the I_d degradation contribution of concerned traps. Since VAS is usually used on the gate oxide and an oxide defect can have a significant effect on the performance of small-area MOSFETs, further consideration should be given to the possible impact the findings may have on the reliability testing of small-area devices.

A study of the gate frequency dependence of shallow-to-deep electron-trapping transformation subjected to DPBTI stress using HfO_2/TiN gate n -MOSFETs uncovered an unexpected positive dependence on the gate frequency, i.e. there are more shallow oxide trap transforming into a more permanent electron trapping at a higher gate frequency. The result from DC stress implemented for the same stressing duration as the total clock period used in DPBTI frequency stress presents a much lower degree of electron-trapping conversion, which is similar to that observed for DPBTI stressing at ultralow gate frequencies. This observation does not favor the initial hypothesis based upon intermediate trapping states. In view of this observation, that initial hypothesis would have to be abandoned and further study is required to shed light on the nature of the underlying mechanism.

7.2 Recommendations for Future Work

A direct examination of the recoverable behavior of individual SOTs in small-area SiON p -MOSFETs subjected to NBTI stress reveals the recoverable-to-permanent hole-trapping transformation of individual SOTs. The emission time constant is not time-invariant but evolves continuously during stress. It is recommended to study the recoverable behavior of individual SOTs in small-area HKMG p/n -MOSFETs subjected to N/PBTI stress in terms of the recoverable-to-permanent hole/electron-trapping conversion. Charge pumping technique can be used to study the interface trap behaviour in small-area device subjected to NBTI. The study of the temperature and voltage dependence of the transformation may be possible as well. The investigation in the correlation of drain and gate currents of small-area HKMG p/n -MOSFETs subjected to N/PBTI stressing and the study in the locations of the SOTs in gate oxide can help in further advance the understanding of the nature SOTs. Also, charge pumping technique can be used to study the interface trap behaviour in small-area device subjected to NBTI.

The likelihood of a time-zero oxide trap in a SiON p -MOSFET capturing a hole at operating condition can be increased or decreased by VAS. Similar experiments are suggested to be performed on HKMG p/n -MOSFETs, and the extent of the capture activity observed in HKMG p/n -MOSFETs can be compared to the one observed in SiON p -MOSFETs. Besides the DC-VAS, AC-VAS with different frequencies and duty-factors can be applied to study the capture activity of individual SOTs. Moreover, it is intriguing to study the higher temperature effect on the capture activity monitored at lower temperatures provided that the equipment can bring down the temperature to the desired monitoring level instantly in order to reduce the recovery effect caused by the delay in temperature cooling. Further, baking with zero and reverse-bias can be carried out on the devices having traps which show the alteration of capture activity after VAS step in order to better elucidate the underlying mechanism of the switching oxide traps.

The evidence from the study of gate frequency dependence of shallow-to-deep electron-trapping transformation subjected to DPBTI stress using large-area HfO_2/TiN gate n -MOSFETs suggests that the linear drain current I_d degrades more when the frequency increases, indicating that the shallow-to-deep electron-traps transformation is more severe at higher frequency. However, unexpectedly, the DC bias results also revealed that the degradation, or shallow-to-deep electron-trap transformation, is similar to that observed for DPBTI stressing at ultralow gate frequencies. Hence, the underlying reason or failure mechanism needs to be further investigated for this unexpected experimental result. This work can be extended to study the frequency duty-factor dependence of electron traps under PBTI in terms of shallow-to-deep electron-traps transformation, which will further advance the understanding of the PBTI study in HKMG transistors. Moreover, further investigations may be needed to improve the fitting equation of the post-stress relaxation curve. Further, locations and possible regions of deep level electron traps may be studied.

In addition to the future works proposed above using the small-area planar HKMG p/n -MOSFETs, it is intriguing if these future works can be extended to FinFETs and the results can be compared to that of the planar transistors owing to the further reduced dimensions, the multiple interfaces, and the increased lateral gate control of the FinFETs.

LIST OF PUBLICATIONS

- **Z. Y. Tung**, D. S. Ang, and Y. Gao, “On electron-trap transformation and its unexpected frequency dependence under dynamic positive-bias temperature stressing,” in Proceedings IEEE International Reliability Physics Symposium (IRPS), Jun. 2014, pp. XT.10.1–XT.10.6, doi: 10.1109/IRPS.2014.6861187.
- **Z. Y. Tung** and D. S. Ang, “Transient to temporarily permanent and permanent hole trapping transformation in the small area SiON *p*-MOSFET subjected to negative-bias temperature stress,” in Proceedings IEEE 21st International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA), Jun./Jul. 2014, pp. 254–257, doi: 10.1109/IPFA.2014.6898197.
- **Z. Y. Tung**, and D. S. Ang, “Alteration of oxide-trap switching activity at operating condition by voltage-accelerated stressing,” in Proceedings IEEE International Reliability Physics Symposium (IRPS), Apr. 2016, pp. XT.01.1–XT.01.3, doi: 10.1109/IRPS.2016.7574643.
- **Z. Y. Tung** and D. S. Ang, “Impact of voltage-accelerated stress on hole trapping at operating condition,” IEEE Electron Device Letters, vol. 37, no. 5, pp. 644–647, May 2016, doi: 10.1109/LED.2016.2543239.
- A. A. Boo, **Z. Y. Tung**, and D. S. Ang, “New observations on the correlation between hole-trapping transformation and SILC generation under NBTI stressing,” IEEE Electron Device Letters, vol. 37, no. 4, pp. 369–372, April 2016, doi: 10.1109/LED.2016.2531752.
- D. S. Ang, C. J. Gu, **Z. Y. Tung**, A. A. Boo, and Y. Gao, “Evolution of oxide charge trapping under bias temperature stressing,” Microelectronics Reliability, vol. 54, no. 4, pp. 663–681, Apr. 2014, doi: 10.1016/j.microrel.2014.01.025.

REFERENCES

- [1] J. Bardeen and W. H. Brattain, "Physical principles involved in transistor action," *Phys. Rev.*, vol. 75, no. 8, pp. 1208–1225, Apr. 1949.
- [2] J. S. Kilby and E. Keonjian, "Design of a semiconductor-solid-circuit adder," *Electron Devices Meeting, 1959 International*, vol. 5. pp. 76–78, 1959.
- [3] J. S. Kilby, "Miniaturized electronic circuits, reprint of U.S. Patent 3,138,743 (Issued June 23, 1964. Filed Feb 6, 1959)," *IEEE Solid-State Circuits Soc. Newsl.*, vol. 12, no. 2, pp. 44–54, 2007.
- [4] J. S. Kilby, "Invention of the integrated circuit," *IEEE Transactions on Electron Devices*, vol. 23, no. 7. pp. 648–654, 1976.
- [5] R. N. Noyce, "Semiconductor device-and-lead structure, reprint of U.S. Patent 2,981,877 (Issued April 25, 1961. Filed July 30, 1959)," *IEEE Solid-State Circuits Society Newsletter*, vol. 12, no. 2. pp. 34–40, 2007.
- [6] D. Kahng, "Electric field controlled semiconductor device," US Patent 3,102,230., 1963.
- [7] G. E. Moore, "Cramming more components onto integrated circuits, reprinted from electronics, volume 38, number 8, April 19, 1965, pp.114 ff.," *IEEE Solid-State Circuits Society Newsletter*, vol. 11, no. 3. pp. 33–35, 2006.
- [8] G. E. Moore, "Progress in digital integrated electronics," *Electron Devices Meeting, 1975 International*, vol. 21. pp. 11–13, 1975.
- [9] K. J. Kuhn, "Moore's Law past 32nm: future challenges in device scaling," in *Computational Electronics, 2009. IWCE '09. 13th International Workshop*, pp. 1–6, 2009; doi: 10.1109/IWCE.2009.5091124.
- [10] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, "Device scaling limits of Si MOSFETs and their application dependencies," *Proceedings of the IEEE*, vol. 89, no. 3. pp. 259–288, 2001.
- [11] B. E. Deal, "Standardized terminology for oxide charges associated with thermally oxidized silicon," *IEEE Transactions on Electron Devices*, vol. 27, no. 3. pp. 606–608, 1980.
- [12] A. B. Joshi, D. L. Kwong, and S. Lee, "Improvement in performance and degradation characteristics of MOSFETs with thin gate oxides grown at high temperature," *IEEE Electron Device Letters*, vol. 12, no. 1. pp. 28–30, 1991.
- [13] H. S. Momose, S. I. Nakamura, T. Ohguro, T. Yoshitomi, E. Morifuji, T. Morimoto, Y. Katsumata, and H. Iwai, "Study of the manufacturing feasibility of 1.5-nm direct-tunneling gate oxide MOSFETs: uniformity, reliability, and dopant penetration of the gate oxide," *IEEE Transactions on Electron Devices*, vol. 45, no. 3. pp. 691–700, 1998.
- [14] W.-K. Shih, E. X. Wang, S. Jallepalli, F. Leon, C. M. Maziar, and A. F. Tasch Jr., "Modeling gate leakage current in nMOS structures due to tunneling through an ultra-thin oxide," *Solid. State. Electron.*, vol. 42, no. 6, pp. 997–1006, Jun. 1998.

