THRESHOLD VOLTAGE INSTABILITIES IN MOS TRANSISTORS WITH ADVANCED GATE DIELECTRICS

SHEN CHEN

(B.Eng. (Hons.), NUS)

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DEPARTMENT OF ELECTRICAL
AND COMPUTER ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE
FEBRUARY 2008

To Sarah

Acknowledgement

After working on other projects for almost a year, I came back to my files on reliability study, to reconstruct my memory on the three years dedicated to it, and to write this thesis. Each graph now tells a story on the guidance, inspiration and support I received from many colleagues. Without their contribution, this study would not reach the depth I had hoped.

First of all, I would like to thank my supervisor, prof. Li Ming-Fu for bringing me to the field of transistor reliability study, and the liberal environment he created in the group. It is difficult to imagine how little I could have done if prof. Li had not encouraged me to attempt on those ideas seemingly beyond reach. He demonstrated to us how a researcher should be confident in his study; how the confidence comes from the pursue of every detail and continual cross-checking; and how one should be open-minded and actively seek criticism. Working with him has been very much a character building process to me. I would like to thank Dr. Yeo Yee-Chia for the many inspiring discussions on a wide range of topics. It has been very beneficial to put things into perspective and see the big picture. It is such a pleasure working with you. Prof. Ganesh Samudra and prof. Kwong Dim-Lee provided many thoughtful suggestions to many of my manuscripts, to which I am very grateful.

I would like to thank Dr. Yu Hong-Yu, who was my mentor when I first joined the group. Many of research plans in the initial year was under his steering. He provided many invaluable advices on tackling the challenges in research. Similarly I owe debt to Dr. Hou Yong-Tian and Dr. Zhu Shi-Yang for their advices. Many of the data in this work are results of collaborations with Ms. Yang Tian and Mr. Wang Xin-Peng. Their incredible dedication to the project really set a standard for all members in the group, and continually spurred me to work harder. I was great pleasure to have your collaboration and friendship. A few undergradu-

ate students participated in the project, including Mr. Yong Chen-Hong, Mr. Foo Che-E and Mr. Lai Cheng-Yi, and their contributions are gratefully appreciated.

The experimental work was carried out in the silicon nano device lab, and the center for IC failure analysis and reliability, both at the National University of Singapore. I received a lot of technical and logistic support from the managers and technicians of both labs. I would like to thank Prof. Byung-Jin Cho for his tremendous contribution in establishing SNDL in both its facilites and traditions. Mrs. Ho Chiow-Mooi, Mr. Yong Yu-Fu, Mr. Patrick Tang, Mr. O Yan Wai Linn and Mr. Abdul Jalil bin Din are gratefully acknowledged for their support.

I have also been working on a few other projects under the collaboration with many colleagues in SNDL, and there are even more general technical discussions on a large variety of topics. This culture of open discussion has been very memorable experience, and I believe, is to some extent a unique character of SNDL. It is impossible to enumerate all, but I cannot fail to mention Jing De, Ren Chi, Wu Nan, Xiong Fei, Qing Chun, Gao Fei, Jing Hao, Wan-Sik, Ying Qian, Pu Jing and He Wei for the numerous discussions over lunch, or while idling in the cleanroom.

There are a few people outside NUS contributed to this work, to whom I owe a big thank you. Prof. A. Alam's pioneering work in the modeling of dynamic NBTI to some extent framed many parts of this thesis. Discussions with him revealed to me many insights in the reaction-diffusion model, and my gratitude to him transcends by many orders, our different views on NBTI. The very inspiring discussions with Mr. Zhao Yue-Gang from Keithley Instruments and H. Reisinger from Infineon provided many of the ingredients in the fast I-V measurement technique presented in this work. I must take this chance to thank them for sharing their insights without reservation.

A special thank goes to my wife, Sarah, whose tremendous understanding and support allowed me to pursue my dream.

Shen Chen
Singapore, Jan 2008

Contents

Abstract . . . viii

List of Figures . . . x

List of Tables . . . xv

List of Symbols . . . xvi

List of Abbreviations . . . xviii

1 Introduction 1

1.1 Imperfections in Gate Dielectrics and Reliability . . . 2

1.2 New Materials in Advanced Gate Dielectrics . . . 5

1.3 Threshold Voltage Instability . . . 6

2 Fast $I_d - V_g$ Characterization for Transistors 11

2.1 Development of the Fast Techniques . . . 12

2.2 Fast Measurement Setup . . . 15

2.3 Source of Errors . . . 20

2.4 Applications To Charge Trapping in High-κ gate dielectrics . . . 23

2.5 Conclusions . . . 24

3 Negative-Bias-Temperature Instability in SiON Gate Dielectrics 27

3.1 A Brief Historical Introduction . . . 27

3.1.1 Dynamic Recovery . . . 30

3.1.2 Role of Hole Trapping . . . 32

3.2 Theories for the Dynamic NBTI ... 34

3.2.1 Reaction-Diffusion Model ... 34

3.2.1.1 Stress ... 34

3.2.1.2 Recovery ... 37

3.2.1.3 Dynamic Stress ... 41

3.2.1.4 Other Diffusion Species ... 44

3.2.1.5 Numerical Solution ... 44

3.2.2 Charge Trapping/De-trapping Model ... 45

3.3 Review on Measurment Techniques ... 51

3.3.1 Fast I_d-V_g Measurement ... 53

3.3.2 Slow On-the-Fly Measurement ... 54

3.3.3 Fast On-the-Fly Measurement ... 60

3.3.4 Single-Pulse Measurement ... 61

3.3.5 Discussion and Summary ... 62

3.4 Hole Trapping and the Fast Transient Component in NBTI ... 63

3.4.1 Fast Recovery and Dependence on Stress Time ... 64

3.4.2 Effect of Measurement Delay ... 66

3.4.3 Frequency Dependence Under Dynamic Stress ... 69

3.4.4 Discussions and Summary ... 72

3.5 Interface States and Slow Component in NBTI ... 74

3.5.1 Existence of Interface Trap Recovery ... 75

3.5.2 Attribution of the Slow Component of NBTI 76

3.6 Conclusions 77

4 Charge Trapping in High-κ Gate Dielectrics **83**

4.1 Introduction 83

4.2 Slow Component of Charge Trapping in HfO_2 85

4.2.1 Dynamic Charge Trapping in HfO_2 and its Frequency Dependence 85

4.2.2 Physical Model of the Frequency Dependent Charge Trapping 88

4.3 Fast Component in Charge Trapping in HfO_2 94

4.3.1 Sample Preparation and Measurement technique 94

4.3.2 Characterization of fast charge traps 98

4.3.2.1 Voltage dependence 98

4.3.2.2 Frequency dependence 100

4.3.2.3 Stress history 102

4.3.2.4 Duty-cycle dependence 102

4.3.2.5 Temperature dependence 106

4.3.3 Modeling of the fast V_{th} instability 106

4.4 Impacts on digital circuits 110

4.4.1 SRAM 111

4.4.2 Logic circuits 112

4.5 Conclusions 116

5 Summary **122**

List of Publications **126**

Curriculum vitae **130**

Abstract

The scaling of MOSFET is not only a geometric shrinkage, but also accompanied by new materials and process technologies. The gate dielectric, as the most critical component in a MOSFET transistor, is undergoing rapid and substantial changes with the adoption of ultra-thin plasma-nitrided oxide, and more recently high-κ dielectrics. The reliability physics of these new gate dielectric materials are important and urgent tasks to the IC industry. One important aspect of the transistor reliability is the threshold voltage instability, which causes degradation of circuit performance, and in some cases, loss of functionality as well.

This thesis examines the dominant V_{th} instability mechanisms in two advanced gate dielectric materials, namely the nitrided silicon oxide (or silicon oxynitride), and the hafnium oxide. Negative bias temperature instability and charge trapping phenomena in these two dielectric films are the focus of this study, and form the main chunk of this thesis.

Since the accurate characterization of threshold voltage instabilities is a prerequisite of the desired study, much effort was spent on developing the fast I_d−V_g measurement technique. The minimum measurement time for an I_d−V_g curve of 100 ns is achieved. The operation principle, circuit construction and sources of errors of this technique are documented in detail. The fast measurement is shown to be indispensable for accurate characterization of threshold instabilities in the advanced gate dielectrics, due to the fast recovery of threshold voltage when stress is removed.

With the accurate measurement technique established, the V_{th} degradation mechanisms are studied in detail. In the case of oxynitride dielectric, the relative importance of interface-state generation and charge trapping is currently under debate in the community. Analytical and numerical calculations are performed on each of the two theories, and compared to the extensive experimental data. It is argued that for the oxynitride dielectric, hole trapping must be present along with

the interface trap generation. More specifically, charge trapping is the dominant mechanism giving rise to the fast transients in NBTI.

In the case of HfO_2 dielectric, it is observed that two distinct charge trapping components exists, with the slower component shows an unexpected dependence on the frequency of the stress signal. A two-step charge trapping model, possibly associated with the negative-U traps in HfO_2 film, is proposed to explain the observed frequency dependence. The faster charge trapping component, which has large magnitude, is modeled with traditional charge trapping theories. The obtained dynamic model of the fast charge trapping is used to predict its impact on digital circuits. It is shown that different circuit topologies have very different sensitivity to the instability of V_{th}.

List of Figures

2.1 Schematic illustrating the fast $I_d - V_g$ measurement setup developed by A. Kerber. 15

2.2 Schematic illustrating the fast $I_d - V_g$ measurement setup utilizing a transimpedance amplifier. 16

2.3 Photographs on the fast measurement setup. 17

2.4 Schematic illustrating the fast $I_d - V_g$ measurement setup utilizing a transimpedance amplifier, with matched impedance and cable delay. 18

2.5 Voltage waveform recorded from the oscilloscope. 19

2.6 $I_d - V_g$ characteristics measured from a short-channel n-MOSFET with SiON gate dielectric, with the V_g waveform shown in the inset. 20

2.7 $I_d - V_g$ characteristics measured from a long-channel n-MOSFET with HfO_2 gate dielectric, with the V_g waveform as in Figure 2.6. 23

2.8 $I_d - V_g$ characteristics after correction for the effect of C_{gd}. 24

2.9 . . . Slow measurements leads to serious underestimation of the threshold voltage shift caused by charge trapping. 25

3.1 Approximate hydrogen concentration profile in the diffusion process. 37

3.2 The solution to the diffusion equation during recovery is approximated by the convolution of ϕ and g, as in (3.15). 39

3.3 Approximate hydrogen concentration profile in diffusion-limited recovery of NBTI. 40

3.4 Approximate hydrogen concentration profile in diffusion-limited dynamic stress of NBTI. 42

3.5 A schematic illustrating the trapping and de-trapping of holes in oxide traps. Several hole traps are assume to distribute in different energy levels. The double-sided arrow indicates the exchange of holes between the silicon substrate and the trap states, through trapping and de-trapping processes. 46

3.6 Distribution function of τ_{C}, τ_{E1} and τ_{E2} used in this study. 50

3.7 The waveform of the gate and drain voltage in a simplest NBTI measure–stress–measure cycle. 52

3.8 ΔV_{th} for identical pMOSFETs is measured with five different measurement techniques, which yield very different NBTI results. 53

3.9 The waveform of the gate voltage using the fast measurement technique. 55

3.10 The V_{th} shift measured after 1s stress, with different measurement time t_{m}. 55

3.11 The waveform of the gate voltage using the *on-the-fly* measurement technique. 56

3.12 Calculation of ΔV_{th} using (3.35) leads to error ε_{r} when ΔV_{th} is large, due to the V_g dependent transconductance. 58

3.13 (a) Conventional (slow) on-the-fly suffers from degradation during the measurement of the inital I_{d0}, and estimates ΔV_{th}. 59

3.14 The waveform of the gate voltage using the single-pulse measurement technique. 62

3.15 Using fast $I_d - V_g$ technique, the dynamics of ΔV_{th} under stress/recovery cycles is studied ($f = 1/2000$Hz is shown). 65

3.16 Hypothetical hydrogen profile necessary to explain the fast V_{th} recoverying after long time (1000s) stress within the reaction-diffusion framework. 66

3.17 After stress is removed, the majority of ΔV_{th} recovers in a very short time. 67

3.18 According to the reaction-diffusion model . . . , recovery occurs mainly between $0.1t_{\text{stress}}$ and $10t_{\text{stress}}$ after stress is removed. 68

3.19 Simulated recovery process according to the hole trapping/de-trapping model. 69

3.20 (a) From the R-D model, ΔV_{th} with measurement delay is simulated. (b) The error due to delay diminishes when stress time is much greater than the delay time. 70

3.21 As measurement time t_{m} increases, V_{th} shift is underestimated, and the power-law exponent increased. 71

3.22 From the trapping/de-trapping model, the trapped charge (and thus ΔV_{th}) is simulated with delay. 71

3.23 From R-D model, ΔV_{th} under dynamic stress is simulated. 72

3.24 Experimental ΔV_{th} data under dynamic stress. 73

3.25 The trapping/de-trapping model predicts that, under dynamic stress, ΔV_{th} at both S_{f} and P points are frequency dependent, and the difference between the two is large. 73

4.1 Time evolution of threshold voltage V_{th} under static and dynamic stresses of different frequencies, for (a) n-MOSFETs, and (b) p-MOSFETs. 87

4.2 V_{th} shift in alternating stress and recovery cycles of period $T = 2000$s. 88

4.3 Three possible cases for the relationship between the number of trapped electrons Δn versus stress time Δt in one cycle of the dynamic stress. 89

4.4 Two-step procedure of capturing two electrons by a negative-U trap. 90

4.5 Calculated time evolutions of V_{th} using equations (4.1) at various frequencies are plotted using lines, showing good agreement with the experimental data (symbols). 93

4.6 Waveform of stress voltage used in a) static stress, and b) dynamic stress. 95

4.7 Threshold Voltage shift under stress/recovery cycles with frequency = 1/2000Hz. 97

4.8 Recovery of linear region drain current I_d with respect to the pre-stress I_{d0}, after stress voltage is removed from the gate of the nMOSFET. 98

4.9 A schematic illustrating the trapping and de-trapping of electrons in HfO_2. Several electron traps are assume to distribute in different energy levels. The double-sided arrow indicates the exchange of electrons between the silicon substrate and the trap states, through trapping and de-trapping processes. 99

4.10 V_{th} shift dependence on static stress voltage and gate overdrive at the end of a one second stress ($V_g - V_{\mathrm{th,1s}}$), measured by fast technique. 101

4.11 Threshold voltage shift under static stress and dynamic stress of different frequencies, measured by fast technique. 103

4.12 Frequency dependence of ΔV_{th} after 100 seconds of dynamic stress, using fast measurement. 104

4.13 Evolution of ΔV_{th} during transition from static stress to dynamic stress (filled symbols), and ΔV_{th} of fresh device under dynamic stress (open symbols). 105

4.14 ΔV_{th} after 100 second dynamic stress of different duty cycle, but same stress voltage, frequency and rise/fall time. 107

4.15 Stress voltage dependence of ΔV_{th} under static and dynamic stress of different duty cycle. 108

4.16 Stress voltage dependence of ΔV_{th}, stressed under different temperatures. 109

4.17 Spectrum of trapping (τ_{C}) and de-trapping time constants (τ_{E1}, τ_{E2}) used in calculation of trapping/de-trapping dynamics for (a) electrons and (b) holes. 110

4.18 Simulation of recovery of linear region drain current I_d after stress voltage is removed from the gate of the nMOSFET, as shown in Figure 4.8. 111

4.19 Schematics of 6T SRAM cell. 112

4.20 Butterfly plot of the SRAM cell showing the transfer characteristics between voltage at the left storage node (V_{L}) and that at the right storage node (V_R). 113

4.21 Static noise margin of SRAM cell with cell ratio $\beta = 2$. 114

4.22 Schematics of a) NAND3 gate implemented with static logic, b) 2-input multiplexer implemented with CMOS transmission gate, and c) NAND3 gate implemented with dynamic logic. 115

4.23 Percentage increase in gate propagation delay due to V_{th} degradation as function of supply voltage. 116

List of Tables

4.1 Time constants used in the model. 92

List of Symbols

τ_{C}	capture time constant
C_{gd}	gate-to-drain capacitance
C_{inv}	capacitance density in inversion
$C_{\mathrm{overlap,d}}$	gate-drain overlap capacitance
D	diffusivity
D_{it}	density of interface states ($\mathrm{cm}^{-2}\mathrm{eV}^{-1}$)
τ_{E}	emission time constant
g_m	transconductance
I_d	drain current
I_{d0}	initial drain current
$I_{d,\mathrm{lin}}$	linear-region drain current
$I_{d,\mathrm{sat}}$	saturation-region drain current
I_{gd}	gate-to-drain current
I_{off}	off-state leakage current
k_{F}	forward reaction coefficient
k_{R}	reverse reaction coefficient
L_{g}	gate length
N_0	total density of defects
N_{H}	concentration of hydrogen
N_{it}	density of interface traps (cm^{-2})
N_{ot}	density of oxide traps
$N_{\mathrm{ot,p}}$	density of oxide hole traps
Q_{f}	density of fixed charge
t_{d}	delay time
t_{m}	measurement time
t_{rise}	rise time
t_{stress}	stress time

V_d	drain voltage
V_{dd}	positive supply voltage
V_{ds}	drain-to-source voltage
V_{fb}	flat-band voltage
V_g	gate voltage
V_{gd}	gate-to-drain voltage
V_{gs}	gate-to-source voltage
V_{meas}	measurement voltage
V_{out}	output voltage
V_{stress}	stress voltage
V_{th}	threshold voltage
V_{th0}	initial threshold voltage

List of Abbreviations

$I_d{-}V_d$	drain current — drain voltage characteristic
$I_d{-}V_g$	drain current — gate voltage characteristic
AC	alternating current
$C{-}V$	capacitance — voltage characteristic
CMOS	complementary metal-oxide-semiconductor device
CMOSFET	complementary metal-oxide-semiconductor field-effect transistor
DC	direct current
DCIV	DC current-voltage method of interface trap measurement
DUT	device under test
DVSS	dynamic voltage scaled system
EOT	equivalent oxide thickness
F-N	Fowler-Nordheim
HCI	hot carrier injection
$I{-}V$	current — voltage characteristic
IC	integrated circuits
ITRS	International Technology Roadmap for Semiconductors
MOCVD	metal-organic chemical vapor deposition
MOS	metal-oxide-semiconductor device, usually the MOS capacitor
MOSFET	metal-oxide-semiconductor field-effect transistor
MUX	multiplexer
n-MOS	n-type MOS device
n-MOSFET	n-type channel MOSFET
NAND	not-and logic gate
NBTI	negative-bias-temperature instability
OPAMP	operational amplifier
p-MOS	n-type MOS device
p-MOSFET	p-type channel MOSFET

PCB	printed circuit board
PVD	physical-vapor deposition
R-D	reaction-diffusion
RDF	random dopant fluctuation
RF	radio frequency
SNM	static noise margin
SOI	silicon on insulator
SRAM	static random-access memory
SS	sub-threshold swing

CHAPTER 1

Introduction

As the author is typing this thesis concerning transistor technology, the 60th anniversary of the invention of transistor (23rd, December 1947) is approaching. It is interesting to note that, though the first practical transistor, demonstrated by John Bardeen and Walter Brattain at Bell Labs, was a point-contact transistor, it was an accidental discovery in the search for a field-effect transistor[1]. Actually, the field-effect transistor was conceived as early as 1928, by Julius Edgar Lilienfeld. William Shockley and others attempted extensively on it, but never succeeded. Bardeen was first to point out that the large amount of surface traps of semiconductor would screen out the desired field-effect almost completely, and any attempt on FET would certainly be futile unless the surface states were tamed. He and Brattain set to fix the surface state problem by, for example, using electrolyte solution as gate electrode, or using germanium oxide as the insulator. Unfortunately, the germanium oxide film on their sample has a hole, and does not insulate. Fiddling this defective sample, they discovered transistor effect in a point contact transistor structure, and not a MOSFET. It was only in the 1960s that the surface states problem saw the first practical solution with thermal oxidation of silicon, and the development of silicon MOSFET gained momentum. However, the quality of the insulator thin-film as well as its interface with the semiconductor is still the key to a successful MOSFET technology.

The current dominance of MOSFET technology is largely associated with the development of integrated circuit (IC) technology. For long time, MOSFET integrated in IC was seen as a cheap alternative to bipolar transistor, which shows inferior performance, but occupy less area on the chip and is cheaper to fabricate. However, the scaling of MOSFET technology improved both the packing density and performance, making it suitable for digital computers and memories. On the other hand, the invention of CMOSFET technology reduced the power consumption of IC by orders of magnitude and made large scale integration practical. The scaling of the CMOSFET transistors since became the driving force to IC industry.

From 1970s to early 90s, the scaling of MOSFET largely followed the constant voltage scaling rule. As the hot-carrier induced reliability issue became difficult to handle, the supply voltage is reduced as MOSFET scales, and the constant field scaling is used. Since mid 90s, the semiconductor industry association (SIA) started to publish technology roadmaps for semiconductor industry, which includes an outlook of future scaling of MOSFET technology. It later became an international effort as the International Technology Roadmap for Semiconductors (ITRS)[2]. The scaling trend projection in ITRS is determined from transistor performance targets and power dissipation constraints, together with a sophisticated compact model of MOSFET transistors. In all scaling schemes, the gate oxide thickness scales down with the transistor feature size at a steady rate. The reduction of oxide thickness is motivated by many device design considerations, including the control of short channel effects, the adjustment of proper threshold voltage and the improvement in drive current.

1.1 Imperfections in Gate Dielectrics and Reliability

As the oxide thickness scales down and the electric field across the oxide increases, the quality of the oxide insulator becomes increasingly a concern. MOSFET operation requires the oxide film to be 1) insulating, 2) free of electric charge and 3) free of interface states. The silicon/silicon dioxide system adopted in modern IC industry was chosen primarily under these criteria. However, under high electric field, all these three properties may degrade.

First of all the gate dielectric film may breakdown under high field, and completely lose the insulating property. Dielectric breakdown is a long-standing subject in the study of IC reliability. It is easy to see that as the oxide thickness scales down the oxide sustains less voltage. In practice, MOSFET operates at voltages much lower than the dielectric breakdown voltage, but there is still a finite probability of breakdown. The lifetime before breakdown is random and follows the Weibull distribution. This time dependent dielectric breakdown lifetime is a major challenge to the oxide scaling.

Secondly, carriers injected into the oxide via hot-carrier injection (HCI) or Fowler-Nordheim (F-N) tunneling create defects in the oxide film which can then trap charged carriers. The hot carrier injection was the most critical reliability issue in the 80s and early 90s, and was extensively studied since then. The most discussed injection mode is due to the channel hot carriers as described below. When high voltage is present on both gate and drain terminal, a lot of carriers are flowing in the channel, and there is high longitudinal electric field near the drain region. Carriers are accelerated in this high-field region, and the carrier temperature increases. If the carriers gain sufficient energy, they can cross the energy barrier of the silicon/oxide interface and get injected into the insulator. Typically the maximum hot carrier generation occurs when gate voltage is around half of the drain voltage. This bias condition only occurs when the transistor is switching from off state to on state or *vice versa*. The attempt to minimize HCI led to the development of lightly doped drain (LDD) structure, the nitrided silicon oxide gate dielectric, and is one of the motivations for the scaling of supply voltage. As the supply voltage has scaled to around 1 V, which is less than the bandgap of silicon, hot carriers is much less a reliability concern to current technologies, though it is still regularly examined.

Lastly, many electrical stress tests generate interface states at the silicon/oxide interface. Although the silicon/silicon dioxide interface is considered among the best interfaces, there is still slight stress in the film, and dangling silicon bonds are present at the interface. These dangling bonds are usually passivated by hydrogen atoms, and are not electrically active. In modern MOSFET technology, the density of unpassivated silicon dangling bonds at the interface is negligibly low, usually below 10^{10} cm^{-2}. However, under electrical stress, the weak Si–H bonds may break, and interface states are created. In addition to HCI and F-N stress, the *negative bias temperature* stress is another important cause of interface states generation. The last (NBT) stress mode, is usually applied to p-MOSFET, where a large negative voltage is applied to the gate, and the temperature is raised above the room temperature. No drain voltage is applied. Interface states are generated under this stress mode, and the transistor degradation under this mode is called the negative bias temperature instability (NBTI). The NBTI degradation occurs even when the circuit is in quiescent, if the p-MOSFET happens to have its gate tied to high voltage. In recent years, NBTI degradation has been found as the most serious reliability concern of all, and attracted a lot of researches. Since the NBTI degradation occurs at the silicon/oxide film, and involves processes inside the insulator film, its physical origin is much less understood, compared to the HCI injection.

In addition to the three defects created by electrical stress, there are a few other imperfections in the oxide that are result of poor fabrication processes and are present before stress. They often appear similar to the stress-induced degradations mentioned earlier, and are usually discussed together. Three important examples of such pre-existing imperfections are the mobile ions, fixed oxide charge, and oxide traps. The mobile ions, such as Na^{+} and K^{+}, come from contaminations during the fabrication process, and were the major obstacles in the development of stable MOSFETs in the 60s. However, after the identification of its origin, it has been eliminated by the combination of cleanroom environment, deionized water and gettering processes. The fixed oxide charge in the oxide, residing close to the interface with silicon substrate, can be minimized by appropriate oxidation recipes, and has been well controlled.

Oxide traps can be pre-existing or generated by stress. We have discussed the HCI-generated oxide traps earlier, and we shall not missed the pre-existing ones. Oxide traps are usually attributed to broken Si–O bonds or oxygen vacancies. It is well known that Si–O bonds are surprisingly flexible, and do not easily break. However, under non-optimal process conditions, or when excessive nitrogen is added to the oxide film, broken bonds and vacancies can be abundant in the oxide film. These process related oxide traps are considered pre-existing to the device, as they are present before any electrical stress. In addition, we shall see that the new gate dielectric materials with higher dielectric constant values contain more pre-existing oxide traps than silicon dioxide does.

Strictly speaking, reliability is a concept associated with long-term effects, and the reliability of gate oxide should include the three stress-induced degradations. However, the pre-existing defects are customarily also included in the domain of reliability study.

1.2 New Materials in Advanced Gate Dielectrics

In addition to the reliability problems associated with the increasing electric field, new dielectric materials used in advanced gate stacks constitute another challenge.

As the thickness of gate dielectrics scales down, new dielectric materials are considered for a few reasons. First encountered was the boron penetration problem in p-MOSFETs, which requires the incorporation of nitrogen in the gate dielectric to suppress boron diffusion.

The other more fundamental problem is the direct tunneling of carriers between substrate and gate. The quantum mechanical tunneling of carriers increases exponentially as the insulator thickness decreases, and becomes a significant portion of the total leakage current as the dielectric thickness scales below about 3 to 4 nanometers. In order to suppress the excessive gate leakage current, while maintaining the scaling of oxide capacitance, it is necessary to increase the dielectric constant of the dielectric film. With higher dielectric constant (κ value), one can increase capacitance, thus reducing electrical thickness of the dielectric layer, without reducing the physical film thickness. One important metric to the advanced gate dielectric materials is the scaling trend of leakage current versus the effective oxide thickness (EOT), which is the thickness of SiO_2 film to achieve the same capacitance. Theoretical calculation of direct tunneling current shows that dielectric materials with higher permittivity offers significant reduction in leakage current at the same EOT. As a result, many dielectric materials with κ value greater than that of SiO_2 (3.9) have been investigated as potential replacement of SiO_2. We shall discuss the two materials with most technical importance studied in this thesis.

First is the nitrided silicon oxide or silicon oxynitride gate dielectric (SiON), which is used in current CMOS technologies. The dielectric constant of the film increases with increasing nitrogen content, up to about 8 for pure Si_3N_4, but the dielectric quality tends to degrade when nitrogen content were too high or non-optimal nitridation processes were used. Due to the sensitivity on process conditions, vast effort is required in the many iterations of process optimization and reliability tests.

Second is the hafnium oxide (HfO_2), which offers a much greater dielectric constant up to 25 and promises the potential of sub-1 nm EOT. However, the process and reliability issues are more serious in this high-κ film. Notably the HfO_2 film contains large number of pre-existing traps, which was one of the show-stoppers of high-κ dielectrics.

1.3 Threshold Voltage Instability

Except for the dielectric breakdown, all other degradations or imperfections described in the previous section result in charge build-up in the oxide or at the interface. This in turn causes the threshold voltage (V_{th}) to deviate from its initial value[3]. Threshold voltage is the most important device parameter of MOSFET, and its stability is a basic assumption in circuit design. If the threshold voltage drifts too much from the designed value, circuits may fail to function.

The exact effect of threshold voltage instability is very specific to the individual circuit design, and should be discussed in two levels, namely the loss of functionality and the degradation in performance. We shall confine our discussion within the domain of digital circuits. Combinational logic can be implemented with many styles of circuits, including the static logic, pass gate logic and dynamic logic, to name a few.

The static CMOS combinational logic is the most common one, and is the most robust circuit. It has large active gain and is level-restoring, and therefore offers very large static noise margin. It could produce the correct output even if the threshold voltage is just a few kT/q away from the supply voltage. At a single gate level, threshold voltage instability usually only results in drift in delay time and leakage current. However, as the threshold voltage deviate from the delicate optimal level[4], the delay time of the switching would increase, and/or the quiescent leakage current would increase. Therefore, the drift of threshold voltage to either higher or lower values is detrimental to the circuit performance in terms of the delay versus power trade-off. In larger circuit with multiple paths, drift in delay time due to V_{th} instability would lead to delay mismatch between different paths (skew). Since the overall delay is limited by the slowest path, the delay degradation in a small number of transistors may plague the entire circuit.

On the other hand, the functionality of sequential circuits and dynamic logic circuits rely heavily on accurate timing, and are more susceptible to delay degradation of transistors. Due to the large number of variants in circuit design, we shall not attempt to enumerate them, but leap to conclude that the threshold voltage instability is detrimental to the performance digital circuits in general.

It may be necessary to mention a few details here. First there are mechanisms not related to gate dielectric that cause the threshold voltage to change with time. The most well-known example of this kind is the floating body effects in partially-depleted SOI MOSFETs. Secondly the threshold voltage also vary spatially from transistor to transistor, due to random dopant fluctuation(RDF) and process variations. The spatial variation is detrimental to the circuit performance similar to the temporal variation discussed earlier. Lastly not all changes in threshold voltage degrades circuit performance. The dynamic threshold voltage MOS (DTMOS), for example, has lower V_{th} as it switches on and high V_{th} as it switches off. The controlled variation of V_{th} in this way help to achieve small delay with low leakage current.

In all cases, the knowledge of the threshold voltage instabilities, in both its magnitude and dynamics, is essential to successful circuit design. Unfortunately, traditional compact models of MOSFETs does not include any V_{th} instabilities, and circuit designers rely on corner-device tests to ensure the robustness of the circuit[5–8]. Circuit designers sets the maximum amount of device degradation allowed, and make sure the circuit works with the worst-case (degraded) devices. On the other hand, reliability assurance engineers qualify devices from a process technology within the bound of the specified maximum degradation. However, as the V_{th} instability is becoming a much important threat to circuit design, this division of task has become increasingly awkward, because little design margin is left. The worst-case design is inherently too conservative, but traditional reliability models were too much a simplification compared to the multitude of V_{th} instabilities with complex dynamics in real devices, hence do not allow a more aggressive design. In response to this gap between reliability study and circuit design, there is growing effort in understanding and modeling the V_{th} instabilities, and the present thesis is a part of this effort.

This thesis examines the dominant V_{th} instability mechanisms in two advanced gate dielectric materials, namely the nitrided silicon oxide (or silicon oxynitride), and the hafnium oxide. Negative bias temperature instability (NBTI)[9] and charge trapping phenomena[10–11] in these two dielectric films are the focus of this study, and form the main chunk of this thesis in chapter 3 and chapter 4, respectively. Since the accurate characterization of threshold voltage instabilities is a pre-requisite of the desired study, much effort was spent on developing the fast

$I_d - V_g$ measurement technique as detailed in chapter 2.

References

[1] W. S. Gorton. The genesis of the transistor. *Proceedings of IEEE*, 86(1):50–52, 1998. Reprinted from W. S. Gorton, “The Genesis of the Transistor,” Bell Telephone Laboratories Memorandum for Record, Dec. 27, 1949..

[2] International technology roadmap for semiconductors. 2005 [Online]. Available: http://www.itrs.net.

[3] Bruce E. Deal. The current understanding of charges in the thermally oxidized silicon structure. *Journal of The Electrochemical Society*, 121(6):198C-205C, 1974.

[4] R. Gonzalez, B.M. Gordon and M.A. Horowitz. Supply and threshold voltage scaling for low power CMOS. *Solid-State Circuits, IEEE Journal of*, 32(8):1210-1216, Aug 1997.

[5] R. Thewes, R. Brederlow, C. Schlunder, P. Wieczorek, A. Hesener, B. Ankele, P. Klein, S. Kessel and W. Weber. Device reliability in analog CMOS applications. In *IEEE International Electron Devices Meeting 1999*, pages 81-84, 1999.

[6] A.T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John and S. Krishnan. NBTI impact on transistor and circuit: models, mechanisms and scaling effects. In *IEEE International Electron Devices Meeting 2003*, pages 14-5, Washington, DC, USA, 2003.

[7] Vijay Reddy, Anand T. Krishnan, Andrew Marshall, John Rodriguez, Sreedhar Natarajan, Tim Rost and Srikanth Krishnan. Impact of negative bias temperature instability on digital circuit reliability. *Microelectronics Reliability*, 45(1):31–38, 2005.

[8] S.V. Kumar, K.H. Kim and S.S. Sapatnekar. Impact of nbti on sram read stability and design for reliability. In *Quality Electronic Design, 2006. ISQED '06. 7th International Symposium on*, pages 6pp., 2006.

[9] Dieter K. Schroder and Jeff A. Babcock. Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing. *Journal of Applied Physics*, 94(1):1-18, 2003.

[10] Sufi Zafar, Alessandro Callegari, Evgeni Gusev and Massimo V. Fischetti. Charge trapping in high-κ gate dielectric stacks. *Technical Digest - International Electron Devices Meeting*, pages 517 - 520, 2002.

[11] A. Kerber, E. Cartier, L. A. Ragnarsson, M. Rosmeulen, L. Pantisano, R. Degraeve, T. Kauerauf, G. Groeseneken, H. E. Maes and U. Schwalke. Characterization of the VT-instability in SiO_2/HfO_2 gate dielectrics. In *Reliability Physics Symposium Proceedings, 2003 IEEE International*, pages 41-45, 2003.

CHAPTER

2

Fast I_d–V_g Characterization for Transistors

The MOSFET transistor operates almost always quasi-statically with respect to the terminal voltages, and the MOSFET device physics evolves around its steady-state output and transfer characteristics. Often people use the term *DC* and *steady-state* interchangeably in this context. However it is important to note that even in AC operation with moderately high frequency, the carrier concentration in MOSFET, and hence the particle current, remains in steady-state and follows that in DC operation. Only the displacement current component is to be added in these AC situations. Non-quasi-static operation is rare for MOSFET, though becoming more important recently. Accordingly, the measurement of the steady-state I_d–V_g and I_d-V_d curves is the most frequent on MOSFETs.

In 1980s, highly integrated and automated electrical measurement instruments became commonly available, which includes Hewlett Packard 4140B pA meter/DC voltage source[1], followed by Hewlett Packard 4145[2], 4155/4156 semiconductor parameter analyzers[3], and similar products from other manufacturers. To many young engineers like the author, the semiconductor parameter analyzers are the *de facto* standard for measuring the $I-V$ characteristics of transistors. In fact, what is accepted is not only the measurement instrument, but also the quasi-static or DC measurement paradigm.

On the other hand, the demand for accurate measurement of small current, which is critical to MOSFET characterization, requires the measurement to be slow. In practice, to achieve sub-pico-ampere accuracy, the measurement time at each bias point (integration time) should be at least 20 ms, or one power-line-cycle, in order to minimize the interference from the power supply. In addition, small bias steps in sweep measurement and a delay time before measurement at each step are recommended, to ensure that all transients caused by the bias step die down and the true steady-state characteristics is measured. As a result, slow DC measurement was quite often equated to accurate measurement.

This assumption on DC measurement saw some challenges in the field of CMOS reliability study. The degradation of MOSFET characteristics under electrical stress can recover after the stress is removed. This recovery produces strong transients in MOSFET device parameters, notably in the threshold voltage V_{th}, in the time scale of micro second to tens of seconds. The typical measurement time with the semiconductor parameter analyzers ranges from milli-seconds to tens of seconds, which overlaps with that of the transient in device parameters. This overlap in time scale leads to the uncertainty as to what extent the transistor parameters (e.g., threshold voltage) have changed during the measurement time. It is therefore required to remove this overlap by a fast measurement of transistor characteristics with measurement time of 1 μs or less. Note again that when compared to the time constant of carrier redistribution (sub-nanosecond) in the transistor, both the parameters shifts (e.g., threshold voltage shift) and the proposed sub-microsecond measurement are slower by orders of magnitude. Therefore the MOSFET transistor proper remains in quasi-static operation during the fast measurement, although the capacitive (displacement) current, as we shall see, plays a role. The implementation and validation of one possible sub-microsecond fast measurement technique is the subject matter of this chapter.

2.1 Development of the Fast Techniques

Our attempt on fast measurement techniques was motivated by the studies on the stress induced threshold voltage (V_{th}) shift of MOSFET transistors, notably PBTI and NBTI. Conventionally one evaluates the V_{th} degradation in measure-stress-measure cycles with DC parametric semiconductor analyzer. The time delay between the end of stress and the I_d–V_g measurement for V_{th} extraction is typically in the order of $0.1 - 10$ seconds, and the degraded V_{th} will possibly recover during this delay. The recovery during the short delay has long been thought negligible until recently. Studies on charge trapping in high-κ and SiON dielectrics in recent years showed that the recovery in V_{th} is significant even within an 1 ms delay [4–7]. In fact, as charge trapping was one of the main show-stoppers of high-κ dielectrics, the accurate measurement of it, without contamination from recovery, is of vital importance. Therefore, a fast and accurate V_{th} measurement technique is required to capture all the fast transient trapping/de-trapping phenomenon.

