MOSFET DEGRADATION DUE TO NEGATIVE BIAS TEMPERATURE INSTABILITY (NBTI) AND HOT CARRIER INJECTION (HCI) AND ITS IMPLICATIONS FOR RELIABILITY-AWARE VLSI DESIGN

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Haldun Kufluoglu

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2007

Purdue University

West Lafayette, Indiana

To my wife, Aysegul, and baby daughter, Kayra, for being with me, yet having patience, tolerance and maintaining their support during the completion of this thesis. And to my parents, for raising me, preparing me for this life, their constant affection and trust despite being far away.

ACKNOWLEDGMENTS

This work could not have been completed without the support of several people. First, I am gratefully indebted to my advisor, Prof. Ashraf Alam, for his insight, guidance and exemplary professionalism. He has never hesitated to share his technical skills, diligence and his approach to problems. He consistently pursues to improve and contribute to the existing knowledge by paying attention to the details many would oversee, and works tirelessly to convey his vision to us. If I ever have a significant technological or scientific contribution in my life, that would be due him, having utilized his view, methods and work ethic. He is not only sourceful in academia, but also a great manager. He has dealt with issues, technical or personal, with extreme professionalism; at times when I struggled, he handled the situation smoothly to a degree that a conscious person would only awe. I feel blessed for having an advisor and colleague like him.

My lovely wife, Aysegül, has and will always have a special role in my life. Her love, support and tolerance for me was endless. She has been a wonderful partner, friend and family in the last two years. Her keenness on every day tasks, household chores and cooking have positively influenced me. I am inspired by her care on quality of work done, timely delivery and interpersonal communication. She has also crowned our family with a beatiful baby, Kayra, whom I enjoy spending time with. Despite the chaotic hassle and added responsibilities, Kayra is playful and very cute. I specifically have to mention her yummy cheeks, arms, legs, and tummy which I occasionally take a bite to relieve my stress. As your name suggests, you are a blessing for us. You also brought in extra motivation to finish my Ph.D.

My close family have been the greatest supporters of me. I don't know how to thank my parents, Mevlide and Mustafa, for bringing me to life. They always cared and loved me. They have provided the best possible education for me with their

limited resources. They wondered why I went to school for so many years, but they never questioned my choices. I know they trusted me and they haven't expected anything in return other than that I become a good individual.

The graduate school and research experience are only meaningful with our friends and colleagues. There are a lot of people in this category, so I hope I don't forget anybody (and spell their names correctly). At the early days of my Ph.D. life, being the first student of my advisor, Anisur Rahman, Jing Wang and Sayed Hasan set the senior-student examples for me, with their discipline, openness and productivity. As the time lapsed, I had the opportunity to work with new members of our group. Particularly with Ahmad E. Islam, I had a lot of interaction on research, courses and also outside the office. He is hard-working and supportive. By looking at his progress, I expect that he will have a bright future. Dhanoop Varghese is another member to whom I am grateful. He provided the expertise on experimental setup and procedures, thus has been a significant contributor on my research. His structured approach and careful analysis have solved several problems in the lab. On the theory and simulations side, Tamer Cakici and Pradeep Nair introduced me to MEDICI and SMC, respectively. I also acknowledge Jeff Bude for allowing us to use the SMC code. Sam Shichijo has provided and patiently helped me on the TSUPREM device process files. Meetings with Dr. Srikanth Krishnan and Dr. Vijay Reddy have demonstrated the strengths of our research. Colloboration with Kun Hyuk Kang and Dr. Bipul Paul on the circuit-level reliability was an integral part of my research and a significant learning experience. Marty Agostinelli, my supervisor from Intel internship, has shown me an objective view of the industry environment. With his help, I could better grasp why reliability is important for the semiconductor industry, what needs to be done and what is feasible to do.

Nauman Butt, Ninad Pimparkar, Neophytos Neophytou, Siyu Koswatta, Sayeef Salahuddin and Lutfe Siddique are among the friends I took courses with and interacted in our office. We have been nearly the same generation of graduate students, so I certainly feel closer to them after sharing the days of graduate life.

The contributions of teachers are not passable. Especially, Prof. Alam, Cooper, Datta and Lundstrom stand out in my graduate studies. They pointed out important concepts in this field and made sure that we learned those. They analyzed the material in depth which gave us more edge in these topics, but at the same time simplified complicated theories for us to comprehend. Their style is definitely influential on the way I think and how I look at problems.

During the preparation of this thesis, I received useful directions and valuable suggestions from my committee members: Prof. Mark Lundstrom, Prof. Kaushik Roy and Dr. Vijay Reddy. They have evaluated this work from their perspectives and specialization. Prof. Lundstrom's comments on device physics and Prof. Roy's experience in circuits enabled a balanced story. Dr. Reddy shared his extensive practical background in MOSFET reliability. Their insight helped me to improve the completeness of the thesis. I particularly thank Dr. Reddy for allotting time in his busy schedule, reading the thesis drafts carefully and providing timely feedback.

Aside from research and courses, the support of administrative staff has been very pleasing. Annie Cheever and Cheryl Haines have always taken care of my paperwork promptly. The lab staff at Birck Nanotechnology Center (formerly known as Solid State Lab) have provided the training, maintenance and logistics of our equipment. The Graduate and Business Offices have made sure that my academic progress and paperwork were on track.

An indispensable part of graduate school and research is funding. I have been lucky to receive funds from Semiconductor Research Corporation (SRC) and National Science Foundation (NSF) during my Ph.D. studies. Not only financial support is maintained but also the networking opportunities, conferences and student meetings have allowed me to improve my skills and meet with influential people, in the current time frame and for the future. The SRC connection opened up several internship and employment doors which I utilized.

I would also like to thank my teachers from previous levels of my education. They are mostly unaware of where I came but I hope to contribute to next generation of

students and thus fulfill their requests. I have to mention that among my teachers, those from my high school -Bursa Fen Lisesi- gave me the strongest training. I still remember and use the fundamentals I learned there. I am certainly overwhelmed with gratitude for their contributions. Without that education, I would not be at this point.

I might have forgotten several others here. I apologize in advance for not recalling those names but it doesn't change the fact that they helped in some ways; my work will always reflect a piece from them.

TABLE OF CONTENTS

Page

Page

Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

ABSTRACT

Kufluoglu, Haldun Ph.D., Purdue University, December, 2007. MOSFET Degradation due to Negative Bias Temperature Instability (NBTI) and Hot Carrier Injection (HCI) and Its Implications for Reliability-aware VLSI Design . Major Professor: Muhammad A. Alam.

The scaling trends in CMOS technology and operating conditions give rise to serious degradation mechanisms such as Negative Bias Temperature Instability (NBTI) and Hot Carrier Injection (HCI) in MOSFETs, threatening the circuit and product lifetimes. The aging phenomena, on top of process variations, translate into complexity and reduced design margin for circuits. International Technology Roadmap for Semiconductors mentions reliability as one of the "Design Technology Challenges" and calls attention to "Design for Reliability." In order to increase the overall design efficiency, it is important to (i) understand MOSFET-level degradation, (ii) develop physically robust compact models compatible with circuit simulators, and (iii) implementing tools that incorporate NBTI and HCI reliability into VLSI design process at an early stage. In this work, NBTI and HCI degradation and their implications on MOSFET and circuit reliability are studied. Transistor-level NBTI degradation is explored and experimentally calibrated voltage, temperature, and time dependences are obtained. Recovery characteristics, degradation under AC and random activity relations are derived. Implications for aggressively-scaled and non-planar MOSFET geometries are discussed. With comprehensive experimental and theoretical tools, HCI degradation is investigated, particularly for short channel MOSFETs with lower operating conditions. HCI issues that are unclear in the literature are resolved. The interface trap generation under HCI is linked to NBTI theory, thus degradation of circuits experiencing both mechanisms can be assessed efficiently. Finally, compact

reliability models for MOSFETs are presented so that this framework can be applied to digital and analog circuits as well as memory circuits.

1. INTRODUCTION

MOSFETs have been the major force behind the tremendous development in the microelectronics industry, yielding unprecedented scientific and technological advancements. This fundamental impact of MOSFETs has been facilitated by the phenomenal properties of both Si and its oxide, SiO_2. The ability to integrate various circuit components on the same substrate and the scalability of the dimensions of the MOSFETs have allowed increasing density of transistors with higher speed and computational capability along with lower power consumption and cost per MOSFET.

The progress in MOSFET based microelectronics has not been without its problems. The operating requirements of ICs put stress on the devices, leading to performance and reliability problems. MOSFETs, and particularly the oxide, degrade during the device operation and cannot retain its original specifications. One general degradation type is defect generation in the oxide bulk or at the Si-SiO_2 interface over time. The defects can increase leakage current through the gate dielectric, change transistor metrics such as the threshold voltage or result in the device failure due to oxide breakdown.

Negative Bias Temperature Instability (NBTI) is one of the most important threats to PMOSFETs in VLSI circuits. The electrical stress ($V_{GS} < 0$) on the transistor generates traps at the $Si - SiO_2$ interface. These defect sites increase the threshold voltage, reduce channel mobility of the MOSFETs or induce parasitic capacitances and degrade the performance. Overall, the significant reduction in the performance metrics implies a shorter lifetime and poses a challenge for the IC manucfacturers. One peculiar property of NBTI is recovery, i.e., the interface traps can be passivated partially when the stress is reduced.

Hot Carrier Injection (HCI), although alleviated for current generation MOSFETs, is another mechanim that can create defects at the $Si - SiO_2$ interface near the

drain edge as well as in the oxide bulk. Similar to NBTI, the traps shift the device metrics and reduce the performance. The damage is due to carrier heating in the high electric field near the drain side of the MOSFET, resulting in impact ionization and subsequent degradation. Historically, HCI has been more significant in NMOSFETs because electrons have higher mobilities (due to lower effective mass) than holes and thus, can gain higher energy from the channel electric field. HCI has a faster rate of degradation compared to NBTI. HCI occurs during the low-to-high transition of the gate of an NMOSFET, therefore the degradation increases for high switching activity or higher frequency of operation. Furthermore, the recovery in HCI is negligible, making it worse for AC stress conditions.

In an IC, the MOSFETs operate under various stress conditions at different times, are, therefore, exposed to different degradation types. For instance, in a CMOS inverter -the fundamental building block of the digital ICs- both the NMOSFET and PMOSFET are tied to the same input voltage. When the input signal is low ($\approx 0V$), the PMOSFET is under NBTI stress and therefore degrades while the NMOSFET is turned off. When the input is pulled to high (V_{DD}), the NMOSFET goes through impact ionization condition and experiences HCI degradation. At the same time, PMOSFET is turned off and some of the NBTI damage relaxes. Because the fact that each degradation mechanism generates defects either in the bulk oxide or at the interface, the overall MOSFET degradation can be very complex.

In this work, NBTI and HCI degradation and their implications on circuit reliability are studied. First, Chapter 2 presents the NBTI background and the theory for NBTI degradation from the perspective of Reaction-Diffusion model. Additionally, alternative NBTI theories are compared in the context of saturation dynamics. NBTI recovery, then geometry-dependent R-D model are provided in Chapters 3 and 4, respectively. Compact models for digital circuits are given in Chapter 5. In Chapter 6, HCI background and existing issues in the HCI literature are discussed. Modeling approach for HCI degradation is given in Chapter 7. Experimental and theoretical HCI characterization of technologically-relevant NMOSFETs are presented in Chap-

ter 8. Then, in Chapter 9 the impact of HCI degradation on MOSFET characteristics is investigated. Future directions are provided in Chapter 10 followed by Summary in Chapter 11.

2. MECHANICS OF NBTI DEGRADATION: A PHENOMENOLOGICAL DESCRIPTION

2.1 NBTI Background

Negative Bias Temperature Instability is a significant reliability concern for digital and analog circuits in current generation CMOS technology [2]- [6]. NBTI occurs in negatively biased ($V_{GS} < 0$V) PMOSFETs at elevated temperatures and is a consequence of interface trap generation at the $Si/oxide$ interface. In conventional Si MOSFETs, the transistors are annealed in hydrogen ambient to passivate the dangling Si bonds during manufacture [7]. This traditional method proved to be an effective solution to the interface trap instabilities for decades; however the continuing MOSFET miniaturization trends, (i.e., aggressive oxide thickness scaling leading to higher oxide field and process modifications such as nitridation of oxides to prevent Boron diffusion from p+ poly gate) and higher operation temperatures (due to power dissipation from the circuits or ambient conditions) accelerate bond-breaking at the interface over time during the device operation [8], [9]. The traps increase the threshold voltage, reduce the channel mobility due to scattering and induce parasitic capacitances in the transistors. Overall, the drain current degrades over time and parametric reliability becomes a significant concern.

NBTI has become a forefront reliability issue recently, however, it has been known for decades [10], [11], [12]. Present day operating conditions of circuits favor significant NBTI related trap generation in PMOSFETs. In NBTI, the interface trap density increases over time in a characteristic power law behavior ($\propto t^n$), and eventually threatens the operational lifetimes of the transistors and the circuits. Therefore understanding the NBTI degradation is of primary importance for existing and near-future technologies.

The seriousness of NBTI has led to intense studies in order to characterize the degradation. Major observations are (a) recovery dynamics after stress [13], [14] (b) robust fractional power-law exponents [15], (c) activation energies of degradation [16], [17], [18], (d) frequency dependence [4], [13], [19], (e) isotope (hydrogen and deuterium) effect [20], (f) lock-in mechanism for recovery [21], (g) material dependence [22], and (h) microscopics of degradation [14], [17], [18], [23] - [31].

One of the most important characteristics of NBTI is the recovery behavior; when the stress is removed, the degradation relaxes. This is attributed to annealing of the interface traps by free hydrogen in the gate dielectric [13], [19]. Recovery can increase the lifetime for AC stress, however it can also distort the characterization techniques [21], [31], [32], [33], [34]. Ultra-fast measurement techniques were developed to overcome the recovery effect [14], [35].

At the circuit level, the degradation can be modeled based on the shifts of threshold voltage, linear and saturation currents of a PMOSFET [2], [3], [36]. The delay or frequency degradation can be mapped to transistor level NBTI degradation. This provides a compact form for circuit simulators and design tools [36], [37]. Memory as well as digital logic circuitry can be treated by this modeling [38], [39]. Additionally, the impact of degradation on leakage current was discussed in [40].

2.2 NBTI Time-dependence

The experimental characterization period of the transistor degradation is very short compared to the desired lifetime (e.g., few weeks for the former, about 10 years for the latter) of a transistor. For a given technology generation, accelerated characterization tests with higher voltages and temperatures with respect to the operating conditions are performed, the characteristic NBTI time-dependence is extracted, and the results are projected to the long device lifetimes. In practise, accurate quantification of the degradation is of great interest because the lifetime of the MOSFETs or circuits are directly related to the time-dependence of damage. As it will be presented

in the next chapters, the observed time behavior strongly depends on the degradation mechanisms, measurement techniques and MOSFET features such as geometry and size. Moreover, the NBTI degradation under the accelerated test conditions and the real operating conditions can vary. Therefore, robust, physics-based, and well-calibrated models are needed to capture these aspects of NBTI for present and future-generation devices and stress conditions.

In this thesis, the NBTI time-dependence is explored by encapsulating the physical mechanisms into a phenomenological framework, i.e., Reaction-Diffusion (R-D) model. Although several models have been proposed for the time-dependence of NBTI, the R-D provides a very robust framework to explain experimental observations.

In this chapter, starting from the standard version, the R-D model will be expanded by considering the role of hydrogen explicitly in NBTI time-dependence during stress. A generalized approach consistent with experimental results will be developed. A distinct evolution in time-dependence, not seen from the standard model, will be discussed. Then, the generalized R-D framework will be compared with alternative NBTI models by studying the saturation phenomena of degradation. Finally, a summary will be provided.

2.3 Reaction-Diffusion (R-D) Framework: A Mathematical Analysis

2.3.1 Standard Reaction-Diffusion Model

The modeling of NBTI through the Reaction-Diffusion (R-D) Framework can successfully explain several experimental observations such as (*i*) fractional time-exponents [15], (*ii*) activation energies [16], (*iii*) relaxation dynamics of degradation [13], (*iv*) frequency dependence under AC stress [13], [19], (*v*) isotope (i.e., deuterium) effects [20], (*vi*) lock-in mechanism [21] and (*vii*) quasi-saturation of NBTI [41].

The model is based on interface trap generation and related hydrogen dynamics [42]. The generation takes place at the semiconductor/oxide interface which is a

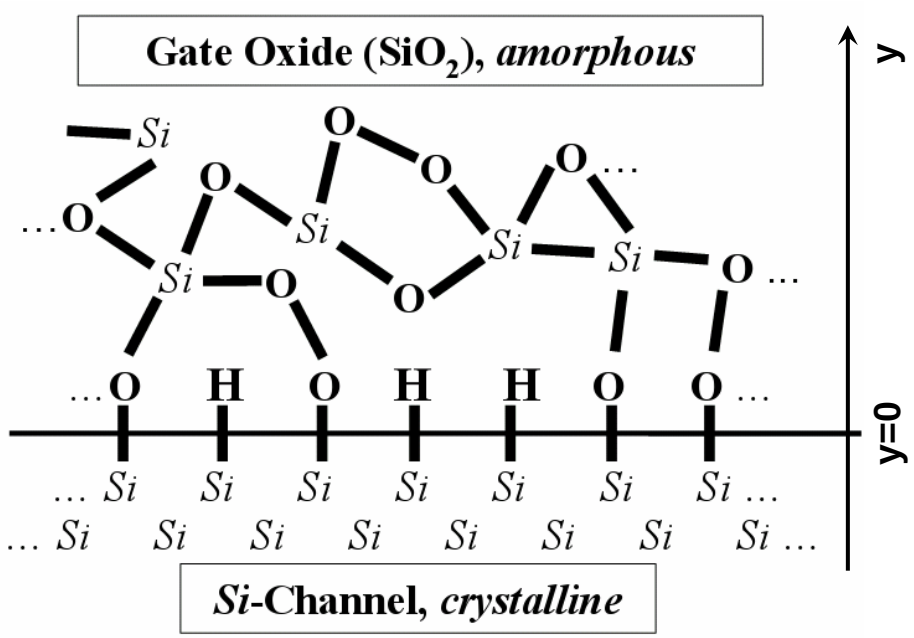


Fig. 2.1. The schematic of the $Si/oxide$ interface of a MOSFET. The dangling Si bonds are present due to the mismatch between the ordered channel and amorphous oxide. These act as interface traps and unless they are passivated by hydrogen annealing. NBTI induces the dissociation of the $Si-H$ bonds causing hydrogen to diffuse away from the interface.

rough surface where the highly ordered crystalline channel and the amorphous SiO_2 dielectric meet. At the junction of these dissimilar materials, some of the Si atoms from the channel remain dangling without satisfied chemical bonds, thus, forming the interface traps. The traps lead to poor device performance; therefore the transistors are annealed in hydrogen ambient during the manufacture. The hydrogen gas diffuses into the gate-oxide and yields passivated Si bonds as shown in Fig. 2.1. In the R-D model, the interface trap generation at the $Si/oxide$ interface is represented as a chemical reaction, i.e.,

$$Si - H + h^+ \longleftrightarrow Si^\star + H^0 \tag{2.1}$$

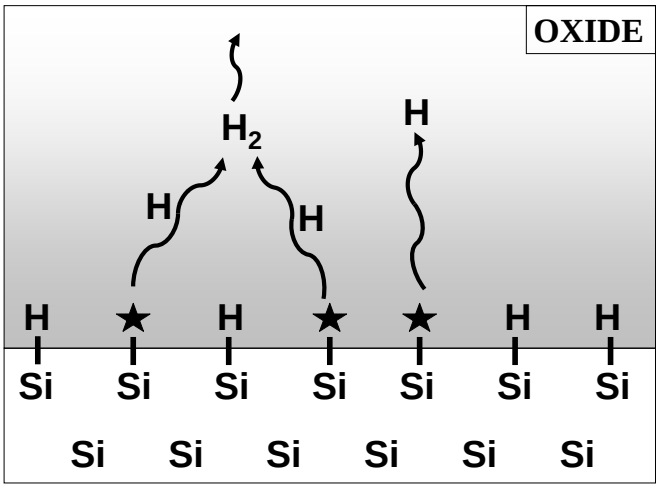


Fig. 2.2. The schematic of hydrogen dynamics after it is released from a $Si-H$ bond. Interface traps are denoted by $\star$. Atomic hydrogen can diffuse away from the interface freely or forms H_2 with another H. Recent experiments suggest H_2 diffusion due to the activation energy of degradation.

in which an inversion layer hole, h^+, weakens a $Si-H$ bond and hydrogen is detached as a result of thermal vibrations of the chemical bond [13], [19], [24], [43]- [46]. The remaining Si dangling bond ($Si^\star$ in Fig. 2.2) acts as a donor-like interface trap [2], [9]. The H released from the bond can diffuse away from the Si/oxide interface or anneal an existing trap. The interface trap density, N_{IT}, increases with the net rate of the reaction given in (2.1), so

$$\frac{dN_{IT}}{dt} = k_F[N_0 - N_{IT}] - k_R N_{IT} N_H^{(0)} \tag{2.2}$$

where k_F, k_R, N_0 and $N_H^{(0)}$ are bond-breaking rate, bond-annealing rate, $Si-H$ bond density available before stress and hydrogen density at the $Si/oxide$ interface ($y = 0$), respectively. When a device is stressed, initially, both N_{IT} and $N_H^{(0)}$ are negligible ($N_0 \gg N_{IT}$), and so is the $k_R N_{IT} N_H^{(0)}$ term in (2.2). Therefore, the increase in N_{IT} is generation-limited as in Fig. 2.3 (region i). When sufficient hydrogen builds up at the interface, (2.1) reaches an equilibrium (region ii in Fig. 2.3). The diffusion of H away from the traps then removes hydrogen from the interface, so the interface

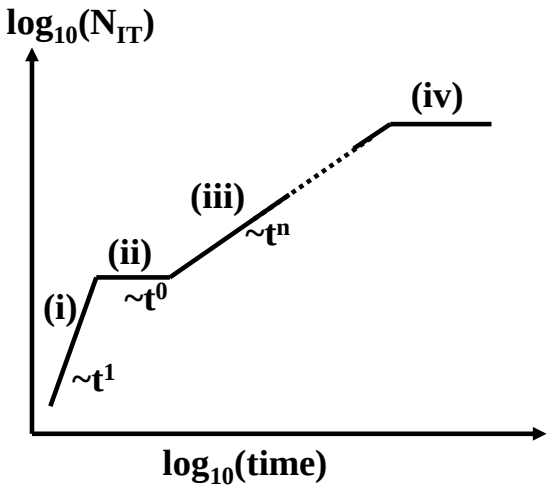


Fig. 2.3. The schematic of time evolution of the classical R-D model. Regions (i) generation-limited, (ii) dynamic equilibrium, (iii) diffusion-limited, (iv) saturation. Region (iii) gives the power law behavior for NBTI. The degradation in the first two regions is very quick and not observable in experimental measurements.

trap generation rate becomes limited to the diffusion of hydrogen (region iii), and thus represents the characteristic time evolution of NBTI degradation. In this period, the diffusion of hydrogen obeys

$$\frac{dN_H}{dt} = D_H \frac{d^2 N_H}{dy^2} \tag{2.3}$$

where D_H is the diffusion constant, and y is the distance into the dielectric (Fig. 2.1). If nearly all the $Si-H$ bonds are broken ($N_{IT} \approx N_0$), then the interface trap generation slows down and saturates (region iv).