- [15] S. H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, "Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET's," *IEEE Electron Device Letters*, vol. 18, no. 5, pp. 209–211, 1997.
- [16] J. Robertson, "High dielectric constant gate oxides for metal oxide Si transistors," *Reports Prog. Phys.*, vol. 69, no. 2, pp. 327–396, 2006.
- [17] M. Depas, B. Vermeire, P. W. Mertens, R. L. Van Meirhaeghe, and M. M. Heyns, "Determination of tunnelling parameters in ultra-thin oxide layer poly-Si/SiO₂/Si structures," *Solid. State. Electron.*, vol. 38, no. 8, pp. 1465–1471, 1995.
- [18] G. D. Wilk, R. M. Wallace, and J. M. Anthony, "High- κ gate dielectrics: current status and materials properties considerations," *J. Appl. Phys.*, vol. 89, no. 10, pp. 5243–5275, May 2001.
- [19] D. A. Muller, T. Sorsch, S. Moccio, F. H. Baumann, K. Evans-Lutterodt, and G. Timp, "The electronic structure at the atomic scale of ultrathin gate oxides," *Nature*, vol. 399, no. 6738, pp. 758–761, Jun. 1999.
- [20] E. Harari, "Dielectric breakdown in electrically stressed thin films of thermal SiO₂," *J. Appl. Phys.*, vol. 49, no. 4, pp. 2478–2489, 1978.
- [21] D. J. DiMaria, E. Cartier, and D. Arnold, "Impact ionization, trap creation, degradation, and breakdown in silicon dioxide films on silicon," *J. Appl. Phys.*, vol. 73, no. 7, pp. 3367–3384, 1993.
- [22] D. J. DiMaria and E. Cartier, "Mechanism for stress-induced leakage currents in thin silicon dioxide films," *J. Appl. Phys.*, vol. 78, no. 6, pp. 3883–3894, 1995.
- [23] B. Y. Kim, I. M. Liu, H. F. Luan, M. Gardner, J. Fulford, and D. L. Kwong, "Impact of boron penetration on gate oxide reliability and device lifetime in p⁺-poly PMOSFETs," in *1997 IEEE International Reliability Physics Symposium Proceedings*, 1997, pp. 287–291.
- [24] Y. Wu, G. Lucovsky, and Y.-. Lee, "The performance and reliability of PMOSFET's with ultrathin silicon nitride/oxide stacked gate dielectrics with nitrided Si-SiO₂ interfaces prepared by remote plasma enhanced CVD and post-deposition rapid thermal annealing," *IEEE Trans. Electron Devices*, vol. 47, no. 7, pp. 1361–1369, 2000.
- [25] A. Karamcheti, T. Y. Luo, G. Bersuker, G. A. Brown, R. W. Murto, M. D. Jackson, H. R. Huff, P. Kraus, D. Lopes, C. Olsen, and G. Miner, "Device performance of in situ steam generated gate dielectric nitrided by remote plasma nitridation," *Appl. Phys. Lett.*, vol. 78, no. 24, pp. 3875–3877, 2001.
- [26] P. E. Nicollian, G. C. Baldwin, K. N. Eason, D. T. Grider, S. V Hattangady, J. C. Hu, W. R. Hunter, M. Rodder, and A. L. P. Rotondaro, "Extending the reliability scaling limit of SiO₂ through plasma nitridation," *Electron Devices Meeting, 2000. IEDM '00. Technical Digest. International*, pp. 545–548, 2000.
- [27] M. Togo, K. Watanabe, T. Yamamoto, N. Ikarashi, K. Shiba, T. Tatsumi, H. Ono, and T. Mogami, "Low-leakage and highly-reliable 1.5 nm SiON gate-dielectric using radical oxynitridation for sub-0.1 μ m CMOS," *2000 Symposium on VLSI Technology. Digest of Technical Papers* (Cat. No.00CH37104), pp. 116–117, 2000.
- [28] L. Kang, K. Onishi, Y. Jeon, B. H. Lee, C. Kang, W.-J. Qi, R. Nieh, S. Gopalan, R. Choi, and J. C. Lee, "MOSFET devices with polysilicon on single-layer HfO₂ high- k dielectrics," *Electron Devices Meeting, 2000. IEDM '00. Technical Digest. International*, pp. 35–38, 2000.

- [29] X. Chen, S. Samavedam, V. Narayanan, K. Stein, C. Hobbs, C. Baiocco, W. Li, D. Jaeger, M. Zaleski, H. S. Yang, N. Kim, Y. Lee, D. Zhang, L. Kang, J. Chen, H. Zhuang, A. Sheikh, J. Wallner, M. Aquilino, J. Han, Z. Jin, J. Li, G. Massey, S. Kalpat, R. Jha, N. Moumen, R. Mo, S. Kirshnan, X. Wang, M. Chudzik, M. Chowdhury, D. Nair, C. Reddy, Y. W. Teh, C. Kothandaraman, D. Coolbaugh, S. Pandey, D. Tekleab, A. Thean, M. Sherony, C. Lage, J. Sudijono, R. Lindsay, J.-H. Ku, M. Khare, and A. Steegen, "A cost effective 32nm high- k /metal gate CMOS technology for low power applications with single-metal/gate-first process," *2008 Symposium on VLSI Technology*, pp. 88–89, 2008.
- [30] V. Huard, M. Denais, and C. Parthasarathy, "NBTI degradation: from physical mechanisms to modelling," *Microelectron. Reliab.*, vol. 46, no. 1, pp. 1–23, Jan. 2006.
- [31] N. Kimizuka, K. Yamaguchi, K. Imai, T. Iizuka, C.-T. Liu, R. C. Keller, and T. Horiuchi, "NBTI enhancement by nitrogen incorporation into ultrathin gate oxide for 0.10- μ m gate CMOS generation," *VLSI Technology Symposium, 2000. Digest of Technical Papers*, pp. 92–93, 2000.
- [32] D. K. Schroder, "Negative bias temperature instability: what do we understand?," *Microelectron. Reliab.*, vol. 47, no. 6, pp. 841–852, Jun. 2007.
- [33] K. Torii, H. Kitajima, T. Arikado, K. Shiraishi, S. Miyazaki, K. Yamabe, M. Boero, T. Chikyow, and K. Yamada, "Physical model of BTI, TDDB and SILC in HfO_2 -based high- k gate dielectrics," *Electron Devices Meeting, 2004. IEDM Technical Digest. IEEE International*, pp. 129–132, 2004.
- [34] J. H. Stathis and S. Zafar, "The negative bias temperature instability in MOS devices: a review," *Microelectron. Reliab.*, vol. 46, no. 2–4, pp. 270–286, Feb. 2006.
- [35] Y. M. and Y. Matukura, "Investigation of silicon-silicon dioxide interface using MOS structure," *Jpn. J. Appl. Phys.*, vol. 5, no. 2, pp. 180, 1966.
- [36] J. P. Campbell, P. M. Lenahan, C. J. Cochrane, A. T. Krishnan, and S. Krishnan, "Atomic-scale defects involved in the negative-bias temperature instability," *IEEE Transactions Device and Materials Reliability*, , vol. 7, no. 4, pp. 540–557, 2007.
- [37] D. S. Ang and S. Wang, "Insight into the suppressed recovery of NBTI-stressed ultrathin oxynitride gate pMOSFET," *Electron Device Letters, IEEE*, vol. 27, no. 9, pp. 755–758, 2006.
- [38] D. S. Ang and K.-L. Pey, "Evidence for two distinct positive trapped charge components in NBTI stressed p-MOSFETs employing ultrathin CVD silicon nitride gate dielectric," *Electron Device Letters, IEEE*, vol. 25, no. 9, pp. 637–639, 2004.
- [39] Y. Mitani, M. Nagamine, H. Satake, and A. Toriumi, "NBTI mechanism in ultra-thin gate dielectric - nitrogen-originated mechanism in SiON ," *Electron Devices Meeting, 2002. IEDM '02. International*, pp. 509–512, 2002.
- [40] S. Zafar, A. Kumar, E. Gusev, and E. Cartier, "Threshold voltage instabilities in high- k ; gate dielectric stacks," *IEEE Transactions Device and Materials Reliability*, vol. 5, no. 1, pp. 45–64, 2005.
- [41] 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 HfO_2 -based high- k metal gate stack reliability and implications on t_{inv} -scaling," *Electron Devices Meeting (IEDM), 2011 IEEE International*, pp. 18.4.1-18.4.4, 2011.