Some attempts on reducing the measurement time with the semiconductor parameter analyzers Agilent 4156 were made initially. The minimum measurement time for a single bias point is 80 μs as specified by the manufacturer, which may not be sufficiently fast, but is much faster than it is usually configured. However, it was soon realized that a lot of overhead time must be added to the quoted minimum. Some examples include:

- Time required to setup the source/meter for an $I_d - V_g$ sweep after the stress is removed. Manual operation at the front panel is clearly not viable option. The at-the-time popular programming interface (SCPI) actually presses the front panel keys internally, and is not much faster than manual operation. The low-level *FLEX* programming interface, though much harder to program, can minimize the setup time by storing all low-level commands in the built-in execution queue. However, the time required to execute these low-level commands is not documented in its manual. This setup delay was estimated from field measurement to be tens of milliseconds.

- Time required to switch to the correct measurement range. Since each measurement range covers current in about one order of magnitude with its best precision, range switching is often required. Again the exact time for selection is not specified.

- Integration time is automatically increased in low current range, which exact figures appeared in the latest manual update[3].

Other source-meters from the major manufacturers shared similar problems, as they are designed towards the accurate but slow DC measurement. Due to these limitations, it is apparent that one has to sought solution in other instruments.

In response to the large charge trapping in high-κ dielectrics, A. Kerber et al. developed a *pulsed IV* method to measure hysteresis in I_d–V_g within $10-100$ μs in 2003[4]. This marked the first demonstration of the short time constant in charge trapping and de-trapping in gate dielectrics, and motivated many researchers in the community to explore the fast measurement techniques.

The measurement method developed by A. Kerber et al. is shown in Figure 2.1. A short trapezoidal pulse is applied to the gate of MOSFET. The oscilloscope measures the voltage drop across the sense resistor R at the drain, and hence the drain current. However, this method suffers from the following limitation for ultra-fast measurement. In this method, the drain voltage of the MOSFET under test is not a constant, but changes with changing drain current. The parasitic capacitor C_0 and C_{gd} must be charged or discharged as the drain voltage and gate voltage change, and the charging current distorts the measured drain current. The parasitic capacitor C_0 consists of the C_{ds} of the MOSFET, the input capacitance of the oscilloscope and the cable capacitance, and ranges from 20 pF to over 100 pF depending on the length of the cables. The distortion is more serious when the voltage pulse's rise or fall time becomes shorter.

Another approach to fast I_d–V_g measurement uses a transimpedance amplifier instead of a sense resistor[5, 8–9]. The circuit adopted by the author is shown in Figure 2.2, while other groups used slightly different circuit configurations. The active gain of the amplifier provides a constant bias on the drain terminal. However, the problems associated with the parasitic components remain in this setup. As was later realized, most measurement difficulties originate from the fact that devices on the wafer are connected to the measurement instrument through

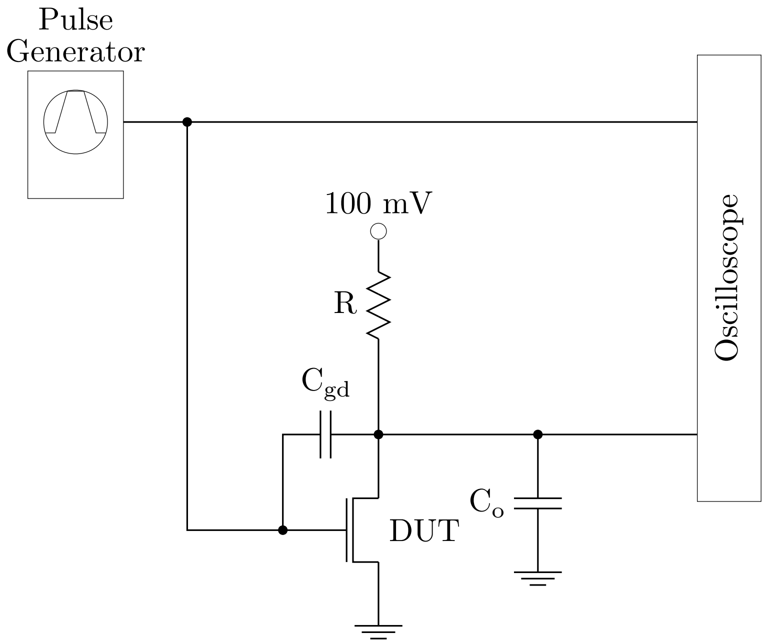


Figure 2.1 Schematic illustrating the fast $I_d - V_g$ measurement setup developed by A. Kerber.

long cables. The signal path is typically greater than 3 m with a common manual probe station from major manufacturers. To resolve the problem, we attempted to reduce the length of cables, and the setup was continually improved. In the following we shall describe a recent version, along with an analysis on source of measurement errors.

2.2 Fast Measurement Setup

We developed an improved pulsed $I_d - V_g$ measurement technique, as shown in Figure 2.3. The circuit schematic diagram is shown in Figure 2.4.

The MOSFET (DUT) is connected to the operational amplifier (OPAMP) configured in transimpedance mode. By virtual short circuit property of OPAMP, the voltage at the two input terminals are approximately equal when negative feedback is present through R. The drain voltage of the MOSFET is thus fixed at V_{ds} supplied by the voltage source. Since the input impedance at the input terminal of OPAMP is very high, the drain current flows entirely through the gain resistor

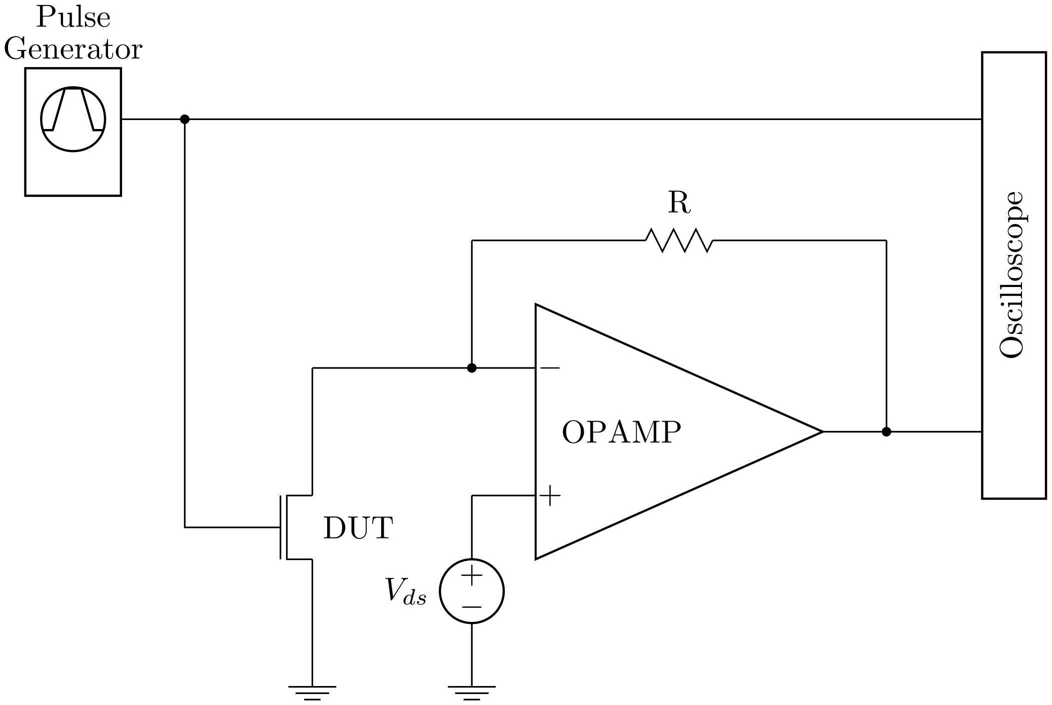


Figure 2.2 Schematic illustrating the fast I_d–V_g measurement setup utilizing a transimpedance amplifier.

R. In other words, the drain current is measured by the gain resistor R. Resistors ranged from $1 - 10$ kΩ are used in this study for different transimpedance gain. The output voltage from the OPAMP is related to the MOSFET drain current by

$$V_{\text{out}} = \left(I_d - I_{\text{gd}}\right) \cdot R + V_{\text{ds}} \tag{2.1}$$

where R is the sense resistance, V_{ds} is the drain voltage, and I_{gd} is the current from gate to drain through the parasitic capacitor C_{gd}. The current I_{gd} is caused by the fast transient at the gate and is given by

$$I_{\text{gd}} = C_{\text{gd}} \cdot \frac{\mathrm{d}V_{\text{gd}}}{\mathrm{d}t} = C_{\text{gd}} \cdot \frac{\mathrm{d}V_{\text{gs}}}{\mathrm{d}t} \tag{2.2}$$

In the measurement, the MOSFET is biased in linear region in $I_d - V_g$ measurements, and C_{gd} is given by

$$C_{\text{gd}} = C_{\text{overlap,d}} + \frac{1}{2} C_{\text{inv}} \tag{2.3}$$

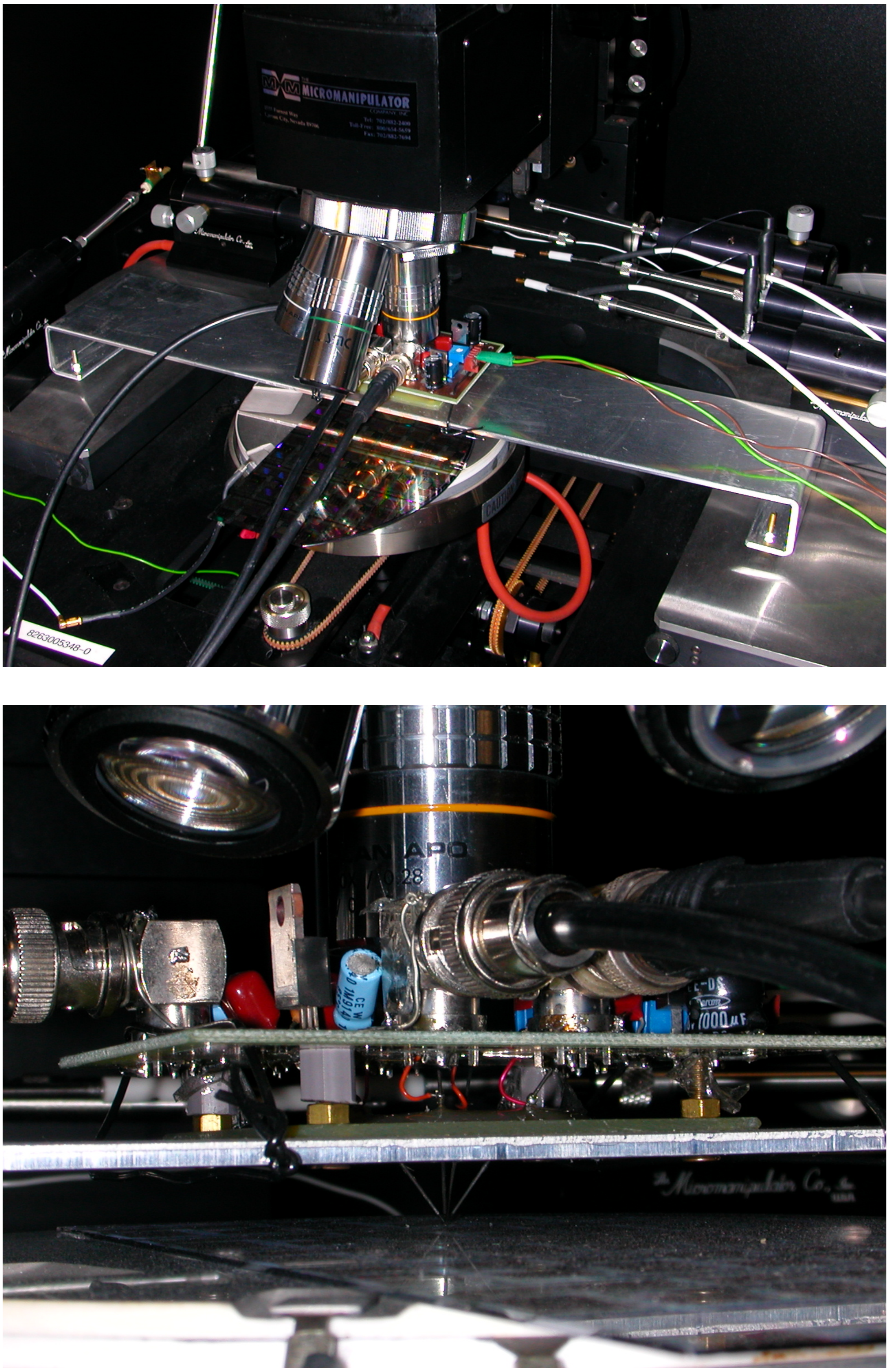

MICROMANIPULATOR
8263005348-0
APO

Figure 2.3 Photographs on the fast measurement setup.

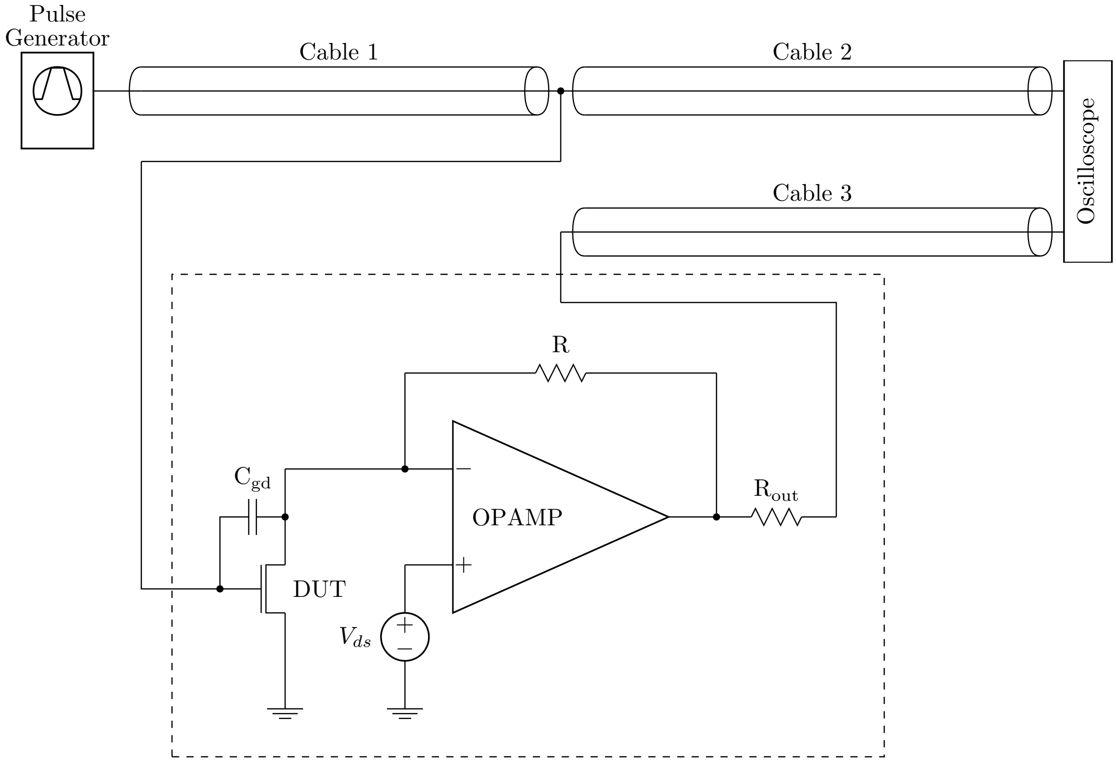


Figure 2.4 Schematic illustrating the fast I_d–V_g measurement setup utilizing a transimpedance amplifier, with matched impedance and cable delay.

where $C_{\mathrm{overlap,d}}$ and C_{inv} are the capacitance of the drain overlap region and the inversion capacitance, respectively.

A high-speed OPAMP (OPA657) with 1.6 GHz gain bandwidth product is used to achieve fast measurement[10]. As will be discussed in more details later, the accurate and fast measurement primarily relies on the minimization of the length of signal paths without impedance control. In Figure 2.4, impedance controlled cables are labeled explicitly, while thin wires represent PCB traces, wires and probe tips. The components enclosed by the dashed box (except the DUT) is on a printed circuit board measuring 10 × 6 cm. The PCB is mounted immediately above the home-brew probe card with four probe tips, as shown in Figure 2.3. The signal path of any non-impedance controlled section (e.g. from the drain of the transistor to the input of OPAMP, or from to the gate to the junction with cable 1 and cable 2) is less than 10 cm, in order to minimize parasitics.

All the transmission lines are 50 Ω co-axial cables. The output impedance of the pulse generator, and the input impedance of the oscilloscope are adjusted to 50 Ω as well. Since the MOSFET gate has very high impedance, therefore for short enough wires, the branch leading to the gate does not break the impedance matching between cable 1 and 2. Resistor $R_{\text{out}} \approx 50\ \Omega$ is used to match the cable impedance, so the voltage recorded by the oscilloscope, through cable 3, is 1/2 of V_{out} from the OPAMP. The minimization of uncontrolled signal path marks the largest difference between the improved setup of Figure 2.4 from the previous attempts as in Figure 2.2.

If one feeds a voltage pulse to the gate of the transistor DUT, and turn on the transistor, the drain current would induce a corresponding voltage pulse in output voltage. Both pulses are recorded by the oscilloscope, as shown in Figure 2.5, and conversion from V_{out} to I_d is possible with (2.1). An parametric plot of $V_g(t)$ and $I_d(t)$ would yield the familiar I_d–V_g curve.

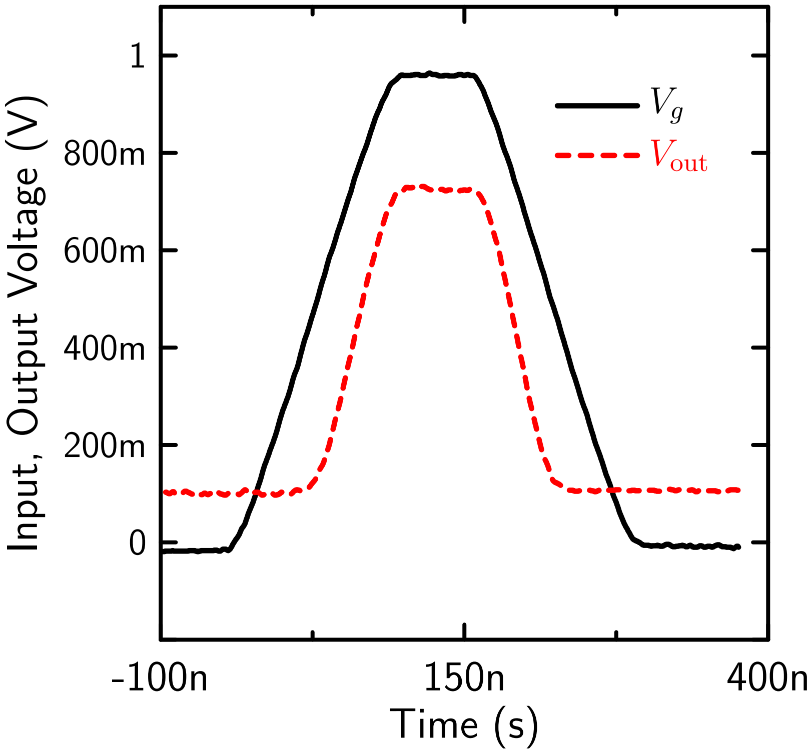


Figure 2.5 Voltage waveform recorded from the oscilloscope.

2.3 Source of Errors

For short-channel devices, C_{gd} is small, and the corresponding I_{gd} is much smaller than the drain current, and therefore the charging current through C_{gd} can be ignored. When a symmetric triangular pulse is applied at the gate as shown in Figure 2.6 inset, I_d–V_g curve can be measured at both the up-trace and down-trace of the pulse. In the two cases, $\mathrm{d}V_g/\,\mathrm{d}t$ are of the same magnitude but of opposite sign. For n-MOSFET with short channel length $L_{\mathrm{g}} = 0.1\ \mu$m, the $I_d - V_g$ curves measured in the up-trace and down-trace of V_{gs} both coincide with that from DC measurement, as shown in Figure 2.6, which indicates negligible effect of charging current through C_{gd}.

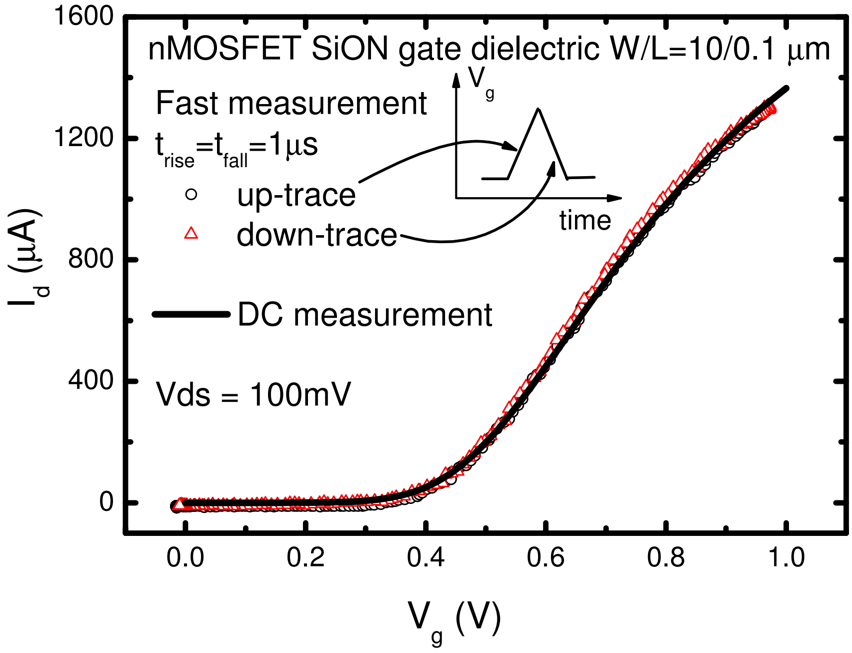


Figure 2.6 $I_d - V_g$ characteristics measured from a short-channel n-MOSFET with SiON gate dielectric, with the V_g waveform shown in the inset. Fast I_d–V_g measurement (1 μs measurement time) result is identical to the conventional DC-ramp measurement result in both up-trace and down trace.

The measurement speed is limited to about 100 ns in this study due to the frequency response of the transimpedance amplifier. As seen from Figure 2.5, V_{out} waveform must be synchronous to V_{gs} waveform in order to get the correct $I_d - V_g$ curve. A delay difference δt between V_{gs} and V_{out} waveforms will generate approximately a horizontal shift of $I_d - V_g$ curve, and is given by

$$\delta V_{\text{gs}} = \frac{\mathrm{d}V_{\text{gs}}}{\mathrm{d}t} \cdot \delta t. \tag{2.4}$$

When measurement time (rise/fall time) is reduced $\mathrm{d}V_{\text{gs}}/\,\mathrm{d}t$ increases, and the distortion of $I_d - V_g$ curve worsens quickly. One source of this delay difference arises from the unmatched signal path length, for example, between the two long cables 2 and 3 in Figure 2.4. One meter of cable length difference causes 5 ns delay time skew. If in the fast measurement V_g ramps from 0 to 1 V in 100 ns, 5 ns introduces 50 mV shift in $I_d - V_g$ curve. Cables of equal length are therefore required.

The other, and more fundamental, source of the delay skew arises from the parasitic capacitance in parallel with the feedback resistor R. For $R = 1$ kΩ typically used in this study, a capacitance as little as 1 pF would cause a delay of 1 ns in the amplifier output waveform. If one requires the distortion of $I_d - V_g$ curve to be less than 10 mV, the maximum ramping rate for V_g would be 1 V over 100 ns. Therefore, the control of this parasitic capacitance is critical to successful fast measurement. In practice, small stray capacitance is unavoidable. Sometimes, it is even necessary to include a small capacitance to improve the stability of the amplifier[10]. The major source of instability originates from the parasitic capacitance seen at the inverting input terminal of the amplifier, which consists of the intrinsic input capacitance of the OPAMP, the capacitance between PCB traces, wires connecting the probe tip to the PCB, and finally the probe tips. Considering a typical coaxial cable with capacitance of 67 pF/m, or parallel wires (5 mm separation) with capacitance around 10 pF/m, it is immediately clear that the extremely short wires in the current setup is absolutely necessary. This is the limit of the current fast measurement technique, arising from the requirement of synchronous V_g and V_{out} waveforms.

On the other hand, if only the drain current at a specific V_g bias is required as in the case of Ref. [11], synchronization of the V_{out} and V_g waveforms is not necessary, and much faster measurement is possible.

For long channel devices, the charging current through the large C_{gd} can be a significant portion of the drain current, and need to be corrected. Figure 2.7 shows that for n-MOSFET with $L_{\text{g}} = 10\ \mu\text{m}$, the I_d-V_g curve measured in the up-trace of V_g does not coincide with that from the down-trace. The difference between the two I_d-V_g curves is inversely proportional the rise time t_{rise} of V_g. This indicates a capacitive branch and is attributed to C_{gd}. This gate-to-drain capacitance can be measured with a slightly modified split-$C-V$ measurement. The source and body of the MOSFET are grounded, and capacitance between gate and drain is recorded, as shown in the inset of Figure 2.8.

This measured $C-V$ curve provides good approximation to C_{gd} when the MOSFET is operating in linear region when V_{ds} is small. The effect of charging and discharging of C_{gd} to the drain current measurement can be corrected using (2.1) and (2.2). Figure 2.8 shows the raw and corrected I_d-V_g curve obtained from a MOSFET with channel length $L_{\text{g}} = 10\ \mu\text{m}$. After correction, the up-trace and down-trace I_d-V_g curves coincide, verifying the above analysis. An first order analysis suggests that the error in drain current due to C_{gd} scales with the channel length as

$$\frac{\delta I_d}{I_d} \propto {L_{\text{g}}}^2 \cdot \frac{\mathrm{d}V_{\text{gs}}}{\mathrm{d}t}. \tag{2.5}$$

Therefore, the effect of C_{gd} quickly vanishes for gate length less than 1 to 2 μm for 1 μs measurement time.

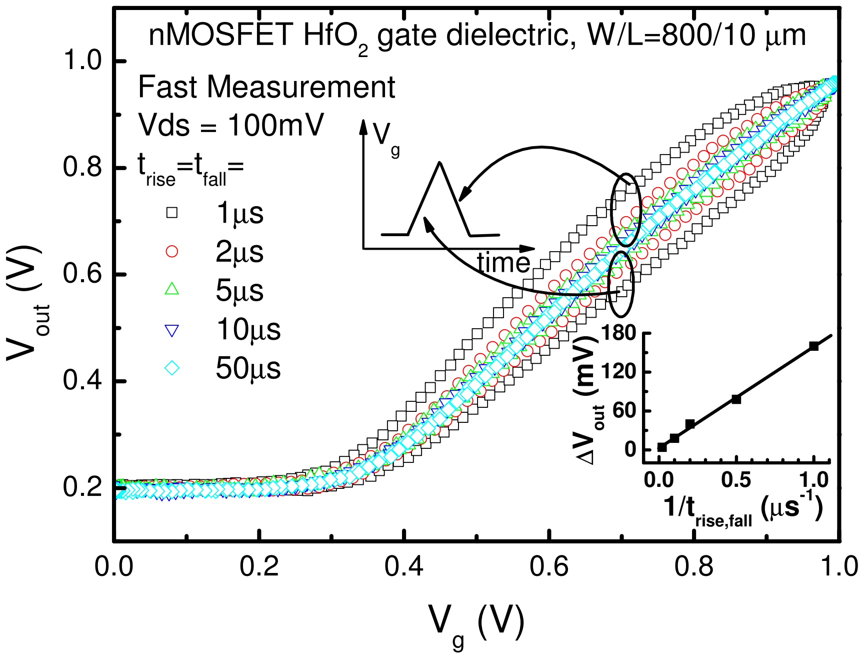


Figure 2.7 $I_d - V_g$ characteristics measured from a long-channel n-MOSFET with HfO$_2$ gate dielectric, with the V_g waveform as in Figure 2.6. As the rise time and fall time of the waveform decreases, the $I_d - V_g$ curves from up-trace and down-trace move apart from each other, due to the charging and discharging of C_{gd} capacitance. The difference between the two curves measured at $V_g = 0.6$ V is plotted in the inset as a function of the reciprocal of rise time.

2.4 Applications To Charge Trapping in High-κ gate dielectrics

We applied this improved fast measurement technique to study the charge trapping in high-κ dielectrics. The MOSFET transistors used in this experiment have MOCVD deposited HfO$_2$ gate dielectric with EOT $\approx$ 1.3 nm[12]. After an initial $I_d - V_g$ measurement, MOSFETs are stressed for 1s, and $I_d - V_g$ is measured again at the falling edge of the stress voltage. Threshold voltage shift is extracted from the horizontal shift of the $I_d - V_g$ curve before and after the stress. In Figure 2.9, the V_{th} shift of an n-MOSFET is plotted with varying falling edge time, i.e. $I_d - V_g$ measurement time, at the end of the 1 s stress. It is observed that as measure-

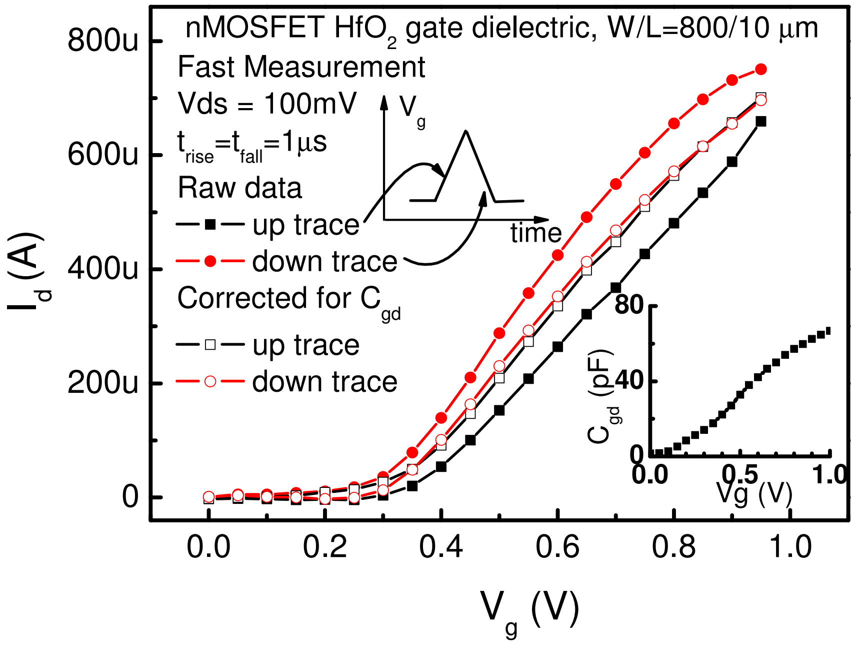


Figure 2.8 I_d-V_g characteristics after correction for the effect of C_{gd}. Correction are based on (Figure 2.1) and (Figure 2.2), for the $t_{\mathrm{rise}} = 1$ μs case. The gate to drain capacitance is measured using modified split-$C-V$ method, and is shown in the inset.

ment time is increased above 100 μs, the measured V_{th} drops due to the significant de-trapping during the measurement. The V_{th} shift measured by conventional DC measurement is only 20% of the true value, and gives a very poor under-estimation of the trapped charge in high-κ dielectrics. On the other hand, from this figure, measurement time of 5 μs to 10 μs is fast enough to evaluate the charge trapping in this particular device adequately.

2.5 Conclusions

In conclusion, we have developed a fast measurement technique to obtain MOSFET I_d-V_g characteristics in 100 ns without using expensive RF measurement setup. The sources of measurement errors are analyzed in detail. Techniques to minimize those errors, or correction procedure are provided to each of the error sources.

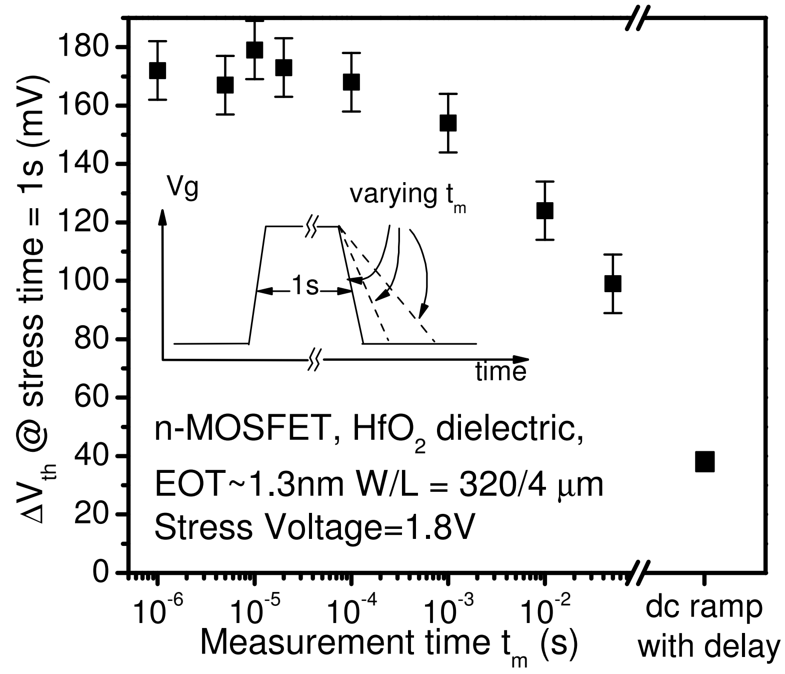


Figure 2.9 NMOSFET transistors with HfO_2 gate dielectrics are stressed for 1 second with $V_g = 1.8$ V, after which $I_d - V_g$ is measured with different measurement time $t_{\rm m}$ at the falling edge of the stress voltage. The threshold voltage shift before and after stress ($\Delta V_{\rm th}$) is plotted against measurement time, and the threshold voltage shift obtained with DC measurement is shown for comparison. Slow measurements leads to serious underestimation of the threshold voltage shift caused by charge trapping.

The limits of such measurement setup is identified, which shows that the current implementation with 100 ns measurement time is close to the performance limit.

References

[1] *Operation and Service Manual: Model 4140B pA Meter/DC Voltage Source.* Hewlett Packard Company, 1980.

[2] *Operation Manual: Model 4145B Semiconductor Parameter Analyzer*. Hewlett Packard Company, 1989.

[3] *Agilent 4155C/4156C Manuals: including User's Guide, Programmer's Guide and Command Reference.* Agilent Technologies, 2003.

[4] A. Kerber, E. Cartier, L. A. Ragnarsson, M. Rosmeulen, L. Pantisano, R. Degraeve, T. Kauerauf, G. Groeseneken, H. E. Maes and U. Schwalke. Characterization of the VT-instability in SiO_2/HfO_2 gate dielectrics. In *Reliability Physics Symposium Proceedings, 2003 IEEE International*, pages 41-45, 2003.

[5] 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 HfO_2 gate dielectrics and frequency dependence of dynamic BTI in MOSFETs. In *IEDM Tech. Digest*, pages 733-736, 2004.

[6] M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, Y. Rey-Tauriac and N. Revil. On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's. In *IEDM Tech. Digest*, pages 109-112, 2004.

[7] T. Yang, M. F. Li, C. Shen, C. H. Ang, Chunxiang Zhu, Y.-C. Yeo, G. Samudra, S. C. Rustagi, M. B. Yu and D. L. Kwong. Fast and slow dynamic NBTI components in p-MOSFET with SiON dielectric and their impact on device life-time and circuit application. In *VLSI Symp.*, pages 92-93, 2005.

[8] C. Leroux, J. Mitard, G. Ghibaudo, X. Garros, G. Reimbold, B. Guillaumot and F. Martin. Characterization and modeling of hysteresis phenomena in high-κ dielectrics. In *IEDM Tech. Digest*, pages 737–740, 2004.

[9] C. T. Chan, C. J. Tang, C. H. Kuo, H. C. Ma, C. W. Tsai, H. C.-H. Wang, M. H. Chi and Taihui Wang. Single-electron emission of traps in HfSiON as high-k gate dielectric for MOSFETs. In *IRPS*, pages 41-44, 2005.

[10] *Datasheet of OPA657: 1.6GHz, Low-Noise, FET-Input Operational Amplifier*. Texas Instruments Inc..

[11] C. D. Young, Y. G. Zhao, M. Pendley, B. H. Lee, K. Matthews, J. H. Sim, R. Choi, G. Bersuker and G. A. Brown. Ultra-short pulse I-V characterization of the intrinsic behavior of high-κ devices. In *SSDM Proc.*, pages 216-217, 2004.

[12] H. Y. Yu, J. F. Kang, J. D. Chen, C. Ren, Y. T. Hou, S. J. Whang, M. F. Li, D. S. H. Chan, K. L. Bera, C. H. Tung, A. Du and D.-L. Kwong. Thermally robust high quality HfN/HfO_2 gate stack for advanced CMOS devices. In *IEDM Tech. Digest*, pages 99-102, 2003.

CHAPTER
3

Negative-Bias-Temperature Instability in SiON Gate Dielectrics

3.1 A Brief Historical Introduction

In the 1960s, the instabilities in thermally oxidized MOS structure was the major inhibiting issue for MOSFETs. In a 1974 review paper [1], Deal recalled a cartoon made by R. P. Donovan, which depicted a bunch of scientists, blindfolded, patting different parts of an elephant, and proposing different theories for instabilities found in MOS. The "blind men and the elephant" era ended, after the alkali ions were eliminated, and the other oxide charges thoroughly investigated and categorized. We have largely followed the classification of Deal[1], with however somewhat different terminologies. As usually appeared in modern textbooks, major source of instabilities of MOS with thermally grown SiO_2 dielectric includes:

- Mobile ions, which are mostly positive alkali ions such as Na^+, K^+ and Li^+.