From (2.2) and (2.3), it is obvious that the temperature and electric-field dependence of NBTI is not considered explicitly in the R-D model. The oxide-field dependence (exponential dependence) is included in the k_F term and the temperature dependence of the degradation is incorporated through the activation energies of k_F, k_R and D_H (Arrhenius activation) [8]. The rates k_F and k_R contain the microscopic details of the $Si-H$ bond breaking and annealing of $Si^{\star}$, and reflect the

associated activation energies as distributions over the energy range [25]. The effective bond-breaking and annealing rates can be written as

$$k_{F,R} = \int_0^\infty k_{F,R}^0 e^{\left(\frac{-E}{k_B T}\right)} \cdot g_{F,R}(E) dE \tag{2.4}$$

where $g_{F,R}(E)$ is the Gaussian distribution (with mean E_A and variance $(\Delta E_A)^2$) of the bond activation energies.

When a $Si - H$ bond breaks, every dangling Si bond is associated with a free H atom in the oxide, therefore $N_{IT}(t) = \int N_H(r,t) d^3r$. The density of H in the oxide during diffusion can be approximated with a triangular profile, broadening as $\sqrt{(D_H \cdot t)}$. Then the interface trap density is given by

$$\begin{aligned} N_{IT}(t) &= \int_0^{\sqrt{D_H t}} N_H^{(0)} \left(1 - \frac{y}{\sqrt{D_H t}}\right) dy \\ &= \frac{N_H^{(0)}}{2} \sqrt{D_H t}. \end{aligned} \tag{2.5}$$

During the diffusion-dominated regime, the $\frac{dN_{IT}}{dt}$ term is negligible compared to the other bond-breaking and annealing terms in (2.2), therefore (2.2) can be simplified as

$$N_{IT} N_H^{(0)} = \frac{k_F N_0}{k_R}. \tag{2.6}$$

Substituting $N_H^{(0)}$ from (2.5) into (2.6), (2.7) is obtained

$$N_{IT}(t) = \sqrt{\frac{k_F N_0}{2 k_R}} (D_H t)^{\frac{1}{4}} \tag{2.7}$$

which gives the time-exponent ($n = 0.25$) of NBTI when only H diffusion is taken into account. The derivations up to (2.7) correspond to H-diffusion branch in Fig. 2.4, therefore generation and diffusion of H_2 are ignored completely.

2.3.2 Reaction-Diffusion Modeling with H_2-only Diffusion

In the standard R-D model, only H diffusion is considered. However, it was shown that measurement induced delays or interruption of stress can yield time-exponents that are higher than the uninterrupted case [21], [22], [32], [34]. Moreover, theoretical

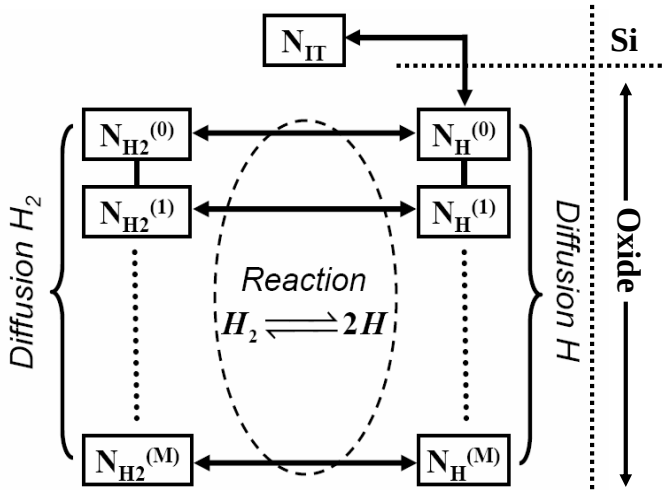


Fig. 2.4. Schematic representation of the numerical implementation for the generalized $H \leftrightarrow H_2$ simulations. Both H and H_2 dynamics are handled explicitly. The standard R-D model with H diffusion ignores H_2 generation. The H_2-only model assumes instantaneous conversion, therefore the H branch is not included. The simulation domain is discretized into M nodes, first node being the N_{IT} term. The rest of the nodes have H and H_2 densities at the same physical location.

calculations suggest that atomic H is unstable and converted into molecular H_2 after it is released from the $Si/oxide$ interface [24], [26]. Furthermore, activation energies extracted from NBTI measurements support that the dominant diffusing hydrogen species is H_2 [17], [18]. Therefore, a modified R-D modeling with H_2 can be obtained based on the principles of the H-only framework. In the H_2-only model, after dissociation of $Si-H$ bonds, released H atoms react and form the hydrogen molecule (see Fig. 2.2) as

$$H + H \longleftrightarrow H_2. \tag{2.8}$$

Analytically, the derivations that yield (2.7) can be repeated when H_2 diffusion is the dominant factor that limits the degradation rate. The fundamental assumption in this modified model is that all atomic H becomes H_2 and only the H_2-diffusion

branch in Fig. 2.4 is taken into account. Once the H atoms react and generate H_2, the change of the interface trap density in (2.2) can be rewritten as

$$\frac{dN_{IT}}{dt} = k_F[N_0 - N_{IT}] - k_R N_{IT} N_{H,eff}^{(0)} \tag{2.9}$$

in which the annealing component has an effective hydrogen density term (and k_F, k_R can be different from H-only model). The $N_{H,eff}$ is obtained by considering the equilibrium dynamics of (2.8), namely

$$N_{H,eff}^{(0)} \propto \sqrt{N_{H_2}^{(0)}}. \tag{2.10}$$

Additionally, corresponding to H_2 diffusion, (2.3) becomes

$$\frac{dN_{H_2}}{dt} = D_{H_2}\frac{d^2 N_{H_2}}{dy^2}. \tag{2.11}$$

The triangle approximation with H_2 diffusion away from the interface results in

$$N_{IT} = 2 \cdot \frac{1}{2} N_{H_2}^{(0)} \sqrt{D_{H_2} t} = N_{H_2}^{(0)} \sqrt{D_{H_2} t}. \tag{2.12}$$

The factor 2 in (2.12) comes from the fact that H_2 contains two H atoms and therefore it is associated with two interface traps. At the *Si/oxide* interface, (2.6) is still valid and inserting (2.12) into (2.10) yields

$$N_{IT}(t) \propto [\frac{k_F N_0}{k_R}]^{\frac{2}{3}} \cdot (D_{H_2} t)^{\frac{1}{6}}. \tag{2.13}$$

The H_2 diffusion and the time-exponent in (2.13) is consistent with the experimental activation energy of NBTI and time-dependences observed in measurements.

2.3.3 Generalized R-D Model with Both H and H_2 Diffusion

The H_2-only R-D model employs assumptions stating that H-to-H_2 conversion is extremely fast and H is consumed totally to generate H_2. However, these assumptions may not be realistic in general for such a conversion reaction and the validity of the robust power-law time-dependences predicted by H-only and H_2-only models (also widely supported by experiments) is of concern. This issue can be addressed by a

generalized modeling approach in which the aforementioned assumptions are relaxed by considering the dynamics of (2.8) [47]. In this approach, both H and H_2 branches in Fig. 2.4 are included and their diffusion and mutual conversion are explicitly accounted for. According to (2.8), the rate of change in N_{H_2} is given by

$$\frac{dN_{H_2}}{dt} = k_{H_1} N_H^2 - k_{H_2} N_{H_2}. \tag{2.14}$$

where k_{H_1} and k_{H_2} are generation and dissociation rates for H_2. The rate for H can be written similarly. After $Si-H$ bonds break, the released H atoms can diffuse away from the interface or get converted into H_2. Although neutral H atom is thought to be unstable compared to the charged H ions in the oxide [48], in this model, the assumption is that the conversion into H_2 takes place very quickly, so neutral H^0 is possible under non-equilibrium conditions. The molecular H_2 can also diffuse or dissociates back to H atoms. The H_2 cannot be formed directly from the $Si-H$ breaking nor it can passivate an interface trap. The complex nature of the dynamics in the generalized approach does not permit understanding the time-dependence of NBTI readily. Therefore, numerical solutions are needed to assess the effects of $H-H_2$ dynamics on NBTI degradation. The solutions are based on (2.2), (2.3), (2.11) and (2.14) and do not contain any approximations regarding the shape of the hydrogen profiles or restrict $H-H_2$ conversion reactions with specific assumptions. The details of the numerical implementation are presented in Appendix A.1.

2.3.4 Simulation Results and Discussion

The result of $H \leftrightarrow H_2$ simulation is compared with those of H-only and H_2-only R-D numerical solutions in Fig. 2.5. As predicted by (2.7) and (2.13), H-only and H_2-only simulations reflect the time-exponents of 1/4 and 1/6, respectively. The H-only and H_2-only curves act as limits to the $H \leftrightarrow H_2$ result, i.e., as k_{H_1} in (2.14) reduces toward 0, $H \leftrightarrow H_2$ solution approaches that of H-only. Similarly, as k_{H_1} is increased further so that more H is consumed, the $H \leftrightarrow H_2$ result approximates the H_2-only solution. In Fig. 2.5, despite the fact that it gives the same result as H_2-only

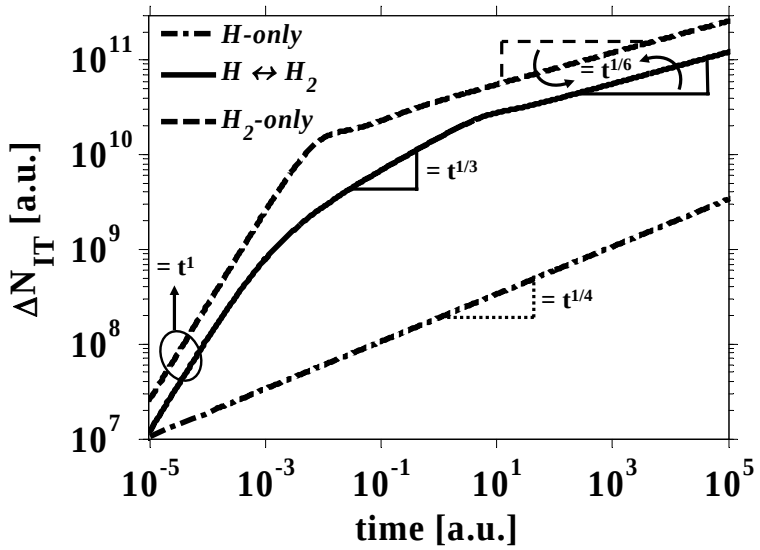


Fig. 2.5. When the conversion of H to H_2 is added into the simulations, the time-behavior changes significantly at earlier times. The diffusion mechanism in H-only and H_2-only implementations cannot foresee this behavior. The H and H_2 results are the limiting cases for the $H \leftrightarrow H_2$ solution when the H_2 generation rate, k_{H_1}, is increased or decreased further.

implementation at later times, the $H \leftrightarrow H_2$ simulation shows a distinct behavior at earlier times. While the H_2-only model shows a sharp transition from t^1 (reaction-dominated) to $t^{1/6}$ (diffusion-dominated), the $H \leftrightarrow H_2$ shows an intermediate soft-transition region with slope $1 < n < 1/6$ ($n \sim 1/3$ as in Fig. 2.5). The time-exponent ($1 > n > 1/6$) predicted by the theory supports experimental data presented in Fig. 2.6 [14], [21]. This change in the time-exponent is observed for a wide range of values of the k_{H1} and k_{H_2} parameters and the transition point where the exponent becomes 1/6 always denotes the beginning of the diffusion-dominated regime. The time-exponent of $\sim$1/3 in the $H \leftrightarrow H_2$ solution can be explained by the aid of Figures 2.7 and 2.8. In this regime, the atomic hydrogen released from the interface is being converted to H_2 ($dN_{H_2}/dt > 0$) according to (2.14) and thus $N_H^{(0)}$ keeps decreasing as $N_{H_2}^{(0)}$ increases over time (see Fig. 2.7). Unlike (2.12), the diffusion in this regime is negligible and since $N_H^{(0)} \ll N_{H_2}^{(0)}$ due to the conversion, $N_{IT} \propto N_{H_2}^{(0)}$ as verified by

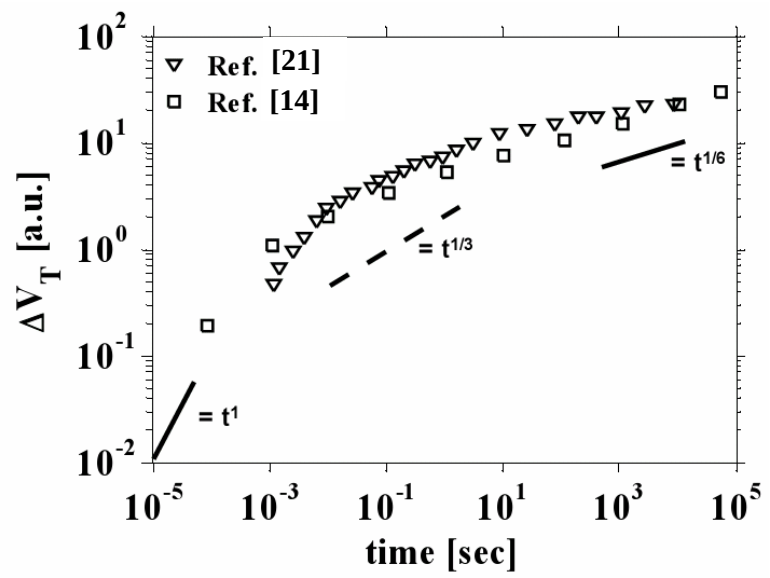


Fig. 2.6. Experimental data published in the literature shows a higher time-slope initially. $H \leftrightarrow H_2$ theory supports the observations. Data from [14], [21].

Fig. 2.8. Therefore, $\frac{dN_{IT}}{dt} \propto \frac{dN_{H_2}^{(0)}}{dt}$ can be obtained. Also, from Fig. 2.8, although $N_H^{(0)} \ll N_{H_2}^{(0)}$ at the $Si/oxide$ interface, due to H_2 generation, $k_{H_1}[N_H^{(0)}]^2 \gg k_{H_2} N_{H_2}^{(0)}$ in (2.14). Substituting $N_H^{(0)}$ from (2.6) gives

$$\frac{dN_{IT}}{dt} \propto \frac{dN_{H_2}^{(0)}}{dt} \propto k_{H_1}[N_H^{(0)}]^2 \propto \frac{1}{(N_{IT})^2}. \tag{2.15}$$

The analytical solution of (2.15) gives $N_{IT} \propto t^{1/3}$, as observed in the numerical simulation results.

As the H_2 density increases at the interface, eventually, it begins to diffuse away. As shown in Fig. 2.9, H also diffuses, however its density is much smaller than H_2, so the interface trap generation rate is governed by H_2 diffusion. Therefore this behavior shifts the time-exponent to 1/6 at later times, in agreement with the analytical solution of (2.13). The $H \leftrightarrow H_2$ model shows a transition behavior as the time-exponent changes from a higher value to 1/6 at earlier times. This early-time transition is different from the later-time quasi-saturation behavior observed in NBTI experiments (detailed discussion in Sec. 2.4). The quasi-saturation is often

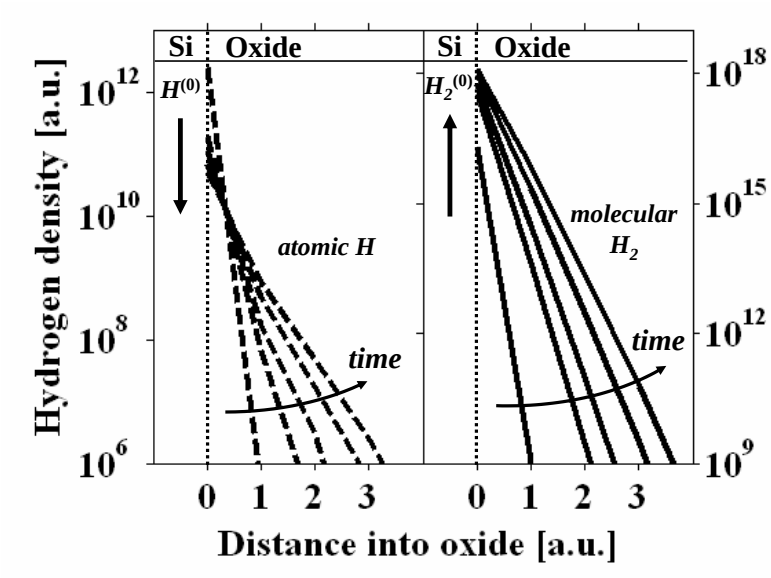


Fig. 2.7. The profiles of H (left panel) and H_2 (right panel) during the $N_{IT} \propto t^{1/3}$ regime. This period is marked by a significant consumption of H and its conversion into H_2. The H_2 density is much higher than that of H. The diffusion is negligible so the profile tips extend only slightly with time and the change in densities of H and H_2 at the interface is much more pronounced. Therefore $N_{IT} \propto N_{H_2}^{(0)}$.

explained by (i) reflection of diffusing species from material boundaries [19], (ii) consumption of all $Si - H$ bonds [29], [41], (iii) distribution of bond-energies [30], and (iv) experimental artifact arising from measurement delay (e.g., charge pumping technique) [27], [34]. Overall, quasi-saturation in NBTI has a direct impact on the device lifetime; it can provide increased lifetime since the time-exponent is reduced at later times. However, in the $H \leftrightarrow H_2$ mechanism, the initial time-slope is higher than the traditional exponents and the model anticipates this change in the time-exponent even when the delay effects are corrected. Saturation effects are not included in this section, however the higher slope would still be apparent in that case since saturation happens at later stages of NBTI. Experimentally, the transition in the exponent may not be visible in the measurement window depending on the rate of the conversion from H to H_2. Even so, the MOSFET would degrade more at earlier times in the operating conditions and thus the benefits of quasi-saturation behavior may not be

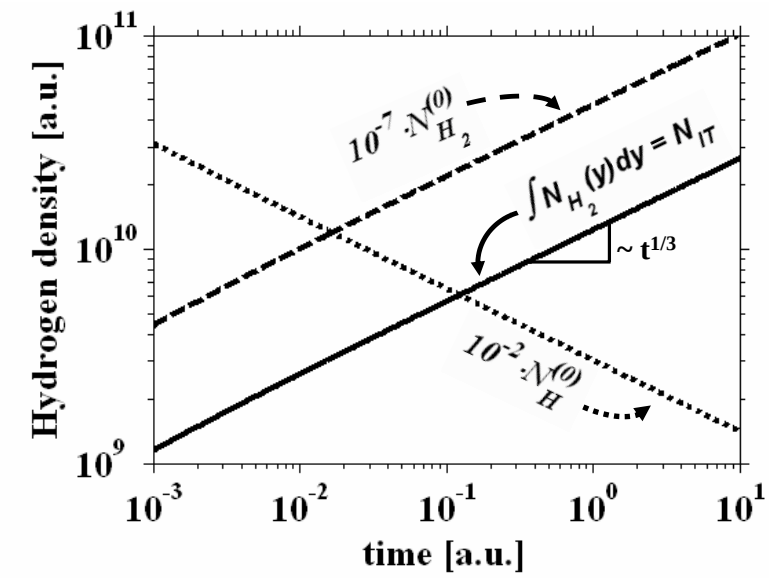


Fig. 2.8. During the $N_{IT} \propto t^{1/3}$ regime, simulations show that the density of $H^{(0)}$ decreases due to conversion into H_2. $N_{H_2} \gg N_H$ and since integrated N_{H_2} and $N_{H_2}^{(0)}$ are directly proportional, $N_{IT} \propto N_{H_2}^{(0)}$.

significant in the $H \leftrightarrow H_2$ model. It was suggested that dispersive diffusion of hydrogen species can explain time-dependent exponents in NBTI [31]. However, once the artifacts due to measurement delays were accounted for, dispersion was found to be negligible at later stages of NBTI time characteristics [34]. At the early stage, dispersion can be effective in addition to the H to H_2 conversion and modify the time exponents.