- [42] K. Zhao, J. H. Stathis, A. Kerber, and E. Cartier, "PBTI relaxation dynamics after AC vs. DC stress in high-k/metal gate stacks," *Reliability Physics Symposium (IRPS), 2010 IEEE International*. pp. 50–54, 2010.
- [43] S. Pae, M. Agostinelli, M. Brazier, R. Chau, G. Dewey, T. Ghani, M. Hattendorf, J. Hicks, J. Kavalieros, K. Kuhn, M. Kuhn, J. Maiz, M. Metz, K. Mistry, C. Prasad, S. Ramey, A. Roskowski, J. Sandford, C. Thomas, J. Thomas, C. Wiegand, and J. Wiedemer, "BTI reliability of 45 nm high-k + metal-gate process technology," *Reliability Physics Symposium, 2008. IRPS 2008. IEEE International*. pp. 352–357, 2008.
- [44] G. Bersuker, J. H. Sim, C. S. Park, C. D. Young, S. Nadkarni, R. Choi, and B.-H. Lee, "Mechanism of electron trapping and characteristics of traps in HfO₂ gate stacks," *IEEE Transactions Device and Materials Reliability*, vol. 7, no. 1. pp. 138–145, 2007.
- [45] H. C.-H. Wang, C.-W. Tsai, S.-J. Chen, C.-T. Chan, H.-J. Lin, Y. Jin, H.-J. Tao, S.-C. Chen, C. H. Diaz, T. Ong, A. S. Oates, M.-S. Liang, and M.-H. Chi, "Reliability of HfSiON as gate dielectric for advanced CMOS technology," *VLSI Technology Symposium, 2005. Digest of Technical Papers*, pp. 170–171, 2005.
- [46] N. Kimizuka, T. Yamamoto, T. Mogami, K. Yamaguchi, K. Imai, and T. Horiuchi, "The impact of bias temperature instability for direct-tunneling ultra-thin gate oxide on MOSFET scaling," *VLSI Technology Symposium, 1999. Digest of Technical Papers*, pp. 73–74, 1999.
- [47] 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," *Reliability Physics Symposium (IRPS), 2010 IEEE International*. pp. 16–25, 2010.
- [48] J. Franco, B. Kaczer, M. Toledano-Luque, P. J. Roussel, J. Mitard, L.-A. Ragnarsson, L. Witters, T. Chiarella, M. Togo, N. Horiguchi, G. Groeseneken, M. F. Bukhori, T. Grasser, and A. Asenov, "Impact of single charged gate oxide defects on the performance and scaling of nanoscaled FETs," *Reliability Physics Symposium (IRPS), 2012 IEEE International*. pp. 5A.4.1-5A.4.6, 2012.
- [49] G. Chen, K. Y. Chuah, M.-F. Li, D. S. H. Chan, C. H. Ang, J. Z. Zheng, Y. Jin, and D. L. Kwong, "Dynamic NBTI of PMOS transistors and its impact on device lifetime," *Reliability Physics Symposium Proceedings, 2003 IEEE International*. pp. 196–202, 2003.
- [50] M. Ershov, S. Saxena, H. Karbasi, S. Winters, S. Minehane, J. Babcock, R. Lindley, P. Clifton, M. Redford, and A. Shibkov, "Dynamic recovery of negative bias temperature instability in p-type metal-oxide-semiconductor field-effect transistors," *Applied Physics Letters*, vol. 83, no. 8. pp. 1647–1649, 2003.
- [51] A. A. Boo and D. S. Ang, "Evolution of hole trapping in the oxynitride gate *p*-MOSFET subjected to negative-bias temperature stressing," *IEEE Transactions Electron Devices*, vol. 59, no. 11. pp. 3133–3136, Nov. 2012.
- [52] A. A. Boo, D. S. Ang, Z. Q. Teo, and K. C. Leong, "Correlation between oxide trap generation and negative-bias temperature instability," *IEEE Electron Device Letters*, vol. 33, no. 4. pp. 486–488, 2012.
- [53] Y. Gao, D. S. Ang, A. A. Boo, and Z. Q. Teo, "On the dynamic NBTI of the HfO₂ and HfSiON *p*-MOSFET," *Microelectron. Eng.*, vol. 88, no. 7. pp. 1392–1395, Jul. 2011.

- [54] Y. Gao, D. S. Ang, C. D. Young, and G. Bersuker, "Evidence for the transformation of switching hole traps into permanent bulk traps under negative-bias temperature stressing of high-k P-MOSFETs," *Reliability Physics Symposium (IRPS), 2012 IEEE International*. pp. 5A.5.1-5A.5.5, 2012.
- [55] Y. Gao, A. A. Boo, Z. Q. Teo, and D. S. Ang, "On the evolution of the recoverable component of the SiON, HfSiON and HfO₂ p-MOSFETs under dynamic NBTI," *Reliability Physics Symposium (IRPS), 2011 IEEE International*. p. XT.8.1-XT.8.6, 2011.
- [56] Y. Gao, D. S. Ang, G. Bersuker, and C. D. Young, "Electron trap transformation under positive-bias temperature stressing," *Electron Device Letters, IEEE*, vol. 34, no. 3. pp. 351–353, 2013.
- [57] T. Grassler, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P. J. Wagner, F. Schanovsky, J. Franco, P. Roussel, and M. Nelhiebel, "Recent advances in understanding the bias temperature instability," *Electron Devices Meeting (IEDM), 2010 IEEE International*. pp. 4.4.1-4.4.4, 2010.
- [58] D.-C. H. and J. G. and C.-F. Huang, "Investigation of positive and negative bias temperature instability of high- κ dielectric metal gate metal–oxide–semiconductor-field-effect-transistors by random telegraph signal," *Jpn. J. Appl. Phys.*, vol. 52, no. 3R, pp. 36503.1-36503.4, 2013.
- [59] T. Grassler, K. Rott, H. Reisinger, M. Waltl, P. Wagner, F. Schanovsky, W. Goes, G. Pobegen, and B. Kaczer, "Hydrogen-related volatile defects as the possible cause for the recoverable component of NBTI," *Electron Devices Meeting (IEDM), 2013 IEEE International*. pp. 15.5.1-15.5.4, 2013.
- [60] N. Tega, H. Miki, M. Yamaoka, H. Kume, T. Mine, T. Ishida, Y. Mori, R. Yamada, and K. Torii, "Impact of threshold voltage fluctuation due to random telegraph noise on scaled-down SRAM," *2008 IEEE International Reliability Physics Symposium*. pp. 541–546, 2008.
- [61] C. L. Chen, Y. M. Lin, C. Wang, and K. Wu, "A new finding on NBTI lifetime model and an investigation on NBTI degradation characteristic for 1.2nm ultra thin oxide," *Reliability Physics Symposium, 2005. Proceedings. 43rd Annual. 2005 IEEE International*. pp. 704–705, 2005.
- [62] V. Reddy, A. T. Krishnan, A. Marshall, J. Rodriguez, S. Natarajan, T. Rost, and S. Krishnan, "Impact of negative bias temperature instability on digital circuit reliability," *Reliability Physics Symposium Proceedings, 2002. 40th Annual*. pp. 248–254, 2002.
- [63] A. Acovic, G. La Rosa, and Y.-C. Sun, "A review of hot-carrier degradation mechanisms in MOSFETs," *Microelectron. Reliab.*, vol. 36, no. 7, pp. 845–869, 1996.
- [64] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J. Appl. Phys.*, vol. 94, no. 1, pp. 1–18, Jun. 2003.
- [65] D. J. DiMaria and J. H. Stathis, "Trapping and trap creation studies on nitrided and reoxidized - nitrided silicon dioxide films on silicon," *J. Appl. Phys.*, vol. 70, no. 3, pp. 1500 – 1509, Aug. 1991.
- [66] C. H. Liu, M. T. Lee, C.-Y. Lin, J. Chen, K. Schroefer, J. Brighten, N. Rovedo, T. B. Hook, M. V Khare, S.-F. Huang, C. Wann, T.-C. Chen, and T. H. Ning, "Mechanism and process dependence of negative bias temperature instability (NBTI)