- Fixed charges (Q_f), which are positively charged silicon species in the oxygen-deficient transition region near the SiO_2/Si interface.

- Interface states (N_{it}), also commonly known as fast states, which are unsaturated silicon bonds at the interface.

- Oxide traps (N_{ot}), which are broken Si—O bonds in the bulk of the oxide film.

Interface states are easily differentiated from the other types of charges because they are in direct communication with both types of carriers in the semiconductor substrate, which leads to fast response to carrier concentrations. As one sweep the gate voltage, the carrier capture/emission processes at the interface states are usually fast enough to align the quasi-fermi level of trap occupancy to the quasi-fermi level of the dominant carriers at the interface. This is also the origin of the synonym *fast states*. Early works established that under combined bias and temperature stress, the density of interface states increases significantly[2], which gave the name of *negative-bias-temperature instability* or NBTI. Only p-channel MOS devices were found susceptible to this type of stress, so only negative voltage bias on the gate terminal is relevant. Jeppson and Svensson first proposed a model to the degradation dynamics in 1977[3]. They recognized the power-law growth of interface states with time ($N_{\mathrm{it}} \sim t^{1/4}$), a signature of diffusion controlled process. They suggested that in addition to the electrochemical process that generates interface states, the resultant species of this reaction diffuse away from the surface, which ultimately controls the speed of interface state generation. This reaction-diffusion model was widely accepted, and with few amendments, remains the most popular theory for NBTI phenomena. At the same time, annealing in H_2 ambient was found to effectively reduce the density of interface states before and after NBT stress[4]. Sintering in H_2 subsequently become part of the standard MOS fabrication process.

However, during this period of initial research on NBTI, n-MOS VLSI process became dominant and replaced the p-MOS process. When CMOS process subsequently gained popularity, it employed buried channel p-MOSFETs which is less susceptible to NBTI. Therefore, it was only in the late 1990s when surface channel p-MOSFET with p^+-polysilicon gate was introduced, that NBTI received attention again. The number of research papers on NBTI exploded since the year 1999. As the industry had just switched to silicon-oxynitride (SiON) for the gate dielectric, much attention was paid to the effect of nitrogen incorporation on NBTI[5–12]. Results varies among groups, due to the diverse techniques employed to incorporate nitrogen in the dielectric. However, it is generally agreed that

- NBTI is worse with higher nitrogen concentration in the oxide.
- Plasma nitridation processes offers significant advantage over thermal nitridation.

Schroder and Babcock gave a comprehensive review on the subject in early 2003[13], which summarized the understanding on NBTI at the time, its origin, its effect on transistor, and most importantly its dependence on process and device parameters. Research on NBTI prior to 2003 largely followed the paradigm set in the 70s, which was reinforced with Schroder's definition for NBTI[13]:

> "NBTI occurs in p-channel MOS devices stressed with negative gate voltages at elevated temperatures. It manifests itself as absolute drain current $I_{d,\mathrm{sat}}$, and transconductance g_m, decrease and absolute "off" current I_{off}, and threshold voltage V_T increases. Typical stress temperatures lie in the $100-250°\mathrm{C}$ range with oxide electric fields typically below 6 MV/cm."

A typical NBTI study involves first stressing a MOS at elevated temperature under large negative gate bias. The gate voltage must be chosen such that the electric field in the oxide is less than about 6 MV/cm to avoid degradation due to Fowler-Nordheim tunneling[3]. After a preset stress time, the sample is cooled to room temperature and its device characteristics measured. Typically the parametric shift in threshold voltage, flat-band voltage and linear drain current are recorded. Generation of interface state density is commonly measured by one of the many available techniques, while the accompanying generation of fixed charge is calculated indirectly. This stress-measure cycle is repeated which extends to, typically, 10^3-10^5 seconds. Simply put, NBTI degradation was thought to be almost permanent, and remains unchanged after the electrical stress is removed. Although back in 1977 Jeppson reported that NBTI degradation can be healed by annealing the stressed sample for many hours at elevated temperature ($125-200°\mathrm{C}$), the recovery seems too slow to appreciably affect the measurement. If we view studies prior to Schroder's 2003 review paper as the old paradigm largely following the Deal-Jeppson tradition, we can identify a few new developments, which include two prominent topics widely studies and hotly debated in the NBTI community: the dynamic recovery effect and the role of hole trapping in NBTI. These two new phenomena are especially important if nitrogen is introduced to the SiO_2 gate dielectric. They are the subject of this chapter, which will un-fold some investigations performed by the author and co-workers. In the remainder of this introductory section, a brief review on each of the two topics is provided.

3.1.1 Dynamic Recovery

Around the time of Schroder's review, the NBTI community started to realize that the recovery effect is not negligible as always assumed [14–19]. It is observed that NBTI degradation has appreciable partial recovery within short time less than 1 second. Therefore, if one applies an AC stress voltage on the p-MOSFETs (e.g. square wave with 50% duty cycle), the sample is stressed for half of the period in each cycle. During the other half-cycle, existing NBTI degradation is being healed. As a result, NBTI degradation is much less for samples under AC stressing than those under constant voltage stressing, for equivalent stress time. Since transistors in logic circuits are quite often under dynamic operation condition, and are not under constant voltage stress, it was claimed that one could exploit the dynamic recovery of NBTI and relax the reliability specifications.

However, it is only recognized later, that there is a second significant implication of the dynamic recovery effect. Since a large portion of the NBTI degradation can recover within a time scale of 1 second, the traditional stress-measure paradigm is in trouble. Cooling wafer from stress temperature to measurement temperature (room temperature) takes minutes, after which the measured NBTI degradation is seriously contaminated by the recovery effect. Even if one eliminates the cooling procedure, and measure right after the electrical stress ends, the commonly used parametric analyzer takes some time (anywhere between 0.5 second to 10 seconds depending on the specific setup). The above cited early reports on dynamic recovery effect of NBTI all use the traditional stress-measure paradigm, and the registered NBTI degradations are always inclusive of some recovery effect. As a result, the relatively slow recovery transient recorded in the > 1 s time scale prompts one to ignore this second effect of dynamic recovery, and concentrates on the desirable relaxation under AC operation. The unexplored sub-one-second regime of dynamic NBTI was for a brief moment overlooked, and alternative characterization paradigm was sought only to avoid contamination from measurement delay soon after.

S. Rangan first described a scheme to investigate the sub-one-second recovery transient[20]. It involves converting the change in drain current ΔI_d at one fixed bias point into the change in threshold voltage ΔV_{th}. Since measuring the drain current at one bias point can be done much more quickly than a sweep measurement of the I_d–V_g characteristics. This provides a much faster way to extract the threshold voltage degradation due to NBTI, and thus reduces the NBTI recovery during the measurement delay. In this scheme, it is also possible to measure ΔI_d at the NBTI stress condition, if one applies a small drain bias during stress. This scheme was later adopted by many researchers under the name *on-the-fly measurement*, with some variations in implementation details [21–26]. The *on-the-fly* measurement technique is commonly referred as delay-free or recovery-free, which is apparently true because the stress voltage is never removed from the gate terminal. However, due to the fact that all these researchers use slow measurement instruments, the non-negligible NBTI degradation during measurement time seriously affects the measurement result. The author believes that some variant of fast measurement techniques must be used to adequately characterize the fast transient in NBTI. A detailed review on the appropriate characterization procedure for NBTI is discussed in section 3.3.

3.1.2 Role of Hole Trapping

It is well known that an SiO_2 gate dielectric thin film can be degraded by carriers passing through it, also well known is that defects in dielectric, generated or pre-existing, can capture carriers and then emit them. The charge trapping/de-trapping dynamics has been extensively studied and modeled in the 1980s[27]. In MOSFET with gate dielectric thicker than 4 nm, gate current is significant only under hot-carrier stress or Fowler-Nordheim stress conditions. In order to separate the charge trapping in the bulk dielectrics and the interface states generation under the classification of Deal[1], NBTI stress is usually performed at zero drain bias to avoid hot-carrier injection and oxide electric field of < 6 MV/cm to avoid Fowler-Nordheim tunneling. Higher electric field was shown to cause threshold voltage shift due to charge trapping in addition to interface state generation[28], which is seen as a pitfall in NBTI characterization for SiO_2 thickness between 2.5 nm and 4 nm. It is commonly accepted that F-N tunneling only causes damage to the region where electron has been injected to the conduction band (or valance band) of the dielectric layer. For ultra-thin gate oxide films, F-N tunneling is less damaging to the dielectric, while at the same time direct tunneling current becomes significant. Little trap generation is expected from direct tunneling carriers, but these tunneling carriers can momentarily get trapped in pre-existing traps in the gate dielectric. On the other hand, as the gate dielectric scales down into direct tunneling regime, a significant content of nitrogen is always added to the SiO_2 dielectric. Since the silicon-nitrogen bond is less flexible than silicon-oxygen bond, it is well known that nitrogen incorporation leads to high defect density in SiON. This charge trapping effect at low electric field occurring at the bias condition of NBTI stress complicates the traditional understanding of NBTI as an interface-states dominated phenomenon.

Since NBTI is found to be worse in transistors with nitrided gate oxide[5–12], the possible contribution from hole-trapping was considered. Ushio, Kushida-Abdelghafar and Watanabe first proposed that hole-trapping near the Si/SiON caused the enhanced NBTI in SiON gate dielectric[7, 29]. On the other hand, alternative theory was suggested by Tan and co-workers, suggesting that hydrogen released from broken Si–H bonds can be trapped at nitrogen sites at the interface and create additional fixed charges[9–10]. Tan's hydrogen-induced-fixed-charge theory appears to explain the nitrogen-enhanced NBTI equally well, and avoided being "contrary to the prevailing NBTI mechanisms (reaction-diffusion)"[10]. These initial discussions started the debate on the role of hole-trapping in SiON gate dielectric. At the core of the debate is the question of whether a hole-trapping model is mandatory to adequately explain the experimental observation of worse NBTI in SiON, or some amendments to the long-standing reaction-diffusion model is sufficient. A few important arguments supporting the hole-trapping model are Ref.[21, 30–32], while examples supporting the reaction-diffusion interpretation are Ref.[24, 33–34].

It was soon realized that trapped hole can easily de-trap from the extremely thin dielectric film, which transient effect contaminates the measurement as earlier described. Therefore, the accurate characterization of NBTI becomes a prerequisite for studying the role of hole-trapping in NBTI. Each argument, either in favor or in disapprove of the hole-trapping proposal, very often starts with a prologue on the "correct" characterization procedure. Our discussion on this subject is no exception, after presenting the theories to NBTI in section 3.2, section 3.3 will first describe the characterization procedure, followed by section 3.4 on the arguments for the hole-trapping model as a necessary supplement. Lastly, in section 3.5, we consider the contribution of interface traps.

The samples measured in this chapter has an equivalent oxide thickness of 1.3 nm. The gate oxide underwent decoupled plasma nitridation (DPN) and subsequent thermal anneal. The pMOSFET devices used in this study has the dimensions $W = 100$ μm and $L = 0.1$ μm. Unless otherwise indicated, the measurements are performed at room temperature.

3.2 Theories for the Dynamic NBTI

In this section, two theories to the dynamic effect of NBTI, namely the reaction-diffusion model and the charge-trapping model, are outlined.

3.2.1 Reaction-Diffusion Model

Since the reaction-diffusion model was first proposed in 1977[3], it has been the most successful and most widely accepted model to NBTI. Although controversies, including the exact form of the diffusing hydrogen species, the exact interfacial chemical reaction and the appropriate boundary conditions, remain topics of debates[13, 29, 35], the general framework of the model is generally agreed upon by the year 2003. When the dynamic behavior was first discovered in 2003[14–17, 36], the reaction-diffusion model was soon used to explain the recovery effect and the NBTI degradation under dynamic stress[37–38].

3.2.1.1 Stress

According to the reaction-diffusion theory, interface traps generated under NBTI stress is result of the electrochemical reaction at the Si/SiO_2 interface. The exact equation of the reaction is unclear, but it is widely accepted

- that a Si—H bond is broken, hydrogen species are released, and leaving behind a silicon dangling bond (interface trap);
- that holes at the interface catalyze or participate the reaction;
- that the forward reaction is accelerated by stress voltage and temperature.

A hypothetic equation of this interface reaction under electrical stress is shown in (3.1), where atomic hydrogen is released from the interface. Assuming ample supply of holes, the rate equation of this reaction can be written as in (3.2). The generation rate of interface traps is $\frac{\mathrm{d}N_{\mathrm{it}}}{\mathrm{d}t}$, and since one atomic hydrogen is released for each generated dangling bond, this is also the generation rate of hydrogen. The released hydrogen diffuses away from the interface. If one assumes homogeneous diffusion in semi-infinite space, the diffusion equation can be written as (3.3), where natural boundary conditions apply.

$$\underset{\text{passivated}}{\equiv\text{SiH}} \quad + \quad \text{h}^{+} \quad \underset{\text{recovery}}{\overset{\text{stress}}{\rightleftharpoons}} \quad \underset{\text{dangling bond}}{\equiv\text{Si}^{+}} \quad + \quad \text{H}^{0} \tag{3.1}$$

$$\frac{\mathrm{d}N_{\mathrm{it}}}{\mathrm{d}t} = k_{\mathrm{F}}\left(N_0 - N_{\mathrm{it}}\right) - k_{\mathrm{R}} \cdot {N_{\mathrm{H}}}^{(0)} \cdot N_{\mathrm{it}}\ , \tag{3.2}$$

$$\frac{\partial N_{\mathrm{H}}}{\partial t} = D\frac{\partial^2 N_{\mathrm{H}}}{\partial x^2} + \delta\left(x\right)\frac{\mathrm{d}N_{\mathrm{it}}}{\mathrm{d}t} \qquad (x > 0). \tag{3.3}$$

N_{it} = interface trap density at time t cm^{-2}

N_0 = total density of Si—H bonds cm^{-3}

k_{F} = forward reaction rate

k_{R} = reverse reaction rate

${N_{\mathrm{H}}}^{(x,t)}$ = hydrogen concentration at position x and time t cm^{-3}

λ = effective thickness of hydrogen layer available to the reaction

D = diffusivity of hydrogen.

It is instructive to look at a few asymptotic solutions to the coupled equations (3.2). First if one assumes infinite diffusivity, the released hydrogen is immediately removed from the interface, and its concentration at the interface is almost zero. The reverse reaction is therefore suppressed, and assuming small N_{it} at $t = 0$, N_{it} would increase linearly with time in the initial stage. The forward reaction proceeds and saturates when all Si—H bonds are broken. On the other hand, if the diffusivity is zero, all released hydrogen piles up at the interface. Equation (3.2) reduces to

$$\frac{\mathrm{d}N_{\mathrm{it}}}{\mathrm{d}t} = k_{\mathrm{F}}\left(N_0 - N_{\mathrm{it}}\right) - k_{\mathrm{R}}{N_{\mathrm{it}}}^2. \tag{3.4}$$

One observes that, at short time, the growth of N_{it} is similar to the first case, but the reverse reaction would become significant as N_{it} increases. The forward reaction and reverse reaction would reach equilibrium at $N_{\mathrm{it,eq}} = \frac{k_{\mathrm{F}}}{2k_{\mathrm{R}}}\left(\sqrt{1+4\frac{k_{\mathrm{R}}}{k_{\mathrm{F}}}}-1\right) < N_0$.

The last, and most important situation arises from introducing a small diffusivity in addition to the fast reactions. It is obvious that the reaction would quickly saturate, as in the zero-diffusivity case, to $N_{\mathrm{it}} \approx N_{\mathrm{it,eq}}$. However, the slow diffusion would remove some hydrogen from the interface as time elapses, and moves the equilibrium of the interface reaction towards higher N_{it}. Asymptotic solution was derived by Jeppson to show that N_{it} grows with time as $N_{\mathrm{it}} \sim t^{1/4}$ in this diffusion-limited situation[3]. Alam, by assuming a triangular hydrogen concentration profile, arrived at the same result[37]. This derivation is briefly repeated as follows, as it leads us to a few insights into the process. As shown in Figure 3.1, the hydrogen concentration away from the interface ($x = 0$) is approximated with a simple triangular function, with characteristic diffusion length $\sqrt{Dt}$. The hydrogen concentration at the interface is ${N_{\mathrm{H}}}^{(0)}$. Since each broken Si—H bond releases one atomic hydrogen, the number of interface traps equals the number of hydrogen atoms in diffusion,

$$\begin{aligned} N_{\mathrm{it}}\left(t\right) &= \int_0^\infty dx{N_{\mathrm{H}}}^{(x,t)} \\ &= \int_0^{\sqrt{Dt}} dx{N_{\mathrm{H}}}^{(0,t)}\left(1-\frac{x}{\sqrt{Dt}}\right) \\ &= \frac{1}{2}{N_{\mathrm{H}}}^{(0,t)}\sqrt{Dt}. \end{aligned} \tag{3.5}$$

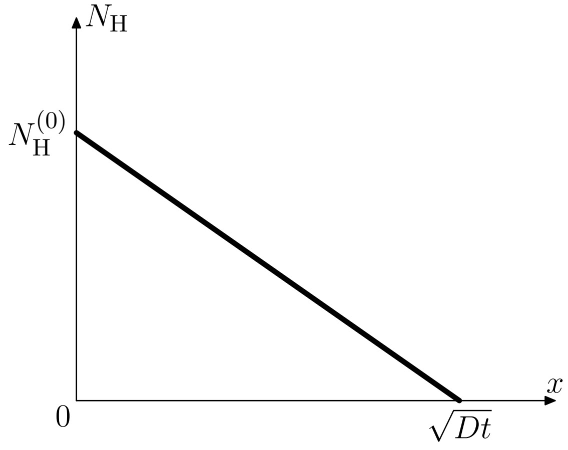


Figure 3.1 Approximate hydrogen concentration profile in the diffusion process.

It is worth repeating that the reaction is much faster than diffusion, and is close to equilibrium even in the presence of a slow diffusion. Therefore, the left-hand side of (3.2) is almost zero, and

$$N_{\mathrm{it}} \cdot N_{\mathrm{H}}{}^{(0)} = \frac{k_{\mathrm{F}}}{k_{\mathrm{R}}} N_0. \tag{3.6}$$

Combining (3.5) and (3.6), one obtains

$$N_{\mathrm{it}} = \sqrt{\frac{k_{\mathrm{F}}}{2k_{\mathrm{R}}} N_0}\,(Dt)^{1/4}\,. \tag{3.7}$$

The power-law dependence with exponent of 0.25 is easily identified. If one allows even longer stress time, the unreacted Si—H bonds become scarce, the forward reaction rate reduces, and the N_{it} generation saturates to N_0. On the other hand, as described earlier, the reaction of interface trap generation had reached an equilibrium at $N_{\mathrm{it}} = N_{\mathrm{it,eq}}$ before diffusion has the time to remove significant amount of hydrogen from the interface. Therefore, the power-law ($t^{0.25}$) relation is valid in the region $N_{\mathrm{it,eq}} \ll N_{\mathrm{it}} \ll N_0$.

3.2.1.2 Recovery

A. Alam developed the dynamic theory of the reaction-diffusion model, which considers the recovery of NBTI degradation[37]. The treatment of the diffusion process during recovery is, however, rather empirical. An analytic treatment is presented as follows.

When the stress voltage is removed from the gate at time t_0, one can assume that the forward reaction in (3.2) is effectively stopped, while the reverse reaction coefficient is not affected by the gate voltage. Silicon dangling bonds are passivated by hydrogen atoms in the vicinity of the interface through the reaction. Assuming again that the diffusion is slow. As a result, after a short recovery time period t after t_0, the reverse reaction would quickly consume all hydrogen species near the interface (${N_\mathrm{H}}^{(0)} = 0$), while hydrogen concentration away from the interface is left unchanged. Any further recovery of interface traps depends on the supply of hydrogen species back-diffused to the interface, and is diffusion limited. This can be modeled by the differential equation

$$\begin{cases} \dfrac{\partial N_\mathrm{H}}{\partial t} = D\dfrac{\partial^2 N_\mathrm{H}}{\partial x^2} \quad x > 0, & (3.8) \\ N_\mathrm{H}\big|_{x=0} = 0, & (3.9) \\ N_\mathrm{H}\big|_{t=0} = \phi(x). \quad x > 0 & (3.10) \end{cases} \tag{3.11}$$

The initial condition $\phi(x)$ is the hydrogen profile after the stress and the brief reaction-limited recovery. To solve this differential equation, one takes the anti-symmetric reflection of $\phi(x)$ to extend the domain of independent variable from $x > 0$ to $-\infty < x < \infty$, as shown in Figure 3.2. The odd symmetry of the new $\phi(x)$ guarantees that the boundary conditions in (3.8) are still satisfied. The solution is expressed with the integration

$$N_\mathrm{H}(x,t) = \int\limits_{-\infty}^{\infty} \phi(\xi)\left[\frac{1}{2\sqrt{\pi}}\frac{1}{\sqrt{Dt}}\mathrm{e}^{-\frac{(x-\xi)^2}{4Dt}}\right]\mathrm{d}\xi. \tag{3.12}$$

To obtain a closed-form formula, the initial hydrogen profile is again approximated by triangular shape

$$\phi(x) = \begin{cases} {N_{\mathrm{H}}}^{(0)}\left(1 - \frac{x}{L}\right) & 0 < x < L_0, \\ -{N_{\mathrm{H}}}^{(0)}\left(1 + \frac{x}{L}\right) & -L_0 < x < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.13}$$

Suppose the prior stress time is t_0, the width of the profile is $L_0 = \sqrt{Dt_0}$, the generated interface trap density is ${N_{\mathrm{it}}}^{(0)} = 1/2\, N_{\mathrm{H}} L_0$ as shown in Figure 3.2. Similarly, the kernel of the integration is approximated by

$$g(\xi) = \frac{1}{L}\left(1 - \frac{x - \xi}{L}\right), \tag{3.14}$$

where $L = \sqrt{Dt}$. The solution to the differential equation is thus the convolution

$$N_{\mathrm{H}}(x,t) = \int\limits_{x-L}^{x+L} \phi(\xi) g(x - \xi)\, \mathrm{d}\xi. \tag{3.15}$$

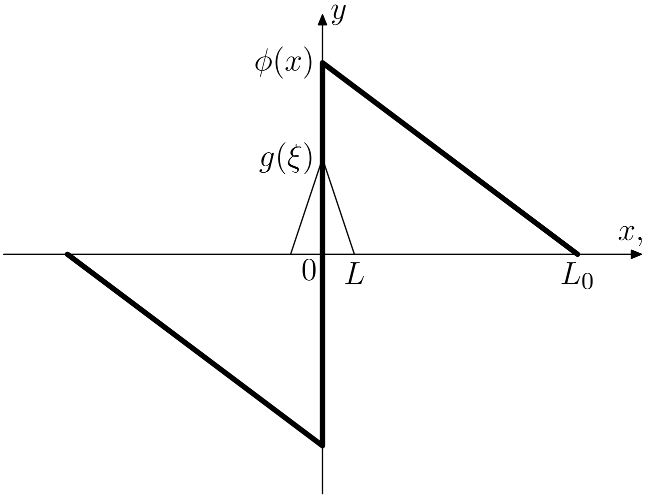


Figure 3.2 The solution to the diffusion equation during recovery is approximated by the convolution of ϕ and g, as in (3.15).

An approximation to the convolution integration is sketched in Figure 3.3 with thick solid line. It is assumed that the far end of the diffusion front is not affected by the new boundary condition at $x = 0$, and has advanced to $\sqrt{D(t + t_0)}$. The hydrogen profile has a peak at $x_{\max}$. The hydrogen lost to the left boundary (recovery of N_{it}) is between

$$\Delta N_{\mathrm{it,min}} = -1/2\, {N_{\mathrm{H}}}^{*} x_{\mathrm{max}} = -1/2\, {N_{\mathrm{H}}}^{(0)} \sqrt{\frac{t}{t+t_0}}\, x_{\mathrm{max}}, \quad \text{and} \tag{3.16}$$

$$\Delta N_{\mathrm{it,max}} = -1/2\, {N_{\mathrm{H}}}^{(0)} x_{\mathrm{max}}. \tag{3.17}$$

To determine the peak position x_{max}, one observes that when L is small compared to L_0, the peak occurs at $x_{\mathrm{max}} \approx L$ to avoid the negative contribution from the region with $\phi(x) < 0$. On the other hand, for large L, the priority is to stay in the region with large positive $\phi(x)$, and the peak position lags behind ($x_{\mathrm{max}} < L$).

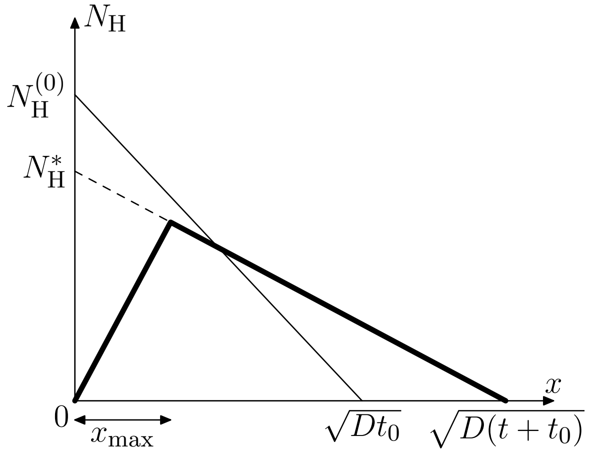


Figure 3.3 Approximate hydrogen concentration profile in diffusion-limited recovery of NBTI. The thin solid line represents the hydrogen profile immediately after stress, while the thick solid line represents that after a recovery of length t. The hydrogen concentration at the interface drops to ${N_{\mathrm{H}}}^{*}$.

Under the condition of $x_{\mathrm{max}} < L$ and $x_{\mathrm{max}} < L_0$, the integration (3.15) is evaluated, and the value of x_{max} is found

$$x_{\mathrm{max}} = \begin{cases} \left(1 - \frac{L}{2L_0}\right) L & L \le \left(2 - \sqrt{2}\right) L_0 \\ \left(3 - 2\sqrt{2}\right) L_0 + \left(\sqrt{2} - 1\right) L & L > (2 - \sqrt{2}) L_0. \end{cases}$$

Alam, in his paper[37], observed that the peak position x_{max} advances according to $\sqrt{\gamma D t}$ with $\gamma \approx 0.5$, which is in fact too slow initially and too fast as L approaches L_0. The passivated interface trap ΔN_{it} is then estimated with (3.16) or (3.17). In particular, at $t = t_0$, $L = L_0$, the peak position $x_{\mathrm{max}} = 0.586 L_0$ and ΔN_{it} is calculated to be between $\Delta N_{\mathrm{it,min}} = -0.414 {N_{\mathrm{it}}}^{(0)}$ and $\Delta N_{\mathrm{it,max}} = -0.586 {N_{\mathrm{it}}}^{(0)}$

of the previously generated interface states. It is easy to observe that at small t, $\Delta N_{\mathrm{it,max}}$ is a good estimation, while at large t, $\Delta N_{\mathrm{it,min}}$ is better. When time $t \to \infty$, $\Delta N_{\mathrm{it,min}}$ approaches $-{N_{\mathrm{it}}}^{(0)}$ implying total recovery of NBTI degradation.

3.2.1.3 Dynamic Stress

When the pMOSFET is under dynamic stress, there is alternating generation and recovery of interface traps. The generation and recovery processes are treated individually in the two proceeding sections. In this section we discuss the repeated process of generation and recovery. We focus on the long-term growth trend under dynamic stress with 50% duty-cycle, i.e. the gate voltage is high (V_{stress}) in half of the period, and is zero in the other half. Alam first discussed this problem through numerical simulation[37], we provide an analytical treatment as follows.

First we discuss the case of dynamic stress of very low frequency, such that in each stress-recovery cycle, the N_{it} generation and recovery processes are limited by diffusion. We consider the asymptote problem depicted in Figure 3.4. If one assumes the dynamic stress time t is much greater than the period T of the stress/recovery cycles, the net generation of N_{it} in each cycle is negligibly small, so that the hydrogen profile in the consecutive cycles are almost periodic. The periodic boundary condition is $N_{\mathrm{H}}(0) = 0$ during recovery and $N_{\mathrm{H}}(0) = N_{\mathrm{H0}}$ during stress. The value of N_{H0} is determined by the interface trap density N_{it} which is assumed to be almost constant during these cycles.

The hydrogen concentration, at the Si/SiO$_2$ interface, fluctuates between 0 and N_{H0}. The fluctuation damps down, however, as one moves away from the interface. At a distance of order $\lambda = \sqrt{DT}$, the fluctuation decays to a level with little importance. We thus approximate the hydrogen concentration at the stationary point $x = \lambda$ with a constant $N_{\mathrm{H}\lambda}$. The hydrogen profile in the region where $x > \lambda$ is approximated by the usual triangular shape. It is easy to see that

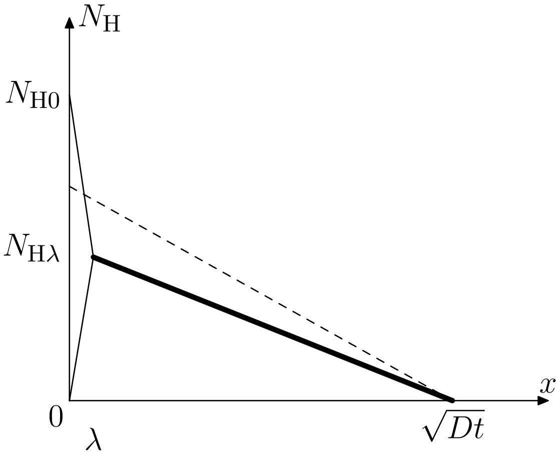


Figure 3.4 Approximate hydrogen concentration profile in diffusion-limited dynamic stress of NBTI. The dotted line represents the hydrogen concentration profile if the sample were under static NBTI stress. Under dynamic stress, however, the hydrogen concentration near the interface fluctuates between 0 and N_{H0}, while beyond a distance λ, the fluctuation decays to almost zero.

$$N_{it} \approx \frac{1}{2} N_{H\lambda} \sqrt{Dt}. \tag{3.18}$$

Since $t \gg T$, $\sqrt{Dt} \gg \lambda$. The remaining task is to find $N_{H\lambda}$ in relation to N_{H0} or N_{it}, which is determined by the diffusion in the region of $x < \lambda$.

As stated above, it is only necessary to solve the diffusion equation from cycle to cycle in the vicinity of the interface, and at distance $x = \lambda$, an approximate boundary condition $N_H = N_{H\lambda}$ exists. In order to maintain the approximate periodicity (in time) of the hydrogen concentration profile, it is easy to see, from the symmetry of the problem, that $N_{H\lambda} = \frac{1}{2} N_{H0}$ if the stress signal has 50% duty cycle. Therefore, assuming fast reaction, one has

$$k_F N_0 - k_R N_{it} \cdot 2 N_{H\lambda} \approx 0. \tag{3.19}$$

Together with (3.18), one obtains

$$N_{it} = \sqrt{\frac{k_F}{4 k_R} N_0} \, (Dt)^{1/4}. \tag{3.20}$$

Compared with (3.7), it is seen that the power-law (with 0.25 exponent) growth is preserved. However, the value of N_{it} in dynamic stress is $1/\sqrt{2} \approx 0.707$ of that in the static stress case. This is in agreement with the result of numerical simulation of the reaction-diffusion equations.

As one increases the frequency of the stress signal, the stationary point moves closer to the interface. As long as $\lambda = \sqrt{DT} \ll \sqrt{Dt}$, or the stress/recovery period is much less than the total stress time, (3.18) remains good approximation, and the N_{it} growth is essentially independent of frequency. Numerical simulation shows frequency independence[37], while experiments show either frequency independency[15] or weak dependency[36].

We further discuss the case of very high-frequency stress where the assumption of diffusion-limited process is no longer applicable. Note that the change of hydrogen concentration in one cycle decreases with increasing frequency, and becomes negligible at very high frequency. In that case, the hydrogen concentration at the interface is determined by the reaction process, and the hydrogen species released in the stress half cycle is simply

$$\left[k_{\mathrm{F}}\left(N_0 - N_{\mathrm{it}}\right) - k_{\mathrm{R}} N_{\mathrm{it}} {N_{\mathrm{H}}}^{(0)}\right] \frac{T}{2},$$

while hydrogen absorbed during recovery is

$$k_{\mathrm{R}} N_{\mathrm{it}} {N_{\mathrm{H}}}^{(0)} \frac{T}{2}.$$

Since the net generation in each stress/recovery cycle is the result of a balance between generation and recovery, and is negligible under a near-equlibrium condition when the stress time is long. One therefore has

$$k_{\mathrm{F}}(N_0 - N_{\mathrm{it}}) \approx 2 k_{\mathrm{R}} N_{\mathrm{it}} {N_{\mathrm{H}}}^{(0)}. \tag{3.21}$$

Combining with (3.18), one obtains an expression for N_{it} identical to (3.20). The frequency independence therefore is preserved even at high frequency.

3.2.1.4 Other Diffusion Species

Discussions prior to this all assume that the released hydrogen species diffuses in atomic form. However, other diffusion species are possible. Chakravarthi enumerated a list of such possibilities with numerical simulation[38], and later Alam provided an analytic treatment with more physical interpretations[39]. One particularly important case is that H_2 is formed from the released hydrogen atom and diffuses away from the interface. The power-law exponent in this case is 0.167, which matches the experimental results from the slow on-the-fly measurements. The mathematical derivation is similar to that presented in earlier sections, and is not repeated here. Readers are referred to the work of M. A. Alam, H. Kufluoglu, D. Varghese and S. Mahapatra for details[34, 39–40]. A few variants exists for the H_2 diffusion model. Early models assumes instantaneous reaction of atomic hydrogen to form H_2 at the interface, while Kufluoglu also considered co-diffusion of atomic and molecular hydrogen. The general conclusion is, however, unchanged.

3.2.1.5 Numerical Solution

Apart from the asymptotic solution in a few special situations described above, the general coupled equations of reaction-diffusion can be solved only numerically. As the reaction equation is non-linear, the coupled equations is solved with finite-difference method. As we intend to simulate large number of cycles, it is critical that conservation laws are strictly observed. Finite volume discretization is chosen to explicitly include conservation laws, and (3.2) and (3.3) are transformed to

$$\left({N_{\mathrm{it}}}^{t} - {N_{\mathrm{it}}}^{t-1}\right)/\Delta T = k_{\mathrm{F}}\left(N_0 - {N_{\mathrm{it}}}^{t}\right) - k_{\mathrm{R}}{N_{\mathrm{H}}}^{0,t}{N_{\mathrm{it}}}^{t} \tag{3.22}$$

$$\frac{{N_{\mathrm{H}}}^{0,t} - {N_{\mathrm{H}}}^{0,t-1}}{\Delta T}\frac{x_1 - x_0}{2} = D\frac{{N_{\mathrm{H}}}^{1,t} - {N_{\mathrm{H}}}^{0,t}}{x_1 - x_0} + \frac{{N_{\mathrm{it}}}^{t} - {N_{\mathrm{it}}}^{t-1}}{\Delta T} \tag{3.23}$$

$$\frac{{N_{\mathrm{H}}}^{i,t} - {N_{\mathrm{H}}}^{i,t-1}}{\Delta T}\frac{x_{i+1} - x_{i-1}}{2} = D\left(\frac{{N_{\mathrm{H}}}^{i+1,t} - {N_{\mathrm{H}}}^{i,t}}{x_{i+1} - x_i} - \frac{{N_{\mathrm{H}}}^{i,t} - {N_{\mathrm{H}}}^{i-1,t}}{x_i - x_{i-1}}\right) \tag{3.24}$$

where superscripts i and t denote space and time mesh indices, and ΔT is the time step. Newton's method is used to solve the system of non-linear equations.

It is easy task to adapt the above equations to the situation of H_2 diffusion, similar to the work of Kufluoglu[40]. The H_2-only diffusion variant is implemented in this work as an example.

3.2.2 Charge Trapping/De-trapping Model

It is well known that carriers can be trapped in an insulator as they transport through it. Charge trapping in dielectrics under Fowler-Nordheim stress was extensively studied in the 1980s. Trapped charge in gate dielectric of MOSFET would shift the threshold voltage V_{th} as well as the flat-band voltage V_{fb}. Positive charge trapping (hole trapping) would increase the V_{th} of p-MOSFET. On the other hand, as trapped hole emits from the trap and escapes from the insulator, the V_{th} of p-MOSFET decreases. The dynamic trapping and de-trapping of holes under alternating high/low gate voltages would lead to threshold voltage instability. This offers an alternative theory that accounts for the dynamic NBTI phenomena in p-MOSFET.

Huard and Denais suggested that the trapping and de-trapping of holes is the main cause of dynamic V_{th} instability observed in p-MOSFET[21, 30]. The author, however, proposed that the hole trapping is the cause of fast V_{th} transient only, while the slow V_{th} transient under NBTI stress is due to the generation and recovery of interface traps[31–32]. In other words, the proposed charge trapping/de-trapping theory is intended to complement the reaction-diffusion theory in the domain of short time scale. The necessity of this complementary theory is the topic of section 3.4, while reasons for the reaction-diffusion model to stay in the scene is presented in section 3.5.

We begin with two comments on the physical picture of the model, before dwelling on the mathematical details. In the first place, as Nissan-Cohen pointed out, it is important to distinguish between trap-generation during stress and the filling of existing traps. In this present study, we are concerned with the filling of pre-existing traps in the SiON dielectrics. The exact structure of the defect leading to hole trapping is not clear at present, and is not the focus of this study.

In the second place, it must be pointed out that in the studies performed in 1980s, the oxide thickness was typically above 10nm. Therefore, the carriers are captured to the traps as they travel in the conduction band or valence band, which is after they have tunneled through the triangular barrier (Fowler-Nordheim tunneling). In the present case of ultra-thin gate dielectrics with physical thickness less than 3 nm, the traps in the insulator directly exchange carriers with the conduction or valence bands through tunneling. Therefore, the charge trapping is not necessarily proportional to the carrier current through the dielectric, although a positive correlation between the two is obvious.