In comparing experiments with the $H - H_2$ theory, one should be careful about artifacts arising from measurements. The degradation is conventionally calculated from $\Delta I_D(t) = |I_D(t) - I_D(t = 0)|$. In reality, the MOSFET can degrade during the first I_D measurement even when measurement duration, t_0, is very short. At very early times $I_D(t) \approx I_D(t = t_0)$, and their difference is nearly zero. As a result, initial $\Delta I_D(t)$ increases rapidly, yielding an artificial time-exponent higher than the long term n=1/6 behavior. After correcting for the t_0 artifact, it was found that the higher time-exponent, n=1/3 is indeed present in the NBTI characteristics [28].

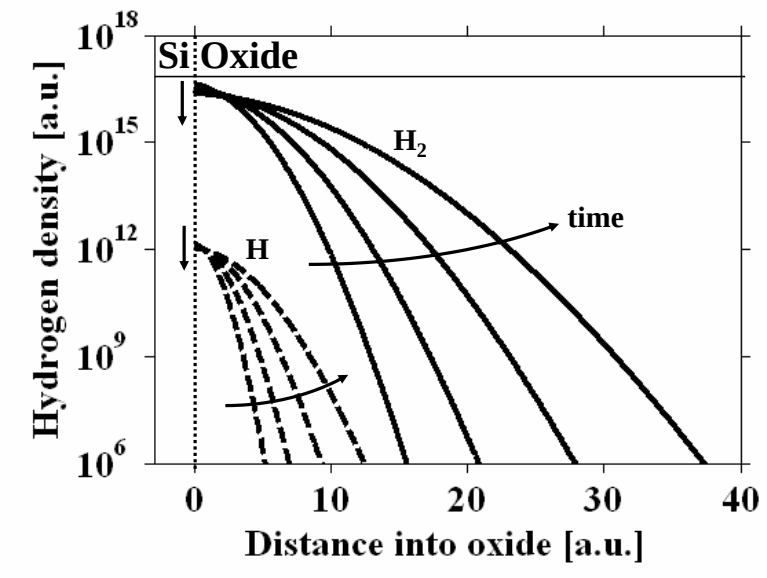


Fig. 2.9. After sufficient H_2 build-up at the interface, diffusion begins to take over. The densities of H and H_2 at the $Si/oxide$ interface decrease with time whereas the profile tips move as $\sqrt{D_{hydrogen} \cdot t}$ (much faster compared to the profiles in Fig. 2.7) in accordance with the diffusion process. Both H and H_2 diffuse but the density of H is much less than that of H_2 and the overall time-dependence of N_{IT} is limited by H_2 diffusion n=1/6 regime.

Moreover, the $H-H_2$ model can also explain the time and voltage dependences better than the $H_2 - only$ approach once the artifact due to t_0 has been corrected for [28].

In this section, a generalized Reaction-Diffusion model for H-to-H_2 conversion with finite transition time was considered. The time-behavior of NBTI is investigated when the dominant diffusing species is H_2 and the conversion of atomic H into H_2 is implemented explicitly in the numerical solutions. It is shown that the conversion of H to H_2 results in time-slopes higher than traditional NBTI time-slopes at earlier stress times. This time-dependence cannot be predicted from diffusion-only implementations. The experimentally observed soft-transition in the time-exponent in NBTI may stem from such conversion-dominated hydrogen dynamics. If the measurement window is not long enough, even with a quasi-saturation in NBTI at later

times, this dynamic behavior may imply shortened lifetimes, and therefore, should be evaluated carefuly.

2.4 Saturation behavior

The NBTI literature until late 1990s generally characterized the interface trap (N_{IT}) generation as a function of time, (t), as a power-law (i.e., $N_{IT} \propto t^n$) with $n \approx 0.3-0.5$. This value of n was assumed robust, that is, independent of stress time. This time-independent exponent had been the unique signature of NBTI degradation and traditional theoretical models of NBTI strived to interpret the constancy of n satisfactorily.

It became increasingly clear in late 1990s, as the semiconductor industry began to struggle with NBTI issue, that the projected IC lifetime based on $n \approx 0.30-0.50$ will be unacceptable. This led to an intense reexamination of the NBTI time-exponent. The essence of many measurements of n done since late 1990s is the following: $n \approx 0.3-0.5$ at the early stage of degradation, however at long stress times, n gradually decreased to $\approx 0.12-0.15$ as in Fig. 2.10. This is good news for semiconductor industry because lower exponents translate to longer extrapolated lifetime for CMOS circuits. However, this time-dependent change in n (in other words, quasi-saturation of trap generation) poses a challenge to old theories of NBTI which focused on the interpretation of time-independent exponents. The old NBTI theories had to be generalized to explain the new phenomena and establish appropriate voltage and temperature scaling laws. Five different mechanisms for NBTI quasi-saturation have been proposed: a) Reflection of H at the poly/oxide interface [8], [55], b) Breaking of all $Si-H$ bonds at the Si/SiO_2 interface [29], c) Variation of bonding strength of the precursors [30], [49] d) Transition from atomic to molecular hydrogen [47], and e) Artifact of finite measurement delay [50]. All these models are based on various refinements of basic Reaction-Diffusion (R-D) model. The basic model has been extensively validated in the literature [8], [13], [15], [31], [42], [55]. In this section,

both analytical formulations as well as numerical solutions of the R-D model to explore the proposed mechanisms of quasi-saturation are used. The voltage, thickness, and temperature dependence of NBTI degradation predicted by these models are different. Then, the predicted differences among the models and propose possible methods for explicit experimental conformation are highlighted.

2.4.1 R-D model vs. Alternative Theories of NBTI

In the standard R-D approximation, one assumes that in (2.2), the trap generation rate is much slower than the dissociation and annealing rates, so that $k_F N_0 \approx k_R N_{IT} N_H$ with neutral H diffusion. Since hydrogen diffusion front moves a distance of $\sqrt{D_H t}$, $\frac{dN_{IT}}{dt} \approx D_H N_H / \sqrt{D_H t} \approx D_H N_H^2 / \sqrt{D_H t}$. Substituting $k_F N_0 \approx k_R N_{IT} N_H$ into this relationship,

$$\frac{N_{IT}}{N_0} \approx a(\frac{t}{\tau})^n; \frac{1}{\tau} \equiv D_{H_2}(\frac{k_F}{k_R N_0})^{1/2n} \tag{2.16}$$

where a is a constant and ($n \approx 1/6$). If diffusion of hydrogen is dispersive, i.e., $D_H \approx D_0 t^{-m}$, the time exponent of (2.16) would change from n to $(1-m)n$ [33]. Similarly, if the diffusing species were a proton, H^+, and the proton drift were dispersive, $\mu_H \approx \mu_0 t^{-m}$, then n would also be $n(1-m)$, as before [29], [31], [51], [52].

Therefore, according to the R-D model of NBTI, the power-law in (2.16) arises from the interplay of generation, annealing, and (either normal or dispersive) diffusion of H at the $Si/oxide$ interface. In R-D model, exponent n is intimately related to the geometry of diffusion, allowing interpretation of NBTI and HCI time-exponents as a consequence of 1-D and 2-D diffusion, respectively [53]. And the finite value of annealing coefficients, k_R , allows interpretation of relaxation experiments [13], [19], [55]. Note that the exponent n in (2.16) is robust and time-independent, that is, given the assumptions, this is an exact solution. One would need to relax some of the assumptions made in deriving (2.16) to interpret the time-dependent exponent n that characterizes the quasi-saturation of NBTI characteristics.

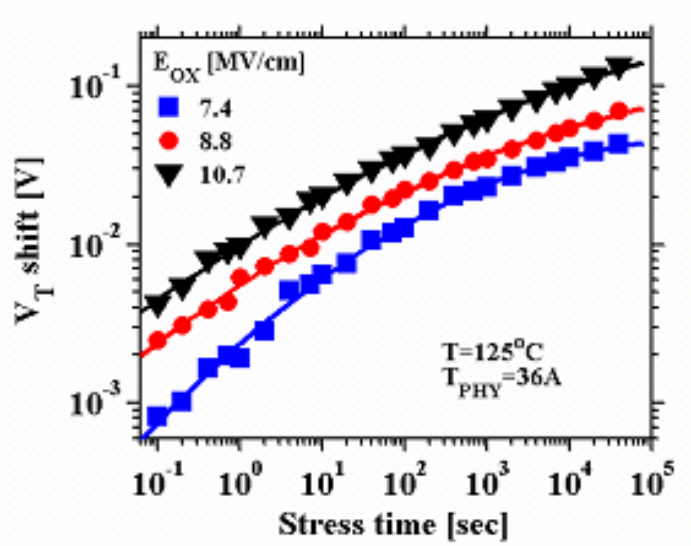


Fig. 2.10. Typical measurement of NBTI shows that the time exponent n, given by the slope of the $\log(N_{IT})$ vs. log(t) curve, is time dependent that can be interpreted as quasi-saturation of interface trap generation [56].

2.4.2 Saturation due to Poly-SiO_2 Interface Reflection (Reflection Model)

In the discussion above, the assumption is that H diffuses in oxide alone, but this is unrealistic for sub-5 nm oxides. If the H diffusion in poly-silicon (which tops the oxide) is slower than that in the oxide, then there will be a build-up of H at the poly-oxide interface. This would increase the reverse reaction and reduce overall trap generation. As seen is Fig. 2.11, numerical solution of (2.2), that allows two different diffusion coefficients for oxides and poly, results in quasi-saturation of N_{IT} vs. time characteristics. However, as shown in Fig. 2.11, this saturation is soft, because given adequate time, the N_{IT} vs. time characteristic would regain the original exponent (except $D_H(oxide)$ will have to be replaced by $D_H(poly)$) [41].

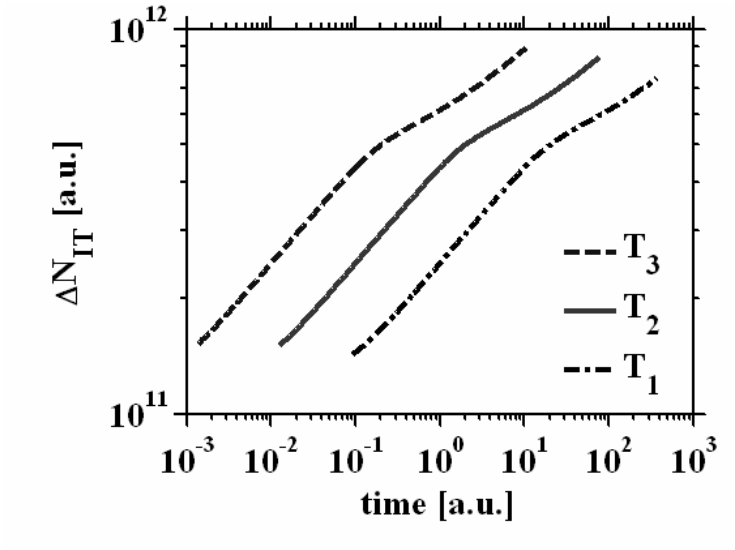


Fig. 2.11. Temperature(T)-dependent reduction in N_{IT} generation due to reflection at the poly-oxide interface ($D_H(T_3) = 10D_H(T_2) = 100{\cdot}D_H(T_1)$) .

2.4.3 Saturation due to Depletion of Precursors (S-E Model):

If the poly and the oxide diffusion coefficients are equal or if diffusion in poly is faster than that in oxide $D_H(oxide) = D_H(poly)$, there would be no build-up of H_2 at the oxide/poly interface, and reflection at that interface would not explain NBTI saturation. Assuming this is the case, the second interpretation of NBTI has been proposed which challenges the assumption that N_{IT} remains negligible compared to precursor density N_0 throughout the measurement window. The distinction of soft saturation due to reflection at poly/oxide interface discussed above with hard saturation due to depletion of all $Si-H$ bonds as $N_{IT} \approx N_0$ allows one to interpret the Stretched-Exponential (S-E) model [33] as an approximation to R-D model. Near the hard saturation limit, $dN_{IT}/dt \cong 0$ (not merely negligible compared to other two terms in (2.2), but $N_{IT} \approx N_0$, therefore, $k_F(N_0 - N_{IT}) \approx k_R N_{IT} N_H$. Inserting this in $dN_{IT}/dt \approx D_H N_H/\sqrt{(D_H t)}$, and integrating,

$$-\frac{N_{IT}}{N_0} - ln(1 - \frac{N_{IT}}{N_0 - N_{IT}(0)}) = b(\frac{t}{\tau})^{2n}, \tag{2.17}$$

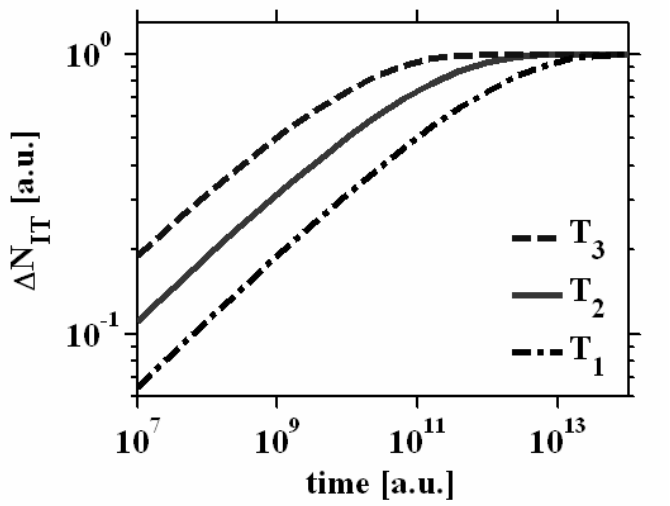


Fig. 2.12. Reduction in N_{IT} generation once $Si-H$ precursors (N_0 in (2.2)) is depleted. The saturation point increases as temperature is reduced.

where b is a constant ($\approx n/2$) and $n = 1/6$. The (2.17) is readily approximated as

$$\frac{N_{IT}}{N_0} \approx (1 - exp[-(\frac{t}{\tau})^n]) \tag{2.18}$$

by Taylor series expansion (correct to the 2nd order, i.e. error $\approx 0.2 \cdot (N_{IT}/N_0)^2$, and absolute error bounded by a factor of 2 as N_{IT}/N_0 approaches 1). Again, if the transport is dispersive, i.e. $D_H \approx D_0 t^{-m}$, the $n(1-m)$ replaces time-exponent n. In addition, if one assumes H^+ drift rather neutral H diffusion, (2.17) and (2.18) remain unchanged except $n = 1/6$ is replaced by $n = 1/3$ and D_H replaced by $(2E_{ox})$. Fig. 2.12 shows that this model would also lead to quasi-saturation of the NBTI characteristics. Since the saturation point is determined by certain density of traps, it will be reached at different times at different temperatures.

2.4.4 Saturation due to Dispersion in Bond Energies (B-D Model)

The third model assumes that $D_H(ox) = D_H(poly)$ and $N_0 \gg N_{IT}$ within the experimental window, so that neither of the above models can explain quasi-saturation.

In the bond dispersion model (B-D) model, quasi-saturation arises from the well known fact that $Si-H$ bonds have a distribution of bonding energies ($\approx 0.1eV$). The weaker bonds are rapidly broken as soon as the stress is applied, so that initial N_{IT} generation is high. The remaining bonds, on the average, are stronger so their dissociation rate continues to fall with time, leading to quasi-saturation of interface trap generation. The B-D models assumes H diffuses away from the interface faster than N_{IT} generation ($D_H/T_{ox} \gg k_F$). The model also assumes that interface annealing is negligible ($k_R \approx 0$) making second term of (2.2) irrelevant. Since B-D model assumes that the $Si-H$ bonding or activation energy, E_A, is not unique, so that (2.2) now needs to be rewritten as $\frac{dN_{IT}}{dt} = \int_{E_A} k_F(E_A)D_0(E_A)dE_A$ with $\int_{E_A} D_0(E_A)dE_A \equiv N_0$ and $N_0 \gg N_{IT}$, D_0 being the bond-energy distribution. Assuming that

$D_0(E_A) = \frac{1}{\sigma}\frac{exp(-\Delta\epsilon)}{[1+exp(-\Delta\epsilon)]^2}$ with $\Delta\epsilon \equiv \frac{(E_A - <E_A>)}{\sigma}$ and $k_F = k_0 \cdot exp(\frac{-E_A}{k_B T})$. ($< E_A >$ is the average activation energy, σ is) the standard-deviation of bonding energies, and T is temperature), trap generation is given by [49] $\frac{N_{IT}}{N_0} = 1 - \frac{1}{1+(t/\tau)^n}$ and $n = (\frac{k_B T}{\sigma})^p$. The exponent p depends on the range of fit to the experimental data. The numerical results for different temperatures are shown in Fig. 2.13. While B-D model provides one possible interpretation of NBTI saturation, since annealing term in (2.2) is neglected, the B-D model (in the current form) can not explain the relaxation experiments. However, like S-E and Reflection model, the T-dependence of time-exponent n arises from dispersion (bond dispersion for B-D model, diffusion or drift-dispersion in S-E and R-D models). In the absence of bond-dispersion, the other models would still predict fractional time-exponents, but the B-D model would not. Finally, since B-D model does not consider H diffusion or H^+ drift explicitly, therefore unlike R-D model, exponent n does not depend on geometry and only hard saturation of N_{IT} is possible (i.e. n approaches zero for long stresses).

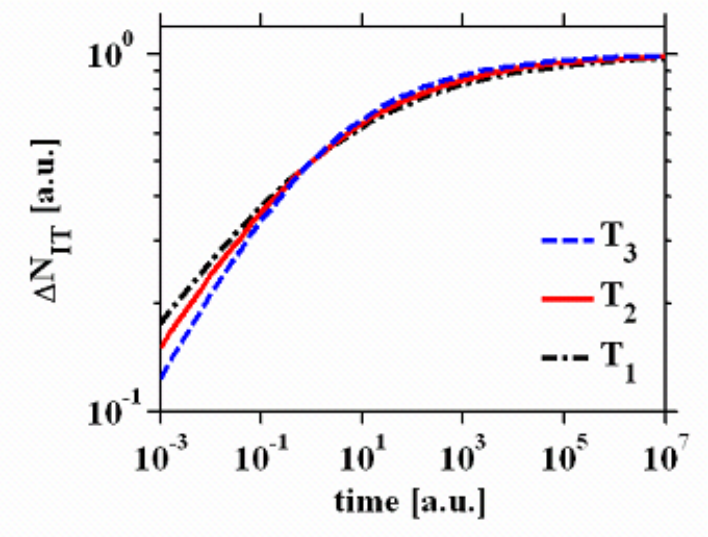


Fig. 2.13. Reduction in interface trap generation dispersion in $Si-H$ bonding energies, $\sigma = 0.1eV$

2.4.5 Saturation Due to H to H_2 Transition ($H - H_2$ Model)

As discussed in Sec. 2.3.3, $H-H_2$ model proposes that the quasi-saturation arises from the gradual transition from H to H_2 (rather than the instantaneous transition implied by the use of law of mass-action. Since H diffuses with exponent $n = 1/4$ and H_2 diffuses with exponent $n = 1/6$, the continual reaction from H to H_2 is reflected in time-dependent transition from high to low trap generation exponent (in other words quasi-saturation). This model interesting because it does predict that the slope will saturate towards $n = 1/6$, which is apparently consistent with experimental data. The details of this model are discussed in Section 2.3.3. A representative curve from this model is also shown in Fig. 2.14. The essential aspect of these plots are that since diffusion coefficients of both H and H_2 are temperature dependent, therefore the decay length of H into H_2 depends on temperature. This temperature dependence is the key to validate if H to H_2 transition is the determining factor of quasi-saturation characteristics of NBTI.