- for pMOSFETs with ultrathin gate dielectrics,” in *International Electron Devices Meeting. Technical Digest* (Cat. No.01CH37224), pp. 39.2.1-39.2.4, 2001.
- [67] “The International Technology Roadmap for Semiconductors.” [Online]. Available: <http://www.itrs2.net/>.
 - [68] K. Saminadayar and J. C. Pfister, “Evolution of surface-states density of Si/wet thermal SiO₂ interface during bias-temperature treatment,” *Solid. State. Electron.*, vol. 20, no. 11, pp. 891–896, 1977.
 - [69] S. M. Sze and K. K. Ng, “Physics of Semiconductor Devices.” *Hoboken, UNITED STATES: John Wiley & Sons, Incorporated*, 2006.
 - [70] C. Sah, J. Y. Sun, and J. J. Tzou, “Deactivation of the boron acceptor in silicon by hydrogen,” *Applied Physics Letters*, vol. 43, no. 2, pp. 204–206, 1983.
 - [71] A. E. Islam, H. Kufluoglu, D. Varghese, and M. A. Alam, “Critical analysis of short-term negative bias temperature instability measurements: explaining the effect of time-zero delay for on-the-fly measurements,” *Applied Physics Letters*, vol. 90, no. 8, pp. 83503–83505, 2007.
 - [72] P. V Gray and D. M. Brown, “Density of SiO₂–Si interface states,” *Applied Physics Letters*, vol. 8, no. 2, pp. 31–33, 1966.
 - [73] E. H. Poindexter and P. J. Caplan, “Characterization of Si/SiO₂ interface defects by electron spin resonance,” *Prog. Surf. Sci.*, vol. 14, no. 3, pp. 201–294, 1983.
 - [74] M. L. Reed and J. D. Plummer, “Chemistry of Si–SiO₂ interface trap annealing,” *Journal of Applied Physics*, vol. 63, no. 12, pp. 5776–5793, 1988.
 - [75] A. Stirling, A. Pasquarello, J.-C. Charlier, and R. Car, “Dangling bond defects at Si–SiO₂ interfaces: atomic structure of the P_{b1} center” *Phys. Rev. Lett.*, vol. 85, no. 13, pp. 2773–2776, Sep. 2000.
 - [76] P. M. Lenahan, “Atomic scale defects involved in MOS reliability problems,” *Microelectron. Eng.*, vol. 69, no. 2–4, pp. 173–181, Sep. 2003.
 - [77] T. D. Mishima, P. M. Lenahan, and W. Weber, “Do P_{b1} centers have levels in the Si band gap? Spin-dependent recombination study of the P_{b1} ‘hyperfine spectrum,” *Appl. Phys. Lett.*, vol. 76, no. 25, pp. 3771–3773, Jun. 2000.
 - [78] A. Stesmans and V. V Afanas’ev, “Si dangling-bond-type defects at the interface of (100) Si with ultrathin HfO₂,” *Appl. Phys. Lett.*, vol. 82, no. 23, pp. 4074–4076, Jun. 2003.
 - [79] D. M. Fleetwood, P. S. Winokur, R. A. Reber, T. L. Meisenheimer, J. R. Schwank, M. R. Shaneyfelt, and L. C. Riewe, “Effects of oxide traps, interface traps, and “border traps” on metal-oxide-semiconductor devices,” *J. Appl. Phys.*, vol. 73, no. 10, pp. 5058–5074, May 1993.
 - [80] D. M. Fleetwood, “Border traps and bias-temperature instabilities in MOS devices,” *Microelectron. Reliab.*, vol. 80, pp. 266–277, 2018.
 - [81] D. Ielmini, M. Manigrasso, F. Gattel, and M. G. Valentini, “A new NBTI model based on hole trapping and structural relaxation in MOS dielectrics,” *IEEE Trans. Electron Devices*, vol. 56, no. 9, pp. 1943–1952, 2009.
 - [82] W. L. Warren, D. M. Fleetwood, M. R. Shaneyfelt, P. S. Winokur, and R. A. B. Devine, “Defect-defect hole transfer and the identity of border traps in SiO₂ films,” *Phys. Rev. B*, vol. 50, no. 19, pp. 14710–14713, Nov. 1994.

- [83] D. M. Fleetwood, T. L. Meisenheimer, and J. H. Scofield, "1/f noise and radiation effects in MOS devices," *IEEE Trans. Electron Devices*, vol. 41, no. 11, pp. 1953–1964, 1994.
- [84] W. L. Warren, M. R. Shaneyfelt, D. M. Fleetwood, J. R. Schwank, P. S. Winokur, and R. A. B. Devine, "Microscopic nature of border traps in MOS oxides," *IEEE Trans. Nucl. Sci.*, vol. 41, no. 6, pp. 1817–1827, 1994.
- [85] A. J. Lelis, T. R. Oldham, H. E. Boesch, and F. B. McLean, "The nature of the trapped hole annealing process," *IEEE Trans. Nucl. Sci.*, vol. 36, no. 6, pp. 1808–1815, 1989.
- [86] J. F. Conley, P. M. Lenahan, A. J. Lelis, and T. R. Oldham, "Electron spin resonance evidence for the structure of a switching oxide trap: Long term structural change at silicon dangling bond sites in SiO₂," *Appl. Phys. Lett.*, vol. 67, no. 15, pp. 2179–2181, Oct. 1995.
- [87] J. F. Conley, P. M. Lenahan, A. J. Lelis, and T. R. Oldham, "Electron spin resonance evidence that E'_{γ} centers can behave as switching oxide traps," *IEEE Trans. Nucl. Sci.*, vol. 42, no. 6, pp. 1744–1749, 1995.
- [88] W. L. Warren, J. Kanicki, and E. H. Poindexter, "Paramagnetic point defects in silicon nitride and silicon oxynitride thin films on silicon," *Colloids Surfaces A Physicochem. Eng. Asp.*, vol. 115, pp. 311–317, 1996.
- [89] J. P. Campbell, P. M. Lenahan, A. T. Krishnan, and S. Krishnan, "Location, structure, and density of states of NBTI-induced defects in plasma nitrided pMOSFETs," in *2007 IEEE International Reliability Physics Symposium Proceedings. 45th Annual*, 2007, pp. 503–510.
- [90] Y. H. Poon, H. Wong, and K. F. Man, "Deep level transient spectroscopy of hole emission from nitrided oxide/silicon interface traps," in *Proceedings 1996 IEEE Hong Kong Electron Devices Meeting*, pp. 81–86, 1996.
- [91] D. M. Fleetwood, "'Border traps' in MOS devices," *IEEE Trans. Nucl. Sci.*, vol. 39, no. 2, pp. 269–271, 1992.
- [92] K. O. Jeppson and C. M. Svensson, "Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices," *Journal of Applied Physics*, vol. 48, no. 5, pp. 2004–2014, 1977.
- [93] M. A. Alam, H. Kufluoglu, D. Varghese, and S. Mahapatra, "A comprehensive model for PMOS NBTI degradation: recent progress," *Microelectron. Reliab.*, vol. 47, no. 6, pp. 853–862, Jun. 2007.
- [94] M. A. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," *Microelectron. Reliab.*, vol. 45, no. 1, pp. 71–81, Jan. 2005.
- [95] S. Ogawa, M. Shimaya, and N. Shiono, "Interface-trap generation at ultrathin SiO₂ (4–6 nm)-Si interfaces during negative-bias temperature aging," *J. Appl. Phys.*, vol. 77, no. 3, p. 1137, 1995.
- [96] S. Mahapatra, P. Bharath Kumar, and M. A. Alam, "Investigation and modeling of interface and bulk trap generation during negative bias temperature instability of p-MOSFETs," *Electron Devices, IEEE Transactions on*, vol. 51, no. 9, pp. 1371–1379, 2004.
- [97] M. A. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETs," *Electron Devices Meeting, 2003. IEDM '03 Technical Digest. IEEE International*, pp. 14.4.1–14.4.4, 2003.

- [98] B. Kaczer, V. Arkhipov, M. Jurczak, and G. Groeseneken, "Negative bias temperature instability (NBTI) in SiO₂ and SiON gate dielectrics understood through disorder-controlled kinetics," *Microelectron. Eng.*, vol. 80, no. 0, pp. 122–125, Jun. 2005.
- [99] D. Varghese, D. Saha, S. Mahapatra, K. Ahmed, F. Nouri, and M. Alam, "On the dispersive versus Arrhenius temperature activation of NBTI time evolution in plasma nitrided gate oxides: measurements, theory, and implications," *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*. pp. 684–687, 2005.
- [100] S. Zafar, B. H. Lee, and J. Stathis, "Evaluation of NBTI in HfO₂ gate-dielectric stacks with tungsten gates," *Electron Device Letters, IEEE*, vol. 25, no. 3. pp. 153–155, 2004.
- [101] L. Tsetseris, X. J. Zhou, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, "Physical mechanisms of negative-bias temperature instability," *Appl. Phys. Lett.*, vol. 86, no. 14, pp. 142103, Mar. 2005.
- [102] C. Shen, M.-F. Li, C. E. Foo, T. Yang, D. M. Huang, A. Yap, G. S. Samudra, and Y.-C. Yeo, "Characterization and physical origin of fast V_{th} transient in NBTI of pMOSFETs with SiON dielectric," *Electron Devices Meeting, 2006. IEDM '06. International*. pp. 1–4, 2006; doi: 10.1109/IEDM.2006.346776.
- [103] H. Reisinger, O. Blank, W. Heinrigs, A. Muhlhoff, W. Gustin, and C. Schlunder, "Analysis of NBTI degradation- and recovery-behavior based on ultra fast V_T -measurements," *Reliability Physics Symposium Proceedings, 2006. 44th Annual, IEEE International*. pp. 448–453, 2006.
- [104] Z. Q. Teo, D. S. Ang, and K. S. See, "Can the reaction-diffusion model explain generation and recovery of interface states contributing to NBTI?," *Electron Devices Meeting (IEDM), 2009 IEEE International*. pp. 1–4, 2009. doi: 10.1109/IEDM.2009.5424233.
- [105] Z. Q. Teo, D. S. Ang, and C.-M. Ng, "Non-hydrogen-transport characteristics of dynamic negative-bias temperature instability," *Electron Device Letters, IEEE*, vol. 31, no. 4. pp. 269–271, 2010.
- [106] D. S. Ang, Z. Q. Teo, T. J. J. Ho, and C. M. Ng, "Reassessing the mechanisms of negative-bias temperature instability by repetitive stress/relaxation experiments," *Device and Materials Reliability, IEEE Transactions on*, vol. 11, no. 1. pp. 19–34, 2011.
- [107] Z. Q. Teo, D. S. Ang, and C. M. Ng, "Separation of hole trapping and interface-state generation by ultrafast measurement on dynamic negative-bias temperature instability," *Electron Device Letters, IEEE*, vol. 31, no. 7. pp. 656–658, 2010.
- [108] T. Grassler, H. Reisinger, W. Goes, T. Aichinger, P. Hehenberger, P.-J. Wagner, M. Nelhiebel, J. Franco, and B. Kaczer, "Switching oxide traps as the missing link between negative bias temperature instability and random telegraph noise," *Electron Devices Meeting (IEDM), 2009 IEEE International*. pp. 1–4, 2009; doi: 10.1109/IEDM.2009.5424235.
- [109] C. Z. Zhao and J. F. Zhang, "Effects of hydrogen on positive charges in gate oxides," *J. Appl. Phys.*, vol. 97, no. 7, p. 73703, Mar. 2005.
- [110] J. H. Stathis, "Physical and predictive models of ultra thin oxide reliability in CMOS devices and circuits," in *2001 IEEE International Reliability Physics Symposium Proceedings. 39th Annual (Cat. No.00CH37167)*, pp. 132–149, 2001.