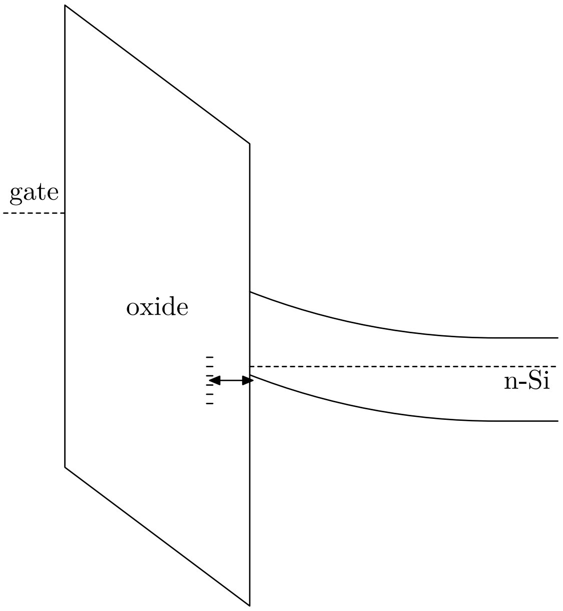


Figure 3.5 A schematic illustrating the trapping and de-trapping of holes in oxide traps. Several hole traps are assume to distribute in different energy levels. The double-sided arrow indicates the exchange of holes between the silicon substrate and the trap states, through trapping and de-trapping processes.

As much of the physical details is uncertain, it is wise to start from a very general form of trapping dynamics. Consider the hypothetical model depicted in Figure 3.5, where several types of pre-existing trap states are distributed at different energy levels in the oxide layer. Holes in the inversion layer in the silicon substrate can tunnel to the states in oxide and get trapped. Conversely, holes trapped at the oxide traps can escape and tunnel back the silicon substrate. We follow the work of Nissan-Cohen[27], and write the rate equations of trapping and de-trapping, respectively, as follows

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{1}{\tau_{\mathrm{C,p}}}(N_{\mathrm{ot,p}} - p) - \frac{1}{\tau_{\mathrm{E1,p}}}p, \tag{3.25}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{1}{\tau_{\mathrm{E2,p}}}p. \tag{3.26}$$

p = density of trapped holes cm^{-3}

$N_{\mathrm{ot,p}}$ = total density of hole traps in oxide cm^{-3}

$\tau_{\mathrm{C,p}}$ = hole capture time constant during stress

$\tau_{\mathrm{E1,p}}$ = hole emission time constant during stress

$\tau_{\mathrm{E2,p}}$ = hole emission time constant during recovery.

A few assumptions are used in these two equations. Firstly the population of trapped holes is modeled by a continuos number p, unlike in the theory of telegraph noise a discrete integer. In the study of NBTI the transistor size is typically not as small, and the discrete nature of the traps is not as important. Secondly, traps are assumed to be far from each other, and the capturing/emission processes from any two traps are independent of each other. Thirdly, a few different types of traps may co-exist, each with a different set of capturing/emission time constants. Again, the dynamics of different types of traps are independent of each other. Lastly, the distance from the trap to the gate electrode is assumed to be uniform. Therefore the V_{th} shift due to hole trapping is calculated from the simple aggregation of trapped holes in all types of traps.

We first consider one single type of traps, with a definite set of time constants. In principle, with given initial condition, successively solving (3.25) for each stress/recovery cycle, and repeating the procedure for m cycles yields the trapping/de-trapping behavior in p-MOSFET. However, as millions of such cycles are involved in this study, numerical error can accumulate to an unacceptable level with iterative calculation. Instead a close-form expression for the density of trapped holes after m cycles is derived as follows.

The hole trapping equation (3.25) can be solved analytically, and the solution can be written in matrix form

$$\begin{pmatrix} p \\ N_{\mathrm{ot,p}} \end{pmatrix} = \begin{pmatrix} A & \lambda\left(1-A\right) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ N_{\mathrm{ot,p}} \end{pmatrix} = T \begin{pmatrix} p_0 \\ N_{\mathrm{ot,p}} \end{pmatrix}, \tag{3.27}$$

with

$$\begin{aligned} \lambda &= \frac{\tau_{\mathrm{E1,p}}}{\tau_{\mathrm{C,p}} + \tau_{\mathrm{E1,p}}}, \\ A &= \exp\left[-t_{\mathrm{s}}\left(\frac{1}{\tau_{\mathrm{C,p}}} + \frac{1}{\tau_{\mathrm{E1,p}}}\right)\right], \end{aligned}$$

where t_{s} is the stress time, and p_0 the initial density of trapped holes before stress.

Similarly the solution to the de-trapping equation (3.26) is written as

$$\begin{pmatrix} p \\ N_{\mathrm{ot,p}} \end{pmatrix} = \begin{pmatrix} B & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ N_{\mathrm{ot,p}} \end{pmatrix} = D \begin{pmatrix} p_0 \\ N_{\mathrm{ot,p}} \end{pmatrix} \tag{3.28}$$

with

$$B = \exp\left(-t_{\mathrm{r}}/\tau_{\mathrm{E2,p}}\right), \tag{3.29}$$

where t_{r} is the recovery time, and p_0 the density of trapped holes before recovery.

With the trapping and de-trapping matrices T and D, the trapped charge after m stress/recovery cycles can be obtained from successively applying T and D to the initial condition

$$\begin{pmatrix} p \\ N_{\mathrm{ot,p}} \end{pmatrix} = (D \cdot T)^m \begin{pmatrix} p_0 \\ N_{\mathrm{ot,p}} \end{pmatrix}. \tag{3.30}$$

The matrix $(D \cdot T)$ can be diagonalized to

$$(D \cdot T) = P \cdot G \cdot P^{-1}, \tag{3.31}$$

with

$$\begin{aligned} P &= \begin{pmatrix} \frac{\lambda(1-A)B}{1-AB} & 1 \\ 1 & 0 \end{pmatrix} \\ P^{-1} &= \begin{pmatrix} 0 & 1 \\ 1 & -\frac{\lambda(1-A)B}{1-AB} \end{pmatrix} \\ G &= \begin{pmatrix} 1 & 0 \\ 0 & AB \end{pmatrix}. \end{aligned}$$

With the diagonalization, (3.30) can be evaluated from

$$\begin{pmatrix} p \\ N_{\mathrm{ot,p}} \end{pmatrix} = P G^m P^{-1} \begin{pmatrix} p_0 \\ N_{\mathrm{ot,p}} \end{pmatrix}. \tag{3.32}$$

Let the duty cycle be $\gamma = t_{\mathrm{s}}/(t_{\mathrm{s}} + t_{\mathrm{r}})$, then one finally has, after m stress/recovery cycles,

$$p = \left\{ \frac{\lambda (1-A) B}{1-AB} \left[1 - (AB)^m \right] \right\} N_{\mathrm{ot,p}} + (AB)^m p_0. \tag{3.33}$$

Similarly, one can calculate the trapped charge after $m + 1/2$ cycles, or m full cycles plus a stress half-cycle,

$$p = \left\{ \frac{\lambda (1-A) AB}{1-AB} \left[1 - (AB)^m \right] + \lambda (1-A) \right\} N_{\mathrm{ot,p}} + A (AB)^m p_0. \tag{3.34}$$

One observes that as stress time increases, $(AB)^m$ in (3.33) vanishes, and the density of trapped holes saturates. A trap only contributes to the V_{th} transient when the stress time is comparable to its trapping and de-trapping time constants. As stated earlier, it is assumed that the hole traps in the oxide have a wide-spreading spectrum of trapping and de-trapping time constants. The distribution functions of the time constants are therefore the critical parameter in the charge-trapping model.

The correlation between τ_{E1}, τ_{E2} and τ_{C} for a particular trap can be quite complicated in reality. For simplicity, it is assumed that for any trap with trapping time constant τ_{C}, the de-trapping time constants during stress and recovery are $\tau_{\mathrm{E1}} = a\tau_{\mathrm{C}}$ and $\tau_{\mathrm{E2}} = b\tau_{\mathrm{C}}$, respectively, where a and b are constants for all types of traps. With a hypothetic probability density function for τ_{C}, it is therefore straightforward to calculate the trapped charge for each type of traps, and obtain the total number of trapped charge from the summation. It will be shown in section 3.4 that this simple relationship has enable us to re-produce most of the characteristics observed in experiments. The distribution of τ_{C} is shown in Figure 3.6.

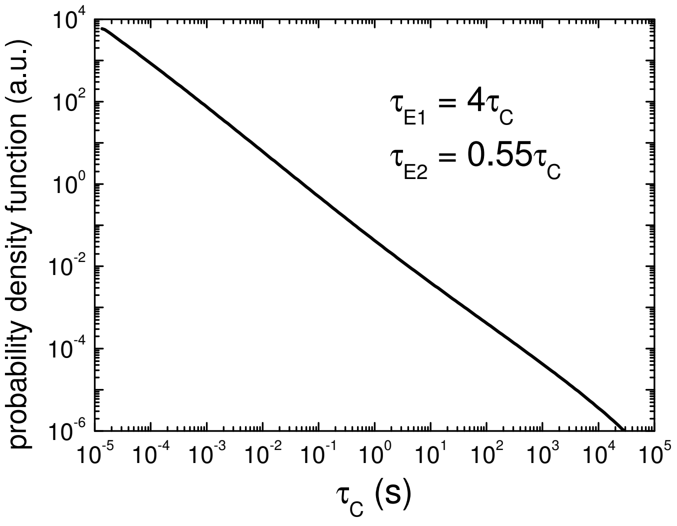


Figure 3.6 Distribution function of τ_{C}, τ_{E1} and τ_{E2} used in this study.

To summarize, in this section, the two theories to NBTI are presented. Analytical treatment is provided to some important cases of the reaction-diffusion model, for greater insights to the physical processes. Appropriate numerical simulation algorithms are laid out briefly for the two theories, respectively, to calculate the theoretical predictions accurately.

3.3 Review on Measurment Techniques

Before 2003, the characterization of NBTI largely followed the paradigm set by the initial researchers in the 1960s and 70s, which consists of a series of measure–stress–measure cycles. The waveform of gate and drain bias voltage during an example measure–stress–measure cycle is shown in Figure 3.7. In this simple example, the measurement phase is an V_g sweep with a constant V_d, and the linear-region transfer characteristics is obtained. From the measured I_d–V_g curve, one can extract transistor parameters such as the linear-region threshold voltage V_{th}, the sub-threshold swing SS, and indirectly calculate the interface-state density D_{it}. In many cases, researchers also perform other measurements to extract V_{th}, D_{it} and other device parameters. Popular choices include charge pumping measurement for D_{it} extraction[41], DCIV measurement for D_{it}[42], drive current $I_{d,\text{sat}}$ at $V_g = V_d = V_{\text{dd}}$, transconductance g_m, terminal capacitances (notably the miller capacitance C_{gd}), and carrier mobility μ. When these additional measurement procedures are added, the corresponding voltage waveforms on gate, drain, and sometimes source and substrate as well, appear much more complicated, but conceptually remain similar to that in Figure 3.7. Note that the most basic parameter of interest in NBTI, the threshold voltage V_{th}, is extracted from the $I_d - V_g$ measurement, and only the segment where $V_g \approx V_{\text{th}}$ is critical to this purpose (highlighted with thick red lines). In a typical wafer-level testing setup, semiconductor parameter analyzers such as the Agilent 4156 are used to perform the stress and measurements. The typical measurement rate for such systems is 20 ms per data point, which adds up to about 1 s for a complete $I_d - V_g$ sweep. The switching time from stress to measurement takes anything from several hundred milliseconds to several seconds, depending on the equipment and setup. Similarly, each additional measurement besides the basic $I_d - V_g$ sweep adds some delay time and measurement time of its own. Furthermore, since NBTI stress is usually performed at elevated temperature, it is sometimes required to cooled down the samples to room temperature before measurement, which leads to longer delay time in the range of several minutes to tens of minutes. Despite the long delay time t_d and measurement time t_m, it is traditionally believed that these delays introduces little error to the NBTI characterization.

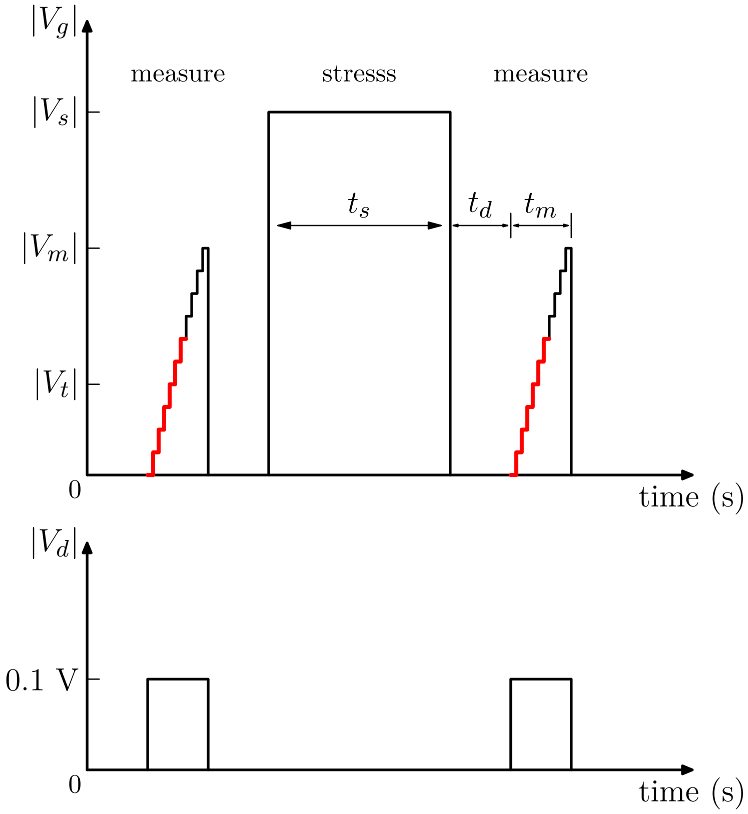


Figure 3.7 The waveform of the gate and drain voltage in a simplest NBTI measure–stress–measure cycle.

Since the discovery of strong dynamic recovery in NBTI, the error due to measurement delay is realized. NBTI degradation measured after the delay time is smaller than the actually degradation. New measurement techniques were sought to eliminate the delay-induced error. As shown in Figure 3.8, different techniques yield drastically different measurement results. We describe a few such measurement techniques below, and analyze their respective strengths and problems in the characterization of NBTI with dynamic recovery.

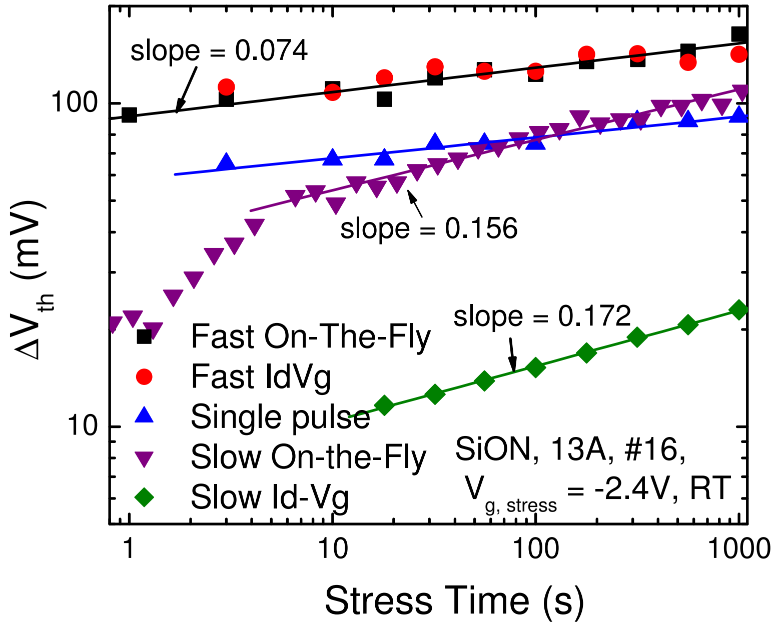


Figure 3.8 ΔV_{th} for identical pMOSFETs is measured with five different measurement techniques, which yield very different NBTI results. The fast $I_d - V_g$(100 ns measurement time) and the fast on-the-fly measurements produce the same and reliable results. All other methods show underestimated ΔV_{th}. The single pulse measurement[43] used in this work requires a 300 ns pulse due to long settling time, which can be improved with proper (but costly) RF measurement setup.

3.3.1 Fast $I_d - V_g$ Measurement

One obvious solution is to reduce the delay time t_{d} and measurement time t_{m}, so that minimum NBTI recovery occurs. As discussed in Chapter 2, the semiconductor parameter analyzer can achieve, at best, delay time close to a hundred milliseconds. On the other hand, the fast measurement setup consists of pulse generator, transimpedance amplifier and oscilloscope is able to complete an I_d–V_g sweep within 100 ns. The gate voltage waveform of NBTI test using fast measurement is shown in Figure 3.9. With a short gate voltage pulse from the pulse generator, an initial $I_d - V_g$ sweep is performed prior to NBTI stress on a fresh device. From this $I_d - V_g$ curve the initial V_{th} is extracted. The pulse generator then toggle its polarity, so that the gate voltage rises to the stress voltage, until at the end of the stress period, a measurement pulse in the downward direction is triggered. A small drain bias voltage is always present throughout the experiment to avoid synchronization between the gate voltage and drain voltage waveform. The time period critical to the V_{th} extraction is highlighted with thick red lines. Some NBTI recovery occurs as the gate voltage falls from the stress voltage, and shall be minimized by reducing t_{m}. The minimum measurement time t_{m} achieved is 100 ns. The limit of measurement speed and the sources of measurement error have been analyzed in Chapter 2, and is not repeated here. To test the effectiveness of short measurement time in reducing NBTI degradation, fresh transistors are stressed for 1 second, and the threshold voltage shift is measured with varying measurement time t_{m}. In Figure 3.10, it is seen that with decreasing t_{m}, the measured ΔV_{th} increases. No sign of saturation is observed even when t_m is reduced to 100 ns. This suggests that shortest possible measurement time is required to adequately study NBTI with minimum recovery-induced error.

Typical measurement result using the fast measurement technique is shown in Figure 3.8. It is seen that the fast $I_d - V_g$ method, along with the fast on-the-fly method (to be described later), yield the highest ΔV_{th}, and agree with each other. All other methods show underestimated V_{th} degradation, arising from different sources of errors.

3.3.2 Slow On-the-Fly Measurement

An alternative measurement technique that aims to tackle the delay-induced error is the *on-the-fly measurement*. A few variants of the on-the-fly technique have

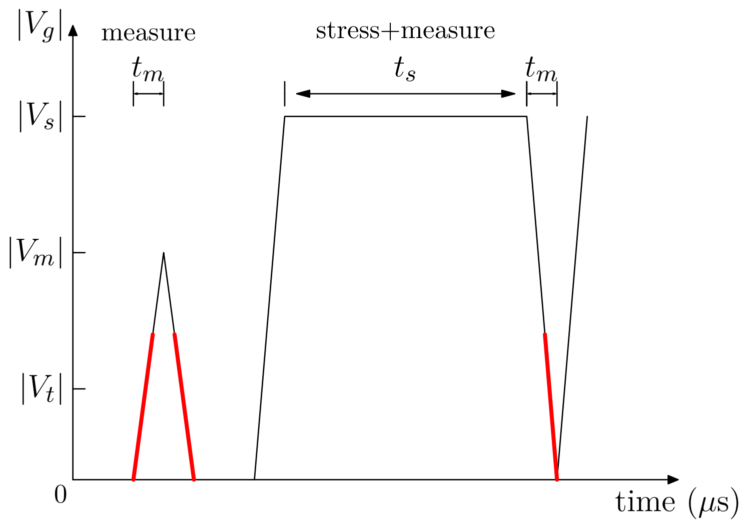


Figure 3.9 The waveform of the gate voltage using the fast measurement technique.

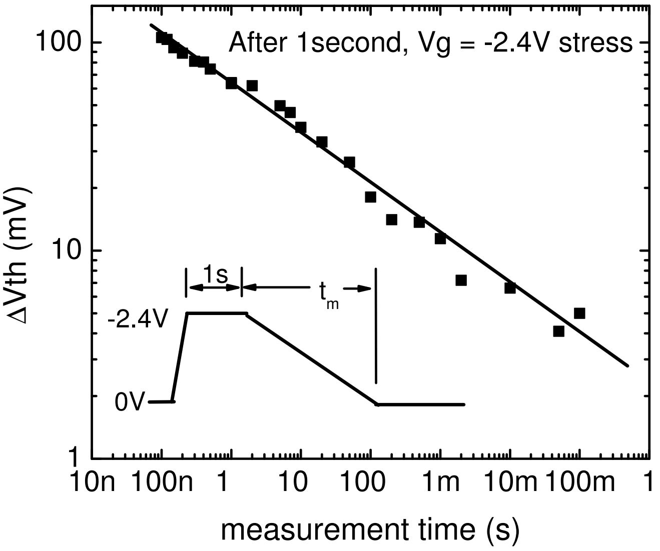


Figure 3.10 The V_{th} shift measured after 1s stress, with different measurement time t_{m}.

been proposed[20–22, 24, 34]. The common feature among all variants is that the stress voltage is always applied on the gate, and that the degradation of drain current (at stress voltage) is measured "on the fly". Since the stress voltage is not removed from the gate, the on-the-fly method is claimed to be free from the fast recovery of NBTI. It soon became popular as it can be performed on common (slow) semiconductor parameter analyzers, without additional equipment requirement. To distinguish it with a variant of on-the-fly technique that uses fast measurements which will be proposed in the next section, we sometimes also explicitly call it the *slow on-the-fly* method.

Described here is one representative implementation of the slow on-the-fly measurement. The gate voltage wave form is shown in Figure 3.9, and a small drain bias is always present. An initial I_d–V_g curve is measured with V_g swept to the stress voltage. Both the drain current I_{d0} and the transconductance g_m at $V_g = V_{\mathrm{stress}}$ are recorded. During the following electrical stress, the drain current is continuously sampled at $V_g = V_{\mathrm{stress}}$, and the change of drain current $\Delta I_d = I_d - I_{\mathrm{d0}}$ is recorded. The threshold voltage shift is then calculated from

$$\Delta V_{\mathrm{th}} = -\Delta I_d / g_m. \tag{3.35}$$

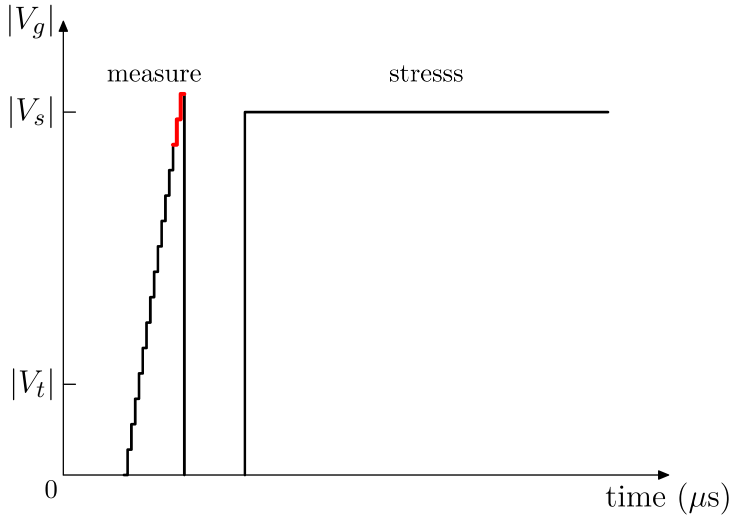


Figure 3.11 The waveform of the gate voltage using the *on-the-fly* measurement technique.

One implicit assumption in (3.35) is that NBTI only causes the I_d–V_g curve to translate horizontally by ΔV_{th}. This assumption is obviously true for an ideal MOSFET with constant mobility, free of parasitic resistance. Some arguments are provided below to show that it remains good approximation in real devices with finite source/drain resistance and E-field dependent mobility. Consider a

MOSFET with relatively small source/drain resistance r_{sd}, the drain current of the intrinsic MOSFET is given as a function of the gate over-drive $V_g - V_{\mathrm{th}}$. Since only the region with large V_g near the stress voltage is important in on-the-fly measurement, the voltage drop on the source resistance is considerably smaller than the gate over-drive $V_g - V_{\mathrm{th}}$ and can be ignored. The parasitic resistance therefore only contributes to the total channel resistance. It can be shown that, in this case, the drain current remains a function of $V_g - V_{\mathrm{th}}$, which means that a V_{th} shift only cause a horizontal translation of the $I_d - V_g$ curve. Similar arguments can be made for the case of E-field dependent mobility[24, 44].

Another error in the on-the-fly method described above arises from using the g_m before stress in (3.35). As illustrated in Figure 3.12, (3.35) only included the first order term in the Taylor expansion of I_d, and leads to error as ΔI_d (ΔV_{th}) increases. An estimate of the relative error in ΔV_{th} can be derived, using the truncation error of Taylor expansion, to be

$$\varepsilon_{\mathrm{r}} = \frac{\mathrm{d}^2 I_d / \mathrm{d}{V_g}^2}{2\,\mathrm{d}I_d / \mathrm{d}V_g} \cdot \Delta V_{\mathrm{th}}. \tag{3.36}$$

This error could be corrected if one measures the transconductance after stress as well as before stress, but the measurement procedure becomes more complicated[21]. For the typical device and stress condition used in this study, the error calculated from (3.36) is less than 10% at $V_{\mathrm{stress}} = -2.4$ V, and $\Delta V_{\mathrm{th}} = 140$ mV. As will be shown later, this is not the major source of error in on-the-fly measurement.

One major source of error in on-the-fly measurement, not recognized in the NBTI community, is the NBTI degradation during the pre-stress measurement of I_{d0} and g_m. If this initial measurement is slow, some V_{th} degradation occur under the high gate voltage ($V_g = V_{\mathrm{stress}}$). The measured I_{d0} is lower than the true value, thus the V_{th} shift after the subsequent stress is underestimated. Typical measurement result using the on-the-fly method is shown in Figure 3.13, whereas the comparison with other measurement techniques is show in Figure 3.8. Obvi-

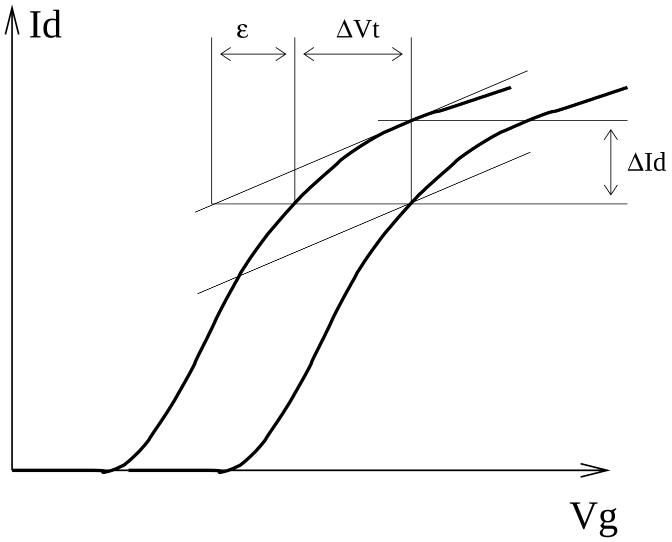
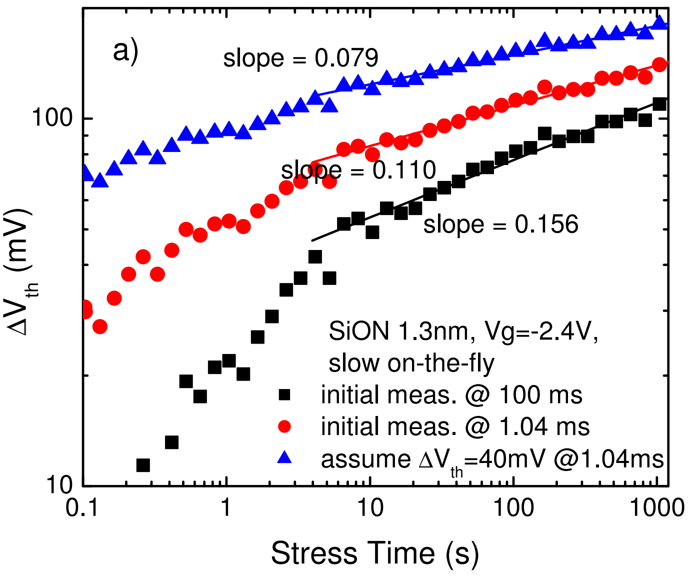


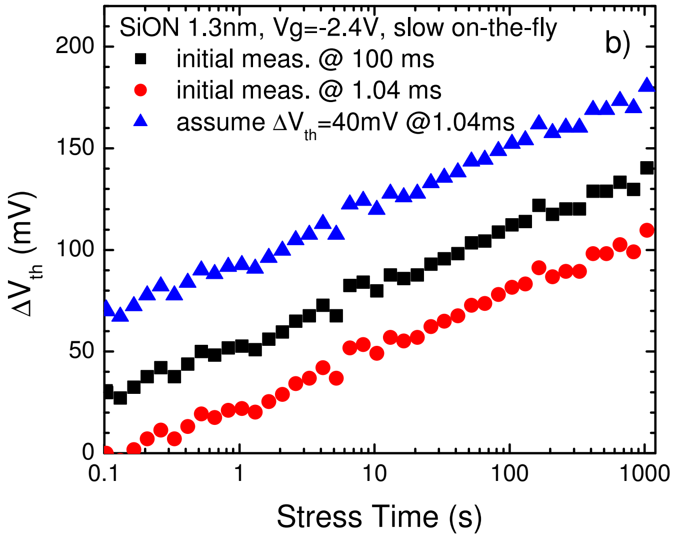
Figure 3.12 Calculation of ΔV_{th} using (3.35) leads to error ε_{r} when ΔV_{th} is large, due to the V_g dependent transconductance.

ous underestimation of NBTI degradation is seen compared with the fast $I_d - V_g$ method.

At short stress time, the stress-induced degradation is not much larger than that in the pre-stress measurement. Therefore, the underestimation of ΔV_{th} is more serious, which results in a kink before 10 sec as observed in the log-log plot as shown in Figure 3.13(a). Similar kink is observed in other reports using on-the-fly measurement[34], similar power-law slope of 0.156 is observed as well. One obvious way to reduce this error is to reduce the measurement time of I_{d0}, so that minimum degradation occurs. With Agilent 4156, 1 ms measurement time is possible if only the I_{d0} at stress voltage is measured (g_m is estimated separately). As shown in Figure 3.13, when the initial measurement time is reduced from 100 ms to 1 ms, the kink becomes less obvious, and the power-law slope decreases. Furthermore, assuming $\Delta V_{\mathrm{th}} = 40$ mV in the 1 ms initial measurement at $V_g = -2.4$ V, one can compensate this initial degradation by adding 40 mV to the subsequent ΔV_{th} values. As a result, the kink is further weakened and the slope reduced. The reduction of V_{th} degradation during the initial pre-stress measurement is therefore the key to improve the on-the-fly technique. In addition, slow measurement under-



(a)



(b)

Figure 3.13 (a) Conventional (slow) on-the-fly suffers from degradation during the measurement of the inital $I_{\mathrm{d}0}$, and estimates ΔV_{th}. When plotted in log-log scale, a characteristic kink is observed in short time. The slower the initial (pre-stress) measurement is, the stronger the kink is. If one compensates for the degradation during the initial (pre-stress) measurement by adding an assumed amount to the value of ΔV_{th}, the kink weakens (triangles). (b) When the same data is plotted in semi-log scale, straight lines are observed showing logarithmic relationship.

estimates V_{th} degradation and yields an errorneous slope in the log-log plot, which is clearly demonstrated by the very different slope values obtained with different initial measurement time.

Therefore, a correct characterization technique for NBTI has to address both the fast recovery after stress and the fast degradation before the stress. The above described on-the-fly method solved the first problem, but not the second. It is believed that fast measurement (in microseconds of faster) has to be used to simultaneously fix both problems.

An interesting observation arises when the same data are plotted on a semi-log scale as in Figure 3.13(b). A logarithmic relation is seen as [21] reported, despite the different initial measurement time. It is easy to see that for processes with logarithmic time dependence, an underestimation of the initial value as in the on-the-fly method does not change its logarithmic behavior. This might suggest that a semi-logarithmic plot is more appropriate to study NBTI in the presence of fast transients, but this possibility is not critical to the present study, and is not pursued further.

3.3.3 Fast On-the-Fly Measurement

To circumvent this problem with slow initial measurement in the on-the-fly measurement, we propose that the pre-stress I_{d0} and V_{th0} to be measured by the fast $I_d - V_g$ technique. Very short measurement time of 1 μs or less is preferred to minimize degradation in measuring I_{d0}. The drain current I_d is then monitored on-the-fly during stress, using either the circuit for fast measurement (as in this study), or the traditional parametric analyzers.

The waveform of the gate voltage in *fast on-the-fly* measurement is similar to that of the slow on-the-fly measurement (see Figure 3.11), except for the much shorter pulse width in the initial $I_d - V_g$ measurement. In this study, both the initial $I_d - V_g$ measurement and the subsequent monitoring of I_d use the fast transimpedance amplifier described in Chapter 2, and the measurement pulse is 200 ns wide. The measured ΔV_{th} is plotted in Figure 3.8, which shows good agreement with the fast $I_d - V_g$ method.

3.3.4 Single-Pulse Measurement

As discussed in Chapter 2, the main limitation to the measurement speed of the fast $I_d - V_g$ technique is the stringent requirement on synchronization. The phase delay of the transimpedance amplifier prevents accurate I_d–V_g measurement faster than 100 ns. On the other hand, the single-pulse measurement does not require synchronization, and can operate with shorter pulses[43]. A scheme for NBTI characterization is devised based on the single-pulse technique, and the gate voltage waveform is shown in Figure 3.14.

A pre-stress measurement pulse is applied to the gate with pulse height V_{meas}, and the drain current I_{d0} is measured at $V_g = V_{\text{meas}}$. During stress, short measurement pulses are inserted in the opposite polarity (down to V_{meas}) to measure the degraded I_d. Similar to the on-the-fly method, the threshold voltage shift is calculated from ΔI_d. This method is different from the fast on-the-fly method in that the drain current is measured at a lower gate voltage V_m. As discussed in section 3.3.2, the measurement error of on-the-fly method increases with decreasing g_m at high gate voltages. The single-pulse measurement thus offers some promise to improve the fast on-the-fly method in accuracy and the fast $I_d - V_g$ method in speed.

In this study, the same transimpedance amplifier approach is used in the single-pulse measurement, though alternative setup with sense resistor is possible[43]. However, long settling time (> 100 ns) in the amplifier output is observed with the current setup. Consequently, I_d readings are taken at 200 ns for accurate and repeatable results. This long waiting time obviously introduced error due to recovery. As shown in Figure 3.8 on page 53, the single-pulse method also underestimates the ΔV_{th}. It should be pointed out that, 3% or better settling is needed for accurate extraction of ΔI_d. The setup used in this study is not designed for RF performance, and the long settling time should be improved with proper RF setup.

There is an alternative single pulse measurement method[45–46], using an operational amplifier to find the threshold voltage defined by a constant drain current. The measurement time of 1 μs has been achieved. However, during its switching from stress mode to measurement mode, the operational amplifier is in open-loop, and likely saturated. Among other consequences, measurement time is limited by the saturation recovery time of the amplifier, and is difficult to scale down further.

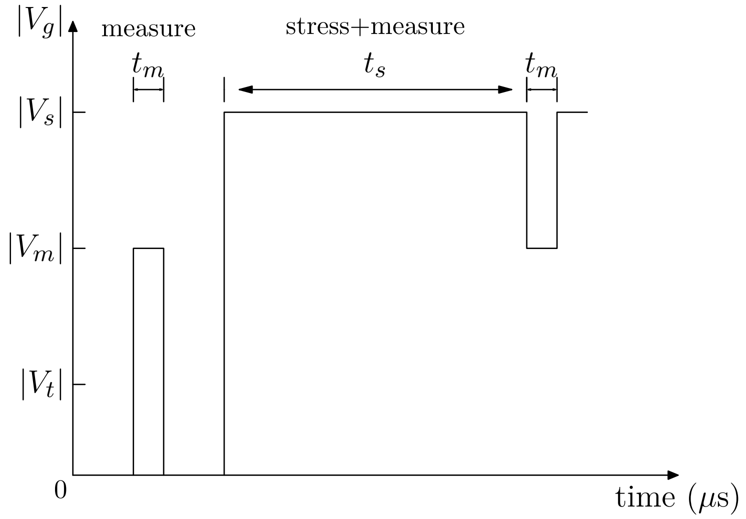


Figure 3.14 The waveform of the gate voltage using the single-pulse measurement technique.

3.3.5 Discussion and Summary

In short, the apparent inconsistencies in the NBTI literature is a result of imperfect measurement setups. The proposed power-law exponent of 0.16 is specific to a particular measurement setup instead of universal. The fast on-the-fly method and the fast I_d–V_g method are less affected by the V_{th} transient artifact, and are the better choices for NBTI characterization.

It may be important to point out at this moment, that the traditional slow measurement is still valuable, in the author's opinion. The recovery of NBTI after the removal of stress voltage, which leads to measurement error, is initially very fast, but becomes much slower after 1 second or so. This prompts one to differentiate the fast-recovering part of NBTI from the slow-recovering part. As will be shown in later sections, there is significant distinction between the two, and separate discussion is needed. We shall call the former component the fast transients in NBTI or the *fast component* of NBTI, and the latter the *slow component* of NBTI. The slow measurement, with measurement delay close to 1 s, is used to measure the slow component. The fast component, strictly speaking, should be the difference (in ΔV_{th}) between the fast and slow measurements, but the fast measurement is used as a close approximate.