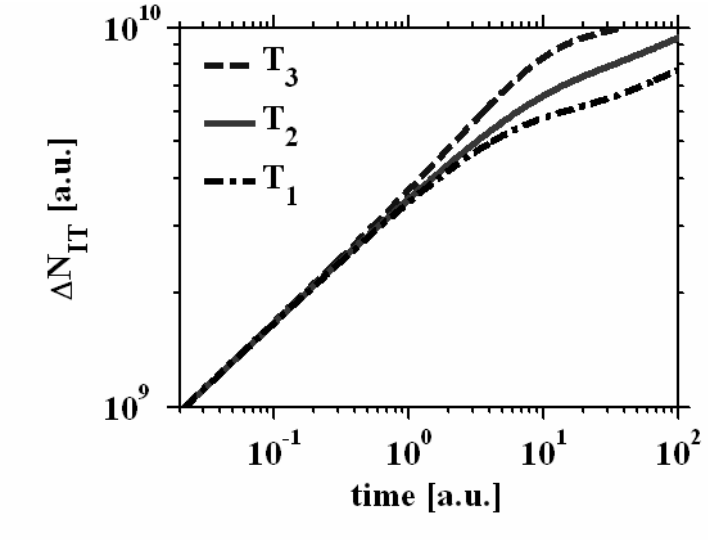


Fig. 2.14. Since the transition from H (which dominates early N_{IT} generation) to H_2 (dominates trap generation at later time) is temperature-dependent, the times for quasi-saturation also depends on temperature.

2.4.6 Saturation as an Artifact of Measurement Delay (Delay Model)

Finally, it has been argued that quasi-saturation in N_{IT} generation does not reflect any actual reduction of trap generation rate (as has been discussed above), but rather this is a consequence of limitations of N_{IT} measurement techniques. Specifically, in order to measure N_{IT} by standard techniques like charge-pumping or threshold voltage shift, the NBTI stress is temporarily interrupted, the N_{IT} measurement is done, and stress is reapplied (details will be discussed in Sec. 3.1). During this measurement interval (after the stress has been turned off and before the measurement has began, i.e. measurement delay) a significant fraction of N_{IT} is annealed because, although once the stress is removed, the field dependent k_F term in (2.2) becomes zero, so that no new N_{IT} generation is possible, however, the field independent reverse reaction term (k_R) remains nonzero and drives the reverse annealing of H and N_{IT}, with a net reduction in N_{IT}. The subsequent measure of N_{IT} reflects this annealing-induced

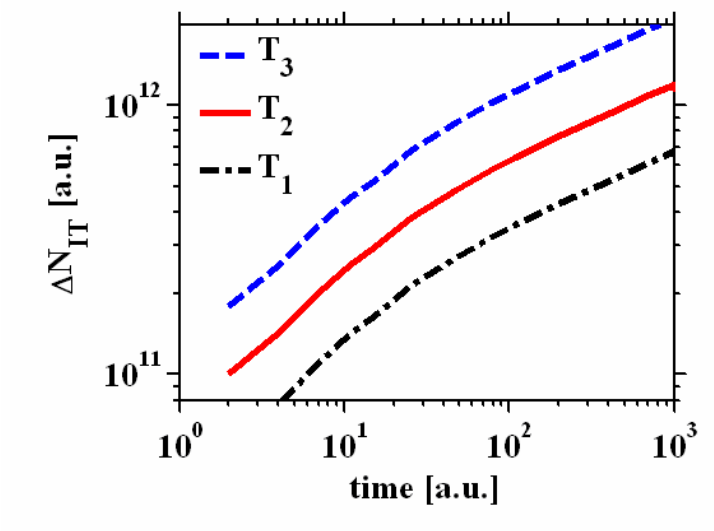


Fig. 2.15. Numerical simulation of measurement delay also gives rise to quasi-saturation behavior in N_{IT} vs. time characteristics ($T_3 > T_2 > T_1$ and delay=1sec).

reduction in actual trap density. The exact loss of N_{IT} due to this measurement issue depends on the relative magnitude of stress interval and measurement delay. Generally, since stress intervals are set to increase geometrically with time while the measurement delay remains constant, N_{IT} loss is initially significant, but it continues to reduce with time so that eventually delay-interrupted N_{IT} approaches the uninterrupted N_{IT} values (Fig. 2.15).

Fig. 2.16 and Fig. 2.17 summarize the results of this section. Fig. 2.16 plots the temperature dependence of saturation point or equivalently, the time to quasi-saturation (Fig. 2.17). Considering the Fig. 2.16, a strong temperature dependence of the quasi-saturation point would support the Reflection model or the S-E (Hard Saturation) model. On the other hand, if the quasi-saturation is temperature-insensitive, one would consider the Delay model or $H - H_2$ model (assuming k_{H_1} and k_{H_2} have weak temperature-dependences) as possible explanation of quasi-saturation. In addition to temperature as being of measurement variable, one can also study the predictions of the models for thickness and voltage dependence.

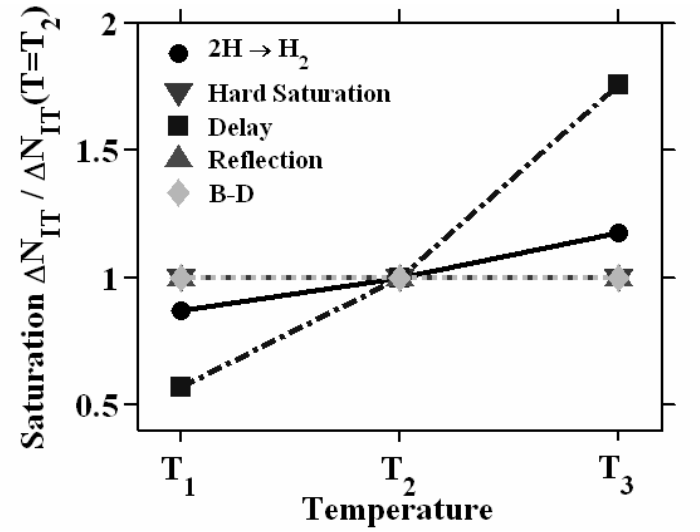


Fig. 2.16. Summary of results from previous figures. The interface trap density at which quasi-saturation point is reached.

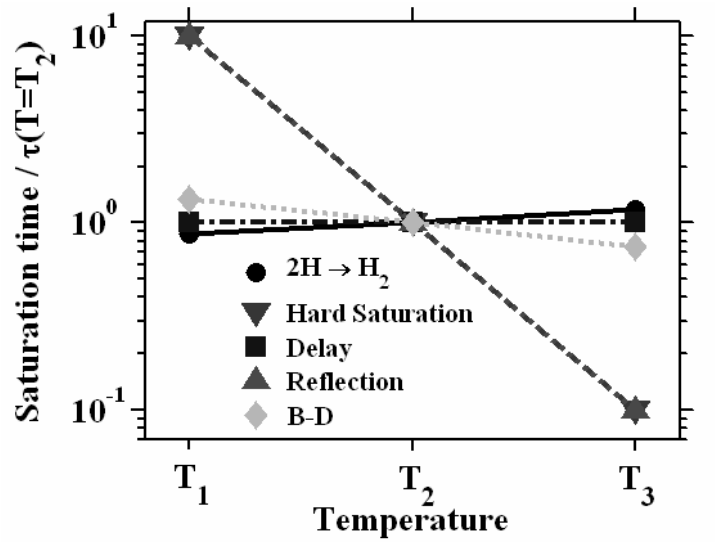


Fig. 2.17. The time (τ) at which the quasi-saturation point is reached depends on the model. Each of these models predicts very different temperature dependence which can be compared against experiments.

Given that the delay model explores measurement artifacts, this has to be an important component of any comprehensive model for NBTI saturation. Once corrected

for measurement delay, the data will then have to be analyzed for consistency by the remaining four models.

Among these four models, B-D model, in the form described above, is an unlikely candidate because the model is not consistent with annealing data. Of the remaining three models, in analysis of temperature, thickness and stress dependence seems to support the Reflection model over the S-E model. Both the $H - H_2$ model and the B-D model highlight important points (that $H - H_2$ transformation is not instantaneous and that bond-strength is not unique), but their contributions may not be the most important factors in determining NBTI quasi-saturation behavior. Overall, the measurement delay and subsequent recovery are the key factors to evaluate the saturation behavior and mechanisms correctly. Recovery effects will be discussed further in the next chapter.

2.5 Summary

The standard Reaction-Diffusion model based on H diffusion is extended for H_2 and $H - H_2$ in the light of experimental observations and theoretical consistency. The generalized $H - H_2$ framework can explain experimental trends better than previous models. Alternative NBTI models are compared in the context of saturation for degradation. Temperature, voltage, time, and recovery characteristics were summarized in order to identify the most plauible theory for NBTI. R-D model is established among others because of its strengths.

3. NBTI RECOVERY: R-D PERSPECTIVE

The discussion on NBTI so far focused on DC degradataion condition, however in most circuits, the MOSFETs experience AC stress. One peculiarity of NBTI is the recovery of degradation when the stress is reduced or turned off. The improvement in MOSFET characteristics are attributed to the reduction of interface trap density during recovery. According to the Reaction-Diffusion model, the interface traps are annealed by the hydrogen released during stress. This can be seen from the interface trap generation rate of (2.2)(for H diffusion, H_2 case is conceptually the same),

During recovery, with the reduction of $|V_{GS}|$, the bond-breaking rate, k_F, is reduced significantly since k_F has exponential voltage dependence [8]. The right-hand side term in (2.2) dominates, therefore the interface trap density decreases over time.

3.1 Impact of measurement delay

One of the most important signatures of recovery is the time-slope of NBTI. Earlier NBTI experiments such as charge pumping employed measurement delay inherently, that is, the stress on MOSFETs were interrupted periodically to measure the characteristics. The delay is also common in industrial practice; for characterizing various PMOSFETs fabricated on the same wafer, the most straightforward method is to stress and measure in a serial fashion because most characterization systems cannot stress the PMOSFETs on the same wafer in parallel. Due to the power-law behavior of NBTI, the measurement times are usually chosen at logarithmically-spaced intervals (data are assessed in $log_{10}(time)$ scale) during stress. At early stages of the experiment, the delay time is comparable to the stress time and therefore the impact of recovery is more. At later times, recovery time is much less than the stress times so the delay has a negligible effect. As a result of this experimental procedure, the un-

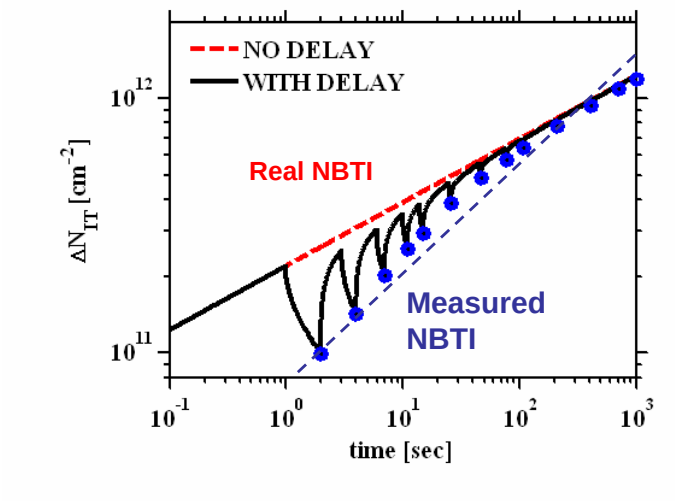


Fig. 3.1. With interruption of stress, NBTI time-exponent increases due to recovery effects from R-D theory. Resulting exponent and extrapolation for device lifetime can be misleading.

avoidable delay yields a higher time-exponent compared to the DC stress conditions, as depicted in Fig. 3.1. As the delay time increases, the time-exponent increases. This is because of the fact that an increase in delay causes more recovery, since recovery time is comparable to the stress times at earlier periods. An experimental manifestation is given in Fig. 3.2 following [32]. Also, the DC-stress lifetime projection in the presence of delay is going to be different from the case of 0-delay due to the recovery effects therefore device and circuit lifetime methods should be carefully evaluated.

3.2 AC stress

NBTI under AC stress is also affected by recovery. During the OFF cycles, the degradation relaxes. However since the subsequent ON and OFF cycles have comparable durations and recovery takes place at linearly-spaced time points, the time-slope of the overall degradation is not higher than the DC-stress case, but rather DC and AC degradation are parallel as in Fig. 3.3. The magnitude of degradation is of course

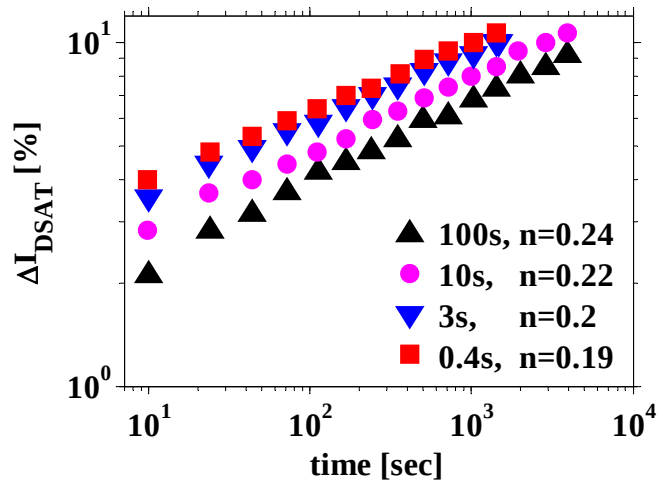


Fig. 3.2. As the measurement delay increases, time exponent of experimental data increases. Data from [32].

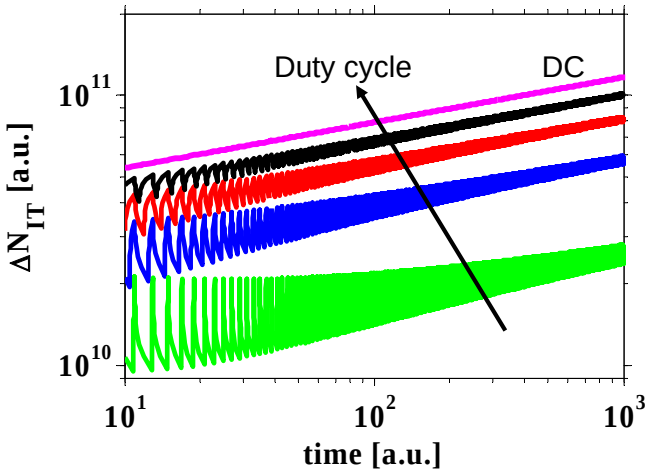


Fig. 3.3. Degradation under AC stress with square waveforms (shape of pulse) having different duty cycles: 100(=DC), 75, 50, 25, and 5% duty cycles were used in R-D simulations.

less than the DC case since effectively, the MOSFET experiences less stress under AC bias. This can also be seen in Fig. 3.3 as for different duty cycles of square-waveform gate bias. As the duty cycle decreases, so does the overall AC degradation. The

asymptotic evolution of degradation can be predicted by a simple relationship in this condition:

$$\Delta N_{IT}(AC) = \alpha_{effective} \cdot \Delta N_{IT}(DC) \tag{3.1}$$

where $\alpha_{effective}$ is the effective activity factor which can be written as

$$\alpha_{effective} = \sqrt{\frac{1}{t_0}\int_0^{t_0} \frac{V_{GS}^2(t)}{V_{GS,max}^2} dt}. \tag{3.2}$$

Essentially, (3.2) is the normalized root mean square of the input signal, where t_0 is sufficiently long time so that $t_0 \gg T_{cycle}$ and T_{cycle} is the period of the waveform. For the square-waveform, $\alpha_{effective} = \sqrt{\frac{\%duty\ cycle}{100\%}}$.

3.3 Random activity

In circuits, generally, the MOSFETs don't undergo square waveforms since the switching activity can be completely random. For that case, it is questionable if the above analysis and (3.1) will hold. Fig. 3.4 illustrates degradation under random activity. The signal was taken to be a uniformly distributed random sequence. The percentange ON time was taken as a parameter. As it can be seen, even in this case, the relationship is valid, the asymptotic degradation under random bias can be predicted by (3.1)(dot-dashed lines in Fig. 3.4) with $\alpha_{effective} = \sqrt{\frac{\%ON\ time}{100\%}}$. These relationships constitute a compact way of predicting AC degradation under arbitrary switching conditions and suitable for circuit simulators.

3.4 Frequency dependence

Another important question in AC degradation is the frequency degradation. Since frequency determines the ON and OFF times, the interface trap generation and annealing time intervals will vary with it. For low frequencies, the recovery time will be higher. Also, if these durations become very short with high frequencies, the rate of mechanisms could be affected. Fig. 3.5 compares AC degradation with several frequencies with the DC case.

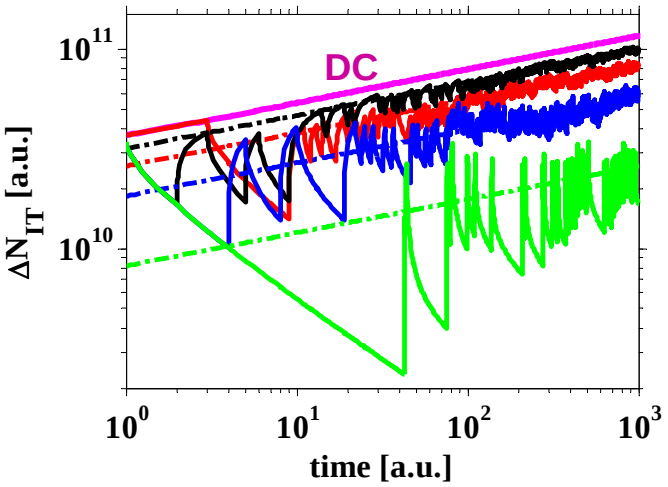


Fig. 3.4. Degradation under AC stress with arbitrary switching activity. 100(=DC), 75, 50, 25, and 5% ON-time were used in R-D simulations. Random sequence is assumed for gate bias. All waveforms have the same pulse frequency.

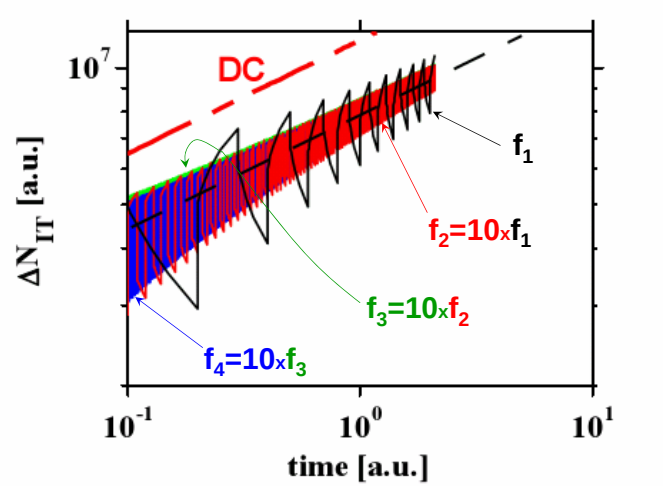


Fig. 3.5. Degradation under AC stress with several different frequencies, compared to DC degradation. Square waveform (50%) is assumed for gate bias. Simulated frequency range is 3 orders of magnitude.

Asymptotically, all the frequencies behave very similar, their degradations are about the same and also obey (3.1). The frequency independence can be derived

analytically from the R-D model, as originally done in [57]. The analysis is repeated here for completeness of the R-D theory. The MOSFET is assumed to experience AC stress (in the order of stress-recovery-stress-recovery...) with square waveforms of 50% duty-cycle and the stress/recovery duration is t_s in one cycle. After the first stress period, the recovery takes place during t, where $t_s < t \leq 2{\cdot}t_s$. The interface trap density in the recovery can be found from

$$N_{IT}(t_s + t) = \frac{N_{IT}(t_s)}{1 + \sqrt{\frac{\xi t}{t_s + t}}}, \tag{3.3}$$

where ξ is the geometric factor for the hydrogen diffusion [13]. The term $\sqrt{\frac{\xi t}{t_s+t}}$ represents the ratio of hydrogen that has reached the interface in time t to hydrogen density that has escaped from it in t_s+t. The hydrogen (ξt) near the interface diffuses back and anneals dangling bonds. Nearly half of the hydrogen keeps diffusing away from the interface, thus ξ=0.5.