- [111] P. E. Blöchl and J. H. Stathis, "Hydrogen electrochemistry and stress-induced leakage current in silica," *Phys. Rev. Lett.*, vol. 83, no. 2, pp. 372–375, Jul. 1999.
- [112] V. Huard, F. Monsieur, G. Ribes, and S. Bruyere, "Evidence for hydrogen-related defects during NBTI stress in *p*-MOSFETs," in *2003 IEEE International Reliability Physics Symposium Proceedings, 2003. 41st Annual.*, 2003, pp. 178–182.
- [113] S. Tsujikawa and J. Yugami, "Evidence for bulk trap generation during NBTI phenomenon in *p*-MOSFETs with ultrathin SiON gate dielectrics," *IEEE Trans. Electron Devices*, vol. 53, no. 1, pp. 51–55, 2006.
- [114] S. Ogawa and N. Shiono, "Generalized diffusion-reaction model for the low-field charge-buildup instability at the Si-SiO₂ interface," *Phys. Rev. B*, vol. 51, no. 7, pp. 4218–4230, Feb. 1995.
- [115] T. J. J. Ho, D. S. Ang, A. A. Boo, Z. Q. Teo, and K. C. Leong, "Are interface state generation and positive oxide charge trapping under negative-bias temperature stressing correlated or coupled?," *Electron Devices, IEEE Transactions on*, vol. 59, no. 4, pp. 1013–1022, 2012.
- [116] K. Hess, I. C. Kizilyalli, and J. W. Lyding, "Giant isotope effect in hot electron degradation of metal oxide silicon devices," *IEEE Trans. Electron Devices*, vol. 45, no. 2, pp. 406–416, 1998.
- [117] J. W. Lyding, K. Hess, and I. C. Kizilyalli, "Reduction of hot electron degradation in metal oxide semiconductor transistors by deuterium processing," *Appl. Phys. Lett.*, vol. 68, no. 18, pp. 2526–2528, Apr. 1996.
- [118] M. Houssa, V. V Afanas'ev, A. Stesmans, M. Aoulaiche, G. Groeseneken, and M. M. Heyns, "Insights on the physical mechanism behind negative bias temperature instabilities," *Appl. Phys. Lett.*, vol. 90, no. 4, pp. 43505, Jan. 2007.
- [119] J. I. Pankove, D. E. Carlson, J. E. Berkeyheiser, and R. O. Wance, "Neutralization of shallow acceptor levels in silicon by atomic hydrogen," *Phys. Rev. Lett.*, vol. 51, no. 24, pp. 2224–2225, Dec. 1983.
- [120] D. S. Ang and S. Wang, "Insight into the suppressed recovery of NBTI-stressed ultrathin oxynitride gate pMOSFET," *Electron Device Letters, IEEE*, vol. 27, no. 9, pp. 755–758, 2006.
- [121] D. S. Ang and S. Wang, "Recovery of the NBTI-stressed ultrathin gate p-MOSFET: the role of deep-level hole traps," *Electron Device Letters, IEEE*, vol. 27, no. 11, pp. 914–916, 2006.
- [122] D. S. Ang, S. Wang, G. A. Du, and Y. Z. Hu, "A consistent deep-level hole trapping model for negative bias temperature instability," *IEEE Transactions Device and Materials Reliability*, vol. 8, no. 1, pp. 22–34, 2008.
- [123] C. D. Y. and Y. Z. and M. P. and B. H. L. and K. M. and J. H. S. and R. C. and G. A. B. and R. W. M. and G. Bersuker, "Ultra-short pulse current-voltage characterization of the intrinsic characteristics of high- κ devices," *Jpn. J. Appl. Phys.*, vol. 44, no. 4B, pp. 2437–2440, 2005.
- [124] T. Grassler, B. Kaczer, T. Aichinger, W. Goes, and M. Nelhiebel, "Defect creation stimulated by thermally activated hole trapping as the driving force behind negative bias temperature instability in SiO₂, SiON, and high-*k* gate stacks," in *2008 IEEE International Integrated Reliability Workshop Final Report*, pp. 91–95, 2008.

- [125] J. P. Campbell, P. M. Lenahan, A. T. Krishnan, and S. Krishnan, "Observations of NBTI-induced atomic scale defects," in *2005 IEEE International Integrated Reliability Workshop*, 2005, doi: 10.1109/IRWS.2005.1609551.
- [126] A. J. Lelis and T. R. Oldham, "Time dependence of switching oxide traps," *IEEE Trans. Nucl. Sci.*, vol. 41, no. 6, pp. 1835–1843, 1994.
- [127] H. Reisinger, O. Blank, W. Heinrigs, W. Gustin, and C. Schlunder, "A comparison of very fast to very slow components in degradation and recovery due to NBTI and bulk hole trapping to existing physical models," *IEEE Trans. Device Mater. Reliab.*, vol. 7, no. 1, pp. 119–129, 2007.
- [128] V. Huard, "Two independent components modeling for negative bias temperature instability," in *2010 IEEE International Reliability Physics Symposium*, pp. 33–42, 2010.
- [129] Z. Q. Teo, A. A. Boo, D. S. Ang, and K. C. Leong, "On the cyclic threshold voltage shift of dynamic negative-bias temperature instability," *Reliability Physics Symposium (IRPS), 2011 IEEE International*. p. XT.10.1-XT.10.5, 2011.
- [130] T. Grassler and B. Kaczer, "Evidence that two tightly coupled mechanisms are responsible for negative bias temperature instability in oxynitride MOSFETs," *IEEE Transactions Electron Devices*, vol. 56, no. 5. pp. 1056–1062, 2009.
- [131] T. Grassler, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel, "A two-stage model for negative bias temperature instability," *Reliability Physics Symposium, 2009 IEEE International*. pp. 33–44, 2009.
- [132] A. G. Marinopoulos, I. Batyrev, X. J. Zhou, R. D. Schrimpf, D. M. Fleetwood, and S. T. Pantelides, "Hydrogen shuttling near Hf-defect complexes in Si/SiO₂/HfO₂ structures," *Appl. Phys. Lett.*, vol. 91, no. 23, p. 233503, Dec. 2007.
- [133] T. Grassler, B. Kaczer, H. Reisinger, P.-J. Wagner, and M. Toledano-Luque, "On the frequency dependence of the bias temperature instability," *Reliability Physics Symposium (IRPS), 2012 IEEE International*. p. XT.8.1-XT.8.7, 2012.
- [134] C. Shen, M. F. Li, X. P. Wang, H. Y. Yu, Y. P. Feng, A. T. L. Lim, Y. C. Yeo, D. S. H. Chan, and D. L. Kwong, "Negative U traps in HfO₂ gate dielectrics and frequency dependence of dynamic BTI in MOSFETs," in *IEDM Technical Digest. IEEE International Electron Devices Meeting, 2004.*, pp. 733–736, 2004.
- [135] C. Gu, D. S. Ang, Y. Gao, R. Gu, Z. Zhao, and C. Zhu, "A vacancy-nterstitial defect pair model for positive-bias temperature stress-induced electron trapping transformation in the high-*k* gate n-MOSFET," *IEEE Trans. Electron Devices*, vol. 64, no. 6, pp. 2505–2511, 2017.
- [136] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant, "Discrete resistance switching in submicrometer silicon inversion layers: individual interface traps and low-frequency (1/f) Noise," *Phys. Rev. Lett.*, vol. 52, no. 3, pp. 228–231, Jan. 1984.
- [137] J. Franco, B. Kaczer, M. Toledano-Luque, P. J. Roussel, G. Groeseneken, B. Schwarz, M. Bina, M. Waltl, P. J. Wagner, and T. Grassler, "Reduction of the BTI time-dependent variability in nanoscaled MOSFETs by body bias," in *2013 IEEE International Reliability Physics Symposium (IRPS)*, p. 2D.3.1-2D.3.6, 2013.
- [138] V. Huard, C. Parthasarathy, C. Guerin, T. Valentin, E. Pion, M. Mammasse, N. Planes, and L. Camus, "NBTI degradation: from transistor to SRAM arrays," in *2008 IEEE International Reliability Physics Symposium*, pp. 289–300, 2008.