3.4 Hole Trapping and the Fast Transient Component in NBTI

When the dynamic recovery of NBTI was discovered, the reaction-diffusion model, as the most accepted theory for NBTI, was naturally chosen to interpret the new findings[37], and achieved much success. However, as researchers realized the importance of measurement error due to the fast recovery of NBTI during measurement delay, and attempted to reduce or eliminate this error, the new experimental results posed some challenges to the reaction-diffusion theory.

In the first place, all new measurement results show that the power-law exponent, after reducing the effect of measurement delay, is much smaller than the traditional value of 0.25. The slow on-the-fly measurement shows a power-law slope of $0.15 - 0.16$, while the fast $I_d - V_g$ measurement gives a smaller slope of $0.07 - 0.10$ if plotted in log-log scale. As discussed in section 3.2, a slope of 0.167 is expected in the reaction-diffusion model, if one assumes that molecular hydrogen (H_2) instead of atomic hydrogen is the diffusion species. This agreement with the results of on-the-fly measurement forms the basis of the reaction-diffusion theory[34, 39].

We demonstrated in section section 3.3 that the slow on-the-fly measurement over-estimated the power-law slope, and a more accurate slope in the range of $0.07 - 0.10$ was obtained from the fast $I_d - V_g$ measurement. A new puzzle thus arises to the reaction-diffusion model. Within the framework of the model, it is still possible to assume more complex diffusion species consisting of three or more hydrogen atoms, which would further reduce the power-law exponent from 0.16. This, however, is by no means an attractive theory unless there is strong motivation in favor of this assumption.

In the second place, fast recovery of NBTI is observed immediately after stress is removed. This is not expected in a diffusion controlled process. The author believes that this is a more fundamental challenge to the reaction-diffusion interpretation, which leads to the belief that a different mechanism is responsible to part of the dynamic NBTI. In the following we present a few arguments to demonstrate this inadequacy of the reaction-diffusion model, and the necessity of a new model to account for the fast component in dynamic NBTI. We further show that the hole-trapping model is a good candidate to the desired new model. The observation of fast recovery of NBTI forms the motivation and basis of discussion to follow, while a few related observations support or complement the main argument.

3.4.1 Fast Recovery and Dependence on Stress Time

Figure 3.15 shows the recovery of V_{th} after NBTI stress is removed, as measured using the fast $I_d - V_g$ technique. One first notices that, even after long time stress (1000 s), the V_{th} shift still shows dramatically fast recovery (more than 60%) within the first second after the stress is removed. This fast recovery is however not expected in the reaction-diffusion model. As illustrated in Figure 3.16, an approximate triangular hydrogen concentration profile is gradually set-up during the 1000-second stress, and the total amount of the released hydrogen species (area under the curve) must equal to the total amount of interface traps, which is in turn proportional to the V_{th} shift. After the stress is removed, if the ΔV_{th} were to recover by 60% within the first 1 second, 60% of the hydrogen species must have diffused back to the interface in this short time to react with the interface traps. In other words, one must assume a much faster backward diffusion than the forward diffusion, which is against one's understanding about a diffusion process.

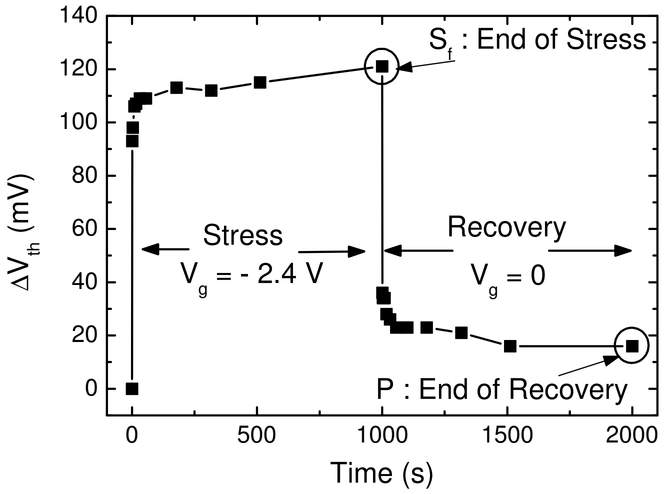


Figure 3.15 Using fast $I_d - V_g$ technique, the dynamics of ΔV_{th} under stress/recovery cycles is studied ($f = 1/2000$Hz is shown). S_{f} denotes the ΔV_{th} at the end of the stress half-cycle, and P denotes that at the end of recovery half-cycle.

In fact, in the reaction-diffusion model, as hydrogen species diffuse out to a greater distance from the interface during a longer stress time, it accordingly takes longer time to diffuse back after stress is removed. Therefore, the recovery after a 1000-second stress is expected to be appreciably slower than the recovery after a 1-second stress. However, the observed fast recovery does not show dependence on the stress time, as shown in Figure 3.17, for the case of both V_{th} recovery and $I_{d,\text{lin}}$ recovery.

To verify the above arguments, the reaction-diffusion equations are solved numerically for both cases of H^0 and H_2 diffusion[37–38], as described in section 3.2.

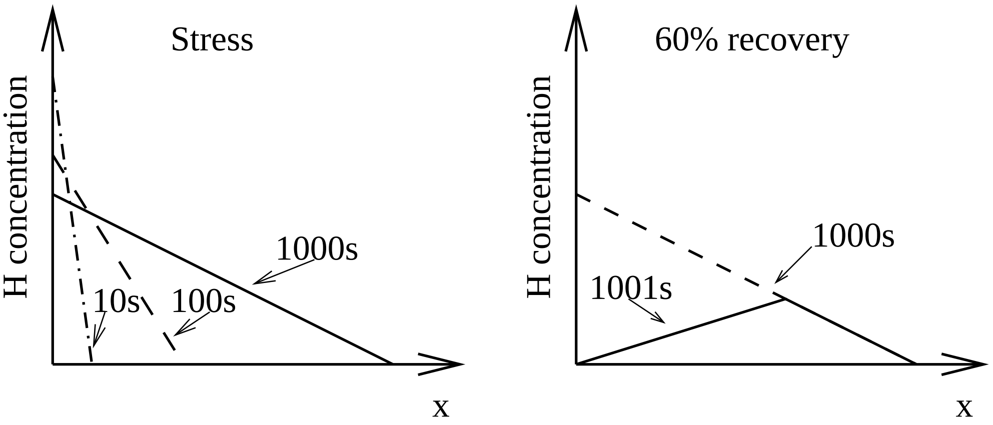


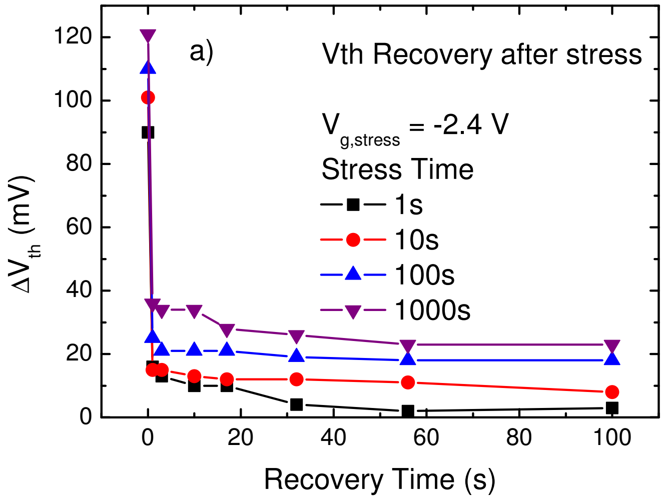
Figure 3.16 Hypothetical hydrogen profile necessary to explain the fast V_{th} recoverying after long time (1000s) stress within the reaction-diffusion framework. For 60% of the ΔV_{th} to recover within 1 sec, 60% of the hydrogen species would have to diffuse back to the interface in this short time, traveling a distance that took $\sim$ 1000 s in out-diffusion.

It is found that after being stressed for long enough time, where hydrogen diffusion is the limiting process, the time for recovery is closely related to the stress time t_{stress}. As shown in Figure 3.18, recovery mainly occurs between $0.1 \times t_{\mathrm{stress}}$ to $10 \times t_{\mathrm{stress}}$, and it takes a time approximately equal to t_{stress} for 50% recovery. This agrees with our analytical results in section 3.2, and the above arguments. However, this obviously contradicts with the fast V_{th} recovery observed experimentally.

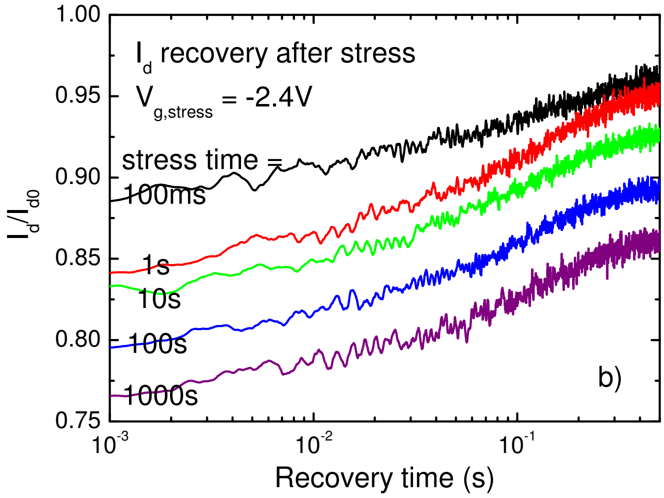
On the other hand, simulation with the hole-trapping model equations is qualitatively consistent with the observed fast recovery for all stress times, as shown in Figure 3.19. When holes get trapped in a certain kind of oxide trap states during stress, their de-trapping time-constant during recovery is determined by the trap states, but does not depend on how long holes have been trapped in the hole traps. This marks a fundamental difference between the dynamics of trapping/de-trapping process and the reaction-diffusion process, which suggests that the experimental data favors the charge trapping model on a fundamental instead of coincidental basis.

3.4.2 Effect of Measurement Delay

The reaction-diffusion model simulation predicts that the relative error caused by



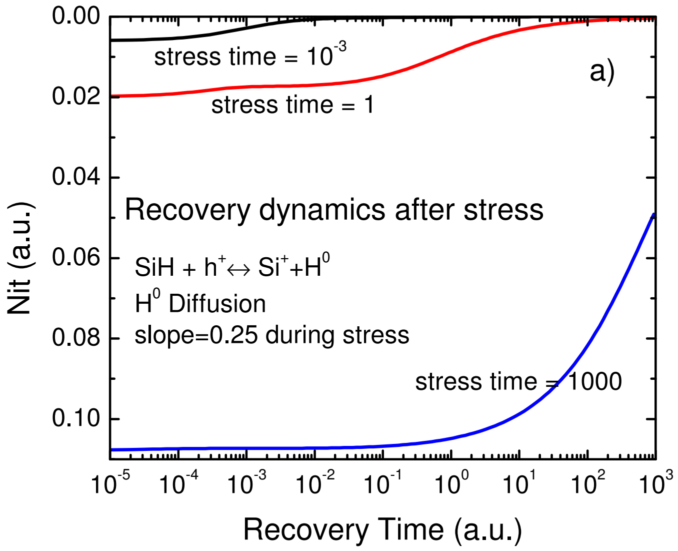
(a)



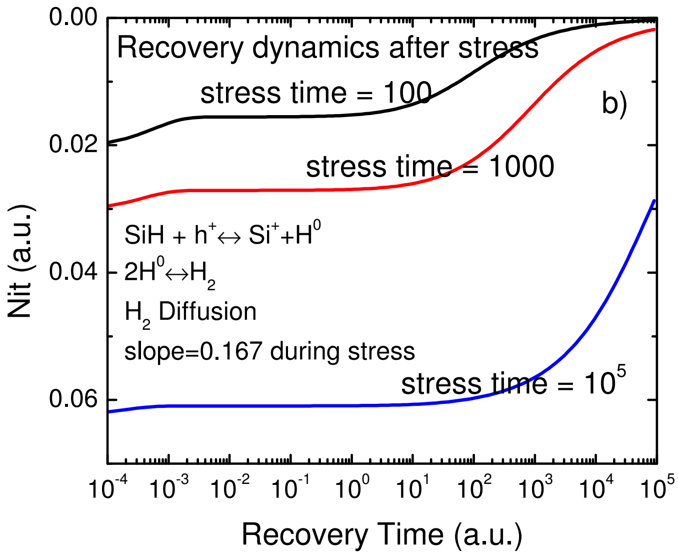
(b)

Figure 3.17 After stress is removed, the majority of ΔV_{th} recovers in a very short time. (a) Even after 1000 s stress, $> 60\%$ of the ΔV_{th} recovers within 1 s after the stress is removed. (b) Similar fast recovery in drain current is seen after stress is removed, despite of the long stress time.

measurement delay diminishes after long time stress, as shown in Figure 3.20.



(a)



(b)

Figure 3.18 According to the reaction-diffusion model ((a)H^0 diffusion and (b) H_2 diffusion), after stressed for a time t_{stress} long enough so that diffusion dominates, recovery occurs mainly between $0.1t_{\text{stress}}$ and $10t_{\text{stress}}$ after stress is removed.

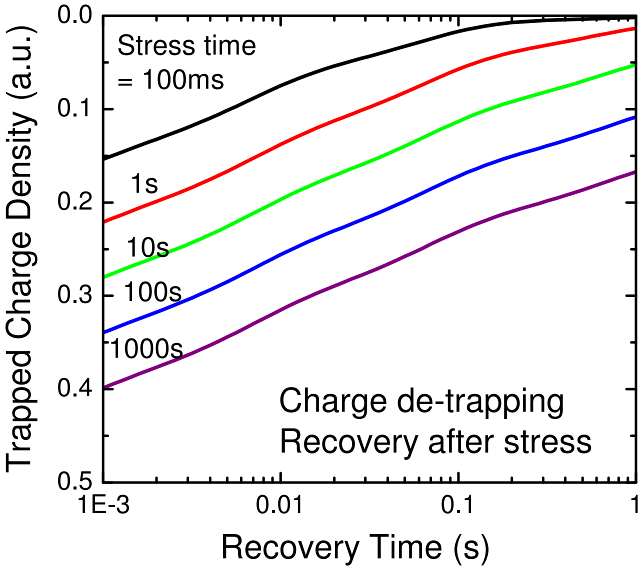


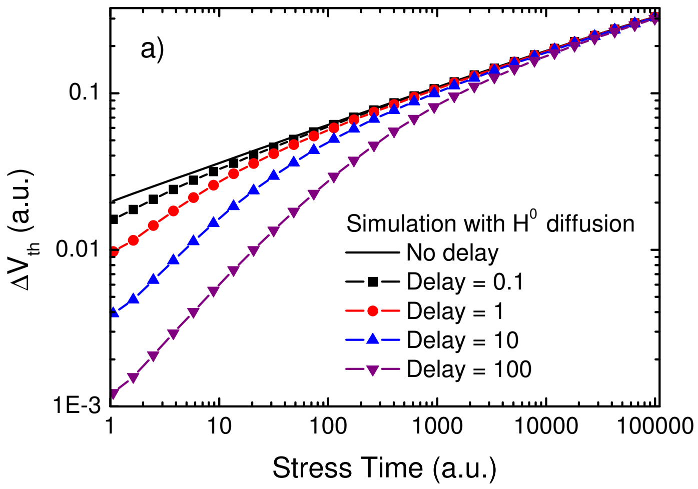
Figure 3.19 Simulated recovery process according to the hole trapping/de-trapping model. Fast V_{th} recovery is expected even for long time stress, which qualitatively agrees with Figure 3.17.

When the stress time exceeds 10 times the measurement delay, the relative error due to delay becomes negligible. This is in contrast with the experimental data shown in Figure 3.21, where a short delay causes large error even after long-time stress. The large delay-induced error is closely related to the fast recovery property described earlier, and it is easy to see that the reaction-diffusion model would not explain it.

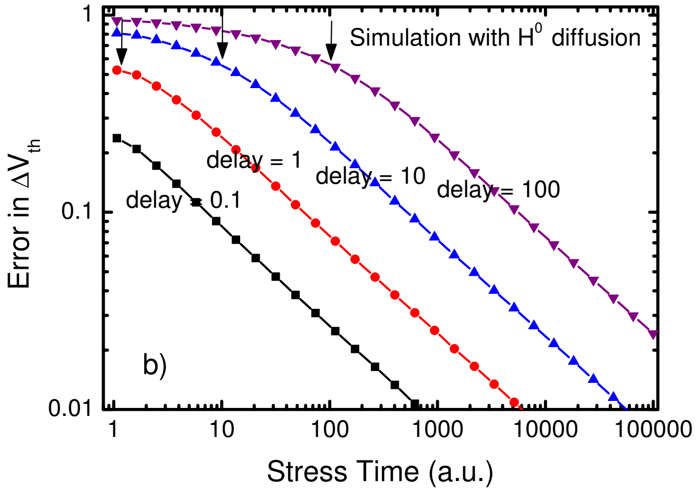
The simulation with the hole trapping model, on the other hand, agrees with the experiments qualitatively, as shown in Figure 3.22. The absolute value of delay-induced error is largely invariant of the stress time, because the amount of recovery during the delay is mainly determined by the amount of traps with de-trapping time constants less than the delay time.

3.4.3 Frequency Dependence Under Dynamic Stress

The reaction-diffusion equations are solved repeatedly to simulate the V_{th} degra-



(a)



(b)

Figure 3.20 (a) From the R-D model, ΔV_{th} with measurement delay is simulated. (b) The error due to delay diminishes when stress time is much greater than the delay time.

dation under dynamic stress, as described in section 3.2. The amplitude of V_{th} transient between the end of stress half-cycle ($\mathrm{S_f}$ point in Figure 3.15) and the

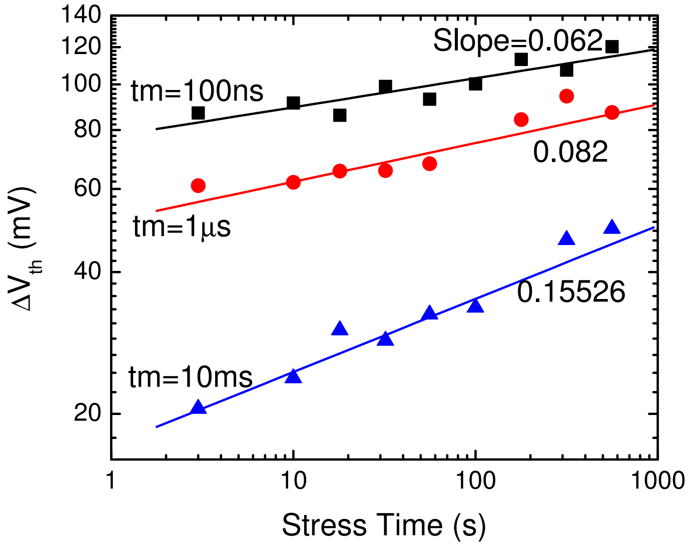


Figure 3.21 As measurement time t_{m} increases, V_{th} shift is underestimated, and the power-law exponent increased. A short delay leads to large underestimation of V_{th} even after 1000 s of stress.

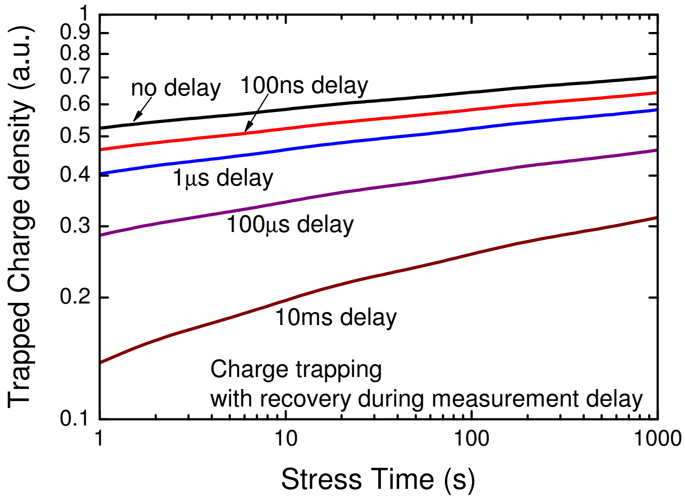


Figure 3.22 From the trapping/de-trapping model, the trapped charge (and thus ΔV_{th}) is simulated with delay. A short delay causes large error, which persists even after long-time stress.

end of recovery half-cycle (P point) would diminish when dynamic stress time exceeds 100× of the period of one stress/recovery cycle (Figure 3.23). Again, this is a direct result of the diffusion-limited character of this process. After long time stress, V_{th} shift under dynamic stress shows very little oscillation, and converges to a trend-line that is 1) parallel to that under static stress, and 2) frequency independent, agreeing with our prior analytical solution.

However, as shown in Figure 3.24, one experimentally observes significant transient amplitude after a 1000 s, 100 kHz dynamic stress, which is drastically different from the predictions from the R-D model. On the other hand, the hole-trapping model shows agreement with frequency-dependence experiments as detailed in the previous work[31, 47], and shown in Figure 3.25.

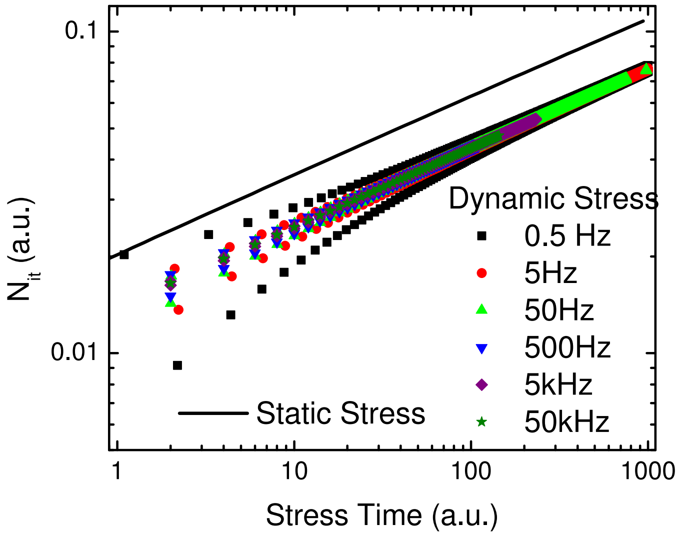


Figure 3.23 From R-D model, ΔV_{th} under dynamic stress is simulated. The S_{f} point and P points (see Figure 3.15) should merge to a frequency-independent trend-line after long time stress.

3.4.4 Discussions and Summary

We presented experimental results and physical considerations to show that the reaction-diffusion model is not adequate to account for the observed fast transients in NBTI, or more specifically, the fast recovery of NBTI. This difficulty necessarily

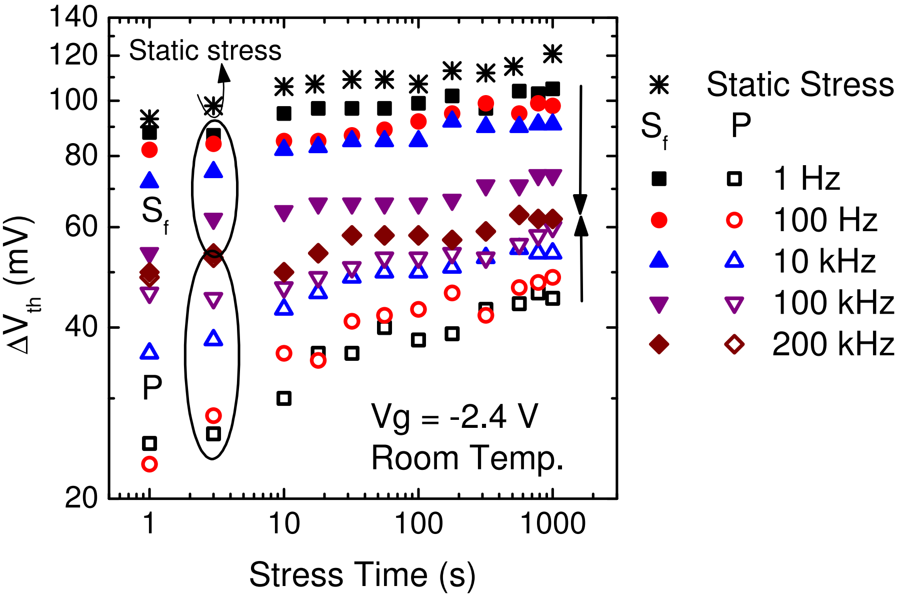


Figure 3.24 Experimental ΔV_{th} data under dynamic stress. The difference between $\mathrm{S_f}$ and P points is large and not closing after long time stress.

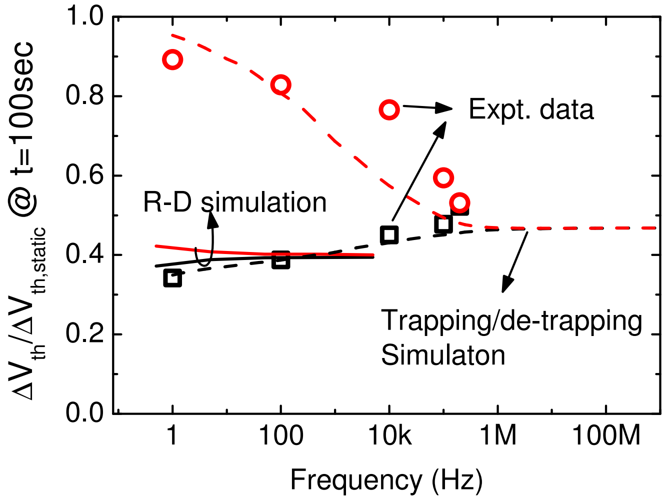


Figure 3.25 The trapping/de-trapping model predicts that, under dynamic stress, ΔV_{th} at both $\mathrm{S_f}$ and P points are frequency dependent, and the difference between the two is large.

prompted us to look for an alternative theory. The hole trapping/de-trapping model, incidentally, explains the observation fairly well. One certainly has noticed that, the above argument is largely based on the dynamics of the experimental observation. No attempt was made to reveal the exact physical/chemical structure involved. The attribution to hole trapping, therefore, should be taken as suggestive rather than conclusive. We discuss, in the following, two possibilities other than the hole trapping inside SiON insulator.

First of all, there might be interface traps not related to the diffusion of hydrogen. These hypothetical interface traps are probably pre-existing before stress. As the fast component of NBTI has very strong voltage dependence[47], these traps necessarily lie beyond the valence band of silicon, so that a higher stress voltage makes available more states for trapping. With these clauses, this kind of hypothetical interface traps is not essentially different from the traps inside the insulator bulk.

Secondly, H. Reisinger and co-authors proposed dielectric relaxation, whereby electrons hops within the insulator in response to applied electric field, as the cause of instabilities in high-κ dielectrics[48]. Similar dielectric relaxation might be present in SiON as well. Dielectric relaxation depends on the magnitude of the E-field, while the direction is irrelevant. However, positive bias on the gate does not lead to fast V_{th} instability on both n-MOSFET and p-MOSFET in experiments, which renders the dielectric relaxation theory improbable.

Therefore, the author is convinced that the hole-trapping theory is the most attractive in explaining the experimental observations of fast transients in NBTI, at present.

3.5 Interface States and Slow Component in NBTI

Having shown the existence of hole-trapping in the NBTI of SiON dielectrics, it is then required to consider the importance, if any, of the interface traps in NBTI.

3.5.1 Existence of Interface Trap Recovery

Although it is generally agreed that there is significant interface trap generation during NBTI stress, whether there is recovery of interface traps is being debated. Some researchers, through charge pumping or sub-threshold swing measurement, claimed that interface trap density does not decrease in NBTI recovery, and that the V_{th} recovery is purely due to the de-trapping of holes[30, 49]. T. Yang, et al., showed that significant D_{it} recovery can be observed with the DC-IV measurement[50]. It was pointed out that the charge pumping measurement, which involves large gate voltages to bias the MOSFET to both strong inversion and accumulation, suffers from strong recovery of interface traps. This recovery causes underestimation of D_{it} during stress, which makes the D_{it} recovery during the recovery phase less obvious. On the contrary, the DC-IV measurement works in the weak inversion and depletion region, and takes less measurement time, thus leading to more accurate D_{it} measurement with less recovery effect.

L. Jin further pointed out that the sub-threshold swing measurement after stress is susceptible to hole de-trapping, and the direction of voltage sweep is critical[51]. A forward (low-to-high) sweep would underestimate interface trap density, which possibly led to the remark of no D_{it} recovery from D.S. Ang[49]. A reverse sweep would overestimate interface trap density, but does not affect the conclusion on the presence of D_{it} recovery.

Additionally, electron paramagnetic resonance/spin-dependent recombination (EPR/SDR) measurements showed that in SiON gate oxide, there is also generation and recovery of recombination centers, though not the usual $\mathrm{P_{b0}}$ or $\mathrm{P_{b1}}$ centers[52].

It is well-known that the Shockley-Reed-Hall recombination process, which is the basis of DC-IV and SDR measurement, requires defects located deep in semiconductor band-gap. Hole traps, which is assumed to be near or beyond the valence band edge, does not contribute to DC-IV or SDR measurement, if the measurement is performed slowly in steady-state. With the above understanding, the author is convinced that recovery of interface traps does exist in NBTI.

3.5.2 Attribution of the Slow Component of NBTI

Unlike the hole-traps discussed in section 3.4, which are assumed to be pre-existing prior to stress, the interface traps are generated during stress.

The dynamics of interface trap generation and recovery is, to date, not well understood. No interface trap measurement technique can work at stress gate voltage. Moreover, the popular measurement techniques are all slow in nature. The charge pumping method requires multiple V_g pulses to the accumulation region; the DC-IV method requires steady-state condition. At the present time, all D_{it} measurements suffer from substantial delay induced errors.

In practice, the D_{it} measurement is often compared with the slow I_d–V_g measurement, which typically has similar measurement time. T. Yang established that, the increase in interface trap density ΔD_{it} and the threshold voltage shift ΔV_{th} show good linear correlation, if one only consider data collected after the initial fast recovery of both[50]. This correlation suggests that the generation/recovery of interface traps are responsible for the slow component in NBTI. Of course, the slow I_d–V_g measurement is inevitably contaminated by some trapping and de-trapping with time constants comparable to the measurement time, i.e. an overlapping of fast and slow components in terms of time constants. Some other distinctions between the fast and slow components of NBTI are in their different dependence on stress voltage and temperature[31], which further supports the attribution of the two components to different physical processes.

Whether the generation and recovery of interface trap is limited by diffusion is another open question. Although the recovery of D_{it} appears to be slower than that of V_{th}, the recovery measured by DC-IV method still seems a bit too fast to fit into the diffusion-reaction model[15, 37, 50]. Moreover, as all current D_{it}

measurements are slow, we may have missed some fast transient in D_{it}, which would be more challenging to a diffusion-limited theory. Despite these potential problems, there is at present no alternative theory proposed, and the reaction-diffusion model remains the best candidate.

In this short section, it is argued that the generation and recovery of interface traps contributes to the degradation and recovery in NBTI, respectively, and manifests as the slower components in NBTI.

3.6 Conclusions

In this chapter, the current understandings on NBTI in SiON gate dielectric is presented. Much emphasis was put on firstly the search of an appropriate measurement technique to characterize the fast transient component in NBTI, and secondly an appropriate theory to explain the fast transients. It is found that fast $I_d - V_g$ measurement is indispensable in the study of NBTI at present time to minimize the delay-induced measurement error. The fast recovery property of this transient component, along with some other experimental observations, contradicts the reaction-diffusion model in a fundamental manner, and an alternative theory is needed. It is proposed that the hole-trapping model is a good candidate that explains a large set of observations. Nevertheless, the generation and recovery of interface traps, is still an important component in NBTI.

References

[1] Bruce E. Deal. The current understanding of charges in the thermally oxidized silicon structure. *Journal of The Electrochemical Society*, 121(6):198C-205C, 1974.

[2] A. Goetzberger, A. D. Lopez and R. J. Strain. On the formation of surface states during stress aging of thermal Si-SiO$_2$ interfaces. *Journal of The Electrochemical Society*, 120(1):90-96, 1973.

[3] Kjell O. Jeppson and Christer M. Svensson. Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices. *Journal of Applied Physics*, 48(5):2004-2014, 1977.

[4] A. K. Sinha, H. J. Levinstein, L. P. Adda, E. N. Fuls and E. I. Povilonis. Effect of high-temperature H_2-anneals on the slow-trapping instability of MOS structures. *Solid-State Electronics*, 21(3):531–535, 1978.

[5] C.H. Liu, M.T. Lee, Chih-Yung Lin, J. Chen, K. Schruefer, J. Brighten, N. Rovedo, T.B. Hook, M.V. Khare, Shih-Fen Huang, C. Wann, Tze-Chiang Chen and T.H. Ning. Mechanism and process dependence of negative bias temperature instability (NBTI) for pMOSFETs with ultrathin gate dielectrics. In *IEDM Tech. Digest*, pages 39.2.1–39.2.4, 2001.

[6] Y. Mitani, M. Nagamine, H. Satake and A. Toriumi. NBTI mechanism in ultra-thin gate dielectric — nitrogen-originated mechanism in SiON. In *Electron Devices Meeting, 2002. IEDM '02. Digest. International*, pages 509–512, 2002.

[7] Keiko Kushida-Abdelghafar, Kikuo Watanabe, Jiro Ushio and Eiichi Murakami. Effect of nitrogen at SiO_2/Si interface on reliability issues — negative-bias-temperature instability and Fowler-Nordheim-stress degradation. *Applied Physics Letters*, 81(23):4362-4364, 2002.

[8] L. Duong, E. Li, V. Gopinath, S. Prasad, J. Lin, D. Pachura and V. Hornback. Effect of nitrogen incorporation on PMOS negative bias temperature instability in ultrathin oxi-nitrides. In *Integrated Reliability Workshop Final Report, 2003 IEEE International*, pages 128–130, 2003.

[9] Shyue Seng Tan, T. P. Chen, C. H. Ang and L. Chan. Relationship between interfacial nitrogen concentration and activation energies of fixed-charge trapping and interface state generation under bias-temperature stress condition. *Applied Physics Letters*, 82(2):269-271, 2003.

[10] Shyue Seng Tan, T. P. Chen, Jia Mei Soon, Kian Ping Loh, C. H. Ang and L. Chan. Nitrogen-enhanced negative bias temperature instability: An insight by experiment and first-principle calculations. *Applied Physics Letters*, 82(12):1881-1883, 2003.

[11] Shyue Seng Tan, T. P. Chen, Jia Mei Soon, Kian Ping Loh, C. H. Ang, W. Y. Teo and L. Chan. Linear relationship between H^+-trapping reaction energy and defect generation: Insight into nitrogen-enhanced negative bias temperature instability. *Applied Physics Letters*, 83(3):530-532, 2003.

[12] Shyue Seng Tan, Tu Pei Chen, Chew Hoe Ang and Lap Chan. Atomic modeling of nitrogen neighboring effect on negative bias temperature instability of pmosfets. *IEEE Electon Device Letters*, 25(7):504–506, 2004.

[13] Dieter K. Schroder and Jeff A. Babcock. Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing. *Journal of Applied Physics*, 94(1):1-18, 2003.

[14] G. Chen, M.F. Li, C.H. Ang, J.Z. Zheng and D.L. Kwong. Dynamic NBTI of p-MOS transistors and its impact on MOSFET scaling. *IEEE Electon Device Letters*, 23(12):734–736, 2002.

[15] 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. In *Reliability Physics Symposium Proceedings, 2003. 41st Annual. 2003 IEEE International*, pages 196–202, 2003.

[16] B. Zhu, J.S. Suehle, Y. Chen and J.B. Bernstein. Negative bias temperature instability of deep sub-micron p-MOSFETs under pulsed bias stress. In *Integrated Reliability Workshop Final Report, 2002. IEEE International*, pages 125–129, 2002.

[17] M. Ershov, R. Lindley, S. Saxena, A. Shibkov, S. Minehane, J. Babcock, S. Winters, H. Karbasi, T. Yamashita, P. Clifton and M. Redford. Transient effects and characterization methodology of negative bias temperature instability in pMOS transistors. In *Reliability Physics Symposium Proceedings, 2003. 41st Annual. 2003 IEEE International*, pages 606–607, 2003.

[18] 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*, 83(8):1647-1649, 2003.

[19] H. Usui, M. Kanno and T. Morikawa. Time and voltage dependence of degradation and recovery under pulsed negative bias temperature stress. In *Reliability Physics Symposium Proceedings, 2003. 41st Annual. 2003 IEEE International*, pages 610–611, 2003.

[20] S. Rangan, N. Mielke and E.C.C. Yeh. Universal recovery behavior of negative bias temperature instability PMOSFETs. In *IEDM Tech. Digest*, pages 14.3.1–14.3.4, 2003.

[21] M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, Y. Rey-Tauriac and N. Revil. On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's. In *IEDM Tech. Digest*, pages 109-112, 2004.

[22] M. Denais, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, N. Revil and A. Bravaix. New methodologies of NBTI characterization eliminating recovery effects. In *Solid-State Device Research conference, 2004. ESSDERC 2004. Proceeding of the 34th European*, pages 265–268, 2004.

[23] D. Varghese, S. Mahapatra and M.A. Alam. Hole energy dependent interface trap generation in MOSFET Si/SiO_2 interface. *IEEE Electron Device Letters*, 26(8):572–574, 2005.

[24] A. T. Krishnan, C. Chancellor, S. Chakravarthi, P. E. Nicollian, V. Reddy and A. Varghese. Material dependence of hydrogen diffusion: Implications for NBTI degradation. In *IEDM Tech. Dig.*, pages pp. 345-358, 2005.

[25] M. A. Alam. A review of new characterization methodologies of gate dielectric breakdown and negative bias temperature instability. In *Physical and Failure Analysis of Integrated Circuits, 2006. 13th International Symposium on the*, pages 25–32, 2006.

[26] V. Huard, M. Denais and C. Parthasarathy. NBTI degradation: From physical mechanisms to modelling. *Microelectronics and Reliability*, 46(1):1–23, 2006.

[27] Y. Nissan-Cohen, J. Shappir and D. Frohman-Bentchkowsky. Dynamic model of trapping-detrapping in SiO_2. *J. Appl. Phys.*, 58(6):2252–2261, 1985.