A fraction, R_N, is defined as the ratio between N_{IT} after $N{\cdot}t_s$ and that of t_s (for N^{th} cycle, R_1=1). After the first recovery,

$$R_2 = \frac{N_{IT}(2t_s)}{N_{IT}(t_s)} = \frac{1}{1 + \sqrt{\frac{0.5t}{2t_s}}} = \frac{2}{3}R_1. \tag{3.4}$$

During $2t_s < t \leq 3{\cdot}t_s$, the MOSFET is again under stress. The analyis of (3.3) can be repeated, however there is non-zero N_{IT} after $2t_s$ in this case. Instead, an effective stress time can be define so that

$$N_{IT}(2t_s) = R_2 A t_s^n = A t_e^n, \tag{3.5}$$

where A and n are extracted from characteristic NBTI time-dependence ($N_{IT}(t) = At^n$). From (3.5), $t_e = (R_2)^{1/n} t_s$. Then the interface trap density after the second stress period can be written as $N_{IT}(2t_s + t) = N_{IT}(t_e + t) = A(t_e + t)^n$. Therefore

$$R_3 = \frac{N_{IT}(3t_s)}{N_{IT}(t_s)} = \frac{(t_e + t_s)^n}{t_s^n} = (1 + (R_2)^{1/n})^n. \tag{3.6}$$

or $(R_3)^{1/n} = 1 + (R_2)^{1/n}$. This can be extended to any stress duration, $(R_{2k-1})^{1/n} = 1 + (R_{2k-2})^{1/n}$, $k > 1$.

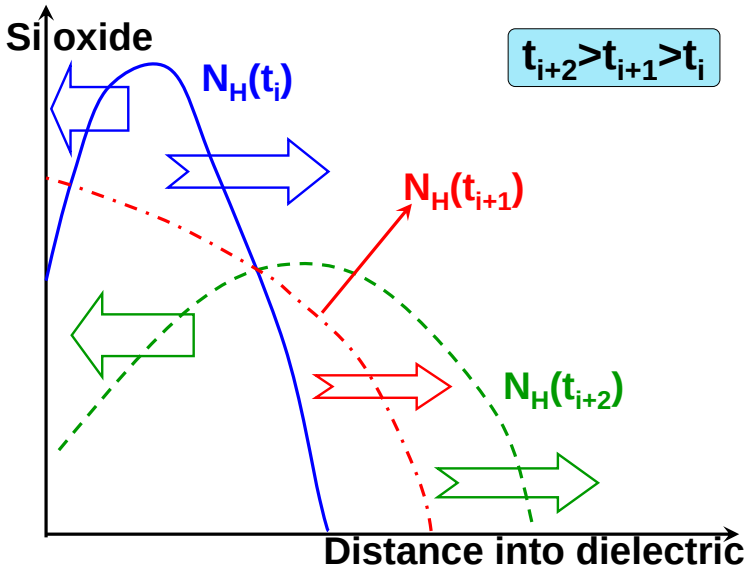


Fig. 3.6. Schematic of the hydrogen profiles during the stress and recovery cycles with $t_{i+1} - t_i = t_s$. Subsequent recovery, stress and recovery are shown.

For the second recovery, $3t_s < t \leq 4{\cdot}t_s$, the N_{IT} decreases similar to (3.3), however the hydrogen consumed in the first recovery does not contribute as in Fig. 3.6. Since $R_2 = \frac{2}{3}R_1$, $\frac{1}{3}R_1$ is already passivated. Therefore $R_4 = \frac{2}{3}(R_3 + 1/3R_1) = \frac{2}{3}R_3 + \frac{1}{3}R_2$. This can also be generalized for any recovery period: $(R_{2k}) = 2/3R_{2k-1} + 1/3R_{2k-2}$. The overall degradation can be considered as a sum of the stress and recovery cycles. For sufficiently large number of cycles,

$$\begin{aligned}(R_{2k-1})^{1/n} &= 1 + (R_{2k-2})^{1/n} \\ R_{2k-2} &= 2/3R_{2k-3} + 1/3R_{2k-4} \\ (R_{2k-3})^{1/n} &= 1 + (R_{2k-4})^{1/n}. \end{aligned} \tag{3.7}$$

The R_{2k-2} and R_{2k-4} can be eliminated to get a relationship between R_{2k-1} and R_{2k-3}, thus degradation can be traced back to R_1. Assuming n=1/6,

$$R^6_{2k-1} = 1+\left[2/3R_{2k-3} + 1/3(R^6_{2k-3} - 1)^{1/6}\right]^6 = 1+R^6_{2k-3}\left[2/3 + 1/3(1 - \frac{1}{R^6_{2k-3}})^{1/6}\right]^6. \tag{3.8}$$

The R_{2k-3} is much greater than unity for large number of cycles, therefore (3.8) can be approximated as

$$R_{2k-1}^6 \approx 1 + R_{2k-3}^6 \left[1 - \frac{1}{12R_{2k-3}^6}\right]^6 \approx \frac{2}{3} + R_{2k-3}^6. \tag{3.9}$$

From (3.9), R_{2k-1} can be obtained in terms of R_1 by expanding intermediate terms, R_{2k-3}, R_{2k-5}, R_{2k-7}, ... and summing all. As a result,

$$R_{2k-1}^6 \approx \frac{2}{3}(k-1) + R_1^6 \approx \frac{2}{3}(k-1). \tag{3.10}$$

Considering two waveforms with different frequencies, at a given stress time, $t_{str} = k_1T_1 = k_2T_2$ where $k_{1,2}$ are the number of cycles in t_{str} and $T_{1,2}$ are the durations of each degradation and recovery periods for the two waveforms. Using (3.10),

$$\frac{N_{IT1}(t)}{N_{IT2}(t)} = \frac{N_{IT1}(k_1T_1)}{N_{IT2}(k_2T_2)} = \frac{R_{k_1}N_{IT1}(t_1)}{R_{k_2}N_{IT2}(t_2)} \approx \frac{(k_1t_1)^{1/6}}{(k_2t_2)^{1/6}} \approx 1. \tag{3.11}$$

Thus, (3.11) implies frequency independence as shown by the R-D simulations in Fig. 3.5 and is consistent with experimental observations upto 2GHz [58].

3.5 Ultrafast degradation and recovery

As pointed out earlier, conventional NBTI characterization involves measurement delay which can cause artifacts. Conventional characterization equipments also limited by the measurement time in the order of milliseconds. Recent studies with dedicated fast measurement setups (Appendix A.8) showed that even in milliseconds, there can be significant degradation or recovery [14], [35]. It is therefore important to take ultrafast NBTI behavior into consideration for accurate characterization.

Although degradation part of NBTI is established, the recovery is not well understood. Experimental observations show that after DC stress, NBTI recovery has a $log(time)$ trend which cannot be explained by the conventional R-D theory (this issue will be revisited in Sec. 3.6) [14]. Moreover, the PMOSFETs in ICs rarely experience DC stress during operation, therefore it is not clear how the experimental recovery would behave under random activity. Understanding such a recovery behavior can

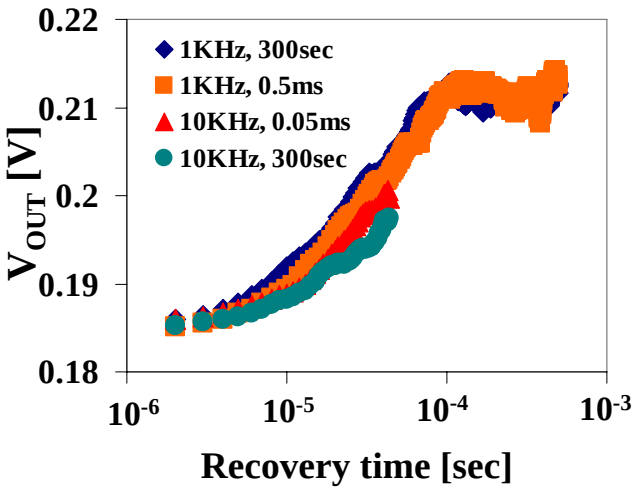


Fig. 3.7. In each cycle of AC stress, the I_D=V_{OUT}/R_F (see A.8) recovers. The shape of the recovery segments for different frequencies and stress times suggest that recovery is universal.

facilitate NBTI modeling in terms of circuit degradation when the switching activity of MOSFETs are considered.

NBTI recovery, particularly under AC stress, is examined by the ultrafast setup. The emphasis is on the history dependence of NBTI which can appear under random switching activity. In order to assess history dependence, AC and combinations of DC and AC stress were performed [59]. The devices used in this study are from 65 nm technology with physical oxide thickness of 1.2nm. Stress was done at V_G=-1.6V and recovery was measured at V_G=-0.6V with V_D=-0.2V, all at T=150C.

Fig. 3.7 shows recovery behavior for AC stress (50% duty cycle) for different frequencies after different stress times. The trends imply that the recovery is universal under AC stress supporting [60]. The shape of the recovery is very similar after a short term (0.05 and 0.5 msec.) as well as long term (300 sec.). Despite the recovery, the overall $|I_D|$ decreases over time under AC stress as given in Fig. 3.8. This is consistent with the R-D model (Fig. 3.3).

Under alternating DC and AC stress, the recovery behavior differs from AC-only stress. Fig. 3.9 compares the two conditions after equivalent stress durations. The

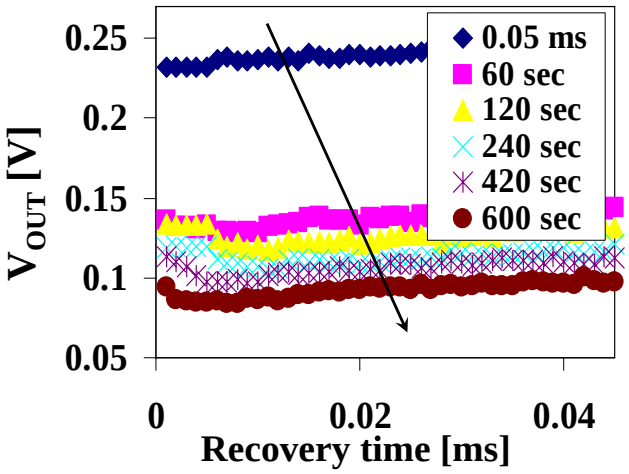


Fig. 3.8. Overall, the degradation continues in spite of the recovery in each cycle under AC-only stress.

AC-only recovery saturates since the degradation in one cycle is small (in 0.5 msec.), however under the DC+AC stress, the recovery continues because the initial 120 sec. DC stress degrades the MOSFET significantly. This result clearly signals history dependence in NBTI recovery. Similar characteristics were obtained for DC stress and recovery in [14].

The history dependence is also apparent from Fig. 3.10. The recovery segments during the AC after DC stress are shown for different frequencies and stress durations. Unlike the AC-only stress, the MOSFET characteristics improve under stress. The initial DC degradation is larger and the short recovery durations in AC cannot compensate for the degradation, therefore the recovery continues in the long term.

The history dependence is consistent with the R-D model as shown in Fig. 3.11. The DC+AC and AC-only stress schemes are simulated. In the AC-only condition, the overall $|V_T|$ increases however the situation reverses for DC+AC stress. The short relaxation in AC cycles are not long enough to alleviate the initial DC degradation, and the R-D model predicts the history effect.

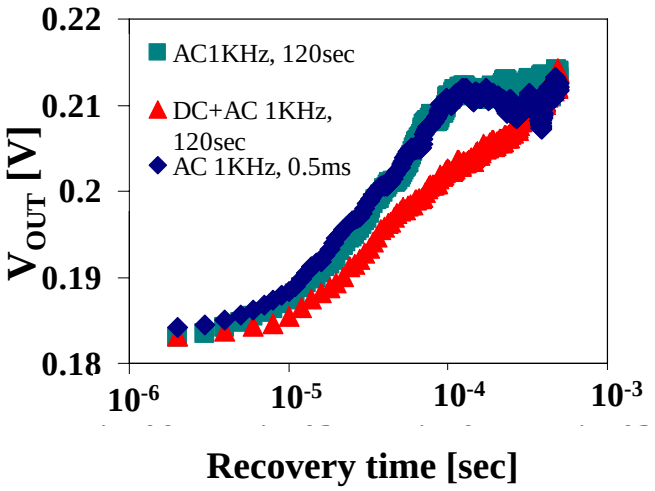


Fig. 3.9. Compared with AC-only stress, the shape of recovery under DC+AC stress is different. Under random activity, this suggests history dependent NBTI recovery.

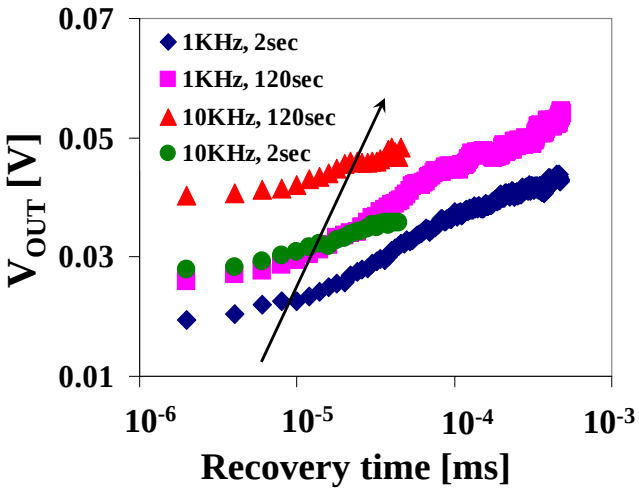


Fig. 3.10. Under DC+AC stress, the recovery continues in the longer term during the AC part of the stress. This result is the opposite for AC-only stress and supports the history dependence.

3.6 Experimental recovery vs. R-D behavior

The $log(time)$ trend of experimental NBTI recovery was posed to criticize the R-D model since the conventional theory, including H_2-only and $H-H_2$, is not satisfactory

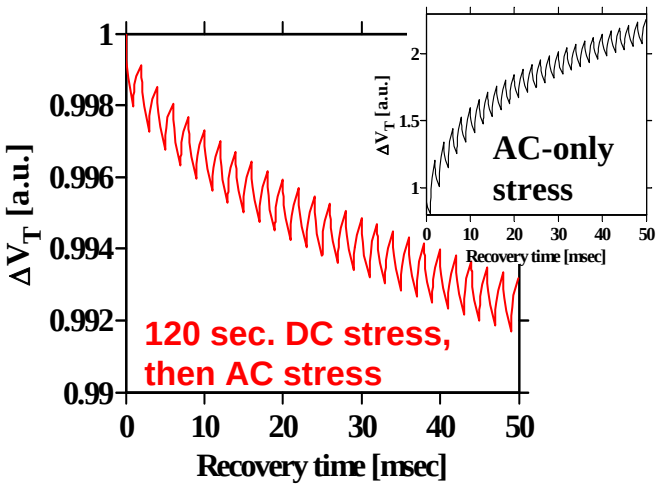


Fig. 3.11. The history dependence is observed from the R-D theory. The experimental AC-only (inset) and DC+AC stress are simulated.

to explain this time dependence. Hole trapping/detrapping besides interface trap generation were invoked to account for NBTI degradation and recovery [14], [61], [62], [63].

The inconsistency of the R-D model for recovery can be resolved by taking a closer look at the theory. As given in (2.2), the recovery increases as the dangling Si bonds are passivated. It is assumed that the interface trap disappears as soon as it is passivated by hydrogen. However, in reality, the interface trap is a donor-type state (dangling Si bond) not occupied by an electron. It requires an electron to completely passivate the interface trap despite the fact that hydrogen can be abundant in the vicinity of the dangling bond. This electron dynamics is not included in the conventional R-D model. When it is explicitly considered, the R-D model can result in the $log(time)$ consistent with the experiments [64]. An alternative model implementing dispersive reaction rates was also proposed to explain the recovery in the R-D framework [59]. The dispersion arises from the density of electrons available for passivation of the interface states [65].

3.7 Summary

Recovery is an important part of NBTI that needs to be considered carefully. Due to recovery, measurements can yield artificial characteristics leading to incorrect lifetime extrapolation. Under AC stress where the lifetime can improve significantly, the degradation is linked to its DC magnitudes by exploring the impact of duty cycle from R-D simulations. Ultra-fast measurements revealed that modeling of recovery is not straightforward under random activity stress because of strong history dependence, however a realistic random switching scenario can be modeled analytically with the same principles of AC stress. As for frequency dependence of NBTI, R-D theory predicts the experimental trends, i.e., weak frequency dependence.

4. GEOMETRIC IMPLICATIONS OF R-D MODEL FOR FUTURE-GENERATION MOSFETS

4.1 Introduction

A realistic NBTI model should also explain the degradation for future-generation transistors. Current technology node involves planar MOSFETs with considerably large dimensions along the width direction of the transistors. However, going down the technology roadmap, alternative device geometries such as narrow-width planar transistors, FINFET/triple-gate MOSFETs and surround-gate structures such as Vertical Replacament Gate (VRG) and nanowire transistors are considered for high density and high performance circuits [1]. Extrapolation of the NBTI reliability from the existing MOSFET generations to the ultra-scaled future technologies may lead to erroneous device lifetimes since the scaling of the dimensions and the structural changes can alter the degradation rates of the transistors. In order to fully evaluate the promises of such next-generation transistors, a solid prediction methodology which can extend the current NBTI characterization techniques to future technologies is needed. Numerical simulations based on the Reaction-Diffusion (R-D) framework as well as analytical calculations show that both the device geometry and the geometry of the degradation can alter the time behavior of the trap generation rate. In this chapter, the geometric interpretation of the R-D model to predict the NBTI reliability for ultra-scaled narrow-width planar, triple-gate and nanowire MOSFETs and VRG transistors is extended. First, the geometry-dependent R-D model is summarized and a straight-forward analysis method to capture the effective trends of scaling such device structures is introduced. Then, the numerical R-D simulation results are compared with the analysis developed. These formulas can be used to quickly estimate

the reliability performance of a device for specific operating conditions without the need for relatively time-consuming simulations.

4.2 Geometry Dependent Reaction-Diffusion Model

The time exponent, n, of the interface trap generation can be obtained by solving (2.2) and (2.3) simultaneously subject to appropriate boundary conditions. Before considering the NBTI degradation for narrow-width or surround gate devices numerically, it is instructive to consider the problem analytically. For a finite size device, the hydrogen diffusion will first begin as 1-D diffusion in bulk of the oxide. Over time, however the diffusion through the edges will become increasing important making the diffusion a 2-D problem, and eventually when the diffusion radius becomes much larger than the gate area of the device, the diffusion becomes three dimensional. Before imposing this transformation of the diffusion problem from 1-D to 2-D to 3-D as a general problem, the asymptotic limits of these diffusion cases are considered individually. For clarity, the analysis discusses only H diffusion, however the geometry dependence is applicable to H_2 and $H - H_2$.

4.2.1 Asymptotic diffusion in 1-D, 2-D and 3-D

One can readily estimate the exponent in diffusion-limited region by using the fact that every free H is associated with one interface trap from (2.1), i.e., $N_{IT}(t) = \int N_H(r,t)d^3r$. The release rate of hydrogen at the $Si/oxide$ interface is much higher than the diffusion velocity and the hydrogen profile broadens as $(D_H t)^{0.5}$ into the oxide. In NBTI, assuming the hydrogen profile in Fig. 4.1(b) due to the 1-D nature of hydrogen diffusion, During the diffusion-dominated regime, the $\frac{dN_{IT}}{dt}$ term is negligible compared to the other bond-breaking and annealing terms in (2.2), therefore (2.2) can be simplified into (2.6) Substituting $N_H^{(0)}$ from (2.5) into (2.6), (2.7) is obtained.

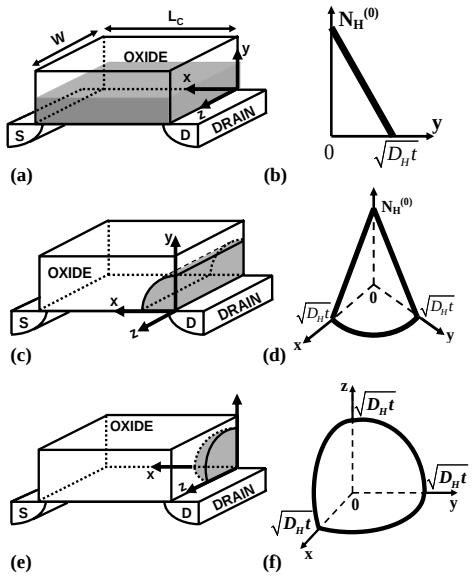


Fig. 4.1. 1-D, 2-D, 3-D diffusion of hydrogen are illustrated in the gate-oxide of planar MOSFETs. Shaded regions in (a,c,e) represent hydrogen. (a) NBTI, trap distribution is uniform over the channel, (b) Hydrogen profile for 1-D diffusion, (c), (d) degradation near the edge and 2-D diffusion profile, (e) Degradation localized at a point, (f) 3-D hydrogen diffusion profile.