- [139] A. Asenov, R. Balasubramaniam, A. R. Brown, and J. H. Davies, "RTS amplitudes in decanometer MOSFETs: 3-D simulation study," *Electron Devices, IEEE Transactions on*, vol. 50, no. 3. pp. 839–845, 2003.
- [140] T. Grassler, K. Rott, H. Reisinger, M. Waltl, F. Schanovsky, and B. Kaczer, "NBTI in nanoscale MOSFETs — the ultimate modeling benchmark," *Electron Devices, IEEE Transactions on*, vol. 61, no. 11. pp. 3586–3593, 2014.
- [141] M. J. Kirton and M. J. Uren, "Noise in solid-state microstructures: a new perspective on individual defects, interface states and low-frequency ($1/f$) noise," *Adv. Phys.*, vol. 38, no. 4, pp. 367–468, Jan. 1989.
- [142] B. Kaczer, T. Grassler, 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 *2009 IEEE International Reliability Physics Symposium*, pp. 55–60, 2009.
- [143] B. Kaczer, T. Grassler, 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," *Reliability Physics Symposium (IRPS), 2010 IEEE International*. pp. 26–32, 2010.
- [144] T. Grassler, K. Rott, H. Reisinger, M. Waltl, J. Franco, and B. Kaczer, "A unified perspective of RTN and BTI," in *2014 IEEE International Reliability Physics Symposium*, pp. 4A.5.1-4A.5.7, 2014.
- [145] M. Toledano-Luque, B. Kaczer, P. Roussel, M. J. Cho, T. Grassler, and G. Groeseneken, "Temperature dependence of the emission and capture times of SiON individual traps after positive bias temperature stress," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, vol. 29, no. 1. pp. 01AA04-1-01AA04-5, 2011.
- [146] M. Toledano-Luque, B. Kaczer, E. Simoen, P. J. Roussel, A. Veloso, T. Grassler, and G. Groeseneken, "Temperature and voltage dependences of the capture and emission times of individual traps in high-k dielectrics," *Microelectron. Eng.*, vol. 88, no. 7, pp. 1243–1246, Jul. 2011.
- [147] T. Grassler, M. Waltl, W. Goes, Y. Wimmer, A.-M. El-Sayed, A. L. Shluger, and B. Kaczer, "On the volatility of oxide defects: Activation, deactivation, and transformation," *Reliability Physics Symposium (IRPS), 2015 IEEE International*. pp. 5A.3.1-5A.3.8, 2015.
- [148] T. Grassler, 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," *2015 IEEE International Electron Devices Meeting (IEDM)*. pp. 20.1.1-20.1.4, 2015.
- [149] T. Grassler, M. Waltl, G. Rzepa, W. Goes, Y. Wimmer, A. M. El-Sayed, A. L. Shluger, H. Reisinger, and B. Kaczer, "The 'permanent' component of NBTI revisited: saturation, degradation-reversal, and annealing," in *2016 IEEE International Reliability Physics Symposium (IRPS)*, 2016, pp. 5A–2–1–5A–2–8.
- [150] T. Grassler, M. Waltl, K. Puschkarsky, B. Stampfer, G. Rzepa, G. Pobegen, H. Reisinger, H. Arimura, and B. Kaczer, "Implications of gate-sided hydrogen release for post-stress degradation build-up after BTI stress," in *2017 IEEE International Reliability Physics Symposium (IRPS)*, pp. 6A–2.1–6A–2.6, 2017.
- [151] T. Grassler, B. Stampfer, M. Waltl, G. Rzepa, K. Rupp, F. Schanovsky, G. Pobegen, K. Puschkarsky, H. Reisinger, B. O'Sullivan, and B. Kaczer, "Characterization and

- physical modeling of the temporal evolution of near-interfacial states resulting from NBTI/PBTI stress in nMOS/pMOS transistors,” in *2018 IEEE International Reliability Physics Symposium (IRPS)*, pp. 2A.2-1-2A.2-10, 2018.
- [152] T. Grasser, W. Goes, Y. Wimmer, F. Schanovsky, G. Rzepa, M. Watzl, K. Rott, H. Reisinger, V. V Afanas'ev, A. Stesmans, A.-M. El-Sayed, and A. L. Shluger, “On the microscopic structure of hole traps in pMOSFETs,” *Electron Devices Meeting (IEDM), 2014 IEEE International*. p. 21.1.1-21.1.4, 2014.
 - [153] Y. Wimmer, W. Goes, A. M. El-Sayed, A. L. Shluger, and T. Grasser, “A density-functional study of defect volatility in amorphous silicon dioxide,” in *2015 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2015, pp. 44–47.
 - [154] T. Yang, M. F. Li, C. Shen, C. H. Ang, C. Zhu, Y. C. Yeo, G. Samudra, S. C. Rustagi, and M. B. Yu, “Fast and slow dynamic NBTI components in p-MOSFET with SiON dielectric and their impact on device life-time and circuit application,” *VLSI Technology, 2005. Digest of Technical Papers. 2005 Symposium on*. pp. 92–93, 2005.
 - [155] D. S. Ang, “Observation of suppressed interface state relaxation under positive gate biasing of the ultrathin oxynitride gate p-MOSFET subjected to negative-bias temperature stressing,” *Electron Device Letters, IEEE*, vol. 27, no. 5. pp. 412–415, 2006.
 - [156] JEDEC STANDARD JESD90, “A procedure for measuring p-channel MOSFET negative bias temperature instabilities,” 2004.
 - [157] M. Denais, C. Parthasarathy, G. Ribes, Y. Rey-Tauriac, N. Revil, A. Bravaix, V. Huard, and F. Perrier, “On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's,” *Electron Devices Meeting, 2004. IEDM Technical Digest. IEEE International*. pp. 109–112, 2004.
 - [158] J. F. Zhang, Z. Ji, M. H. Chang, B. Kaczer, and G. Groeseneken, “Real V_{th} instability of pMOSFETs under practical operation conditions,” *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*. pp. 817–820, 2007.
 - [159] H. Reisinger, U. Brunner, W. Heinrigs, W. Gustin, and C. Schlunder, “A comparison of fast methods for measuring NBTI degradation,” *IEEE Trans. Device Mater. Reliab.*, vol. 7, no. 4, pp. 531–539, 2007.
 - [160] A. Kerber, E. Cartier, L. Pantisano, M. Rosmeulen, R. Degraeve, T. Kauerauf, G. Groeseneken, H. E. Maes, and U. Schwalke, “Characterization of the V_T -instability in $\text{SiO}_2/\text{HfO}_2$ gate dielectrics,” in *2003 IEEE International Reliability Physics Symposium Proceedings, 2003. 41st Annual.*, 2003, pp. 41–45.
 - [161] C. Shen, M.-F. Li, X. P. Wang, Y.-C. Yeo, and D.-L. Kwong, “A fast measurement technique of MOSFET I_d - V_g characteristics,” *Electron Device Letters, IEEE*, vol. 27, no. 1. pp. 55–57, 2006.
 - [162] G. A. Du, D. S. Ang, Z. Q. Teo, and Y. Z. Hu, “Ultrafast measurement on NBTI,” *Electron Device Letters, IEEE*, vol. 30, no. 3. pp. 275–277, 2009.
 - [163] Y. Z. Hu, D. S. Ang, and Z. Q. Teo, “Threshold voltage and mobility extraction by ultrafast switching measurement on NBTI,” *Electron Devices, IEEE Transactions on*, vol. 57, no. 8. pp. 2027–2031, 2010.
 - [164] “Keithley's Model 4200-SCS Semiconductor Characterization System Reference Manual,” 2013. [Online]. Available:

<https://doc.xdevs.com/doc/Keithley/SCS/4200/4200-901-01%20%28P%20-%20Feb%202013%29%28Ref%29.pdf>