[28] S. Mahapatra, P.B. Kumar and M.A. Alam. Investigation and modeling of interface and bulk trap generation during negative bias temperature instability of p-MOSFETs. *IEEE Transactions on Electron Devices*, 51(9):1371–1379, 2004.

[29] Jiro Ushio, Takuya Maruizumi and Keiko Kushida-Abdelghafar. Interface structures generated by negative-bias temperature instability in Si/SiO_2 and Si/SiO_xN_y interfaces. *Applied Physics Letters*, 81(10):1818-1820, 2002.

[30] V. Huard and M. Denais. Hole trapping effect on methodology for DC and AC negative bias temperature instability measurements in PMOS transistors. In *Reliability Physics Symposium Proceedings, 2004. 42nd Annual. 2004 IEEE International*, pages 40–45, 2004.

[31] T. Yang, M. F. Li, C. Shen, C. H. Ang, Chunxiang Zhu, Y.-C. Yeo, G. Samudra, S. C. Rustagi, M. B. Yu and D. L. Kwong. Fast and slow dynamic NBTI components in p-MOSFET with SiON dielectric and their impact on device life-time and circuit application. In *VLSI Symp.*, pages 92-93, 2005.

[32] C. Shen, M.-F. Li, C. E. Foo, T. Yang, D. M. Huang, T. C. Chua, F. Nouri, G. S. Samudra and Y.-C. Yeo. Characterization and physical origin of fast V_{th} transient in NBTI of pMOSFETs with SiON dielectric. In *IEDM Tech. Digest*, pages pp. 333-336, 2006.

[33] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken and M. Goodwin. Temperature dependence of the negative bias temperature instability in the framework of dispersive transport. *Applied Physics Letters*, 86(14):143506, 2005.

[34] 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. In *IEDM Tech. Digest*, 2005.

[35] S. Tsujikawa, K. Watanabe, R. Tsuchiya, K. Ohnishi and J. Yugami. Experimental evidence for the generation of bulk traps by negative bias temperature stress and their impact on the integrity of direct-tunneling gate dielectrics. In *VLSI Technology, 2003. Digest of Technical Papers. 2003 Symposium on*, pages 139–140, 2003.

[36] W. Abadeer. Behavior of NBTI under ac dynamic circuit conditions. In *IEEE IRPS*, pages 17-22, 2003.

[37] M.A. Alam. A critical examination of the mechanics of dynamic NBTI for PMOSFETs. In *IEDM Tech. Digest*, pages 14.4.1–14.4.4, 2003.

[38] S. Chakravarthi, A. Krishnan, V. Reddy, C.F. Machala and S. Krishnan. A comprehensive framework for predictive modeling of negative bias temperature instability. In *Reliability Physics Symposium Proceedings, 2004. 42nd Annual. 2004 IEEE International*, pages 273–282, 2004.

[39] M.A. Alam, H. Kufluoglu, D. Varghese and S. Mahapatra. A comprehensive model for PMOS NBTI degradation: Recent progress. *Microelectronics Reliability*, 47(6):853–862, 2007.

[40] Haldun Kufluoglu and Muhammad Ashraful Alam. A generalized reaction-diffusion model with explicit H-H_2 dynamics for negative-bias temperature-instability (NBTI) degradation. *IEEE Transactions on Electron Devices*, 54(5):1101–1107, 2007.

[41] G. Groeseneken, H. E. Maes, N. Beltran and R. F. De Keersmaecker. A reliable approach to charge-pumping measurements in MOS transistors. *IEEE Transactions on Electron Devices*, 31(1):42, 1984.

[42] A. Neugroschel, C.-T. Sah, K. M. Han, M. S. Carroll, T. Nishida, J. T. Kavalieros and Y. Lu. Direct-current measurement of oxide and interface traps on oxidized silicon. *IEEE Transactions on Electron Devices*, 42(9):1657, 1995.

[43] C. D. Young, Y. G. Zhao, M. Pendley, B. H. Lee, K. Matthews, J. H. Sim, R. Choi, G. Bersuker and G. A. Brown. Ultra-short pulse I-V characterization of the intrinsic behavior of high-κ devices. In *SSDM Proc.*, pages 216-217, 2004.

[44] Dieter K. Schroder. Negative bias temperature instability: What do we understand?. *Microelectronics Reliability*, 47(6):841–852, 2007.

[45] H. Reisinger, O. Blank, W. Heinrigs, A. Muhlhoff, W. Gustin and C. Schlunder. Analysis of NBTI degradation and recovery behavior based on ultra fast VT measurements. In *IEEE International Reliability Physics Symposium Proceedings*, pages 448–453, 2006.

[46] Hee-Gook Lee, Soo-Young Oh and G. Fuller. A simple and accurate method to measure the threshold voltage of an enhancement-mode mosfet. *Electron Devices, IEEE Transactions on*, 29(2):346–348, 1982.

[47] T. Yang, C. Shen, M.F. Li, C.H. Ang, C.X. Zhu, Y.-C. Yeo, G. Samudra, S.C. Rustagi, M.B. Yu and D.-L. Kwong. Fast DNBTI components in p-MOSFET with SiON dielectric. *IEEE Electron Device Letters*, 26(11):826–828, 2005.

[48] H. Reisinger, G. Steinlesberger, S. Jakschik, M. Gutsche, T. Hecht, M. Leonhard, U. Schroder, H. Seidl and D. Schumann. A comparative study of dielectric relaxation losses in alternative dielectrics. *International Electron Devices Meeting. Technical Digest*, pages 12 - 2, 2001.

[49] 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. *IEEE Electron Device Letters*, 27(5):412–415, 2006.

[50] T. Yang, C. Shen, M.F. Li, C.H. Ang, C.X. Zhu, Y.-C. Yeo, G. Samudra and D.-L. Kwong. Interface trap passivation effect in NBTI measurement for p-MOSFET with SiON gate dielectric. *IEEE Electon Device Letters*, 26(10):758–760, 2005.

[51] Lei Jin and Mingzhen Xu. Impact of positive charges on NBTI degradation and characterization in ultra-thin SiON gate dielectrics. *ECS Transactions*, 6(3):371-382, 2007.

[52] P.M. Lenahan. Deep level defects involved in MOS device instabilities. *Microelectronics Reliability*, 47(6):890–898, 2007.

CHAPTER

4

Charge Trapping in High-κ Gate Dielectrics

4.1 Introduction

Threshold voltage instability due to charge trapping is one of the most challenging problems for incorporating high-κ gate dielectrics in CMOSFET technology[1–7]. Early studies[1–3] on threshold voltage instability in HfO_2 and other high-κ gate dielectric used the traditional measure-stress-measure scheme to characterize the evolution of threshold voltage shift, with dc parametric analyzers such as HP4156. These studies suggested that the V_{th} shift is due to electron and hole trapping in the high-κ film, and the V_{th} degradation is moderate, although much larger than that in SiO_2. However, A. Kerber et al. revealed that there is significant V_{th} shift within the time frame of tens of microseconds[4], using a fast $I_d - V_g$ measurement technique. When the stress is removed, the V_{th} degradation quickly recovers. This fast degradation/recovery in threshold voltage shift has a magnitude of a few hundred milli-volts, which is much larger than previously observed with slow measurement. This large instability was soon recognized as the major show-stopper of HfO_2 as a viable gate dielectric material, and extensive research was directed into the charge trapping in HfO_2 and other high-κ gate dielectric films.

Most authors agree that in HfO_2, electron/hole trapping and de-trapping in pre-existing traps are responsible for the threshold voltage shift observed in both fast and slow measurements [1–8]. There was proposals that dielectric relaxation could be an important cause of the fast instability[9–10], but later reports[11] pointed out a series of disagreements of this model with experimental data. Some other groups reported the contribution from the generation of interface states to the slow V_{th} instability[12–13]. However, these samples feature either thick SiO_2

buffer layer or high silicon composition in the dielectric, which make the film more SiO_2 like. For films with higher Hf composition, and hence higher permittivity, the contribution of interface trap generation is found less important than charge trapping.

The author suggested that distinctive fast and slow charge traps exist in HfO_2 gate dielectric, based on the observation that charge trapping measurement using fast and slow methods yields opposite dependence on the frequency of stress voltage signal[6]. One may infer that the fast and slow charge trapping components may be associated with different defects or trapping mechanisms, and can be studied separately. The fast charge trapping causes threshold voltage to shift over 100 mV, and leads unacceptably large drive current degradation, which renders the fast charge trapping the main show-stopper for HfO_2 as gate dielectric material. The slow charge trapping component is a reliability problem, while the fast charge trapping is both a time-zero and a reliability problem. Some authors[14] alternatively term the fast charge trapping component hysteresis effect, and the the slow component BTI (bias temperature instability) effect. This alternative categorization, although not precise physically, shared the recognition on the differences between time-zero and long-term instabilities.

Vast effort have been dedicated to reducing the charge trapping in HfO_2, and several approaches have been demonstrated to reduce the slow charge trapping component[2, 15–21], or to reduce the fast charge trapping[8, 22]. However, as significant fast hole trapping was also recently discovered in the SiON dielectric[23–25], we may not be able to totally eliminate the fast charge trapping in high-κ film. We likely have to accept a certain amount of fast V_{th} instability in MOSFETs with high-κ dielectric. As a result, accurate modeling of the fast charge trapping component is mandatory to assess its impact on circuit performance, and to determine the maximum allowable fast charge trapping as the target of process improvement effort.

In this chapter, we systematically characterized the slow and fast charge trapping component in HfO_2 gate dielectrics in the next two sections, respectively. In each of the two cases, a model of the respective charge trapping dynamics was proposed. It is found that the charge trapping dynamics in the slow and fast components must be explained by models with different structures. In particular, the peculiar frequency dependence of the slow components is not expected in the classical charge trapping model, and a new two-step model is necessary. Possible physical defect structures behind the dynamic model is proposed. The fast component, on the other hand, is modeled with the same classical charge trapping model described in section 3.2. Due to the dominance of the fast component, its impact on digital circuits is analyzed in the last section, with proposals of possible circuit design techniques to reduce the detrimental effect of the fast V_{th} transients.

4.2 Slow Component of Charge Trapping in HfO_2

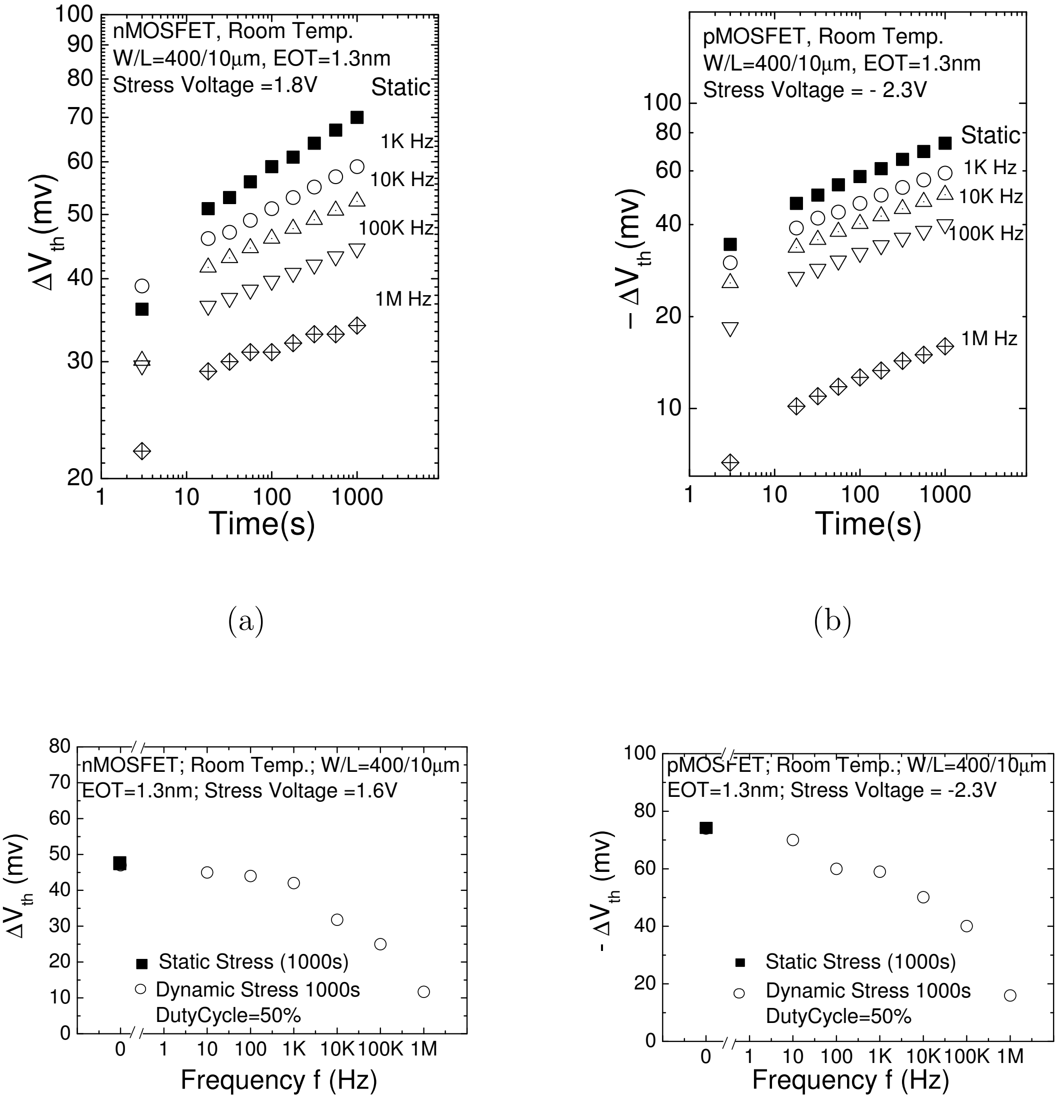
HfO_2 gate dielectric is known to suffers from problems related to charge trapping, resulting in degradation of n-MOSFETs under uniform static stress[1–3]. In this section, we shall focus on the frequency dependent dynamic charge trapping effect for n- and p-MOSFETs with MOCVD deposited HfO_2 gate dielectrics[5]. A Similar frequency dependent positive charge trapping degradation has been reported for n-MOSFETs with PVD grown HfO_2 gate dielectric[26]. Additionally, we report a physical model to explained the frequency dependent dynamic charge trapping effect, based on the negative-U property of the charging traps in the HfO_2 dielectric.

4.2.1 Dynamic Charge Trapping in HfO_2 and its Frequency Dependence

Both n-MOSFETs and p-MOSFETs with HfO_2 gate dielectric and HfN (capped with TaN) metal gate stack were fabricated[27]. The HfO_2 dielectric has an equivalent silicon oxide thickness (EOT) of ~ 1.3 nm, determined by high frequency C–V curve fitting using a numerical simulator that accounts for quantum effect in the Si inversion layer. The charge trapping effect in HfO_2 was investigated by measuring the threshold voltage V_{th} shift under stress for both n- and p-MOSFETs at room temperature using a computer-controlled HP4156A semiconductor parameter analyzer. The delay between the removal of stress and the start of I_d–V_g measurement is about 1 s, so the fast component of charge trapping has almost completely recovered, while the slow component largely remains. The sub-threshold swing (SS) was also monitored during the stress. For both n- and p-MOSFETs, there is no observable change in SS during stress, indicating that interface state generation plays no significant role, and that charge trapping in the dielectric is the primary mechanism responsible for the V_{th} shift in the MOSFET high-κ gate dielectric[1, 3]. For n- and p-MOSFETs under dynamic stress, the rate of V_{th} shift is strongly dependent on the frequency, as shown in Figure 4.1. The V_{th} shift under dynamic stress is reduced as compared to that under static stress, and the reduction becomes larger as the stress frequency is increased (up to 1 MHz as demonstrated in this work).

We shall discuss the model for the case of n-MOSFET. This model can be extended to p-MOSFETs in a complementary way. For static stress, it is assumed that pre-existing electron traps in the dielectric are responsible for the V_{th} shift in n-MOSFETs[1–3]. Under static positive stress voltage, electrons are captured by the traps distributed in the HfO_2 dielectric, leading to a positive shift of V_{th}. When dynamic stress is applied, there is a stress phase with positive gate voltage and a recovery phase with zero gate voltage[28]. In the recovery phase, the trapped electrons can be emitted from the dielectric, leading to a recovery of the V_{th} shift. Figure 4.2 illustrates the V_{th} shift and recovery behavior in stress and recovery cycles.

It is observed from Figure 4.1 that under stress, ΔV_{th} has a power law dependence on stress time, giving rise to a straight line in $\log(V_{\mathrm{th}}) - \log(t)$ plot. This



(a) (b)

(c) (d)

Figure 4.1 Time evolution of threshold voltage V_{th} under static and dynamic stresses of different frequencies, for (a) n-MOSFETs, and (b) p-MOSFETs. The V_{th} evolution has a power law dependence on stress time. (c) Frequency dependence of V_{th} degradation in n-MOSFETs after dynamic stress of 1000 s, and (d) for p-MOSFETs.

power-law dependence can be obtained by assuming that electron traps have a distribution $N(\tau_{C2})$ in the domain of trapping time constant[1]. The mathematical formulation in this work, detailed in the next section, allows a more general distribution function for time constants, but is essentially equivalent to that of S.

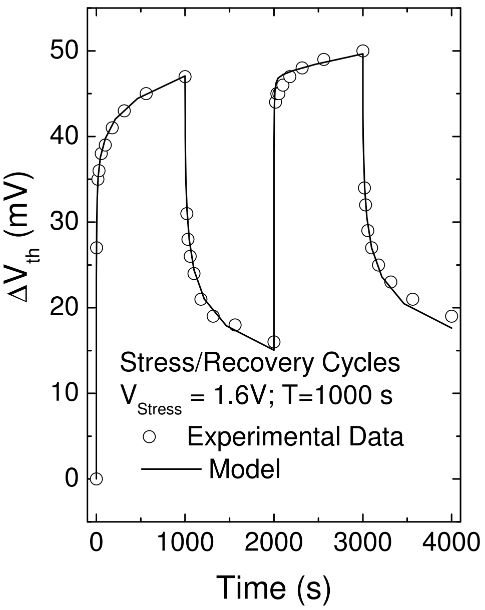


Figure 4.2 V_{th} shift in alternating stress and recovery cycles of period $T = 2000$s. Symbols are experimental data, lines are from model simulation.

Zafar. Good agreement between the model and the static and dynamic charge trapping and de-trapping experiments, as shown in Figure 4.2, can be obtained.

4.2.2 Physical Model of the Frequency Dependent Charge Trapping

Next, we extend the above model to account for the frequency dependence of the dynamic charge trapping effect. The number of electrons Δn being trapped during time Δt in one stress cycle in the dynamic stressing experiment is considered in Figure 4.3. The frequency f is related to Δt by $f = 1/\Delta t$. When t increases from T (where $T = 1/f$, and f is the stress frequency) to $2T$, the number of trapped electrons increases from Δn_T to Δn_{2T}. If the relationship between Δn and Δt is linear (curve L), then $\Delta n_{2T} = 2\Delta n_T$. Therefore, the number of trapped electrons during the same stressing time would be the same for two frequencies and the accumulative V_{th} shift would be frequency independent. In order to explain the frequency dependence observed in dynamic charge trapping, Δn_{2T} must be larger than $2\Delta n_T$. To explain the experimentally observed reduction in trapping at higher frequencies, the $\Delta n - \Delta t$ curve must have a concave-up shape. However, the concave-up characteristics cannot be obtained by simple first order capture equation, which always concaves down, or approximate a linear curve for small Δt [1–2].

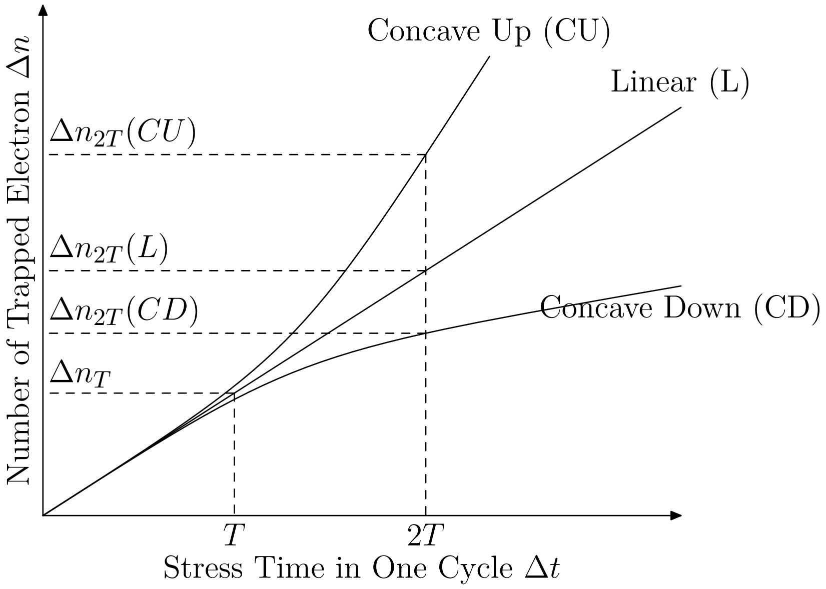


Figure 4.3 Three possible cases for the relationship between the number of trapped electrons Δn versus stress time Δt in one cycle of the dynamic stress.

In order to explain the observed frequency dependence, a higher-order dynamics is required. In this work, we consider a second-order, two-step trapping/de-trapping model, where two electrons (or holes) can be trapped at the same defect successively. Such two-step trapping is characteristic of negative-U traps present in semiconductors[29–31]. Each trap can capture two electrons in two steps, as shown in described in Figure 4.4.

When the first electron is trapped at the defect, the electrostatic potential due to this electron would normally repulse subsequent electrons that come by. However, in some materials, the first captured electron interacts with neighboring atoms, and the lattice relaxes by small displacements to minimize the total energy

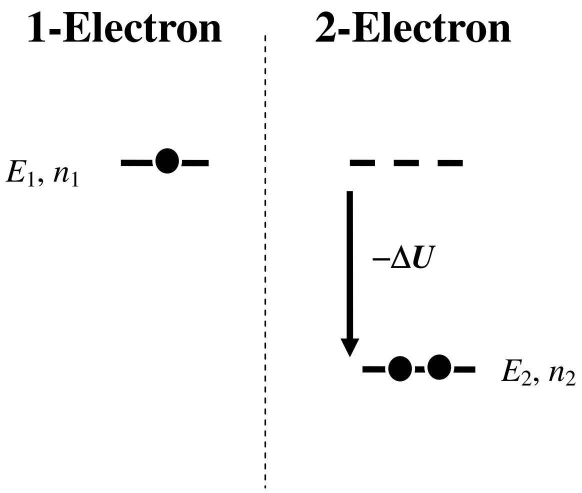


Figure 4.4 Two-step procedure of capturing two electrons by a negative-U trap. E_1 is the trap energy capturing one electron and n_1 is the number of those traps. E_2 is the trap energy capturing 2 electrons and n_2 is the number of those traps. When an additional electron is trapped, energy of the trap is lowered from E_1 to E_2 due to lattice distortion. This energy lowering favors two electrons to occupy on the trap.

of the electron and lattice. The new defect structure now energetically favors the capture of a second electron. The negative total energy, or *negative-U*, due to the electron-lattice interaction and lattice relaxation, gives rise to the name of negative-U traps. Negative-U traps are mostly found in ionic materials (known as DX centers) with strong electron-lattice interaction[32]. Recently, *ab initio* calculations have shown that in SiO_2, the hydrogen and oxygen vacancy related traps are negative-U centers[33–34]. Very recently , *ab initio* calculations also showed the negative-U property of oxygen vacancy defects and hydrogen bridge defects in HfO_2 [6, 35–37]. This result is quite reasonable since HfO_2 is a more ionic material compared with SiO_2, and the strong ionic polarization contributes to more interaction between electron and the lattice.

The rate equations describing the trapping behavior are as follows. During the stress phase

$$\frac{\mathrm{d}n_1}{\mathrm{d}t} = \frac{1}{\tau_{\mathrm{C1}}}\left(N - n_1 - n_2\right) - \frac{\mathrm{d}n_2}{\mathrm{d}t} \tag{4.1}$$

$$\frac{\mathrm{d}n_2}{\mathrm{d}t} = \frac{1}{\tau_{\mathrm{C2}}} n_1 \tag{4.2}$$

where N is the total density of traps, n_1 is the density of traps occupied by one electron, n_2 is the density of traps occupied by two electrons, τ_{C1} is the capture time constant for the first electron, and τ_{C2} is the capture time constant for the second electron. During the recovery phase, when stress is removed, the de-trapping behavior is described by

$$\frac{\mathrm{d}n_1}{\mathrm{d}t} = -\frac{1}{\tau_{\mathrm{E1}}} n_1 - \frac{\mathrm{d}n_2}{\mathrm{d}t} \tag{4.3}$$

$$\frac{\mathrm{d}n_2}{\mathrm{d}t} = -\frac{1}{\tau_{\mathrm{E2}}} n_2 \tag{4.4}$$

where τ_{E1} and τ_{E2} are the emission time constants for the first and second electron, respectively. Since the trap energy E_2 for a trap with two electrons is lower than the trap energy E_1 for a trap with one electron (negative-U[29, 31]), the capture and emission time constants of the E_1 state are much faster than those of the E_2 state. Solving equations (4.1)–(4.4) yields the static and dynamic time evolutions under different frequencies.

The distribution functions of the time constants are listed in Table 4.1, which are obtained from data fitting. First a double-peak distribution is adopted for τ_{C2}, to fit the data from static stress. It is then assumed that the time constant τ_{E2} and τ_{C1} for a certain trap is directly proportional to the corresponding τ_{C2}, so the distribution functions assume the same shape. The ratio between these three time constants is obtained by fitting the data from dynamic stress. The emission of the first electron is assumed to be faster than all other processes, and τ_{E1} is set to 0.1 μs. The accurate calculation of the solution with large number of stress/recovery cycles is achieved by the construction of stress and recovery operator and the diagonalization technique described in section 3.2.

Symbol	Description	Value
$N(\tau_{C2})$	Distribution function of capture time constant τ_{C2} of the second electron. It is given by a sum of two log-normal distributions.	peaks are 40 μs and 0.1 s, width(σ) are 3.1 and 3.5, respectively.
τ_{E2}	Emission time constant of the second electron	$100 \cdot \tau_{C2}$
τ_{C1}	Capture time constant of the first electron	$10^{-5} \cdot \tau_{C2}$
τ_{E1}	Emission time constant of the first electron	0.1 μs

Table 4.1 Time constants used in the model.

Since the de-trapping of the first electron is very fast, almost all traps with only one electron emit that electron when the stress is removed during the recovery phase, and only those traps with two captured electrons retain the electrons. Therefore, only n_2 is responsible for the cumulative V_{th} shift. On the other hand, at each initial time of the stress phase ($\Delta t = 0$ in Figure 4.3), n_1 is almost zero and then increases versus time. Using this initial condition, the solution of n_2 from (4.2) gives a concave-up curve, and therefore, the cumulative increase of n_2 is reduced when the frequency increases. Results of calculations using (4.1)–(4.4) are in excellent agreement with all experiment data, as shown in Figure 4.5. For the case of static stress, since τ_{C1} and τ_{E1} are much smaller than τ_{C2} and τ_{E2}, respectively, the solution of (4.1)–(4.4) converges to that of the conventional first order trapping equation at long times. Therefore, our model is consistent with both static and dynamic charge trapping experiments in HfO$_2$ dielectric.

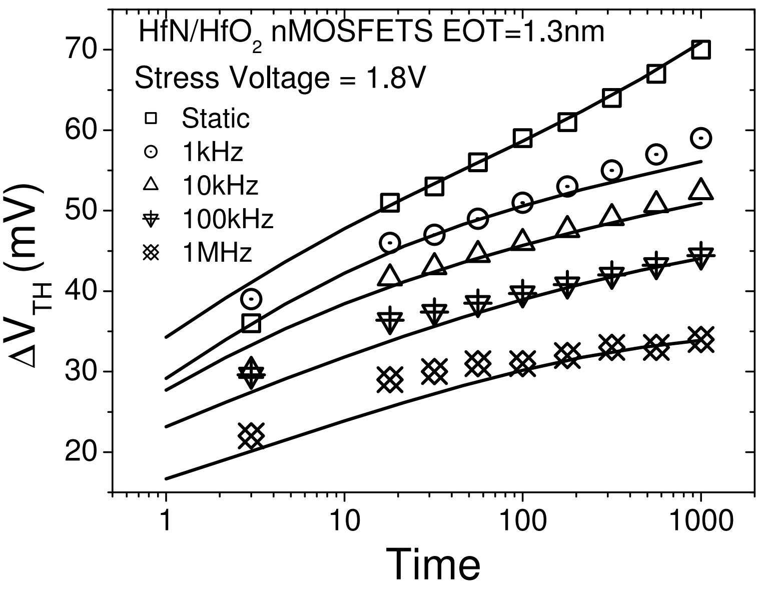


Figure 4.5 Calculated time evolutions of V_{th} using equations (4.1) at various frequencies are plotted using lines, showing good agreement with the experimental data (symbols).

For p-MOSFETs, we can use similar model that is complementary to that for n-MOSFETs. The traps contributing to V_{th} shift in p-MOSFETs are also negative-U traps, but capture two holes instead of two electrons. Negative-U traps for holes have been discussed in negative-U center literature for ionic II-VI semiconductors[30].

In conclusion, the slow component of the dynamic charge trapping and V_{th} shift in MOSFETs with MOCVD HfO_2 gate dielectrics under dynamic stress is reported. A strong frequency dependence of dynamic charge trapping is observed in the experiment. For an ac stress of given gate voltage amplitude, reduction of V_{th} shift for both n- and p-MOSFETs are observed with increasing stress frequency. A two-step trapping model based on negative-U trap centers in HfO_2 dielectric is proposed to explain the above-mentioned phenomena, giving excellent agreement with all experimental data.

4.3 Fast Component in Charge Trapping in HfO_2

In this section, we move the focus to the fast component of the dynamic charge trapping, which is only capture by the fast measurement technique.

4.3.1 Sample Preparation and Measurement technique

Both n- and p-MOSFETs with 4.0 nm MOCVD HfO_2 gate dielectric and HfN/TaN metal gate stack were fabricated, with the process flow detailed in ref.[27]. The equivalent electrical thickness (EOT) of the HfO_2 gate dielectric is 1.3 nm after S/D anneal at 950°C.

As mentioned earlier, a fast characterization technique is required to study the fast charge trapping in HfO_2. The single-pulse measurement originally proposed by Kerber et al.[4] was not sufficiently quantitative, and has difficulty in short measurement time down to 1 μs. Young et al. used RF measurement technique and a multi-pulse approach[38] to measure the intrinsic properties of high-κ MOSFET with ultra-short pulses (35 ns each). However the multi-pulse scheme needs a series of pulses to obtain the $I_d - V_g$ characteristics, which takes long time and is not suitable in studying the charge trapping characteristics. In this work, we use an improved single-pulse scheme with the best measurement speed reduced to 1 μs, noise suppressed, and sources of error analyzed. Measurement time in this work is 10 μs, which has been shown to be fast enough for evaluating the fast de-trapping states in these samples[6]. Both static and dynamic stress are possible with this setup. In the case of static stress, dc stress voltage is applied to the gate of the MOSFET, with pulses down to 0 volt intermittently inserted, as shown in Figure 4.6a. $I_d - V_g$ curve is measured as the gate voltage is dropping from stress voltage to zero, and the threshold voltage is extracted. In the case of dynamic stress, square-wave stress voltage is applied on the gate of the MOSFET, and $I_d - V_g$ characteristics can be measured at both the rising-edge or the falling-edge of the waveform. In the case of nMOSFET, the threshold voltage extracted from the $I_d - V_g$ curve obtained at the falling-edge represents the V_{th} degradation right after the stress phase (S_{f} point), while the V_{th} measured at the rising-edge represents the partially recovered V_{th} after the recovery phase (P point).

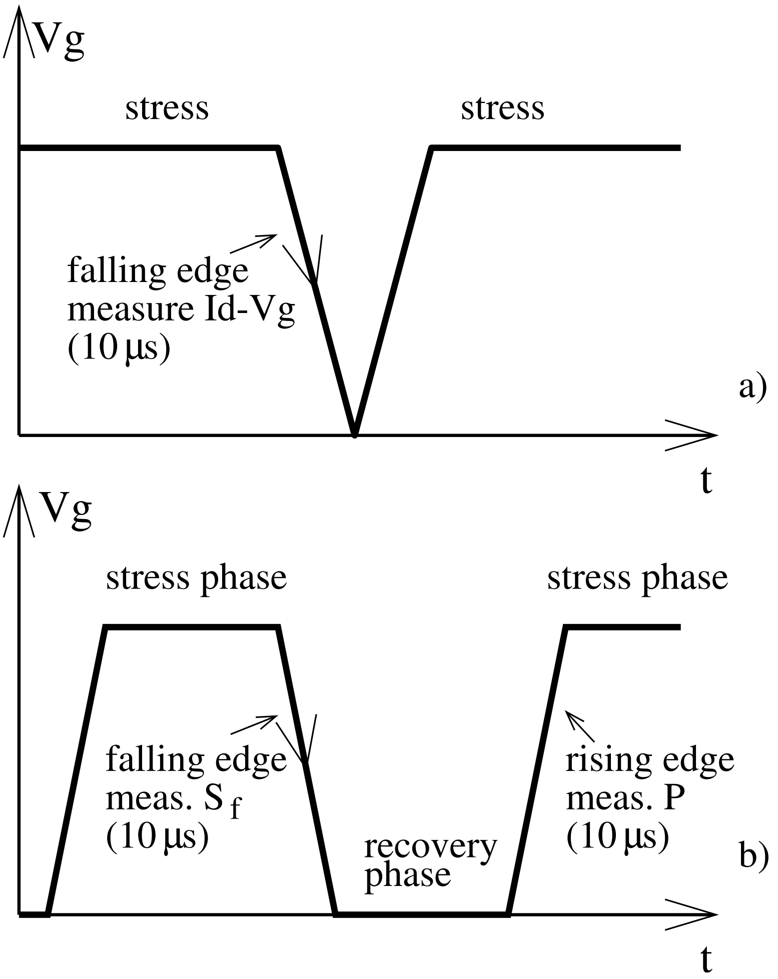
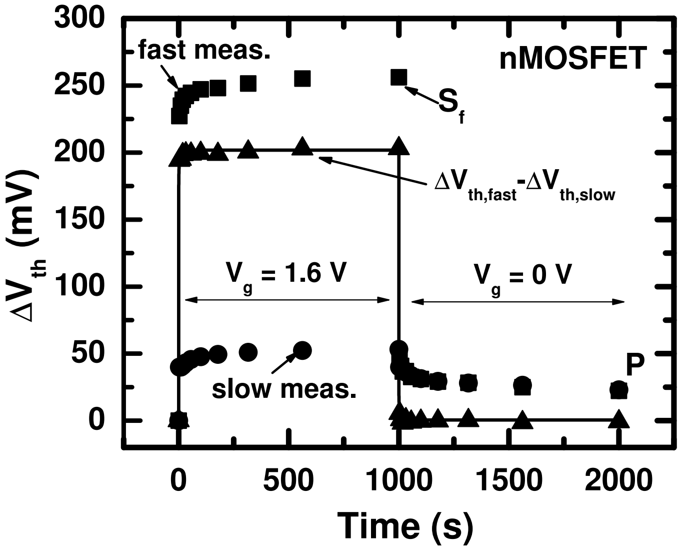

Vg
stress
stress
falling edge
measure Id-Vg
(10 µs)
a)
t
Vg
stress phase
stress phase
falling edge
meas. S f
(10 µs)
recovery
phase
rising edge
meas. P
(10 µs)
b)
t

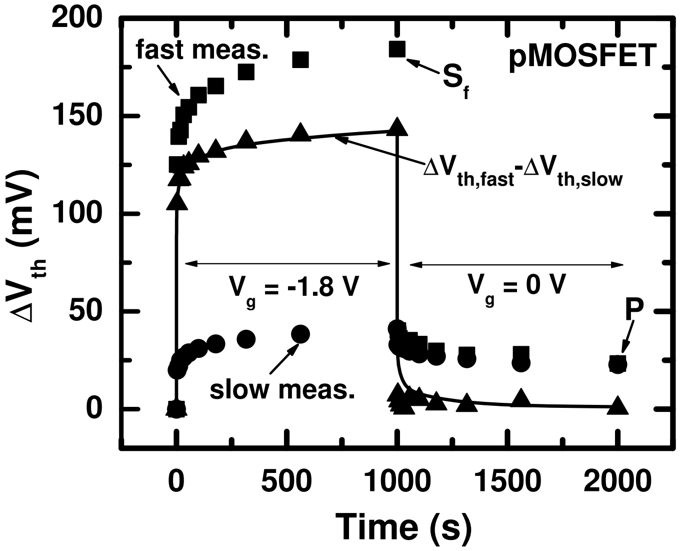
Figure 4.6 Waveform of stress voltage used in a) static stress, and b) dynamic stress.

Figure 4.7 shows time evolution of V_{th} degradation in n- and p-MOSFETs under 1000 s of static stress followed by 1000 s of recovery, using the V_g waveform shown in Figure 4.6(b). In a way, this can be viewed as the first stress/recovery period of a dynamic stress with frequency = 1/2000 Hz. As a comparison, slow measurement with HP4156A parametric analyzer is also used in this work to study the slow component of charge trapping in HfO_2. A large difference in ΔV_{th} measured by fast technique ($\Delta V_{\text{th,fast}}$) and slow technique ($\Delta V_{\text{th,slow}}$) is observed. This difference ($\Delta V_{\text{th,fast}} - \Delta V_{\text{th,slow}}$) is plotted with triangles in the figure, and reflects the contribution of fast charge trap component to the total V_{th} degradation. However, the time constants of de-trapping for fast and slow traps do not have a clear-cut boundary. Therefore, the value of $\Delta V_{\text{th,slow}}$ from slow measurement on one hand is the slow charge trapping component with some amount of recovery (de-trapping), and on the other hand may contain some fast charge trapping component. As the fast component is observed to be much larger than the slow component, we decide to use the threshold voltage shift measured by fast measurement ($\Delta V_{\text{th,fast}}$) to estimate the fast component in this work, unless otherwise stated. Comments on the effect of subtracting the contribution of the slow component is provided when it is necessary. The fast charge trapping component is huge after only 1 second of stress, but the growth of ΔV_{th} after that is slow. When stress voltage is removed, the fast component of threshold voltage shift quickly recovers to almost zero.