Similarly, for 2-D diffusion from a line source, the hydrogen profile can be assumed as in Fig. 4.1(d) and repeating (2.5-2.6) as below gives (4.2)

$$N_{IT}(t) = \frac{1}{4}\int_0^{\sqrt{D_H t}} N_H^{(0)}(1 - \frac{r}{\sqrt{D_H t}})\cdot 2\pi r dr = \frac{\pi\cdot N_H^{(0)}}{12\cdot W}\sqrt{D_H t}. \tag{4.1}$$

$$N_{IT}(t) = \sqrt{\frac{\pi\cdot k_F N_0}{12\cdot W k_R}}(D_H t)^{\frac{1}{2}}. \tag{4.2}$$

Finally, the diffusion from a point source as illustrated in Fig. 4.1(c) is of 3-D and applying the same procedure through (4.3), the R-D model predicts (4.4) assuming the hydrogen profile shown in Fig. 4.1(f)

$$N_{IT}(t) = \frac{1}{8}\int_0^{\sqrt{D_H t}} N_H^{(0)}(1 - \frac{r}{\sqrt{D_H t}})\cdot 4\pi r^2 dr = \frac{\pi\cdot N_H^{(0)}}{24\cdot W}(D_H t)^{3/2}. \tag{4.3}$$

Using (2.6) and (4.3), (4.4) is found as

$$N_{IT}(t) = \sqrt{\frac{\pi\cdot k_F N_0}{24\cdot k_R}}(D_H t)^{\frac{3}{4}}. \tag{4.4}$$

Above estimates show that regardless the physics of k_F, k_R, and D_H, the R-D model predicts that the geometry of the problem determines the rate of interface trap generation (n, the time-exponent) and this rate increases with the dimension of the hydrogen diffusion. Although (4.4) gives a time-exponent of 3/4, numerically, the maximum value it can reach is about 1/2 for 3-D geometry. This can be due to the limited supply of hydrogen release from the interface or to the maximum rate diffusion can remove hydrogen within that volume.

The NBTI time-dependence is governed by the slower diffusion process, therefore relatively small distribution (spread of about 0.1 eV and mean energy of 2.8 eV from [66]) in the activation energies of k_F and k_R does not affect the time-exponent appreciably [41].

4.2.2 Diffusion length, λ_D , and finite size effects

The discussion in the previous section is based on the idealized assumption that for 1-D diffusion the $x-z$ plane (Fig. 4.1(a)) has infinite extent, similar for 2-D diffusion, the assumption is that the length of the cylinder is infinite, and that for 3-D diffusion, the source is a geometrical point. In reality of course, since the channel dimensions of a MOSFET are finite, they can affect the degradation rates of NBTI. In this section, these additional geometrical considerations are discussed and the corresponding rates are derived. The effect of the geometry of hydrogen diffusion on degradation rate can be analyzed with the concept of the diffusion length, $\lambda_D = (D_H \cdot t)^{0.5}$. Fig. 4.2 depicts the diffusion of hydrogen from a planar degraded region. Over the region, the diffusion is of 1-D, however, at the edges, the hydrogen profile extends outward in a cylindrical form. At the corners, similarly, the diffusion becomes 3-D due to the finite dimensions of the degraded area. The exact volume occupied by hydrogen contains the 1-D rectangular box, the 2-D cylindrical side regions and the 3-D spherical corners.

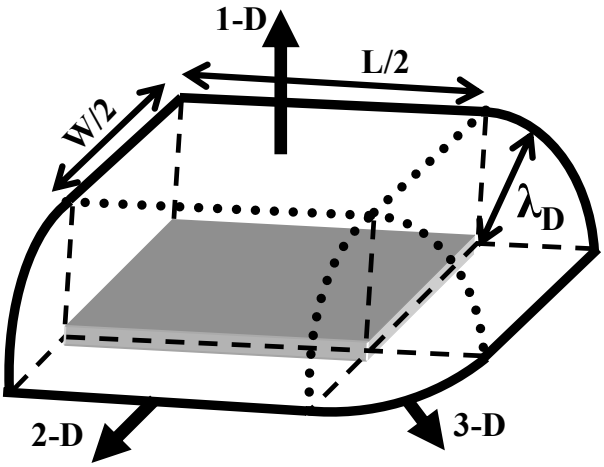


Fig. 4.2. The diffusion of hydrogen from a planar semiconductor/oxide interface. The rectangular plate (shaded region) shows the degraded interface from which H is released. The wire frame represents the hydrogen profile. At the edges and corners, 2-D and 3-D profiles add to the volume occupied by hydrogen. As the structure is scaled, first the 2-D and then the 3-D components begin to dominate the volume. Due to symmetry, only one quadrant of the gate oxide is drawn.

The volume used in the derivation of (2.7) assumes $W, L \gg \lambda_D$, but if the contribution of the edges and corners are considered, the trap density can be generalized into (4.5)

$$N_{IT}^{PL}(t) = \sqrt{\frac{k_F N_0}{W \cdot L \cdot k_R}} [W \cdot \lambda_D \cdot (L + \frac{\pi \cdot \lambda_D}{2}) + \frac{2\pi \cdot \lambda_D^3}{3}]^{1/2}. \tag{4.5}$$

(for detailed derivations of (4.5-4.7) see Appendix A.2). From (4.5), when $W, L \gg \lambda_D$ (Fig. 4.1(a)), $N_{IT} \sim t^{0.25}$ is obtained as in the case of NBTI time dependence in (2.7), and if only L is scaled such that $L \sim \lambda_D$ (Fig. 4.1(c)), then $N_{IT} \sim t^{0.5}$. Furthermore, if $W, L \sim \lambda_D$, N_{IT} approaches $\sim t^{0.75}$ as in (4.4), signaling increased degradation rates for aggressively scaled planar MOSFETs. During the degradation, the hydrogen can diffuse into the Si channel, surrounding isolation or poly-gate. In this condition, the diffusion constant of hydrogen in other materials should be considered and D_H is replaced with the effective diffusion constant (e.g., $D_{EFF} = D_H$ before hydrogen reaches poly gate, $D_{EFF} = D_{POLY}$ afterwards). The D_{Si} is about the same as D_H [67]. The technologically-important poly-gates are highly doped, therefore the diffusion of

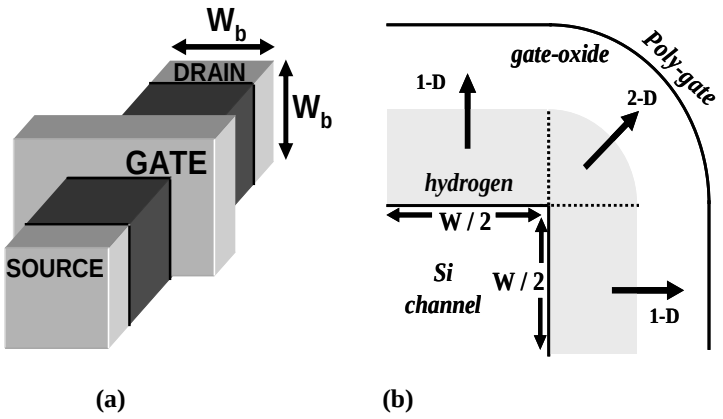


Fig. 4.3. (a) Schematic view of triple-gate MOSFET, (b) the square cross section of the body of the transistor. The planar channels allow 1-D hydrogen diffusion in the dielectric; however, the corners of the triple-gate transistor add 2-D regions and enhance the overall diffusion rate. The effect is pronounced when the transistor is scaled. One quadrant of the cross section is shown due to symmetry.

hydrogen is not affected significantly by the grain-boundaries [68]. The D_{EFF} does not change time-exponent of NBTI except for brief transients when the hydrogen meets with material boundaries. The diffusion length is modified accordingly, i.e., $\lambda_D \sim (D_{EFF} \cdot t)^{0.5}$, and the geometry effect will still increase the degradation rate. The $Si/oxide$ interface roughness has negligible effect on the geometry analysis in the R-D model. In modern MOSFETs, the roughness is at atomistic dimensions (few Angstroms) which is much smaller compared to λ_D. The diffusion length analysis can also be extended for non-planar MOSFET geometries. Fig. 4.3(a) shows the cross section of the body of a triple-gate transistor. The schematic in Fig. 4.3(b) reflects the hydrogen diffusion when NBTI is considered. The profiles over the three planar channels are 1-D, but the corners contribute 2-D cylindrical volumes. Similar to the

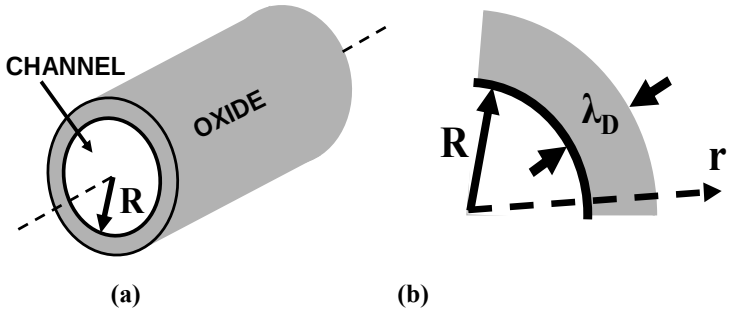


Fig. 4.4. (a) Cylindrical MOSFET with channel radius, R, the gate oxide surrounds the channel, (b) the hydrogen diffusion occupies a cylindrical shell in the gate oxide. The shaded volume corresponds to hydrogen diffusing away from the semiconductor/oxide interface

planar MOSFET case, the degradation can be derived as (4.6) if the gate length is sufficiently long compared to λ_D,

$$N_{IT}^{TRI}(t) = \sqrt{\frac{k_F N_0}{W \cdot k_R}}[W \cdot \lambda_D + \frac{\pi \cdot \lambda_D^2}{4}]^{1/2}. \tag{4.6}$$

When $W \gg \lambda_D$, (4.6) approaches to (2.7) since the diffusion is mainly 1-D. The time-exponent, n, increases as W is scaled and results in a higher degradation rate. Another non-planar structure is the surround-gate, cylindrical MOSFET with channel radius R depicted in Fig. 4.4(a). For NBTI degradation, the hydrogen diffusing is confined within the distance $R < r < R + \lambda_D$, and shown as the cylindrical shell in Fig. 4.4(b). The interface trap density can be calculated from the geometry dependent R-D relation as

$$N_{IT}^{SPH}(t) = \sqrt{\frac{k_F N_0}{R \cdot k_R}}[\lambda_D \cdot (1+\frac{R}{\lambda_D}) \cdot (2R+\lambda_D) - \frac{1}{3} \cdot \{R^2 + R(R+\lambda_D) + (R+\lambda_D)^2\}^{1/2}]. \tag{4.7}$$

In the limit $R \gg \lambda_D$, the hydrogen diffusion is effectively of 1-D nature and the cylindrical channel resembles that of a planar MOSFET. Therefore $N_{IT}(t) \sim t^{0.25}$

is obtained under NBTI stress. For $R \ll \lambda_D$, the hydrogen diffusing into the Si channel can be considered with the D_{EFF}. Eventually, the density of hydrogen inside the channel will be equal (and still time-dependent, inherently in the R-D model) to that of the $Si/oxide$ interface and the diffusion will only evolve outward in the radial direction. In this case, n approaches 1/2, as expected from (4.7). Similar to the narrow-width planar MOSFET, the time exponent increases with time for each radius. As the devices are scaled to very small dimensions, the area of the Si/SiO_2 interface is reduced. Therefore, individual transistor may contain only a few interface traps and the number of traps can vary significantly from one transistor to the next. In this context, the results in this paper based on R-D model should be interpreted as being the median NBTI degradation averaged over a large number of transistors (e.g. millions of transistors in an IC). The fluctuation of NBTI degradation for individual transistors (around the median degradation predicted by the R-D model) can then be easily calculated by statistical models [49], [69]. This is analogous to gate oxide degradation models, where average degradation are analyzed by physical models like Anode Hole Injection or Hydrogen Release, while the fluctuation effects are accounted for by statistical models based on percolation theory [70]. After establishing how geometry of diffusion is related to trap generation rate, specific technology examples where this could be relevant will be considered and the validity of the analysis by detailed numerical simulation will be explored.

4.2.3 Simulation Results and Comparison with Analysis

Since the '90s, the continued miniaturization of MOSFETs has allowed unprecedented scaling of planar devices now reaching down to 90nm node. According to the ITRS roadmap, the oxide scaling is about to reach its technological limits due to the exponential increase in leakage current [1]. This leaves the burden of scaling mainly to the Si body of the transistors. Until the 45 nm node, the employment of SOI and ultra-thin body (UTB) architecture can maintain the miniaturization

trends by reducing short channel effects (SCE) and capacitive losses, and blocking most of the leakage paths [71]. Afterwards, multiple-gate MOSFET structures such as FINFETs and tri-gate transistors will be needed to improve the SCE further and control the leakage better. Scaling below channel lengths of 20 nm will require non-classical architectures and materials. VRG and nanowire devices are considered as promising ultra-scaled options that can offer both high performance and high density circuits [72], [73]. Self-heating observed in the UTB and SOI MOSFETs may speed up the temperature-activated NBTI degradation; therefore NBTI can become a more important concern for the aforementioned devices [74], [75]. Scaling the dimensions of these future generation devices can also affect the NBTI reliability due to the geometric dependence as developed in the previous section. In this section, numerical simulation results for several MOSFET geometries that are candidates mentioned in the ITRS Roadmap are presented. The details of the numerical implementation can be found in Appendix A.2. The analyses based on the diffusion length are also compared here with simulations for each device structure. Throughout this work, only the degradation under DC stress is considered for the simulations to compare the worst-case NBTI. AC degradation is proportional to that of DC and asymptotically gives the same time-exponent, as obtained from experiments and R-D simulations [13], [19]. Another important feature of NBTI is saturation characteristics. The decrease in the time-exponent during saturation can be attributed to build-up of hydrogen at the material interfaces, consumption of Si-H bonds or experimental artifacts such as the delay and subsequent recovery in charge pumping measurements. The devices studied in the literature so far are relatively large-area transistors from the perspective of this work. At this point, it is not clear whether the saturation or geometry effects will dominate for ultra-scaled devices, therefore saturation effects are not included in the numerical solutions. As for the parameters, the D_H (at room temperature) is taken to be $10^{-15} cm^2/sec$ and N_0 as $10^{14} cm^{-2}$ for all the simulations [16], [31], [76]. In the experimentally-feasible time period (up to 10^5-10^6 sec.), $\lambda_D = (D_H \cdot t)^{0.5}$ grows to about 100nm. The k_F and k_R were assumed to be the same for transistors be-

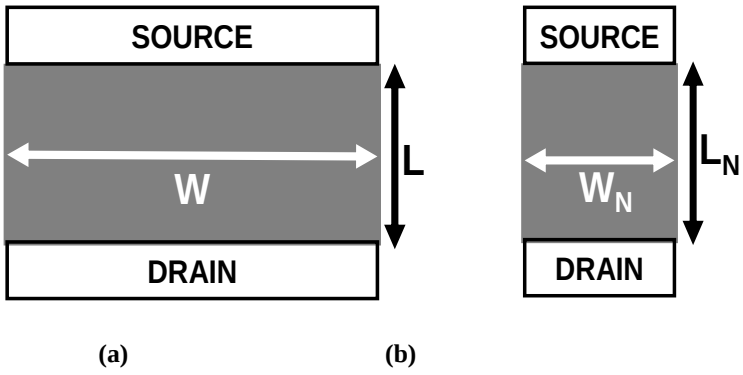


Fig. 4.5. Top view of planar MOSFETs with shaded area representing the channel, (a) conventional wide-width MOSFET ($W > L$), (b) Narrow-width MOSFET ($W_N \approx L_N$). Narrow-width geometry implies higher NBTI degradation in the R-D model.

longing to a particular geometry. To reduce the simulation time, one-sided diffusion (see Fig. 4.1(a,b)) is implemented in the numerical solutions, and since $D_{EFF} \approx D_H$. The distribution in the activation energies of k_F and k_R are not implemented in the simulations because anticipated effects are negligible.

A. Narrow-width Planar MOSFET

Shrinking the channel lengths of planar MOSFETs for each technology node also scales the channel widths. Additionally, the channel widths can be narrowed further for applications that require higher packing densities on chips [77], [78]. A wide-width ($W_W > L$) and a narrow-width ($W_N \approx L$) planar MOSFETs are illustrated in Fig. 4.5(a) and (b), respectively. Such channel dimensions can have significant impact on the device reliability. The numerical simulation result for NBTI degradation is displayed in Fig. 4.6 for the narrow-width and wide-width cases. The narrow

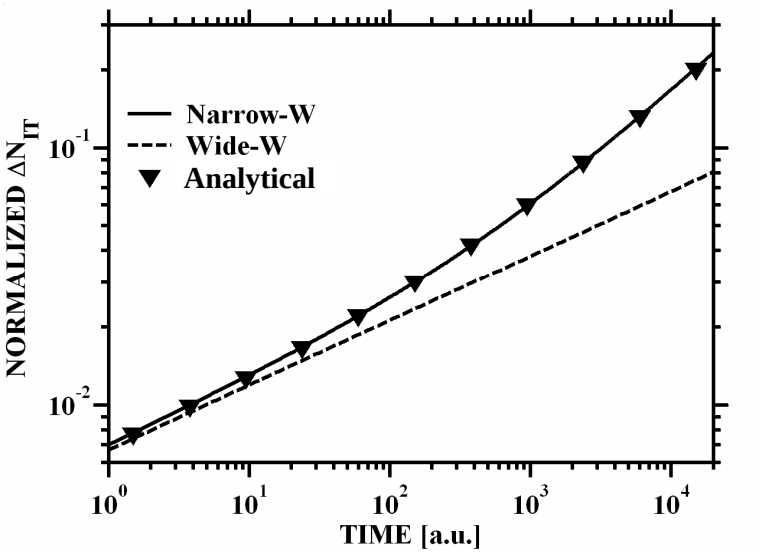


Fig. 4.6. NBTI induced interface trap density for wide ($\approx$ 10 m, dashed line) and narrow-width ($\approx$ 25 nm, solid line) MOSFETs. The time behavior is obtained from the geometry-dependent R-D simulations. The interface trap density increases when the width is scaled. The symbols from (4.5) agree remarkably well with NBTI simulations. Since λ_D increases with time, the degradation becomes faster as λ_D/W grows.

channel width enhances the diffusion of hydrogen released from the Si-H bonds at the Si/oxide interface; therefore NBTI induced interface trap densities increase. The diffusion length analysis obtained from (4.5) agrees well with the R-D model. Since the analysis involves the geometry of the diffusion, the degradation trends for both wide and narrow width transistors can be obtained through (4.5) quickly without time-demanding NBTI simulations. The time-exponent, n, can range from 1/4 to 1/2 for the narrow width MOSFET unlike the traditional NBTI exponent of 1/4. As the channel width gets narrower, the exponent increases as discussed earlier. Also the time-exponent in the narrow-width case is increasing further with time because the diffusion length, λ_D is increasing as $(D_H \cdot t)^{0.5}$.

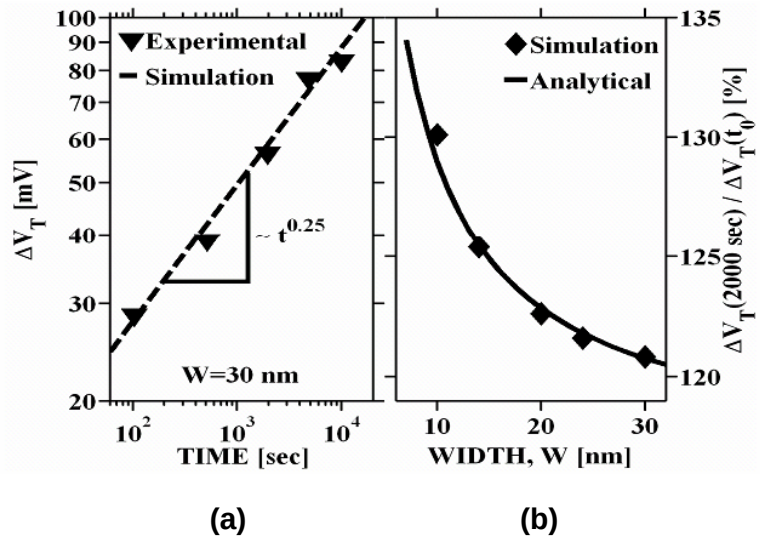


Fig. 4.7. (a) Experimental data from [88] and numerical simulation results for a triple-gate MOSFET with W = 30 nm. The two agree. (b) NBTI degradation for scaled triple-gate transistors. The relation deduced from (4.6) and numerical simulation compares well. Same oxide capacitance is assumed for all widths. The t_0 is the time when the first data point is measured in the experiment.