- [165] Z. Q. Teo, "Characterization of dynamic NBTI by ultra-fast charge pumping and ultra-fast switching method," 2015.
- [166] "Keithley's Automated Characterization Suite (ACS) Reference Manual," 2015. [Online]. Available: download.tek.com/manual/ACS-901-01_J_Sep_2015_Ref.pdf.
- [167] H.-S. Huang, P. Samanta, T.-J. Tzeng, S.-Y. Chen, and C.-H. Liu, "Electron detrapping in thin hafnium silicate and nitrided hafnium silicate gate dielectric stacks," *Appl. Phys. Lett.*, vol. 100, no. 2, p. 23501, Jan. 2012.
- [168] Z. Q. Teo, D. S. Ang, and C.-M. Ng, "'Non-hydrogen-transport' characteristics of dynamic negative-bias temperature instability," *Electron Device Letters, IEEE*, vol. 31, no. 4, pp. 269–271, 2010.
- [169] Z. Y. Tung and D. S. Ang, "Transient to temporarily permanent and permanent hole trapping transformation in the small area SiON P-MOSFET subjected to negative-bias temperature stress," *Physical and Failure Analysis of Integrated Circuits (IPFA), 2014 IEEE 21st International Symposium on the*, pp. 254–257, 2014.
- [170] T. Grassler, M. Waltl, W. Goes, Y. Wimmer, A.-M. El-Sayed, A. L. Shluger, and B. Kaczer, "On the volatility of oxide defects: Activation, deactivation, and transformation," *Reliability Physics Symposium (IRPS), 2015 IEEE International*, pp. 5A.3.1-5A.3.8, 2015.
- [171] S. Sahhaf, R. Degraeve, R. O'Connor, B. Kaczer, M. B. Zahid, P. J. Roussel, L. Pantisano, and G. Groeseneken, "Evidence of a new degradation mechanism in high-k dielectrics at elevated temperatures," *2009 IEEE International Reliability Physics Symposium*, pp. 493–498, 2009.
- [172] D. J. DiMaria and J. H. Stathis, "Non-Arrhenius temperature dependence of reliability in ultrathin silicon dioxide films," *Appl. Phys. Lett.*, vol. 74, no. 12, p. 1752-1754, 1999.
- [173] B. Kaczer, R. Degraeve, N. Pagon, and G. Groeseneken, "The influence of elevated temperature on degradation and lifetime prediction of thin silicon-dioxide films," *IEEE Transactions on Electron Devices*, vol. 47, no. 7, pp. 1514–1521, 2000.
- [174] T. Nigam, R. Degraeve, G. Groeseneken, M. M. Heyns, and H. E. Maes, "A fast and simple methodology for lifetime prediction of ultra-thin oxides," in *1999 IEEE International Reliability Physics Symposium Proceedings. 37th Annual* (Cat. No.99CH36296), 1999, pp. 381–388.
- [175] D. Goguenheim, A. Bravaix, B. Ananou, C. Trape, F. Mondon, and G. Reimbold, "Temperature and field dependence of stress induced leakage currents in very thin (<5 nm) gate oxides," *J. Non. Cryst. Solids*, vol. 280, no. 1, pp. 78–85, 2001.
- [176] A. Ghetti, J. Bude, and G. Weber, " T_{BD} prediction from low-voltage near-interface trap-assisted tunneling current measurements," *IEEE Trans. Electron Devices*, vol. 48, no. 7, pp. 1354–1359, 2001.
- [177] C. Petit and D. Zander, "Low voltage stress induced leakage current and time to breakdown in ultra-thin (1.2–2.3nm) oxides," *Microelectron. Reliab.*, vol. 47, no. 2, pp. 401–408, 2007.
- [178] S. Chakravarthi, A. Krishnan, V. Reddy, C. F. Machala, and S. Krishnan, "A comprehensive framework for predictive modeling of negative bias temperature

- instability,” *Reliability Physics Symposium Proceedings, 2004. 42nd Annual. 2004 IEEE International*. pp. 273–282, 2004.
- [179] P. M. Lenahan, J. P. Campbell, A. T. Krishnan, and S. Krishnan, “A model for NBTI in nitrided oxide MOSFETs which does not involve hydrogen or diffusion,” *IEEE Transactions on Device and Materials Reliability*, vol. 11, no. 2. pp. 219–226, 2011.
 - [180] T. Grassler, H. Reisinger, P.-J. Wagner, and B. Kaczer, “Time-dependent defect spectroscopy for characterization of border traps in metal-oxide-semiconductor transistors,” *Phys. Rev. B*, vol. 82, no. 24, pp. 245318-1-245318-10, Dec. 2010.
 - [181] B. Kaczer, T. Grassler, J. Franco, M. Toledano-Luque, P. J. Roussel, M. Cho, E. Simoen, and G. Groeseneken, “Recent trends in bias temperature instability,” in *J. Vac. Sci. Technol. B*, 2011, vol. 29, no. 1, pp. 01AB01-1-01AB01-7, 2011.
 - [182] T. Grassler, K. Rott, H. Reisinger, M. Waltl, and W. Goes, “Evidence for defect pairs in SiON pMOSFETs,” *Physical and Failure Analysis of Integrated Circuits (IPFA), 2014 IEEE 21st International Symposium on the*. pp. 258–263, 2014.
 - [183] J. K. Rudra and W. B. Fowler, “Oxygen vacancy and the E'_{1} center in crystalline SiO_2 ,” *Phys. Rev. B*, vol. 35, no. 15, pp. 8223–8230, May 1987.
 - [184] A. Kimmel, P. Sushko, A. Shluger, and G. Bersuker, “Positive and negative oxygen vacancies in amorphous silica,” *ECS Trans.*, vol. 19, no. 2, pp. 3–17, May 2009.
 - [185] Y. P. Feng, A. T. L. Lim, and M. F. Li, “Negative-U property of oxygen vacancy in cubic HfO_2 ,” *Appl. Phys. Lett.*, vol. 87, no. 6, pp. 062105-1-062105-3, Aug. 2005.
 - [186] D. J. Chadi, “Negative-U property of the oxygen vacancy defect in SiO_2 and its implication for the E'_1 center in α -quartz,” *Appl. Phys. Lett.*, vol. 83, no. 3, pp. 437 – 439, Jul. 2003.
 - [187] H. D. Xiong, D. Heh, S. Yang, X. Zhu, M. Gurfinkel, G. Bersuker, D. E. Ioannou, C. A. Richter, K. P. Cheung, and J. S. Suehle, “Stress-induced defect generation in $\text{HfO}_2/\text{SiO}_2$ stacks observed by using charge pumping and low frequency noise measurements,” *Reliability Physics Symposium, 2008. IRPS 2008. IEEE International*. pp. 319–323, 2008.
 - [188] Y. Shi, T.-P. Ma, S. Prasad, and S. Dhanda, “Polarity dependent gate tunneling currents in dual-gate CMOSFETs,” *Electron Devices, IEEE Transactions on*, vol. 45, no. 11. pp. 2355–2360, 1998.
 - [189] L. Z. and C. L. and R. W. and R. H. and T. Y. and J. Z. and P. K. and H.-H. T. and Y. Wang, “Characteristics of gate current random telegraph signal noise in $\text{SiON}/\text{HfO}_2/\text{TaN}$ p -type metal–oxide–semiconductor field-effect transistors under negative bias temperature instability stress condition,” *Jpn. J. Appl. Phys.*, vol. 49, no. 4S, p. 04DC08-1-04DC08-4, 2010.
 - [190] O. Baumgartner, M. Bina, W. Goes, F. Schanovsky, M. Toledano-Luque, B. Kaczer, H. Kosina, and T. Grassler, “Direct tunneling and gate current fluctuations,” *Simulation of Semiconductor Processes and Devices (SISPAD), 2013 International Conference on*. pp. 17–20, 2013.
 - [191] X. Ji, Y. Liao, C. Zhu, J. Chang, F. Yan, Y. Shi, and Q. Guo, “The physical mechanisms of I_G random telegraph noise in deeply scaled pMOSFETs,” *Reliability Physics Symposium (IRPS), 2013 IEEE International*. pp. XT.7.1-XT.7.5, 2013.
 - [192] W. Goes, M. Toledano-Luque, O. Baumgartner, M. Bina, F. Schanovsky, B. Kaczer, and T. Grassler, “Understanding correlated drain and gate current fluctuations,”

Physical and Failure Analysis of Integrated Circuits (IPFA), 2013 20th IEEE International Symposium on the. pp. 51–56, 2013.