Alternatively, Chan et al. proposed to monitor the recovery of the linear-region drain current after stress voltage is removed[39–40]. After high voltage stress is removed, a voltage slightly higher than the threshold voltage is applied on the gate, and the drain current is plotted against time. The recovery characteristics of an nMOSFET is plotted in Figure 4.8. The linear-region drain current is proportional to the gate overdrive ($V_g - V_{\text{th}}$), and the threshold voltage shift is approximated from the relation $\Delta I_d = g_m \cdot \Delta V_{\text{th}}$. In the initial period of the recovery, the slope of the recovery characteristics is similar. As the recovery progresses, the slope reduces and recovery saturates, with a break point that depends on the time of stress.



n-MOSFET



p-MOSFET

Figure 4.7 Threshold Voltage shift under stress/recovery cycles with frequency = 1/2000Hz. Results of fast ($\Delta V_{\mathrm{th,fast}}$, squares) and slow ($\Delta V_{\mathrm{th,slow}}$, dots) measurements are compared. The difference between fast and slow measurement is plotted with triangles, representing the fast trap contribution to the total V_{th} degradation. Solid line shows simulation data as described in section 4.3.3.

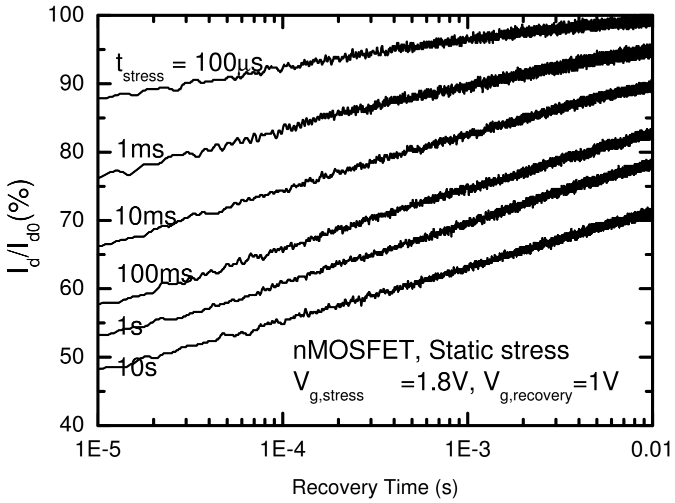


Figure 4.8 Recovery of linear region drain current I_d with respect to the pre-stress I_{d0}, after stress voltage is removed from the gate of the nMOSFET. Recovery after different stress time is shown. The drain current during recovery is measured by the fast measurement, at $V_g = 1.0$ V, and $V_d = 0.1$ V.

4.3.2 Characterization of fast charge traps

4.3.2.1 Voltage dependence

To investigate the voltage dependence of the fast charge trapping component, transistors are stressed at different voltage levels for 1 second (static stress), and the change in V_{th} compared to the fresh device is plotted against the stress voltage, or alternatively against the gate over-drive voltage at the end of the stress, as shown in Figure 4.10. As the fast charge trapping component builds up in very short time, the ΔV_{th} after 1 second provides a good estimate on the steady-state ΔV_{th}.

It is proposed that the observed exponential dependence of $\Delta V_{\mathrm{th,1s}}$ on the gate overdrive is characteristic of the fast charge trapping component. In the case of nMOSFET, assume that the fast electron traps in $\mathrm{HfO_2}$ are near or above the conduction band of Si, and are distributed in space scale[41] or in energy scale, as illustrated in Figure 4.9. These pre-existing trap sites can be occupied by electrons only if the trap level is below or near the energy of electron injection. Qualitatively, under larger positive gate stress voltage, trap levels in HfO_2 are moved downwards with respect to the substrate Fermi level, therefore more traps become available for charge trapping. However, as the V_{th} increases under stress while V_g is kept constant, the electric field across dielectric decreases, and fewer fast traps are available. Therefore, the amount of fast charge trapping is dependent on the gate overdrive (or electric field) at the end of the stress ($V_g - V_{\mathrm{th,1s}}$), as shown in Figure 4.10. The voltage dependence of $\Delta V_{\mathrm{th,1s}}$ in n-MOSFET has a larger slope (0.46 V/decade) than that of p-MOSFET (0.93 V/decade), showing a stronger voltage dependence. The voltage dependence obtained from slow measurement is weaker (close to 1.2 V/decade) for both n- and p-MOSFET. As a result, if the slow component were to be subtracted off from the voltage dependence shown in Figure 4.10, the actual slope would be slightly steeper.

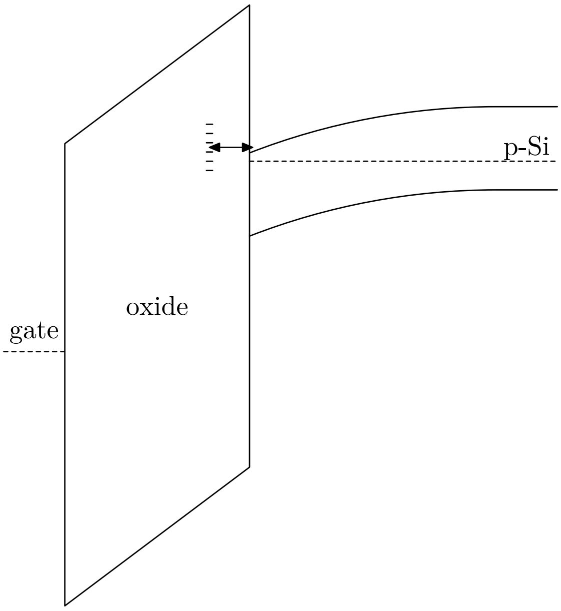


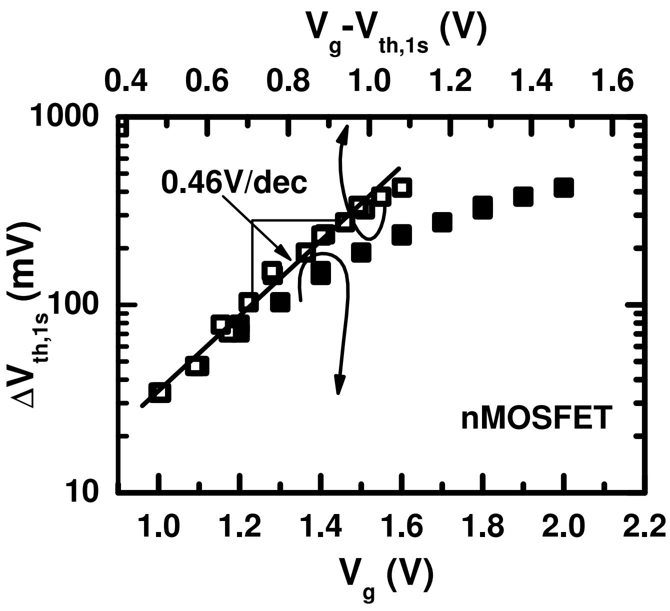
Figure 4.9 A schematic illustrating the trapping and de-trapping of electrons in HfO_2. Several electron traps are assume to distribute in different energy levels. The double-sided arrow indicates the exchange of electrons between the silicon substrate and the trap states, through trapping and de-trapping processes.

It should be noted that, although the amount of trapped charge (or ΔV_{th}) shows exponential dependence on electric field, the density of available trap sites does not necessarily have the same field dependence. As Nissan-Cohen et al. pointed out[42], the density of trapped charge in steady state is determined by the balance between charge trapping and de-trapping processes, the ratio of trapped charge density to trap density depends on the relative strength of trapping versus de-trapping, which in turn may be field dependent. This gives rise to a new dimension of complications, and a quantitative physcial model of the exponential voltage depedence could not be reached in this work. Nevertheless, the empirical relationship obtained above is used in later sections.

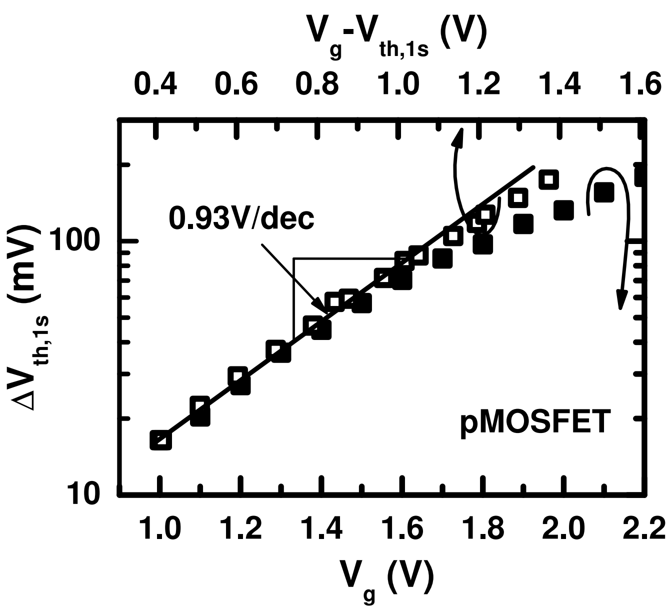
4.3.2.2 Frequency dependence

Figure 4.11 and Figure 4.12 show the frequency dependence of V_{th} shift under dynamic stress. As an example with extremely low frequency, in Figure 4.7, the ΔV_{th} after stress phase is labeled $\mathrm{S_f}$ point, and the ΔV_{th} after recovery phase is labeled P point. As shown in Figure 4.11, when the stress frequency is increased, the maximum dynamic V_{th} degradation ($\mathrm{S_f}$ point) decreases, while the cumulative dynamic V_{th} degradation (P point) measured by the fast technique increases.

The frequency dependence of the slow component, which was discussed in [5–6], is opposite to the frequency dependence of the P point described above. Therefore, even after subtracting the slow charge trapping component from the total V_{th} shift measured here, the V_{th} degradation at P point still increases with frequency. The opposite frequency dependence of the fast and slow components in charge trapping is one important clue suggesting that the two components are distinctive and may have different physical origins. In previous work, we have shown that a two-step trapping/de-trapping model is required to explain the peculiar frequency dependence of the slow component [5, 43]. The fast component, on the other hand, can be explained with conventional charge trapping dynamics, as will be discussed in section 4.3.3.



n-MOSFET



p-MOSFET

Figure 4.10 V_{th} shift dependence on static stress voltage and gate overdrive at the end of a one second stress $(V_g - V_{\text{th,1s}})$, measured by fast technique. Exponential dependence of $\Delta V_{\text{th,1s}}$ on the gate voltage overdrive $V_g - V_{\text{th,1s}}$ is observed.

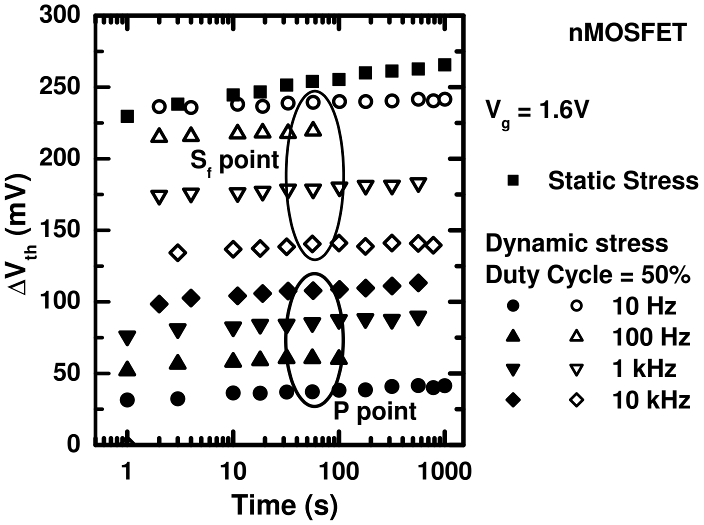
The difference between the ΔV_{th} of S_{f} point and P point reflects the transient amplitude as illustrated in Figure 4.7. In the high frequency limit, S_{f} point and P point converges, which is most evident in Figure 4.12. As frequency increases, in one period of dynamic stress, the degradation in the stress phase and the recovery in recovery phase both decreases, so the threshold voltage transient amplitude reduces, and approaches zero at high frequency. As a result, V_{th} of the MOSFET with HfO_2 dielectric, although exhibiting large instability, can be predicted when it is switching at high-frequency, if the operation voltage and duty cycle is known. This property may have important implication to digital circuits, as we attempt to explore in section 4.4.

4.3.2.3 Stress history

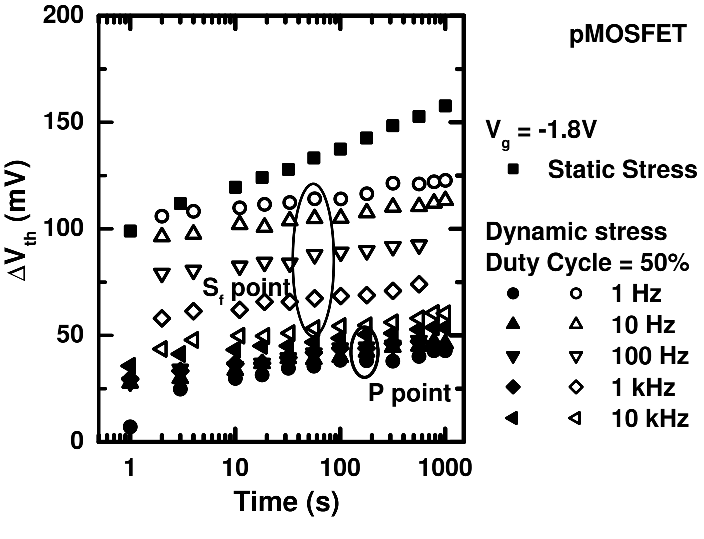
Figure 4.13 shows that evolution of ΔV_{th} under dynamic stress with two different prior conditions: (1) 100 seconds static stress and (2) zero stress applied, on two identical transistors. It is obvious that prior condition or stress history does not affect the steady-state V_{th} shift of dynamic stress. This independence on prior stress history agrees with the assumption that all the traps in the dielectric are pre-existing and are not generated during the stress. The steady-state ΔV_{th} is determined by stress condition (voltage, duty cycle, etc.), but not by stress history. The transition from previous steady-state ΔV_{th} to a new equilibrium takes less than 1 s as observed in Figure 4.13.

4.3.2.4 Duty-cycle dependence

Figure 4.14 and Figure 4.15 show the dependence of ΔV_{th} on the duty cycle of the dynamic stress. In Figure 4.14, the duty cycle dependence is rather weak in the range of $20\% - 80\%$. Due the the limitation of the measurement scheme, it is difficult to investigate the situation for duty cycle $< 20\%$ or $> 80\%$. We are not



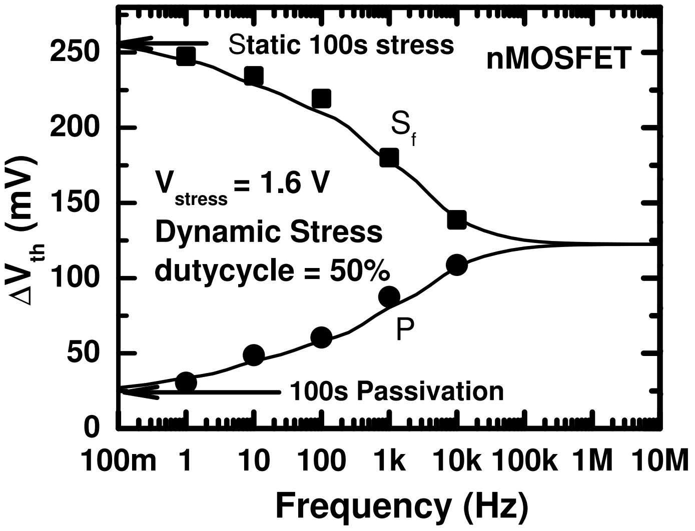
n-MOSFET



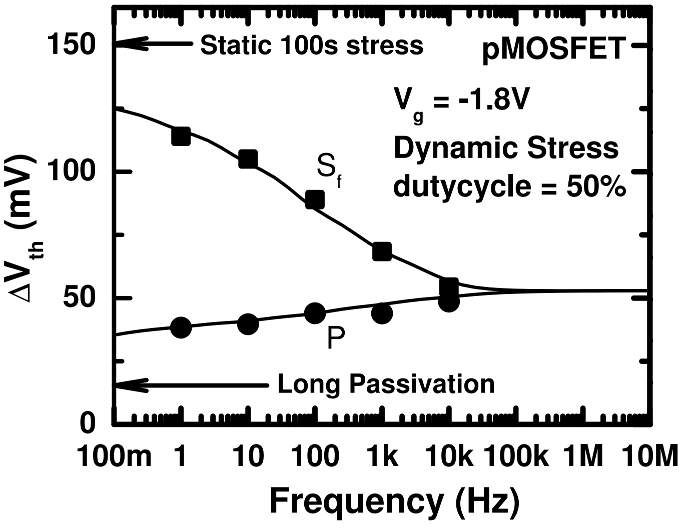
p-MOSFET

Figure 4.11 Threshold voltage shift under static stress and dynamic stress of different frequencies, measured by fast technique. For dynamic stress, both the ΔV_{th} at the end of stress phase (S_f) point and at the end of recovery phase (P) point are plotted.

yet able to understand the weak duty cycle dependence in the middle range and

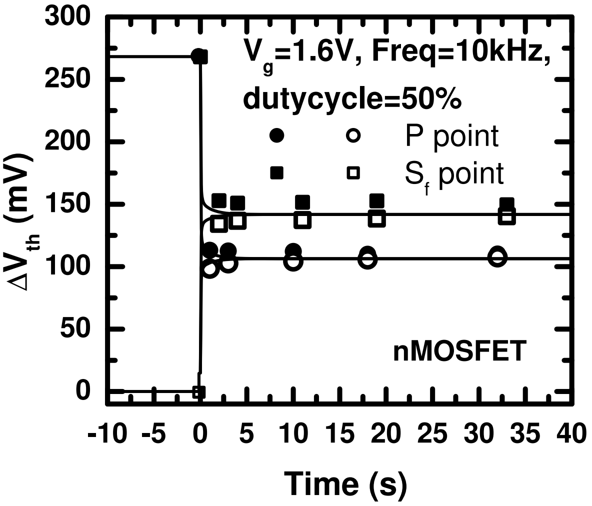


n-MOSFET

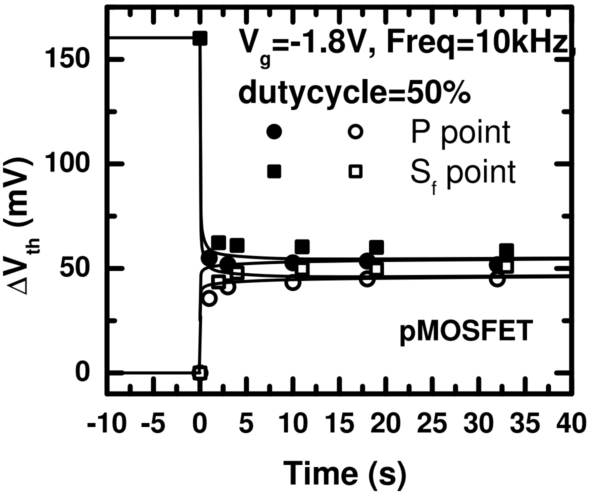


p-MOSFET

Figure 4.12 Frequency dependence of ΔV_{th} after 100 seconds of dynamic stress, using fast measurement. Squares represents the ΔV_{th} at the end of the stress phase in a stress/recovery cycle, while dots represents the ΔV_{th} at the end of the recovery phase. As frequency increases, the amplitude of ΔV_{th} in a stress/recovery cycle (difference between $\mathrm{S_f}$ point and P point) reduces, and the accumulated ΔV_{th} (P point) increases. Solid lines shows simulaton result.



n-MOSFET



p-MOSFET

Figure 4.13 Evolution of ΔV_{th} during transition from static stress to dynamic stress (filled symbols), and ΔV_{th} of fresh device under dynamic stress (open symbols). The steady-state V_{th} shift of dynamic stress does not depend on the prior stress history, which is expected if all the traps are pre-existing. Transition time from static to dynamic stress is in the 100 ms time scale. Solid lines shows simulation result.

the steep dependence at the low and high ends. We suspect that the contribution from the slow charge component may have distorted the duty cycle dependence. From Figure 4.15, the voltage dependence of the dynamic stress of different duty cycle is identical (same slope) to that observed in static stress.

4.3.2.5 Temperature dependence

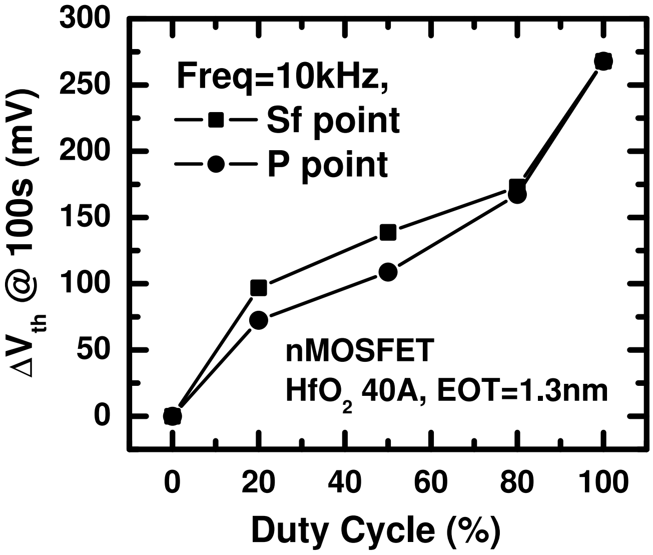
Figure 4.16 shows the temperature dependence of $\Delta V_{\mathrm{th,1s}}$. Negative temperature dependence is observed in the V_{th} shift of n-MOSFETs, while p-MOSFETs show almost no temperature dependence. Therefore, discussion based on room temperature situation represents the worst case.

The net temperature dependence depends on the temperature dependence of both trapping and de-trapping processes. The negative temperature dependence observed in n-MOSFETs may imply that the de-trapping process is more strongly thermal activated.

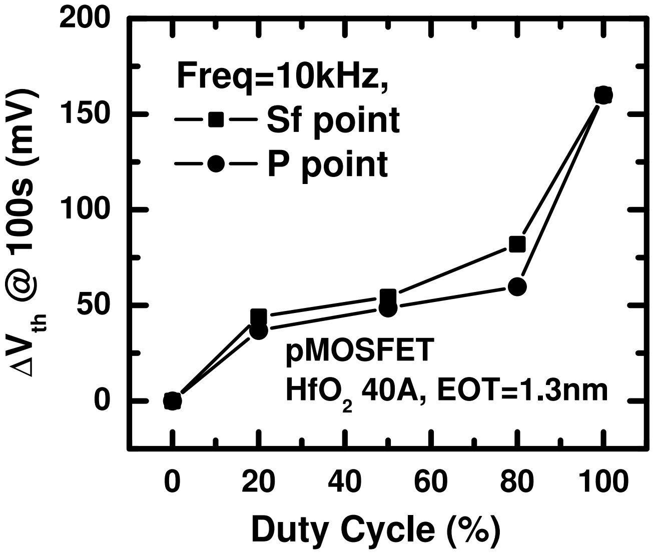
4.3.3 Modeling of the fast V_{th} instability

Empirical models of the voltage and temperature dependence of charge trapping/de-trapping process could be directly obtained from the characterization described above.

On the other hand, the dynamic behavior of the fast component of charge trapping is modeled identical to the hole-trapping model described in section 3.2. However, both electron and hole trapping are considered here. We assumed that N_{ot} has a wide-spreading spectrum of trapping and de-trapping constants τ_{C} and τ_{E} [1], as shown in Figure 4.17. This distribution is obtained by fitting of simulation result to the experimental results. The correct frequency dependence is obtained and is in good agreement with experiments, as shown in Figure 4.7, Figure 4.12, and Figure 4.13 as solid lines.

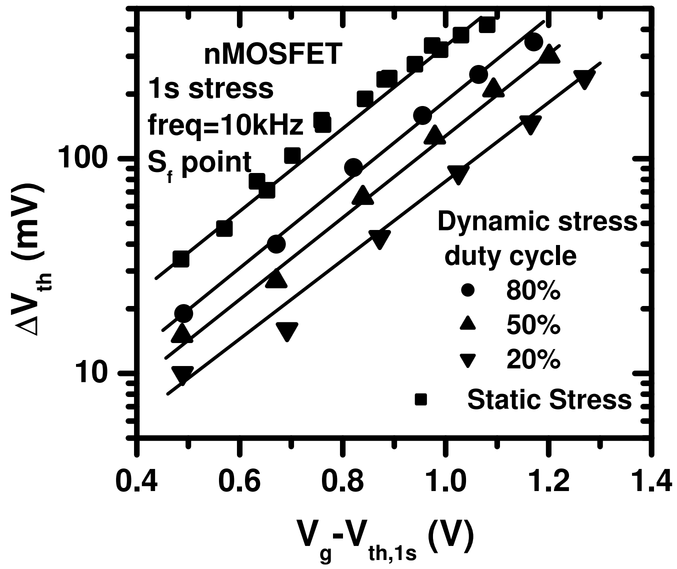


n-MOSFET

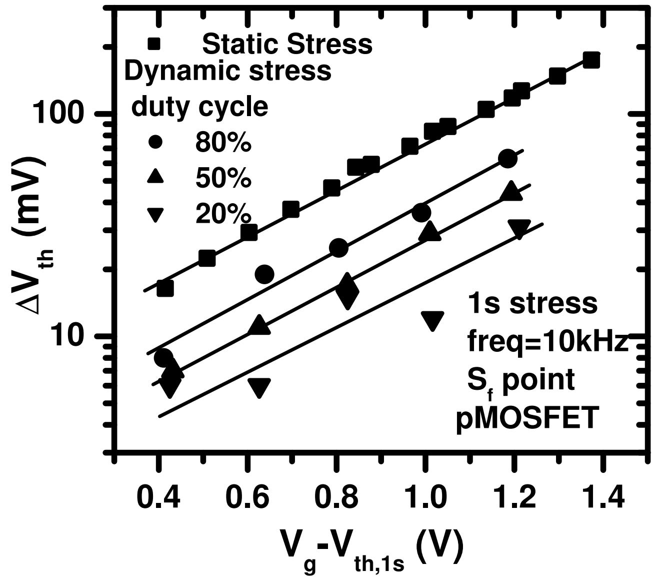


p-MOSFET

Figure 4.14 ΔV_{th} after 100 second dynamic stress of different duty cycle, but same stress voltage, frequency and rise/fall time.

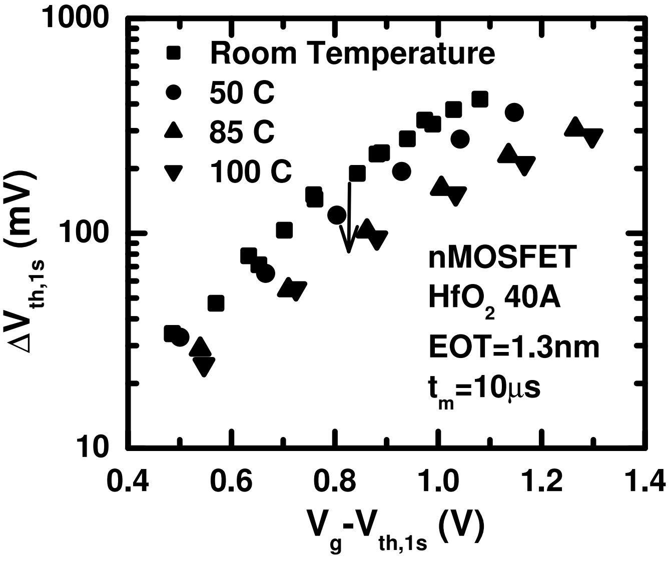


n-MOSFET

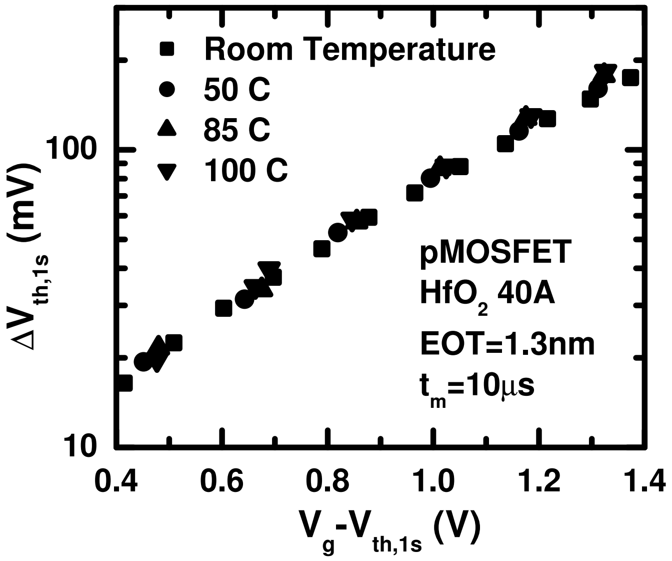


p-MOSFET

Figure 4.15 Stress voltage dependence of ΔV_{th} under static and dynamic stress of different duty cycle. The slope in voltage dependence is the same for static stress and dynamic stress of different duty cycles.

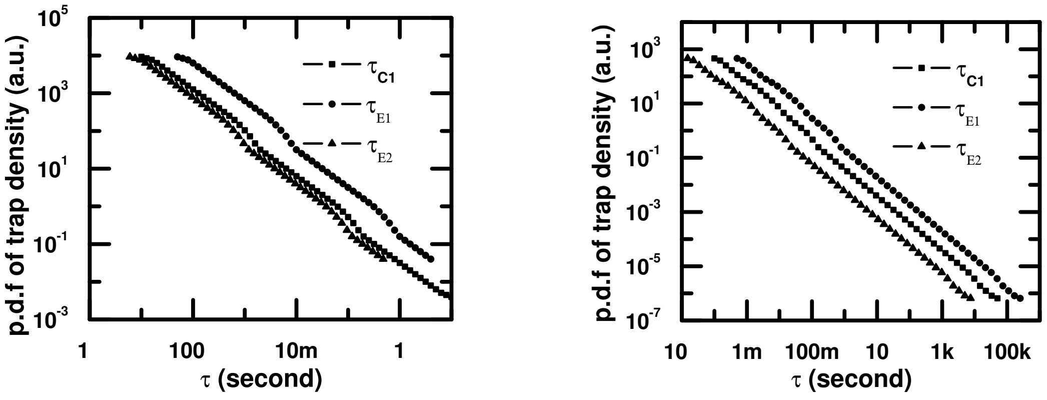


n-MOSFET



p-MOSFET

Figure 4.16 Stress voltage dependence of ΔV_{th}, stressed under different temperatures. NMOSFET shows negative temperature dependence, while pMOSFET shows zero temperature dependence.



(a) (b)

Figure 4.17 Spectrum of trapping (τ_{C}) and de-trapping time constants (τ_{E1}, τ_{E2}) used in calculation of trapping/de-trapping dynamics for (a) electrons and (b) holes. Broad distribution on time constants must be used to explain the gradual frequency dependence observed in Figure 4.12

Under very high frequency, fast charge trapping is predicted to be frequency independent, as shown in Figure 4.12. This prediction enables us to use the experimental data obtained at low frequency to assess the effect of charge trapping in real digital circuit, which operates at much higher frequency.

The recovery in drain current, after removal of stress, is also simulated with the de-trapping time constant distribution obtained above, and is shown in Figure 4.18, which also agrees with the experimental observations qualitatively (Figure 4.8). Note that in Figure 4.8 the gate voltage during recovery is 1 V, while the time constants used here is obtained from the case of 0 V recovery voltage. This difference may partially explain the apparent faster recovery in simulation than in experiment.

4.4 Impacts on digital circuits

After careful characterization and modeling of the fast traps, we are ready to look at its impact on digital circuits. Our discussion is based on both the experimentally measured ΔV_{th} and the high frequency dynamic stress data projected by the charge

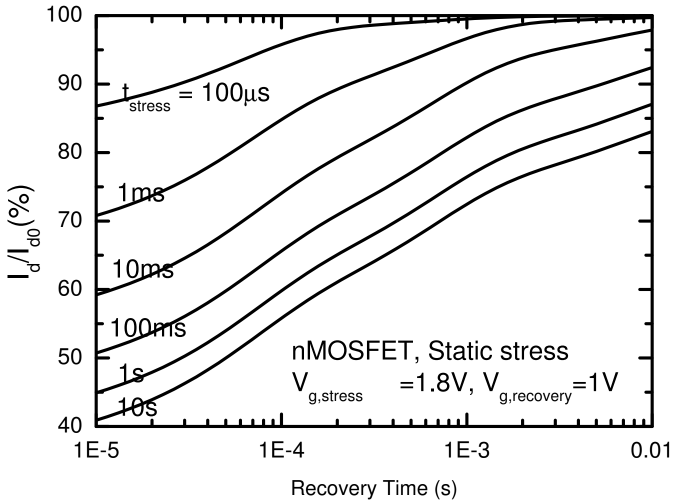


Figure 4.18 Simulation of recovery of linear region drain current I_d after stress voltage is removed from the gate of the nMOSFET, as shown in Figure 4.8.

trapping model. We used HSpice and a 65nm predictive technology model (`vth0` = 0.22 V) [44] for circuit simulation. In the simulation, the parameter `vth0` is changed to reflect the V_{th} shift due to trapping [45]. However, the effect of mobility degradation is ignored.

4.4.1 SRAM

In the widely used dynamic voltage scaled system (DVSS), V_{dd} can be regulated to lower values for low power operation [46]. The static noise margin (SNM) of SRAM can limit the minimum allowed operation voltage. In a 6T SRAM cell shown in Figure 4.19, the worst case for SNM occurs when the storage node x previously stored "1" for long time, and is reading the stored "1". Therefore, the gate voltage for transistors M_{DR} and M_{UL} have been at V_{dd} for long time (statically stressed), and have high V_{th}. The other four transistors in the cell have not been in recovery for long time, and thus have normal V_{th}.

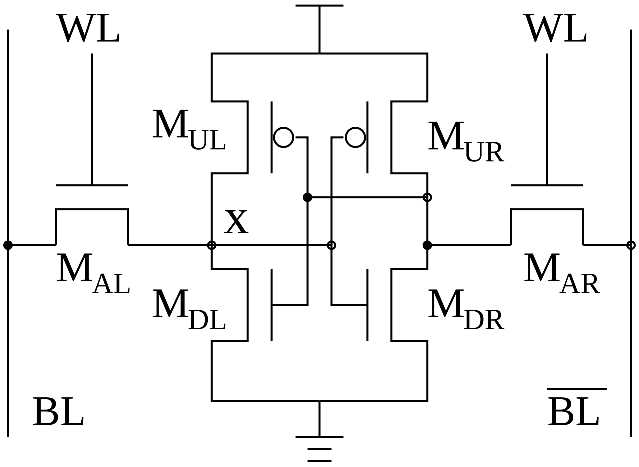


Figure 4.19 Schematics of 6T SRAM cell.

Using the V_{th} shift value extracted from in Figure 4.10, SNM can be obtained from DC circuit simulation. In the butterfly plot shown in Figure 4.20, the increased V_{th} of M_{DR} causes one curve to shift up, and the increased V_{th} of M_{UL} causes one curve to shift to the left. The static noise margin, measured by the maximum square enclosed by the butterfly plot, decreases when the worst case threshold voltage shift is considered.

In Figure 4.21, SNM calculated with and without considering V_{th} degradation is compared, for an SRAM cell with cell ratio $\beta = 2$. The dimensions (W/L) for the cell n-MOSFET is 140/70 nm, p-MOSFET is 90/70 nm, and access transistor is 90/90 nm. As V_{dd} increases, the percentage loss in SNM increases, due to the exponential voltage increase of V_{th} shift. Under reduced V_{dd}, the much reduced fast V_{th} instability causes little SNM degradation. In this low V_{dd} situation, other factors, such as random dopant fluctration and process variation, will dominate the SNM degradation. Therefore, the fast V_{th} instability may not be the limiting factor to the minimum operating voltage of SRAM cell.

4.4.2 Logic circuits

Considering the impact of fast V_{th} shift on ring oscillators, the circuit enjoys the benefit of being dynamically stressed (less ΔV_{th}), and shows improved performance

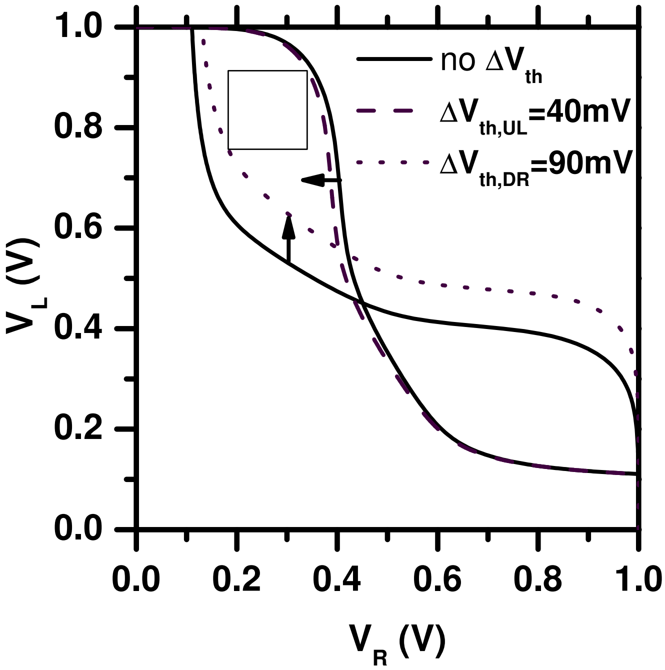


Figure 4.20 Butterfly plot of the SRAM cell showing the transfer characteristics between voltage at the left storage node (V_{L}) and that at the right storage node (V_R). V_{th} increase of $\mathrm{M_{UL}}$ causes the $V_{\mathrm{L}} - V_{\mathrm{R}}$ curve to shift to the left, and V_{th} increase of $\mathrm{M_{DR}}$ causes the $V_{\mathrm{R}} - V_{\mathrm{L}}$ curve to shift up. The static noise margin measured by the maximum square enclosed in the butterfly plot, is therefore reduced.

at high frequency [47]. However, in the most general case in logic circuits, ΔV_{th} under static stress should be used to determine the worst case propagation delay. This is because 1) certain circuit input may be static; 2) When one switches from static stress to dynamic stress, it takes relatively long time for the V_{th} of the MOSFET to transit from the steady-state value under static stress to that under dynamic stress. This transition time is estimated to be about 100 ms from the simulation performed in Figure 4.13, which is many orders of magnitude slower compared with the switching frequency in circuits.