B. Triple-gate MOSFET

Below the 45 nm node, multiple-gate MOSFET structures offer an alternative to the conventional planar transistors. FinFET and tri-gate MOSFETs allow better short channel effect control and since they have multiple channels, lower operating voltages and mobilities are acceptable compared to the single-gate MOSFETs. Decreasing the body thickness, W_b, allows improved electrostatics, therefore it is an important parameter from the scaling perspective [79]. The scalability range of the body thickness is currently approaching down to sub-10 nm regime [80]- [87]. The consequence of such dimensions in terms of NBTI reliability is considered in Fig. 4.7(a) and (b). The numerical NBTI simulation is compared with the experimental NBTI degradation of a triple-gate MOSFET with W_b=30 nm in Fig. 4.7(a) and it agrees well with the data. In Fig. 4.7(b), a scaling scenario for the triple-gate transistor is

illustrated. As the body thickness is reduced, the 2-D components at the corners of the body facilitate the diffusion of hydrogen and the degradation predicted by the NBTI simulation increases significantly. The estimation from (4.6) also confirms the trend of the simulations. Again, the compact relation derived from the concept of diffusion length provides a time-efficient approximation. Besides the scaling, facet dependence of NBTI can induce a problem for the triple-gate MOSFETs as well. If the channels have different crystal orientations, NBTI can degrade these surfaces in dissimilar amounts, thus the electrical parameters of the three channels can shift differently over time [88]. Since the gate control will change, the electrostatic advantages of the triple-gate geometry can be lost during device operation.

C. Surround-gate MOSFET

Further going down the ITRS roadmap below the 20 nm node, the multiple-gate geometry is expected to lead into the all-around-gate structures for ultimate gate control. Surround-gate MOSFETs with Si channels were fabricated in [89] and recently, cylindrical Si-nanowires with diameters down to 5 nm were realized [90], [91]. Both these experimental work and numerical studies reflect that such devices can potentially continue the scaling of transistors [92]- [94]. However, the geometry dependent R-D model predicts worsened NBTI reliability for the ultra-scaled cylindrical MOSFET. Fig. 4.8 shows that the lifetime of the devices decrease significantly as the channel radius is shrunk. The lifetimes are obtained through asymptotic extrapolation of the numerical simulation results, namely the N_{IT} vs. time behavior, as it would be done in experimental characterization. The lifetime criterion for the MOSFETs is when the trap density increases to a certain amount at the Si/oxide interface, thus a critical V_T shift is reached ($V_T \approx \alpha \cdot q \cdot N_{IT}/C_{OX}$, q electron charge, C_{OX} oxide capacitance and α $(0.1 - 1.0)$ is the fraction of N_{IT} that contributes as positive interface charges [9]). For the scaled radii, the curvature of the interface increases the diffusion rate of hydrogen as well as slowing the annealing mechanism [53].

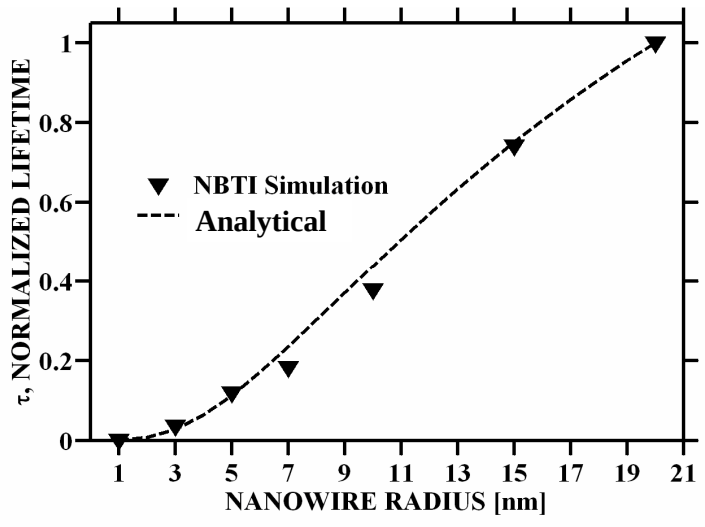


Fig. 4.8. The lifetime of cylindrical MOSFETs as a function of channel radius, R. The lifetimes are normalized to that of R=20 nm. As the device radius is scaled, the hydrogen diffusion is enhanced and the lifetime decreases. The symbols are obtained through the geometry-dependent R-D simulations. The dashed line represents the compact form of (4.7) and it compares well with the simulation.

The interface trap generation accelerates and thus the lifetime of the surround-gate MOSFETs are reduced. The analysis of (4.7) compares well with the results of the numerical solution; therefore an accurate prediction of the device lifetime can be obtained without the involved R-D simulations. For the narrow-width, triple-gate and surround-gate MOSFETs discussed, the NBTI degradation rate, n, increases as the device geometry is scaled. The overall NBTI damage, however, is determined with the additional parameters, k_F, k_R, N_0 and D_H which depend on the operating voltage, oxide thickness, strain and also the transistor geometry [9], [8]. Therefore, since reduced cross-section devices have better electrostatic control at lower operating voltages, one may be able to reduce k_F of these transistors to compensate the increase in NBTI exponent, n, discussed above. Establishing the necessity as well as the possibility of such co-optimization of performance and reliability is key goal of this work.

4.3 Discussion

In this work, the geometric implications of the theory on the time-dependence of interface trap generation were considered: the device structures were simplified to facilitate modeling of the fundamental mechanism arising from the size and shape of the MOSFETs. The devices were composed of Si and oxide only, nitrided spacers or cap-layers are not included. These type of materials are known to block hydrogen diffusion. In that case, the time-dependence is expected to saturate (section 2.4.2) since diffusion is limited. However, the hydrogen can diffuse into the Si channel and still lead to to higher time-exponents if there are no sidewalls or isolation materials that can decelerate the diffusion. Furthermore, the future generation devices may not contain such structures because of ultrascaled dimensions and fabrication techniques. The geometry-dependent analysis with simplified device structures provides general principles; other mechanisms can be included when necessary. The theory for ultrascaled geometries was implemented for atomic H diffusion, however it is also applicable to H_2 because the enhanced diffusion comes from the geometry and not from hydrogen species. For H-only R-D model, the theory predicts that critical MOSFET dimensions (i.e., width) should be scaled down to 20 nm or smaller in order to see the effects. For H_2, the critical dimensions need to be even smaller because the time-exponent is lower than that of H-only. Geometry-dependent R-D model has been evaluated in the literature by various groups. The theory was utilized to explain physical observations or criticized as being insufficient for interpreting experimental results. This section discusses as to whether geometry dependence can be assessed in the ways those references have claimed. In [95], the increase of NBTI for a FINFET was found to be consistent with the theory: as the dimensions scaled, degradation is increased. However, the authors attribute the increased degradation to higher hole density in the channel and larger oxide electric field with scaling. Therefore, the effect is not from the geometry of the device and the theory is misinterpreted. The increase in NBTI stems from the $k_F \propto \sqrt{p} \cdot exp(E_{OX})$ where p is the hole density in the chan-

nel. The geometric dependence was used to extend the R-D model for explaining Hot Carrier Injection (HCI) [53], [96]. In HCI the degradation which is related to $Si-H$ bond breaking, is localized near the drain, therefore 2-D R-D can be applicable to stress and recovery characteristics. HCI will be discussed in more detail in chapters 8 and 9. The HCI time-dependence with 2-D R-D model was criticized in [97] based on the argument that transition behavior from $t^{0.5}$ to t^0 (saturation) in 2-D R-D is not seen experimentally. Instead, the n=0.5 exponent was investigated through Multi Vibrational Hydrogen Release (MVHR) model by the authors. First, the saturation phenomena which is the basis of the argument, is not a sole result of 2-D R-D theory, it is a physical consequence. When all the available $Si-H$ bonds break, the degradation has to saturate. This can be achieved by applying a large stress voltage or degrading the MOSFET for a long time. In most experiments, the stress conditions are such that the saturation (t^0) is not reached. Moreover, the MVHR model was ruled out for HCI in [98] therefore its use is not appropriate. Another study examined the voltage and channel length dependence of HCI and compared with the 2-D R-D model [99]. They claimed that the model is not alone satisfactory for HCI because hot-hole induced degradation does not comply with the theory. The 2-D R-D model, as discussed in this chapter, deals with broken $Si-H$ bonds only. In the subsequent chapters, 8 and 9, the role of hot holes will be presented. One outcome is that, for technologically relevant stress voltages and MOSFET dimensions, 2-D R-D is applicable since the degradation due to hot holes will be negligible. Similar trend was also shown in [99], and those findings are consistent with the 2-D R-D theory.

4.4 Summary

The geometric interpretation of the hydrogen diffusion in the gate oxide allows extending the NBTI modeling from current-generation devices to the ultra-scaled and future-generation MOSFETs. The geometry of the MOSFETs shapes that of hydrogen diffusion and thus determines the NBTI-induced interface trap generation rate.

Assuming that the other parameters remain the same, numerical simulations and a time-efficient, straight-forward theoretical analysis imply that for narrow-width planar, triple-gate and surround-gate MOSFETs, the NBTI degradation rate increases significantly as the devices are scaled down. This signals that the expected performance enhancement of such future generation devices must be tempered by reliability considerations.

5. COMPACT NBTI MODELS FOR DIGITAL CIRCUITS

5.1 Introduction

As seen in the previous chapters, NBTI causes significant degradation for PMOSFETs. In order to investigate the impact on circuits, transistor-level degradation has to be translated into suitable set of parameters. Furthermore, for practical applications, the modeling of MOSFET degradation has to be compact so the circuit analysis such as SPICE can be performed in a time-efficient manner.

5.2 Circuit Delay Degradation

The interface traps generated during NBTI lead to parametric shifts at the MOSFET level. However, their impact can be in multiple ways. First, the interface states, as electrical charges can shift the threshold voltage of the PMOSFET:

$$\Delta V_T(t) = \frac{q \cdot \Delta N_{IT}(t)}{T_{OX}}. \tag{5.1}$$

However, the threshold voltage shift alone is not sufficient for circuit-level degradation modeling. For example in digital circuits, circuit delay is an important metric for circuit operation and performance. The delay is related to the drain current of the MOSFETs from

$$\tau = \frac{C_G V_{DD}}{I_D}. \tag{5.2}$$

The degradation in delay follows that of the drain current. The linear drain current can be written as

$$I_{DLIN} = K \cdot [(V_G - V_T)V_D - \frac{{V_D}^2}{2}]. \tag{5.3}$$

where $K = \frac{\mu W C_G}{L}$. If the degradation is solely due to threshold voltage, then

$$\frac{\Delta I_{DLIN}(t)}{I_{DLIN}} = \frac{-\Delta V_T(t)}{V_G - V_T}. \tag{5.4}$$

Similarly for the saturation current, $I_{DSAT} \propto (V_G - V_T)^m$ $(1 < m < 2)$ [3],

$$\frac{\Delta I_{DSAT}(t)}{I_{DSAT}} = \frac{-m\Delta V_T(t)}{V_G - V_T}. \tag{5.5}$$

For the delay degradation, going back to (5.2),

$$\tau \propto \frac{C_G V_{DD}}{(V_G - V_T)^m}. \tag{5.6}$$

Then,

$$d\tau = \frac{mBdV_T}{(V_G - V_T)^{m+1}} \tag{5.7}$$

where B has the constants (C_G, V_{DD}, μ, W, L). From (5.7) and (5.2),

$$\frac{d\tau}{\tau} = \frac{mdV_T}{(V_G - V_T)}. \tag{5.8}$$

Integrating both parts,

$$\int_{\tau_0}^{\tau} \frac{d\tau'}{\tau'} = \int_{V_0}^{V_T} \frac{mdV_T'}{(V_G - V_T')} \tag{5.9}$$

where V_0 and V_T are the threshold voltage of the virgin (corresponging to delay τ_0 at time=0) and stressed device (corresponging to delay τ at time=t), respectively. The result of (5.9) gives

$$log(\frac{\tau}{\tau_0}) = mlog(\frac{V_G - V_0}{V_G - V_T}). \tag{5.10}$$

Since $V_T = V_0 + \Delta V_T$, $\tau = \tau_0 + \Delta\tau$, $V_G \gg V_0$, and $\tau \gg \Delta\tau$, (5.10) can be written as a Taylor expansion,

$$log(1 + x) \approx \sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{i} x^i \tag{5.11}$$

for small x. From (5.10) and (5.11),

$$\frac{\Delta\tau}{\tau_0} \cong m \cdot \frac{\Delta V_T}{V_G} + M' \tag{5.12}$$

where M' contains the higher order terms and $M' \ll \frac{\Delta V_T}{V_G}$. Inserting the NBTI power-law degradation for threshold voltage, $\Delta V_T = At^n$, (5.12) becomes

$$log(\Delta\tau) = n \cdot log(t) + M" \tag{5.13}$$

with M" encapsulating the time-independent constants. From (5.13), it can be seen that the delay degradation also obeys a power law, with the same time-exponent of V_T degradation.

The above analysis considers only threshold voltage shifts arising from interface states. However, interface traps can also affect channel mobility. In the presence of interface traps, the mobility can be modeled as

$$\mu = \frac{\mu_0}{1 + \Theta N_{IT}}, \tag{5.14}$$

where μ_0 is the effective mobility and Θ is an empirical parameter [100]. The form given in (5.14) is also predicted by theoretical analysis of Coulombic scattering from interface states [101].

The degradation in mobility is then

$$\Delta\mu = \mu_0 - \frac{\mu_0}{1 + \Theta N_{IT}} = \frac{\mu_0 \Theta N_{IT}}{1 + \Theta N_{IT}}. \tag{5.15}$$

For small ΔN_{IT}, (5.15) can also be expanded using Taylor series, and it gives

$$\frac{\Delta\mu}{\mu_0} = \Theta N_{IT}. \tag{5.16}$$

It is apparent that the mobility degradation is also proportional to the interface trap density.

The overall degradation in drain current can be modeled with an effective threshold voltage shift since both mobility and ΔV_T are proportional to N_{IT}. This effective $\Delta V_{T,eff}$ can then be used in (5.4), (5.5) and (5.12) for the I_D and delay degradation (Fig. 5.1) [36], [37]. From experimental observations, it was found that $\frac{\Delta\mu}{\mu_0} = c\Delta V_T$, with c in the range of 1.2 to 2 [3]. The c is a function of oxide thickness and increases with thinner oxides.

Although mobility degradation is visible, its impact is not severe. The reason is that at high gate voltage (V_{DD}), the inversion layer holes screen the interface states and the Coulombic scattering is reduced. The degradation is more pronounced for lower voltages ($V_G \approx V_T$). Therefore, in digital circuits, the main degradation is due to threshold voltage shifts arising from the charges at the interface traps.

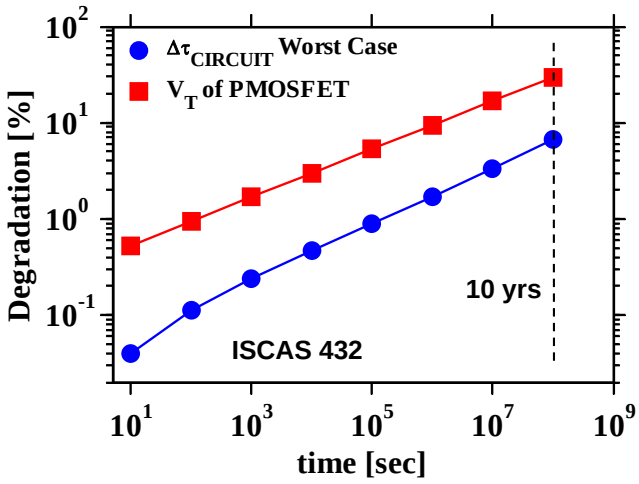


Fig. 5.1. An individual PMOSFET V_T degradation and (ISCAS 432 benchmark) circuit delay degradation shows the same time-exponent of power law. The delay degradation is much less in 10 years.

Another contribution of the interface traps is the increase in capacitance of the gate-drain overlap, C_{GD}. This parasitic component increases the effective load seen between stages in a circuit due to Miller effect. However, for digital circuits, its effect is small and it can be ignored [3]. For analog circuits, C_{GD} has a significant impact.

The discussion so far is based on DC stress conditions, however AC degradation can be easily investigated using the relationships provided in the previous chapter. The duty cycle dependence, random switching activity considerations and frequency dependence terms are all compatible with the effective threshold voltage concept.

The framework developed above for digital logic circuits can be extended to memory as well. The degradation on SRAM cells adds to the process-induced variations of the transistors and cause V_T mismatch, therefore NBTI reliability is one of the paramount concerns for SRAM yield. The effective V_T concept can be used to analyze the effect of NBTI on SRAM memory [38].

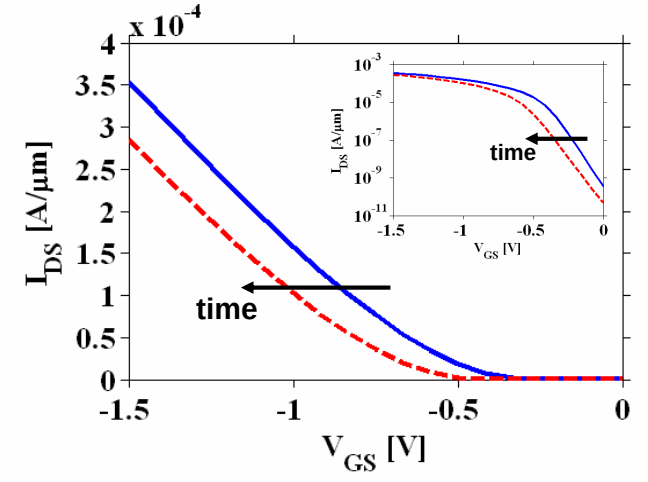


Fig. 5.2. PMOSFET subthreshold current decreases due to interface traps. The OFF current ($V_G = 0$) decreases exponentially since threshold voltage shift dominates the increase in subthreshold slope.

5.3 Source-Drain Leakage Current

The leakage currents in CMOS technology is becoming an increasing concern for power dissipation as the transistor dimensions are shrunk. The gate leakage is nearly constant since the oxide scaling almost reached its limits for the recent technology nodes. Source-drain leakage, on the other hand, increases with decreasing channel lengths. Furthermore, in the subthreshold region, the leakage increases exponentially with voltage, making it worse.

With NBTI, as discussed in the previous section, the ON-current characteristics change. The subthreshold characteristics are also affected due to interface traps. Fig. 5.2 compares ON and subthreshold characteristics of a PMOSFET. For the latter, the subthreshold slope increases due to interface traps and unlike to ON-current, the subthreshold current decreases with NBTI because of the threshold voltage shift. The increase in subthreshold slope can be understood by the following. The slope is

$$S = 2.3k_BT \cdot (1 + \frac{C_{dm} + C_{IT}}{C_{OX}}) \tag{5.17}$$

where $C_{dm} = \frac{\epsilon_{Si}}{W_{dm}}$ is the depletion region capacitance, $C_{IT} = \frac{dQ_{IT}(\Psi_s)}{d\Psi_s}$ is the interface trap capacitance, and $C_{OX} = \frac{\epsilon_{OX}}{T_{OX}}$ is the oxide capacitance [102].

The increase in subthreshold slope over time, $\Delta S = S(t) - S(t=0)$ can be written as

$$\Delta S = 2.3 k_B T \cdot (1 + \frac{C_{IT}(t) - C_{IT}(t=0)}{C_{OX}}). \tag{5.18}$$

By considering $d\Psi_s(t) \approx d\Psi_s(t=0)$, (5.18) becomes

$$\Delta S(t) = 2.3 q k_B T \cdot \frac{d\Delta N_{IT}(t)}{C_{OX} \cdot d\Psi_s(t)} \propto t^n, \tag{5.19}$$

which is proportional to the interface trap density and follows a power law. This derivation assumes uniform interface state density in the Si bandgap.