- [193] M. Toledano-Luque, B. Kaczer, E. Simoen, R. Degraeve, J. Franco, P. J. Roussel, T. Grasser, and G. Groeseneken, “Correlation of single trapping and detrapping effects in drain and gate currents of nanoscaled nFETs and pFETs,” *Reliability Physics Symposium (IRPS), 2012 IEEE International.* pp. XT.5.1-XT.5.6, 2012.
- [194] W. Liu, A. Padovani, L. Larcher, N. Raghavan, and K. L. Pey, “Analysis of correlated gate and drain random telegraph noise in post-soft breakdown TiN/HfLaO/SiO_x nMOSFETs,” *Electron Device Letters, IEEE*, vol. 35, no. 2. pp. 157–159, 2014.
- [195] B. Kaczer, M. Toledano-Luque, W. Goes, T. Grasser, and G. Groeseneken, “Gate current random telegraph noise and single defect conduction,” *Microelectron. Eng.*, vol. 109, pp. 123–125, Sep. 2013.
- [196] C.-Y. Chen, Q. Ran, H.-J. Cho, A. Kerber, Y. Liu, M.-R. Lin, and R. W. Dutton, “Correlation of I_d - and I_g -random telegraph noise to positive bias temperature instability in scaled high- k /metal gate n -type MOSFETs,” *Reliability Physics Symposium (IRPS), 2011 IEEE International.* pp. 3A.2.1-3A.2.6, 2011.
- [197] E. Bury, R. Degraeve, M. J. Cho, B. Kaczer, W. Goes, T. Grasser, N. Horiguchi, and G. Groeseneken, “Study of (correlated) trap sites in SILC, BTI and RTN in SiON and HKMG devices,” *Physical and Failure Analysis of Integrated Circuits (IPFA), 2014 IEEE 21st International Symposium on the.* pp. 250–253, 2014.
- [198] T. Usuki, K. Tsunoda, A. Sato, T. Nakanishi, and H. Tanaka, “Direct tunneling memory,” *FUJITSU Sci. Tech. J*, vol. 39, no. 1, pp. 72–77, 2003.
- [199] G. Kresse and J. Hafner, “Ab initio molecular dynamics for liquid metals,” *Phys. Rev. B*, vol. 47, no. 1, pp. 558–561, Jan. 1993.
- [200] A. Alkauskas, P. Broqvist, and A. Pasquarello, “Defect energy levels in density functional calculations: alignment and band gap problem,” *Phys. Rev. Lett.*, vol. 101, no. 4, pp. 046405-1-046405-4, Jul. 2008.
- [201] E.-C. Lee and H. Ju, “Mutual deactivation of electrically active F interstitials and O vacancies into fluorine-oxygen-vacancy complexes in SiO₂,” *Phys. Rev. B*, vol. 79, no. 19, pp. 193203-1-193203-4, May 2009.
- [202] J. J. Sung and C.-Y. Lu, “A comprehensive study on p⁺ polysilicon-gate MOSFET’s instability with fluorine incorporation,” *Electron Devices, IEEE Transactions on*, vol. 37, no. 11. pp. 2312–2321, 1990.
- [203] D. M. Fleetwood, H. D. Xiong, Z.-. Lu, C. J. Nicklaw, J. A. Felix, R. D. Schrimpf, and S. T. Pantelides, “Unified model of hole trapping, 1/f noise, and thermally stimulated current in MOS devices,” *IEEE Trans. Nucl. Sci.*, vol. 49, no. 6, pp. 2674–2683, 2002.
- [204] D. M. Fleetwood, “1/f noise and defects in microelectronic materials and devices,” *IEEE Trans. Nucl. Sci.*, vol. 62, no. 4, pp. 1462–1486, 2015.
- [205] M. Toledano-Luque, B. Kaczer, J. Franco, P. J. Roussel, T. Grasser, T.-Y. Hoffmann, and G. Groeseneken, “From mean values to distributions of BTI lifetime of deeply scaled FETs through atomistic understanding of the degradation,” *VLSI Technology (VLSIT), 2011 Symposium on.* pp. 152–153, 2011.

- [206] M. Duan, J. F. Zhang, Z. Ji, W. Zhang, B. Kaczer, S. De Gendt, and G. Groeseneken, "Defect Loss: A new concept for reliability of MOSFETs," *Electron Device Letters, IEEE*, vol. 33, no. 4, pp. 480–482, 2012.
- [207] M. F. Bukhori, T. Grasser, B. Kaczer, H. Reisinger, and A. Asenov, "'Atomistic' simulation of RTS amplitudes due to single and multiple charged defect states and their interactions," *Integrated Reliability Workshop Final Report (IRW), 2010 IEEE International*. pp. 76–79, 2010.
- [208] L. Gerrer, R. Hussin, S. M. Amoroso, J. Franco, P. Weckx, M. Simicic, N. Horiguchi, B. Kaczer, T. Grasser, and A. Asenov, "Experimental evidences and simulations of trap generation along a percolation path," in *2015 45th European Solid State Device Research Conference (ESSDERC)*, pp. 226–229, 2015.
- [209] T. Grasser, K. Rott, H. Reisinger, M. Waltl, F. Schanovsky, and B. Kaczer, "NBTI in nanoscale MOSFETs — the ultimate modeling benchmark," vol. 61, no. 11, pp. 3586–3593, 2014.
- [210] S. Realov and K. L. Shepard, "Analysis of random telegraph noise in 45-nm CMOS using on-chip characterization system," *IEEE Transactions on Electron Devices*, vol. 60, no. 5, pp. 1716–1722, 2013.
- [211] K. Sonoda, K. Ishikawa, T. Eimori, and O. Tsuchiya, "Discrete dopant effects on statistical variation of random telegraph signal magnitude," *IEEE Transactions on Electron Devices*, vol. 54, no. 8, pp. 1918–1925, 2007.
- [212] Z. Zhang, S. Guo, X. Jiang, R. Wang, R. Huang, and J. Zou, "Investigation on the amplitude distribution of random telegraph noise (RTN) in nanoscale MOS devices," *2016 IEEE International Nanoelectronics Conference (INEC)*. pp. 1–2, 2016; doi: 10.1109/INEC.2016.7589332.
- [213] N. Tega, H. Miki, Z. Ren, C. P. D'Emic, Y. Zhu, D. J. Frank, M. A. Guillorn, D. G. Park, W. Haensch, and K. Torii, "Impact of HK / MG stacks and future device scaling on RTN," *Reliability Physics Symposium (IRPS), 2011 IEEE International*. p. 6A.5.1-6A.5.6, 2011.
- [214] N. Tega, H. Miki, Z. Ren, C. P. D'Emic, Y. Zhu, D. J. Frank, J. Cai, M. A. Guillorn, D. G. Park, W. Haensch, and K. Torii, "Reduction of random telegraph noise in high- κ /metal-gate stacks for 22 nm generation FETs," *2009 IEEE International Electron Devices Meeting (IEDM)*. pp. 1–4, 2009; doi: 10.1109/IEDM.2009.5424225.
- [215] B. Kaczer, S. M. Amoroso, R. Hussin, A. Asenov, J. Franco, P. Weckx, P. J. Roussel, G. Rzepa, T. Grasser, and N. Horiguchi, "On the distribution of the FET threshold voltage shifts due to individual charged gate oxide defects," in *2016 IEEE International Integrated Reliability Workshop (IIRW)*, pp. 18–20, 2016.
- [216] S. M. Amoroso, C. M. Compagnoni, A. Ghetti, L. Gerrer, A. S. Spinelli, A. L. Lacaita, and A. Asenov, "Investigation of the RTN distribution of nanoscale MOS devices from subthreshold to on-state," *IEEE Electron Device Letters*, vol. 34, no. 5, pp. 683 – 685, 2013.
- [217] T. Nagumo, K. Takeuchi, S. Yokogawa, K. Imai, and Y. Hayashi, "New analysis methods for comprehensive understanding of random telegraph noise," *2009 IEEE International Electron Devices Meeting (IEDM)*. pp. 1–4, 2009; doi: 10.1109/IEDM.2009.5424230.

- [218] K. Fukuda, Y. Shimizu, K. Amemiya, M. Kamoshida, and C. Hu, "Random telegraph noise in flash memories - model and technology scaling," *2007 IEEE International Electron Devices Meeting*. pp. 169–172, 2007.
- [219] A. Ghetti, C. M. Compagnoni, F. Biancardi, A. L. Lacaita, S. Beltrami, L. Chiavarone, A. S. Spinelli, and A. Visconti, "Scaling trends for random telegraph noise in deca-nanometer flash memories," *2008 IEEE International Electron Devices Meeting*. pp. 1–4, 2008; doi: 10.1109/IEDM.2008.4796827.
- [220] K. Takeuchi, T. Nagumo, S. Yokogawa, K. Imai, and Y. Hayashi, "Single-charge-based modeling of transistor characteristics fluctuations based on statistical measurement of RTN amplitude," *2009 Symposium on VLSI Technology*. pp. 54–55, 2009.
- [221] A. Ghetti, C. M. Compagnoni, A. S. Spinelli, and A. Visconti, "Comprehensive analysis of random telegraph noise instability and its scaling in deca-nanometer flash memories," *IEEE Transactions on Electron Devices*, vol. 56, no. 8. pp. 1746–1752, 2009.
- [222] L. Gerrer, J. Ding, S. M. Amoroso, F. Adamu-Lema, R. Hussin, D. Reid, C. Millar, and A. Asenov, "Modeling RTN and BTI in nanoscale MOSFETs from device to circuit: a review," *Microelectron. Reliab.*, vol. 54, no. 4, pp. 682–697, Apr. 2014.
- [223] T. Grassler, W. Gos, V. Sverdlov, and B. Kaczer, "The Universality of NBTI relaxation and its implications for modeling and characterization," *Reliability Physics Symposium, 2007. Proceedings. 45th Annual. IEEE international*. pp. 268–280, 2007.
- [224] T. Grassler, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P.-J. 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," *Electron Devices, IEEE Transactions on*, vol. 58, no. 11. pp. 3652–3666, 2011.