For the 3-input NAND gate (NAND3) implemented with static CMOS logic shown in Figure 4.22a, worst case delay degradation occurs when all three inputs were previously high for long time, all switched to low briefly, and are switching

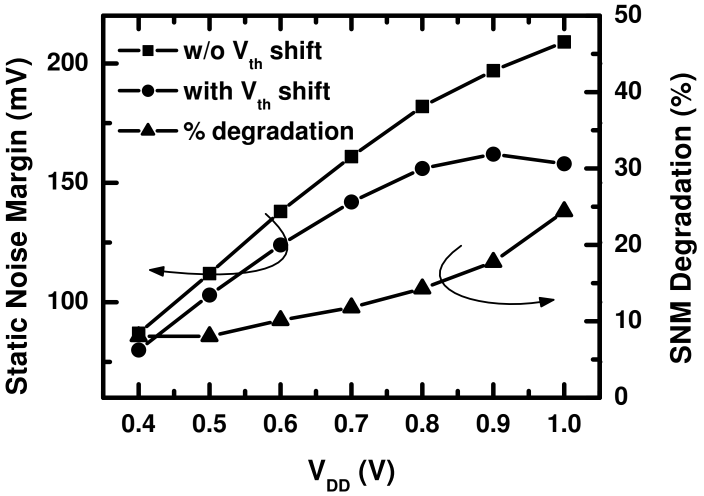


Figure 4.21 Static noise margin of SRAM cell with cell ratio $\beta = 2$. Square symbols show the SNM with no V_{th} degradation, dots show the case with worst case V_{th} degradation, triangles show the percentage loss in SNM due to worst case V_{th} degradation. Under low V_{dd}, the SNM degradation due to fast V_{th} instability is much reduced.

to high at the same time. The V_{th} recovery during the brief low inputs period is negligible. Therefore V_{th} shift under static stress must be used to determine the propagation delay degradation. Delay degradation is plotted as a function of the supply voltage in Figure 4.23. Delay degradation can be well approximated by the degradation in the $I_{d,\text{sat}}$ (due to ΔV_{th}) of the n-MOSFETs[45]. Therefore, other logic gates and circuits implemented with the static logic style would have similar amount of degradation as long as ΔV_{th} of static stress determines the performance.

For CMOS transmission gate, at most one transistor in the pair is under stress at any time, while both transistors conduct when switched on (Figure 4.22b). For example, M_{AP} is under static stress when $\text{S} = 0$, $\text{A} = \text{X} = 1$, but in this situation $\bar{\text{S}} = \text{A} = \text{X} = 1$, so M_{AN} is not under stress. When the transmission gate turns on, both M_{AN} and M_{AP} contribute to the current conduction. If M_{AN} and M_{AP} have equal contribution to the total conductance, the percentage delay degradation is about half of the drive current degradation of M_{AP}. However, the static CMOS inverter for complementary control signal $\bar{\text{S}}$ generation makes the

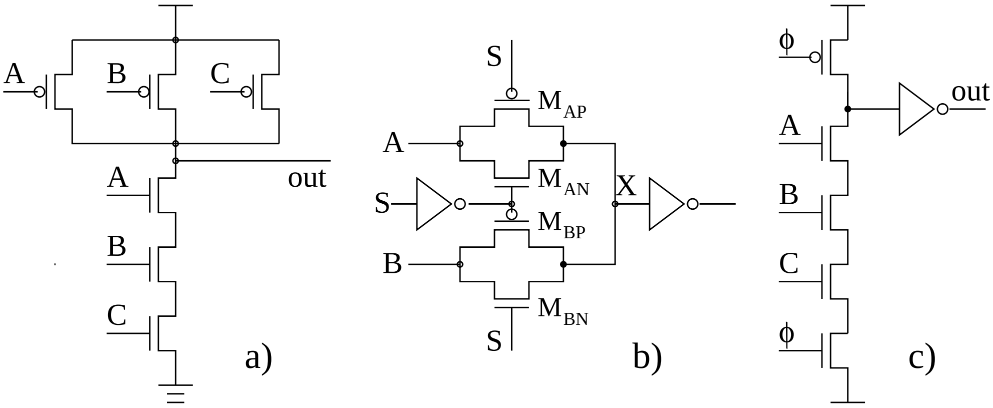


Figure 4.22 Schematics of a) NAND3 gate implemented with static logic, b) 2-input multiplexer implemented with CMOS transmission gate, and c) NAND3 gate implemented with dynamic logic.

delay degradation closer to the static logic case. Overall, CMOS transmission gate logic (MUX2) shows less delay degradation than NAND3 (Figure 4.23), although ΔV_{th} under static stress is considered.

For dynamic logic (Figure 4.22c), transistors are all dynamically stressed at the clock frequency with a pre-defined duty cycle. In pre-charge phase, clock ϕ is low, all the four n-MOSFETs are cutoff (de-trapping), regardless of the values at the three inputs. Therefore, the n-MOSFETs in the pull-down network are stressed dynamically with duty cycle not greater than the clock duty cycle. In the buffer inverter that feeds to the next stage, only low-to-high transition need to be optimized, so the pull-up p-MOSFET is more important. The p-MOSFET is also stressed dynamically, and the stress duty cycle is no worse than the clock duty cycle. Therefore, V_{th} shift under dynamic stress represents the worst case for dynamic logic circuits in terms of propagation delay. With the much lower V_{th} shift under high frequency dynamic stress (Figure 4.12), the delay degradation of the dynamic NAND3 gate is much lower than the static logic. As the most delay sensitive parts of the circuits are often implemented with dynamic logic, the degraded delay of static CMOS logic in non-critical path might not be the limiting factor of overall circuit speed.

It is not possible to exhaustively examine all building blocks in logic circuits in this work. However it is demonstrated that the use of certain circuit designs can take advantage of the dynamic nature of charge trapping effect in a deterministic manner. This potentially allow the performance of logic circuit to be less affected by fast V_{th} instability, and thus impose less stringent target to the optimization of high-κ dielectrics.

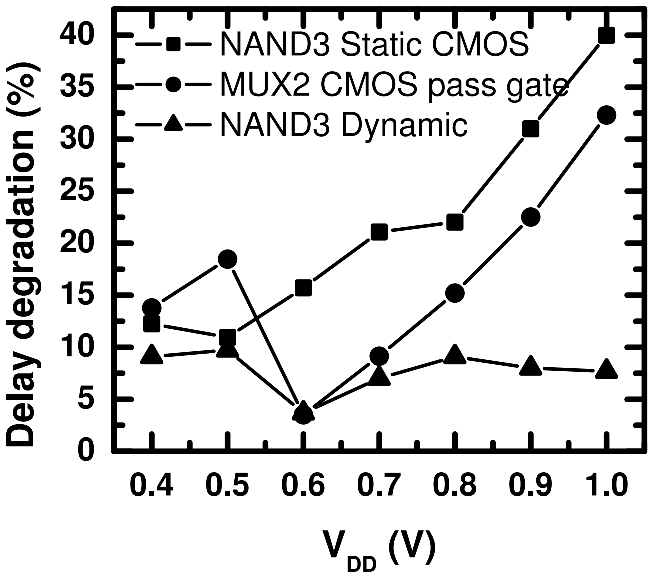


Figure 4.23 Percentage increase in gate propagation delay due to V_{th} degradation as function of supply voltage. Logic gates implemented with transmission gate and dynamic logic shows much reduced degradation in delay.

4.5 Conclusions

Both the slow and fast components of V_{th} in MOSFET with HfO_2 gate dielectric are systematically characterized and modeled. The slow component is shown to require a new two-step charge trapping dynamic model to explain the unexpected frequency dependence. It is proposed that negative-U defects are responsible for such trapping dynamics.

On the other hand, the large ΔV_{th} due to fast charge trapping is shown to be predictable in both static and dynamic stress situations. With the knowledge gained from the quantitative modeling of the fast V_{th} transients, HSpice circuit simulation is performed based on experimental ΔV_{th} to evaluate the impact of V_{th} shift on the performance of digital circuits in more realistic operating conditions. It is shown that one can actively exploit the dynamic nature of the fast V_{th} instability to minimize its effect on circuit performance. Circuit performance should therefore be optimized with both process improvement and circuit design techniques.

References

[1] Sufi Zafar, Alessandro Callegari, Evgeni Gusev and Massimo V. Fischetti. Charge trapping in high-κ gate dielectric stacks. *Technical Digest - International Electron Devices Meeting*, pages 517 - 520, 2002.

[2] A. Shanware, M.R. Visokay, J.J. Chambers, A.L.P. Rotondaro, H. Bu, M.J. Bevan, R. Khamankar, S. Aur, P.E. Nicollian, J. McPherson and L. Colombo. Evaluation of the positive biased temperature stress stability in HfSiON gate dielectrics. In *IRPS*, pages 208 - 213, Dallas, TX, United States, 2003.

[3] K. Onishi, Rino Choi, Chang Seok Kang, Hag-Ju Cho, Young Hee Kim, R.E. Nieh, Jeong Han, S.A. Krishnan, M.S. Akbar and J.C. Lee. Bias-temperature instabilities of polysilicon gate HfO_2 MOSFETs. *IEEE Trans. Electron Devices*, 50(6):1517 - 24, 2003.

[4] A. Kerber, E. Cartier, L. A. Ragnarsson, M. Rosmeulen, L. Pantisano, R. Degraeve, T. Kauerauf, G. Groeseneken, H. E. Maes and U. Schwalke. Characterization of the VT-instability in SiO_2/HfO_2 gate dielectrics. In *Reliability Physics Symposium Proceedings, 2003 IEEE International*, pages 41-45, 2003.

[5] C. Shen, H. Y. Yu, X. P. Wang, M.-F. Li, Y.-C. Yeo, D. S. H. Chan, K. L. Bera and D. L. Kwong. Frequency dependent dynamic charge trapping in HfO_2 and threshold voltage instability in MOSFETs. In *IRPS*, pages 601-602, 2004.

[6] 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 HfO_2 gate dielectrics and frequency dependence of dynamic BTI in MOSFETs. In *IEDM Tech. Digest*, pages 733-736, 2004.

[7] B.H. Lee, C.D. Young, R. Choi, J.H. Sim and G. Bersuker et al.. Intrinsic characteristics of high-k devices and implications of fast transient charging effects (FTCE). *Technical Digest - International Electron Devices Meeting, IEDM*, pages 859 - 862, 2004.

[8] A. Shanware, M.R. Visokay, J.J. Chambers, A.L.P. Rotondaro, J. McPherson, L. Colombo, G.A. Brown, C.H. Lee, Y. Kim, M. Gardner and R.W. Murto. Characterization and comparison of the charge trapping in HfSiON and HfO_2 gate dielectrics. In *IEDM*, pages 939 - 942, Washington, DC, United States, 2003.

[9] H. Reisinger, G. Steinlesberger, S. Jakschik, M. Gutsche, T. Hecht, M. Leonhard, U. Schroder, H. Seidl and D. Schumann. A comparative study of dielectric relaxation losses in alternative dielectrics. *International Electron Devices Meeting. Technical Digest*, pages 12 - 2, 2001.

[10] J.R. Jameson, P.B. Griffin, A. Agah, J.D. Plummer, H.-S. Kim, D.V. Taylor, P.C. McIntyre and W.A. Harrison. Problems with metal-oxide high-κ dielectrics due to 1/t dielectric relaxation current in amorphous materials. *IEEE International Electron Devices Meeting 2003*, pages 4-3, 2003.

[11] Zhen Xu, L. Pantisano, A. Kerber, R. Degraeve, E. Cartier, S. De Gendt, M. Heyns and G. Groeseneken. A study of relaxation current in high-κ dielectric stacks. *IEEE Trans. Electron Devices*, 51(3):402–408, 2004.

[12] S. Zafar, B.H. Lee and J. Stathis. Evaluation of nbti in HfO_2 gate-dielectric stacks with tungsten gates. *IEEE Electon Device Letters*, 25(3):153–155, 2004.

[13] M. Houssa, M. Aoulaiche, S. Van Elshocht, S. De Gendt, G. Groeseneken and M. M. Heyns. Impact of hf content on negative bias temperature instabilities in hfsion-based gate stacks. *Applied Physics Letters*, 86(17):173509, 2005.

[14] G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, E. Vincent and G. Ghibaudo. Review on high-k dielectrics reliability issues. *Device and Materials Reliability, IEEE Transactions on*, 5(1):5–19, 2005.

[15] A. Morioka, H. Watanabe, M. Miyamura, T. Tatsumi, M. Saitoh, T. Ogura, T. Iwamoto, T. Ikarashi, Y. Saito, Y. Okada, H. Watanabe, Y. Mochiduki and T. Mogami. High mobility MISFET with low trapped charge in HfSiO films. *2003 Symposium on VLSI Technology. Digest of Technical Papers*, pages 165 - 6, 2003.

[16] Xiongfei Yu, Chunxiang Zhu, X.P. Wang, M.F. Li, Albert Chin, A.Y. Du, W.D. Wang and Dim-Lee Kwong. High mobility and excellent electrical stability of MOSFETs using a novel HfTaO gate dielectric. In *VLSI Tech. Symp.*, pages 110 - 111, Honolulu, HI, United States, 2004.

[17] X. P. Wang, Ming-Fu Li, C. Ren, X. F. Yu, C. Shen, H. H. Ma, A. Chin, C. X. Zhu, Jiang Ning, M. B. Yu and Dim-Lee Kwong. Tuning effective metal gate work function by a novel gate dielectric HfLaO for nMOSFETs. *Electron Device Letters, IEEE*, 27(1):31-33, 2006.

[18] E.P. Gusev, V. Narayanan, S. Zafar, C. Cabral Jr., E. Cartier, N. Bojarczuk, A. Callegari, R. Carruthers, M. Chudzik, C. D'Emic, E. Duch, P. Jamison, P. Kozlowski, D. LaTulipe, K. Maitra, F.R. McFeely, J. Newbury, V. Paruchuri and M. Steen. Charge trapping in aggressively scaled metal gate/high-κ stacks. *Technical Digest - International Electron Devices Meeting, IEDM*, pages 729 - 732, 2004.

[19] H.-H. Tseng, M.E. Ramon, L. Hebert, P.J. Tobin, D. Triyoso, J.M. Grant, Z.X. Jiang, D. Roan, S.B. Samavedam, D.C. Gilmer, S. Kalpat, C. Hobbs, W.J. Taylor, O. Adetutu and B.E. White. ALD HfO_2 using heavy water (D_2O) for improved MOSFET stability. *IEEE International Electron Devices Meeting 2003*, pages 4 - 1, 2003.

[20] H. H. Tseng, P. J. Tobin, E. A. Hebert, S. Kalpat and M. E. Ramon et al.. Defect passivation with fluorine in a Ta_xC_y/high-k gate stack for enhanced device threshold voltage stability and performance. In *IEDM Tech. Dig.*, 2005.

[21] M. Inoue, S. Tsujikawa, M. Mizutani, K. Nomura, T. Hayashi, K. Shiga, J. Yugami, J. Tsuchimoto, Y. Ohno and M. Yoneda. Fluorine incorporation into HfSiON dielectric for V_{th} control and its impact on reliability for poly-Si gate pFET. In *IEDM Tech. Dig.*, pages 17.1, 2005.

[22] Da-Yuan Lee, Tiao-Yuan Huang, Horng-Chih Lin, Wan-Ju Chiang, Guo-Wei Huang and Tahui Wang. Effects of process and gate doping species on negative- bias-temperature instability of p-channel mosfets. *Journal of The Electrochemical Society*, 151(2):G144-G148, 2004.

[23] T. Yang, M. F. Li, C. Shen, C. H. Ang, Chunxiang Zhu, Y.-C. Yeo, G. Samudra, S. C. Rustagi, M. B. Yu and D. L. Kwong. Fast and slow dynamic NBTI components in p-MOSFET with SiON dielectric and their impact on device life-time and circuit application. In *VLSI Symp.*, pages 92-93, 2005.

[24] T. Yang, C. Shen, M.F. Li, C.H. Ang, C.X. Zhu, Y.-C. Yeo, G. Samudra, S.C. Rustagi, M.B. Yu and D.-L. Kwong. Fast DNBTI components in p-MOSFET with SiON dielectric. *IEEE Electron Device Letters*, 26(11):826–828, 2005.

[25] M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, Y. Rey-Tauriac and N. Revil. On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's. In *IEDM Tech. Digest*, pages 109-112, 2004.

[26] Se Jong Rhee, Young Hee Kim, Chang Yong Kang, Chang Seok Kang, Hag-Ju Cho, Rino Choi, Chang Hwan Choi, Mohammad S. Akbar and Jack C. Lee. Dynamic positive bias temperature instability characteristics of ultra-thin HfO_2 NMOSFET. In *IRPS*, pages 269 - 272, Phoenix, AZ., United States, 2004.

[27] H. Y. Yu, J. F. Kang, J. D. Chen, C. Ren, Y. T. Hou, S. J. Whang, M. F. Li, D. S. H. Chan, K. L. Bera, C. H. Tung, A. Du and D.-L. Kwong. Thermally robust high quality HfN/HfO_2 gate stack for advanced CMOS devices. In *IEDM Tech. Digest*, pages 99-102, 2003.

[28] 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. In *Reliability Physics Symposium Proceedings, 2003. 41st Annual. 2003 IEEE International*, pages 196–202, 2003.

[29] D.J. Chadi and K.J. Chang. Theory of the atomic and electronic structure of DX centers in GaAs and $Al_xGa_{1-x}As$ alloys. *Phys. Rev. Lett.*, 61(7):873 - 876, 1988.

[30] D. J. Chadi. *22nd Int. Conf. on the Physics of Semiconductors*, chapter Vol. 3, pages 2331. World Scientific, Singapore, 1995.

[31] M.-F. Li. *Modern Semiconductor Quantum Physics*, chapter sec.392, pages 316. World Scientific, Singapore, 1994.

[32] D. V. Lang. *Deep Centers in Semiconductors*, chapter Chapter 7, pages 489. Gorden and Breach Sci. Pub., New York, 1986.

[33] P.E. Blochl and J.H. Stathis. Hydrogen electrochemistry and stress-induced leakage current in silica. *Phys. Rev. Lett. (USA)*, 83(2):372 - 375, 1999.

[34] A. Yokozawa, A. Oshiyama, Y. Miyamoto and S. Kumashiro. Oxygen vacancy with large lattice distortion as an origin of leakage currents in SiO_2. In *IEDM*, pages 703 - 706, Washington, DC, USA, 1997.

[35] Joongoo Kang, E.-C. Lee, K. J. Chang and Young-Gu Jin. H-related defect complexes in HfO_2: A model for positive fixed charge defects. *Applied Physics Letters*, 84(19):3894–3896, 2004.

[36] Y. P. Feng, A. T. L. Lim and M. F. Li. Negative-u property of oxygen vacancy in cubic HfO_2. *Applied Physics Letters*, 87(6):062105, 2005.

[37] J. L. Gavartin, D. Munoz Ramo, A. L. Shluger, G. Bersuker and B. H. Lee. Negative oxygen vacancies in HfO_2 as charge traps in high-k stacks. *Applied Physics Letters*, 89(8):082908, 2006.

[38] C. D. Young, Y. G. Zhao, M. Pendley, B. H. Lee, K. Matthews, J. H. Sim, R. Choi, G. Bersuker and G. A. Brown. Ultra-short pulse I-V characterization of the intrinsic behavior of high-κ devices. In *SSDM Proc.*, pages 216-217, 2004.

[39] C. T. Chan, C. J. Tang, C. H. Kuo, H. C. Ma, C. W. Tsai, H. C.-H. Wang, M. H. Chi and Taihui Wang. Single-electron emission of traps in HfSiON as high-k gate dielectric for MOSFETs. In *IRPS*, pages 41-44, 2005.

[40] C. T. Chan, C. J. Tang, Tahui Wang, H. C.-H. Wang and D. D. Tang. Positive bias and temperature stress induced two-stage drain current degradation in HfSiON nMOSFETs. In *IEDM Tech. Dig.*, pages 22.7, 2005.

[41] A. Kerber, E. Cartier, L. Pantisano, R. Degraeve, T. Kauerauf, Y. Kim, A. Hou, G. Groeseneken, H.E. Maes and U. Schwalke. Origin of the threshold voltage instability in SiO_2/HfO_2 dual layer gate dielectrics. *IEEE Electron Device Lett.*, 24(2):87 - 9, 2003.

[42] Y. Nissan-Cohen, J. Shappir and D. Frohman-Bentchkowsky. Dynamic model of trapping-detrapping in SiO_2. *J. Appl. Phys.*, 58(6):2252–2261, 1985.

[43] C. Shen, M. F. Li, H. Y. Yu, X. P. Wang, Y.-C. Yeo, D. S. H. Chan and D.-L. Kwong. Physical model for frequency-dependent dynamic charge trapping in metal-oxide-semiconductor field effect transistors with HfO_2 gate dielectric. *Applied Physics Letters*, 86(9):093510, 2005.

[44] Y. Cao, T. Sato, D. Sylvester, M. Orshansky and C. Hu. New paradigm of predictive MOSFET and interconnect modeling for early circuit design. In *Proc. CICC*, pages 201–204, 2000. [Online]. Available: http://www.eas.asu.edu/~ptm/.

[45] A.T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John and S. Krishnan. NBTI impact on transistor and circuit: models, mechanisms and scaling effects. In *IEEE International Electron Devices Meeting 2003*, pages 14-5, Washington, DC, USA, 2003.

[46] T.D. Burd, T.A. Pering, A.J. Stratakos and R.W. Brodersen. A dynamic voltage scaled microprocessor system. *IEEE J. Solid-State Circuits*, 35(11):1571–1580, 2000.

[47] C.Y. Kang, J.C. Lee, R. Choi, J.H. Sim, C. Young, B.H. Lee and G. Bersuker. Charge trapping effects in HfSiON dielectrics on the ring oscillator circuit and the single stage inverter operation. In *2004 International Electron Devices Meeting*, pages 485–488, San Francisco, CA, USA, 2004.

CHAPTER

5

Summary

In this work, the threshold voltage instabilities in MOSFETs with two kinds of advanced gate dielectrics were investigated. Recounting on the making of this thesis, it is recognized that two research themes evolved in a spiral. One theme is the continual development of electrical measurement techniques to accurately characterize the dynamics of V_{th} instabilities. The other theme, on the other hand, is to compare the observed dynamics with the predictions of existing models, and to build new models when necessary.

As the reader may have noticed, this thesis relies heavily on the fast measurement technique proposed in chapter 2. The development of the fast $I_d - V_g$ measurement technique started in 2004, following the pioneering work of A. Kerber. Initially, the best measurement time was 1 μs, and curve-smoothing is required due to high noise. The measurement setup was then refined, through a few iterations, to the present form, as presented in chapter 2. It was realized that the frequency response of the transimpedance amplifier in our setup is sensitive to the parasitic capacitance connected to the amplifier input. Much effort was therefore spent on minimizing the distance from the probe tip to the amplifier input, which lead to the latest setup using probe-card. The measurement time was reduced to 100 ns, and noise was suppressed. It is now recognized that the fast measurement technique is indispensable to researchers in this field as an accurate and reliable characterization technique. It is the hope of the author that it can be further improved in terms of speed and stability, and made accessible to a wider range of researchers.

With the accurate characterization tool, comprehensive measurements were carried out to characterize the V_{th} instabilities in transistors with both SiON and high-κ gate dielectrics. However, in the interpretation of the experimental data, more attention was paid to the dynamics aspect. In one way, this was motivated by the traditional paradigm of lifetime projection, where one extrapolate from the experimentally measured V_{th} shift and estimate the final V_{th} shift after a ten-year lifetime. A model of the dynamics of the V_{th} instabilities is necessary to make projections of this type. A physical model that predicts a compatible dynamics is sought not only out of our curiosity, but also to rationalize the lifetime projection exercise.

Both analytic and numerical calculations were performed to obtain quantitative predictions from various models of V_{th} instabilities, so that the experimental data can be compared against these models. The coupled reaction-diffusion equation was previously solved for the case of static stress, and in the limit of diffusion-limited process. In this work, an analytic treatment of the reaction-diffusion problem under dynamic stress condition was given under suitable approximations. The dynamic charge trapping problem was also treated analytically, for the second-order two-step trapping model as well as the first-order trapping model.

In our experimental studies, it happened that the observed dynamics do not agree with the most accepted models. In the case of SiON gate dielectrics, very low power-law exponent and very fast recovery of V_{th} shift were observed experimentally, with the help of the newly developed fast measurement technique. Both characteristics are opposite to the predictions of the traditional reaction-diffusion model. We showed that this incompatibility is a fundamental one, not likely to be resolved with simple amendments to the model. It was proposed that in addition to the generation of interface-states, which is a relatively slow process, fast hole trapping and de-trapping is present as well. The model based on hole trapping predicts the correct dynamic behavior as observed experimentally.

In the case of high-κ dielectrics, two distinct components were observed in the charge trapping as well. One component has fast trapping and de-trapping time constants, and is similar to the hole trapping process found in SiON. The slower component, however exhibits a peculiar dependence on the frequency of the dynamic stress voltage. This frequency dependence requires a second-order (or high-order) dynamics, and is not expected in the traditional charge-trapping model. A two-step charge trapping model was proposed to explain this characteristics. It was speculated that negative-U traps exists in the HfO_2 dielectric film, because charge trapping at these defects are naturally a two-step process, and may explain the observed frequency dependence. This hypothesis is partly backed by a few *ab initio* calculations of the oxygen vacancy in HfO_2 films. These new discoveries on the dynamics of V_{th} instabilities should be considered in the projection of V_{th} degradation after aging, and shed new light on the possible physical structures and processes associated with the observed V_{th} degradation.

Another subject studied in this work is the impact of the V_{th} instabilities on circuits. We focused on the fast charge trapping/de-trapping transients found in transistors with high-κ dielectrics, because its fast dynamics and large magnitude were new to the community. Based on the dynamic model we constructed for the fast V_{th} transients, the effect on digital circuits were analyzed, with the dynamic nature of the V_{th} instabilities taken into consideration. It was demonstrated on a few common combinational logic gates that the delay degradation due to V_{th} instability could be minimized with the appropriate choice of logic styles.

The study of NBTI in SiON sees a lot of new findings every year. After the 2007 IEEE Electron Device Meetings at Washington, DC, the author sees that a few topics need to be re-examined, in the light of some new observations. The author regrets for not being able to pursue these new investigations. However, it has been made clear from this thesis that the fast V_{th} instability observed by all researchers can not be explained solely with the reaction-diffusion model. This conclusion is increasingly becoming the current consensus among researchers. The hole trapping model favored by the author seems to provide a good account for a lot of experimental observations, but the arguments is mainly based on its dynamics, while evidence from physical characterization is missing. It is in some way unfair to call it hole trapping model, but the name “a model with trapping-like dynamics” is too awkward, so we caution the readers here, and continue to call it the hole trapping model until the actual physical structure is identified.

As Intel and a few other manufacturers revealed some technical details of the first generation of IC using high-κ and metal gate, the research on reliability of high-κ dielectrics entered a new stage, and would be more aligned to the industry practice. The chemical composition, growth technology and thermal processes of the HfO_2 film studied in this thesis (3 to 4 years ago) is different from what is currently adopted by the industry. Whether the phenomenon and mechanism revealed in our study is relevant to current or future technology is left to be tested. However, much of the methodology should remain valid. In particular, the discussions on the effect of fast V_{th} transients on digital circuit is, to the knowledge of the author, the first investigation of this kind, and the methodology and conclusions hold regardless of the dielectric film actually used.

List of Publications

Regular Paper

[1] M. F. Li, G. Chen, **C. Shen**, X. P. Wang, H. Y. Yu, Y.-C. Yeo and D. L. Kwong. Dynamic bias-temperature instability in ultrathin SiO_2 and HfO_2 metal-oxide-semiconductor field effect transistors and its impact on device lifetime. *Japanese Journal of Applied Physics, Part 1: Regular Papers and Short Notes and Review Papers*, 43(11 B):7807 - 7814, 2004. (invited).

[2] **C. Shen**, T. Yang, M.-F. Li, X. P. Wang, C. E. Foo, G. Samudra, Y.-C. Yeo and D.-L. Kwong. Fast V_{th} instability in HfO_2 gate dielectric MOSFETs and its impact on digital circuits. *IEEE Transaction Electron Devices*, 53:3001–3011, 2006.

[3] **C. Shen**, J. Pu, M.-F. Li and B. J. Cho. P-type floating gate for rentention and P/E window improvement of flash memory devices. *IEEE Transaction Electron Devices*, 54:1910–1917, 2007.

[4] M.-F. Li, D. Huang, **C. Shen**, T. Yang, W. J. Liu and Z. Y. Liu. Understand NBTI mechanism by developing novel measurement techniques. 2007. accepted for publication at IEEE Transaction Devices and Materials Reliability (invited).

Letter or Brief

[1] **C. Shen**, M.-F. Li, H. Y. Yu, X. P. Wang, Y.-C. Yeo, D. S. H. Chan and D.-L. Kwong. Physical model for frequency-dependent dynamic charge trapping in metal-oxide-semiconductor field effect transistors with HfO_2 gate dielectric. *Applied Physics Letters*, 86(9):093510, 2005.

[2] T. Yang, **C. Shen**, M.-F. Li, C. H. Ang, C. Zhu, Y.-C. Yeo, G. Samudra, S. C. Rustagi, M. B. Yu and D.-L. Kwong. Fast DNBTI component in p-MOSFET with SiON dielectric. *IEEE Electron Device Letters*, 26:826–828, 2005.

[3] T. Yang, **C. Shen**, M.-F. Li, C. H. Ang, C. Zhu, Y.-C. Yeo, G. Samudra, S. C. Rustagi, M. B. Yu and D.-L. Kwong. Interface trap passivation effect in NBTI measurement for p-MOSFET with SiON gate dielectric. *IEEE Electron Device Letters*, 26:758–760, 2005.

[4] X. P. Wang, M.-F. Li, C. Ren, X. F. Yu, **C. Shen**, H. H. Ma, A. Chin, C. X. Zhu, J. Ning, M. B. Yu and D.-L. Kwong. Tuning effective metal gate work function by a novel gate dielectric HfLaO for nMOSFETs. *Electron Device Letters, IEEE*, 27(1):31-33, 2006.

[5] **C. Shen**, M.-F. Li, X. P. Wang, Y.-C. Yeo and D.-L. Kwong. A fast measurement technique for MOSFET $I_d - V_g$ characteristics. *IEEE Electron Device Letters*, 27(1):55–57, 2006.

[6] K.-J. Chui, K.-W. Ang, H.-C. Chin, **C. Shen**, L.-Y. Wong, C.-H. Tung, N. Balasubramanian, M.-F. Li, G. S. Samudra and Y.-C. Yeo. Strained-SOI n-channel transistor with silicon-carbon source/drain regions for carrier transport enhancement. *IEEE Electron Device Letters*, 27(9):778–780, 2006.

[7] J. D. Chen, X. P. Wang, M.-F. Li, S. J. Lee, M. B. Yu, **C. Shen** and Y.-C. Yeo. NMOS compatible work function of TaN metal gate with erbium oxide doped hafnium oxide gate dielectric. *IEEE Electron Device Letters*, 28(10):862-864, 2007.

Conference

[1] **C. Shen**, H. Y. Yu, X. P. Wang, M.-F. Li, Y.-C. Yeo, D. S. H. Chan, K. L. Bera and D.-L. Kwong. Frequency dependent dynamic charge trapping in HfO_2 and threshold voltage instability in MOSFETs. In *IEEE International Reliability Physics Symposium*, pages 601–602, 2004.

[2] M. F. Li, H. Y. Yu, Y. T. Hou, J. F. Kang, X. P. Wang, **C. Shen**, C. Ren, Y. C. Yeo, C. X. Zhu, D. S. H. Chan, A. Chin and D. L. Kwong. Selected topics on HfO_2 gate dielectrics for future ULSI CMOS devices. *International Conference on Solid-State and Integrated Circuits Technology Proceedings, ICSICT*, 1:366 - 371, 2004.

[3] **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 HfO_2 gate dielectrics and frequency dependence of dynamic BTI in MOSFETs. In *IEEE International Electron Devices Meeting*, pages 733–736, 2004.

[4] T. Yang, M.-F. Li, **C. Shen**, C. H. Ang, C. Zhu, Y.-C. Yeo, G. Samudra, S. C. Rustagi, M. B. Yu and D.-L. Kwong. Fast and slow dynamic NBTI components in p-MOSFET with SiON dielectric and their impact on device life-time and circuit application. In *VLSI Technology Symposium*, pages 92–93, 2005.

[5] M.-F. Li, C. Zhu, **C. Shen**, X. F. Yu, X. P. Wang, Y. P. Feng, A. Y. Du, Y. C. Yeo, G. Samudra, A. Chin and D. L. Kwong. New insights in Hf based high-k gate dielectrics in MOSFETs. In *208th Electrochecmical Society Meeting*, 2005. (invited).

[6] N. Wu, Q. Zhang, C. Zhu, **C. Shen**, M. F. Li, D. S. H. Chan and N. Balasubramanian. BTI and charge trapping in germanium p- and n-MOSFETs with CVD HfO_2 gate dielectric. In *IEEE International Electron Devices Meeting*, pages 22.5, 2005.

[7] X. P. Wang, M.-F. Li, A. Chin, C. Zhu, C. Ren, X. F. Yu, **C. Shen**, A. Y. D. Du, D. S. H. Chan and D.-L. Kwong. A new gate dielectric HfLaO with metal gate work function tuning capability and superior NMOSFETs performance. In *International Semiconductor Device Research Symposium*, pages 2, 2005.

[8] X. P. Wang, **C. Shen**, M.-F. Li, H. Y. Yu, Y. Sun, Y. P. Feng, A. Lim, H. W. Si, A. Chin, Y.-C. Yeo, P. Lo and D.-L. Kwong. Dual metal gate with band-edge work function on novel HfLaO high-k gate dielectric. In *VLSI Technology Symposium*, pages T2-2, 2006.

[9] **C. Shen**, T. Yang, M.-F. Li, G. Samudra, Y.-C. Yeo, C. Zhu, S. C. Rustagi, M. B. Yu and D.-L. Kwong. Fast V_{th} instability in HfO_2 gate dielectric MOSFETs and its impact on digital circuits. In *IEEE International Reliability Physics Symposium*, pages 653–654, 2006.

[10] **C. Shen**, M.-F. Li, C. E. Foo, T. Yang, D. M. Huang, A. Yap, G. Samudra and Y.-C. Yeo. Characterization and physical origin of fast V_{th} transient in NBTI of pMOSFETs with sion dielectric. In *IEEE International Electron Devices Meeting*, pages 333–336, 2006.

[11] Y. Q. Wang, D. Y. Gao, W. S. Hwang, **C. Shen**, G. Zhang, G. Samudra, Y.-C. Yeo and W. J. Yoo. Fast erasing and highly reliable MONOS type memory with HfO_2 high-k trapping layer and Si_3N_4/SiO_2 tunneling stack. In *IEEE International Electron Devices Meeting*, pages 971–974, 2006.

[12] W. S. Hwang, **C. Shen**, X. P. Wang, D. S. H. Chan and B. J. Cho. A novel hafnium carbide (HfC_x) metal gate electrode for NMOS device application. In *VLSI Technology Symposium*, pages 9A-2, 2007.

[13] **C. Shen**, E.-H. Toh, J. Lin, C.-H. Heng, D. Sylvester, G. Samudra and Y.-C. Yeo. A physics-based compact model for I-MOS transistors. In *International Conference on Solid State Devices and Materials*, pages 608–609.

[14] **C. Shen**, J. Lin, E.-H. Toh, K.-F. Chang, P. Bai, C.-H. Heng, G. Samudra and Y.-C. Yeo. On the performance limit of impact-ionization transistors. In *IEEE International Electron Devices Meeting*, pages 117–120.

[15] E.-H. Toh, G. H. Wang, **C. Shen**, M. Zhu, L. Chan, C.-H. Heng, G. Samudra and Y.-C. Yeo. Silicon nano-wire impact ionization transistors with multiple-gates for enhanced gate control and performance. In *IEEE International Electron Devices Meeting*, pages 195–198.

Curriculum vitae

Shen Chen was born in Shanghai, China, on March 11th, 1981. He graduated from Da-Tong High School, Shanghai, in 1998. From 2000 to 2003, he did his undergraduate study at National University of Singapore, and received the degree of Bachelor of Engineering in electrical engineering, with first class honor. Since 2004, he has been working towards a PhD in electrical engineering at the Silicon Nano Device Laboratory, and the Center for IC Failure Analysis and Reliability, both at National University of Singapore. His doctoral research focuses on the threshold voltage instability of MOS gate dielectrics, including charge trapping and negative bias temperature instability. His recent research focuses on the device physics and modeling of impact-ionization transistors (I-MOS) and tunneling transistors (T-FET).