The $I_{OFF}(V_G = 0)$ also changes with NBTI. If we assume that the subthreshold slope is constant from $V_G = 0V$ to $V_G = V_T$ and the increase in the slope is negligible after stress (parallel shift of the subthreshold region, valid for small N_{IT}), then

$$\frac{1}{S} = \frac{log(I_T) - log(I_{OFF}(t=0))}{V_T(t=0)} = \frac{log(I_T) - log(I_{OFF}(t))}{V_T(t)} \tag{5.20}$$

where I_T is the drain current at $V_G = V_T$ and $V_T(t) = V_T(0) + \Delta V_T$ as in Fig. 5.3. From (5.20), it can be shown that

$$log(\frac{I_{OFF}(0)}{I_{OFF}(t)}) = \frac{\Delta V_T}{V_T(0)} log(\frac{I_T}{I_{OFF}(0)}) \tag{5.21}$$

and since $\frac{I_T}{I_{OFF}(0)}$ is constant, the off-current can be written as

$$I_{OFF}(t) = C \cdot 10^{-t^n} \tag{5.22}$$

where $\Delta V_T \propto t^n$ is assumed and C contains all the time independent constants. As the device is stressed, the leakage decreases exponentially dominating the subthreshold slope increase since the latter changes linearly [40].

5.4 Summary

Compact NBTI models in the form of analytical functions compatible with circuit simulators were described. The MOSFET degradation is converted into metrics

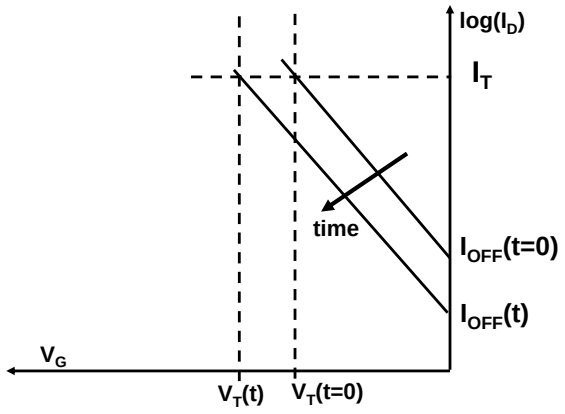


Fig. 5.3. I_{OFF} decreases with time due to the threshold voltage shift. Same subthreshold slope assumed for the virgin and stressed device.

so that the circuit degradation can be assessed. The parametric shifts due to interface traps can be used to derive circuit delay degradation for logic circuits. SRAM threshold voltage mismatch and source-drain leakage current evolution can also be studied with this framework. The model can be calibrated with experimental data and utilized for circuit simulators and design tools.

6. HOT CARRIER INJECTION: DEFINITION OF ISSUES

6.1 HCI Background

Hot carrier damage has been one of the important degradation mechanisms for MOSFETs [103]. Major source of the hot carriers is the electric field inside the channel of a transistor. The energetic carriers themselves or the carriers generated through impact ionization can cause the parametric degradation, i.e., shifts in device characteristics or catastrophic failure such as oxide breakdown.

Significant effort has been focused on understanding the hot carrier phenomena and its implications for circuits. In general, the related work can be categorized into the following categories: i) carrier heating in the channel of the MOSFET, ii) defect generation at the microscopic level, and iii) degradation of transistor characteristics, (especially in terms of time-dependence) and assessment of circuit lifetime . Although there is general consensus regarding the kinetics of carrier heating from experimental and theoretical perspectives, the latter two issues are still in debate and a clear conclusion is not present.

One of the early pioneering work was done in 1980s involved the calculation of HCI lifetime based on experimental device characteristics during hot carrier stressing [104]. In that work, it was assumed that the carriers heated by the channel electric field can lead to impact ionization. For an NMOSFET, the holes generated by ionization flow out of the substrate contact giving rise to substrate current (I_{sub}) whereas the electrons contribute to the drain current and if they are injected into the oxide, constitute the gate current, I_G, as in Fig. 6.1. The hot carrier damage is also attributed to the energetic electrons.

The I_{sub} was conventionally taken as a monitor for the hot carrier damage because it reflects the energy of the hot electrons. The substrate current and degrada-

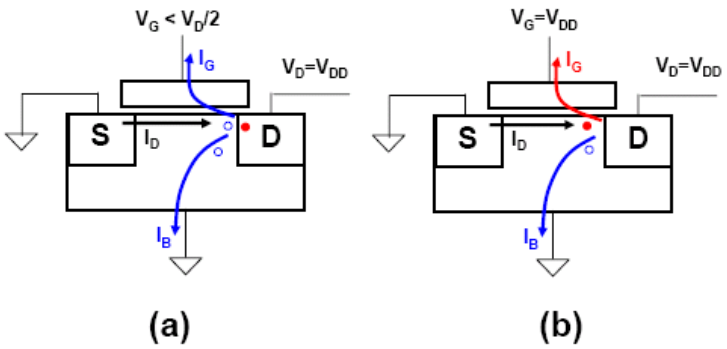


Fig. 6.1. Schematic representation of the substrate and gate currents in an NMOSFET. (a) Holes (open circles) generated by impact ionization flow out of the substrate. Some fraction can be injected into the gate oxide since the vertical electric field favors holes at low V_G. (b) At high V_G, the vertical electric field attracts electrons (filled circles) into the gate oxide and the electrons form the gate current. The substrate current is still due to holes.

tion showed similar bell-shaped trends when plotted with respect to the gate voltage. Their peaks were also at the same V_G value [105]. The bell-shape arises because there exists two competing mechanisms underlying I_{sub} and degradation: the number of electrons in the channel increase with V_G for a fixed drain bias, V_D. Second, the average energy, E_{ave}, of the electrons decreases with increasing gate bias since $V_{D,sat}$ increases and $E_{ave} = \frac{V_D - V_{D,sat}}{\Delta L}$ where ΔL is the pinch-off length. The substrate current is $I_{sub} = \alpha_{II} \cdot I_D$ where α_{II} is the impact ionization efficiency which is proportional to E_{ave}. As a result, I_D increases whereas α_{II} decreases with V_G yielding a bell-shape curve whose peak is around $V_G = V_D/2$. The degradation, similarly, can be interpreted as a flux of energetic carriers directed toward the gate oxide. The energy and density of electrons have opposite trends as a function of V_G, therefore degradation also has a bell-shape peaking around $V_G = V_D/2$. These observations

are valid for older technology MOSFETs with thick gate oxides ($T_{OX} \geq$ 5nm) and long channels. As the MOSFETs are scaled, it was shown that the peak degradation occurs at $V_G = V_D$ case [106], [107].

The correlation between I_{sub} and degradation does not necessarily mean that impact ionization is needed for HCI damage. The I_{sub} is a measure of average carrier energy but only a fraction of carriers in the channel going towards the drain leads to impact ionization ($I_D \gg I_{sub}$). Thus, it is plausible that defects can be generated by channel hot carriers which escape scattering by impact ionization. For older technology generations with higher stress voltages, the carriers generated from impact ionization can gain energy and contribute to overall damage, however, as the technology is scaled, the energy gain after impact ionization is not significant under lower stress voltages. Therefore, the I_{sub} can be viewed as a by-product of channel heating and is not a pre-requisite for HCI degradation.

The other possible candidate of degradation, gate current, shows a different behavior: for a fixed drain bias, the gate current peaks at low V_G, then changes sign and increases towards $V_G = V_D$ condition. Low V_G characteristics were attributed to hole injection into the gate oxide since the electric field near the drain favors holes. The holes at low V_G were coming from impact ionization of channel electrons and this was verified by floating the source contact, thus eliminating the electron heating in the channel electric field. As the V_G is increased with V_D kept fixed at V_{DD}, the perpendicular electric field decreases, the electron component of the gate current increases and the net gate current decreases before changing its sign ($I_G > 0$ for low V_G and $I_G < 0$ as V_G increases).

As the technology is scaled (particularly V_{DD}), the HCI was expected to decrease and ultimately disappear with $V_G = V_D < V_{t,HCI}$, $V_{t,HCI}$ being some threshold value for HCI (3.1 or 1.1 eV, etc). However, this was not the case, the degradation has not vanished for MOSFETs operating at lower voltages. Further experiments and Monte Carlo studies provided insight into the mechanism which lead to high energy carriers for low V_D. It was shown that secondary mechanism on top of heating in the channel

can give rise to carriers populating the high energy tail of the distribution functions. One mechanism responsible for this tail is the electron-electron scattering in which a hot electron, rarely but possibly, gain energy from a cold carrier [108], [109]. Another suggestion is the impact ionization feedback where the secondary impact ionization populates the tail of the distribution function [110], [111]. When the substrate of an NMOSFET is reverse biased, the holes generated by impact ionization near the drain can be accelerated in the substrate-drain field and lead to secondary ionization and then the secondary electrons are directed towards the oxide. By this mechanism, gate current (due to hot electrons) and degradation can be changed significantly whereas the substrate current (due to holes) remains relatively unchanged [112]. This observation further supports the notion that impact ionization is not required for HCI damage. In the absence of these secondary energy-gain events, the high energy part of the distribution functions is characterized by the thermalized tail, with effective temperature equal to the lattice temperature and the tail can play a role in the overall temperature dependence of the substrate current [113], [114], [115].

Although hot carrier degradation has been studied for decades, the physical processes occuring in HCI are still debated. One major difficulty in understanding the degradation is that the MOSFETs and the stress conditions are continually being changed. The supply voltages for the transistors have been scaled significantly since '80s - when HCI was prominent - and the MOSFETs have been technologically advanced, giving rise to utilization different materials and device geometries. The applied voltages to the MOSFET can drastically change the energy and magnitude of hot carriers. Depending on the carrier energy and comparing them with the relevant energy thresholds (i.e., conduction/valence band offsets, semiconductor bandgap energy, damage creation energy), several different physical processes (carrier injection into the oxide, charge trapping, impact ionization, interface and bulk defect creation) can be present at the same time. Scaling the voltages can modify the extent of these mechanisms, increasing or decreasing their relative contributions. Additional consideration of doping profiles of a MOSFET, device geometry (e.g., LDD), and the type

and quality of materials (e.g., gate oxide, spacers), it is plausible that the observations during HCI can vary in the literature.

6.1.1 Underlying mechanisms of hot carrier degradation

The fundamental mechanisms involved in HCI are well known: carrier heating, impact ionization (although not necessary for degradation), charge injection into oxide, defect generation, charge trapping or detrapping. The unclear part is, how much each mechanism occurs for a given technology and operating condition. For a particular technology, even accelerated tests at high voltages might result in incorrect lifetime extrapolations since higher stress voltages can activate some of the aforementioned mechanisms and over-degrade the MOSFETs. It is therefore important to understand the voltage, temperarature dependences of these processes individually and how they behave with scaling operating conditions for realistic estimation of the degradation.

In this section, the debated issues in HCI literature are presented. These topics provide a pre-requisite for the next chapters where experimental and theoretical HCI observations are discussed for various operating conditions.

A. The role of electrons and holes

One of the questions in the hot carrier literature is whether hot electrons or hot holes are the source of damage. Hu's work claimed that hot electrons, not hot holes, were responsible for generating interface traps [116]. This argument was based on hot carrier experiments on NMOS and PMOSFETs. Both transistors had bell shaped I_{sub} vs V_{GS} and ΔV_{TH} vs V_{GS} characteristics. Hu argued that if the holes were responsible for degradation, PMOS would have much higher threshold voltage shift since the channel of the PMOSFET is inverted and full of holes. Even at higher stress voltages, PMOS degradation was less than that of NMOSFET. This is further supported by the fact that electrons have longer mean free path due to their lighter effective mass compared to holes, therefore the electrons are more efficient in impact

ionization [102]. Furthermore, the conduction band offset at the $Si/oxide$ is smaller than the valence band barrier, suggesting that electrons can create interface traps easier by injection over the oxide barrier.

The hypothesis of hot holes as the main source of interface trap (N_{IT}) generation, contrary to Hu et al., was studied by several groups. For both NMOS and PMOSFETs, the N_{IT} density peaked at hot hole injection regimes, $V_G < V_D/2$ for the former and $V_G = V_D$ for the latter [117]. In that study, the lack of correlation between I_{sub} and N_{IT}-peak was attributed to hole induced hot carrier damage (however PMOS NBTI could have also played a role, but was not discussed). Another study used a gated-diode structure to inject either electrons or holes to the gate oxide selectively [118]. By doing so, interface trap generation efficiency for holes and electrons were determined. It was found that the holes were orders of magnitude more efficient for trap creation. Lai proposed a mechanism where holes were the driving force of interface trap generation on devices exposed to radiation. In his model, the hot holes injected into the gate oxide were trapped, then captured electrons, and their recombination dissipated energy sufficient to generate interface traps [119].

Recent experiments also argue that hot holes are the carriers that generate interface traps [99], [120]. It is claimed that channel hot electrons injected through the gate oxide can impact ionize at the anode side. The resulting holes (via Anode Hole Injection, AHI) can tunnel back towards the substrate and break bonds at the interface. Another qualitative mechanism suggested is the Valence Band Hole Tunneling (VBHT). Similar to AHI, holes from poly valence band can tunnel through the thin gate oxide and create interface defects [99].

On a related argument, it is not very clear if the carriers need to be injected into the oxide to generate defects. Hu argued that electron injection must take place for interface trap creation. The original Lucky Electron Model for hot carrier degradation claimed that the maximum energy an electron can gain is limited by the potential variation in the channel, that is $E_{max} = qV_D$ [116]. If the electron energy is less than the conduction band offset, Φ_b, there would be neither electron injection into the

oxide for generating interface traps nor gate current. This implied that if the supply voltages are lower than about $2.5 - 2.7$eV (due to lowering the conduction band barrier, ≈ 3.1eV, by image force effect), hot carrier degradation could be eliminated. Subsequent experiments showed that this is not the case: interface trap creation at V_D as low as 0.9V and gate current (hot carrier induced) for $V_D = 2.35$V were observed [121], [122], [123].

Following studies suggested that electron injection is not necessary for interface trap generation [107], [124]. The results were obtained through selectively injecting hot electrons into the oxide from the substrate of an NMOSFET and monitoring the isotope (hydrogen/deuterium) effects on reliability. The energetic carriers from substrate injection were held at the $Si/oxide$ interface by applying a positive V_G. The energetic carriers gave rise to excess source/drain current, but not appreciable gate current, thus injection into the oxide was ruled out. Both at this condition and at $V_{GS} \approx V_{DS}/2$, the isotope effect was obvious [125], [126]. It was concluded that carrier injection into the gate oxide was not necessary for interface trap generation and hot electrons interacting with $Si-H$ bonds at the $Si/oxide$ interface can directly break the bonds [127].

A further investigation applied negative V_B during hot carrier stress to populate the energetic tail of the electron distribution [147]. The back bias increases the gate current (due to high energy carriers) and degradation significantly but the substrate current (due to low energy carriers) does not change noticeably. This experiment claimed that injection into the oxide increases the interface damage. The authors did not exclude the impact of lower energy electrons on interface degradation but suggested that as V_{DD} scales, the energetic tail becomes more important in HCI [129].

B. Nature of the defect: $Si - H$ or $Si - O$ bonds

Another debated topic in the literature is the chemistry of interface trap generation which is generally attributed to breaking of chemical bonds at the $Si/oxide$ interface

and in bulk oxide. In Hu's model, breaking of $Si-H$ bonds was a hypothesis based on previous work in the literature [116]. A more direct proof of interface degradation due to these bonds came with deuterium (D) desorption from Si surfaces through Scanning Tunneling Microscope (STM) experiments and deuterium-annealed MOSFETs [130], [131], [132]. Deuterium-annealed MOSFETs (deuterium at $Si/oxide$ interface evidenced by SIMS), with deuterium having higher atomic mass, showed a clear isotope effect under hot carrier stress and degradation was less than those of the hydrogen-anneal (H_2) samples.

One theory attributed the desorption process to atomic excitations due to electrons impinging upon $Si-H(D)$ bonds [133]. The STM results show that, at low voltages, multiple energetic electrons can incoherently excite the hydrogen bond to an antibonding state (multiple vibrational excitation model). At high current densities, the energy transfer from multiple electrons can result in desorption by surmounting the decay rate of the antibonding state. Low atomic mass of H increases the lifetime of the excited state and thus makes it more likely to detach. This process requires about 4x10^5 to 2x10^7 electrons to desorb a single hydrogen and deuterium, respectively. At higher electron energies ($< 5-6$eV), single excitations can possibly release the hydrogen atom. This regime does not show any current or voltage dependence unlike the low voltage($<$ 4eV) regime.

Recent theoretical and experimental work have shown that a more physically-accurate explanation is related the coupling of $Si-H/Si-D$ and bulk Si phonon modes [98], [134], [135]. It was found out that the $Si-H$ and $Si-D$ vibrational modes have frequencies of 650 and 460 cm^{-1}, respectively. The frequency of phonon modes of bulk Si is approximately 463 cm^{-1}, therefore the $Si-D$ excitations can be channeled to the phonons effectively. This process reduces the probability of breaking $Si-D$ and increases resiliency of the interface against hot carriers.

In addition to $Si-H$ bonds, $Si-O$ which is abundant at $Si/oxide$ interface, can also be broken during HCI, as suggested by [99], [120]. This mechanism involves particularly energetic holes generated during the stress. Unlike the $Si-H$, recovery

of broken $Si-O$ is negligible since annealing of Si dangling bonds by O is difficult at relevant temperatures. This type of dangling bond adds to the overall interface trap density, however the time-dependence and recovery characteristics are markedly different than the broken $Si-H$ bonds. Mahapatra suggests that at $V_{GS} \approx V_{DS}$ condition, the $Si-H$ are broken based on the evidence of significant HCI recovery and the time-exponent during stress is lower than case of $V_{GS} \approx V_{DS}/2$ which involves $Si-O$ [120].

C. Location of damage

The degradation in HCI occurs near the drain junction of a MOSFET since the electric field peaks around that region. This is evidenced by I_D-V_G characterization with source and drain flipped after a standard drain-sided stress. This flipped source-drain configuration shows higher degradation because it increases scattering for the carriers injected from the source-side (originally drain-side).

The damage region need not be limited to the junction, it can further extend into the drain contact. Since there can be energetic carriers in the gate-drain overlap and underneath the spacers, the overall degradation can increase. The impact of degradation on MOSFET characteristics depends on the extent of the damage [136], [137]. At low V_G, channel resistance of an LDD MOSFET, R_{CH} is larger than the source/drain resistance R_{SD}, therefore the damage in the channel and gate-drain overlap dominates. At higher V_G, $R_{CH} < R_{SD}$, so the R_{SD} has a higher effect. For conventional MOSFETs (no LDD), the degradation is mainly due to the channel damage; the reduction in mobility and interface/bulk charges.

The spatial extent of the interface traps can be investigated by experimental techniques such as charge pumping and DCIV [138], [139]. However, these methods usually require the knowledge of transistors' doping and potential profiles which may not be conveniently accessible. Also, these experimental techniques are not generally

feasible to probe the damage region deeper into the drain. These issues constitute difficulties in understanding the effects of degradation.

Based on the experiments, the location of the interface states was found to be dynamic during stress, particularly for LDD MOSFETs. It is claimed that the damage first appears underneath the spacers, then moves into the channel [140], [141]. The initial damage under the spacers can be attributed to the lower junction field of LDD MOSFETs. Due to the lower field, the electrons gain sufficient energy deeper into the drain. The series resistance increases and potential profiles change over time and the degradation shifts toward the channel at later stages. This behavior gives rise to saturation phenomena of time-dependent degradation, i.e., the rate of degradation slows at later times whereas the initial rate is faster. The substrate current also changes over time, first decreases due to damage under the spacer, then increases as interface traps evolve into the channel. Non-saturating degradation characteristics were also observed for HCI in LDD MOSFETs. This behavior can be explained by scaling of the stress conditions or MOSFET geometry (dimensions, doping etc) [120], [137], [141]. The characteristics for current generation MOSFETs is not yet clear since accelerated tests can cause artifacts not present for operating conditions.

D. Bulk-oxide defect generation

Besides the interface traps, hot carriers can also generate bulk oxide traps. The hot carriers injected into the oxide can get trapped and change the electrostatic potential over the channel of the MOSFET or the spacer region [105], [118], [142]. For older technology MOSFETs with thick gate oxides ($T_{OX} \geq$ 10nm) and high operating voltages, bulk traps can have a significant contribution, however for current technology devices with thin oxides ($T_{OX} \geq$ 3nm), the trapped carriers are likely to tunnel out of the oxide. The bulk defects themselves do not recover, however trapped charges in these defects can lead to recovery due to detrapping. Despite this fact,

the time-dependence of the hot carrier degradation can be affected by the bulk traps which generally evolve faster than the interface traps.

It is believed that the bulk defects are similar to broken $Si-O$ bonds which are generated with energetic carriers [125]. Chen observed that the defects generated when hot electrons or holes were injected deeper into the oxide did not show hydrogen isotope effects. Mahapatra's work claimed that bulk traps were generated also at $V_{GS} \approx V_{DS}/2$ stress condition [120].

6.1.2 Time dependence of HCI degradation

In order to incorporate hot carrier degradation into device lifetime projections and circuit design tools, time-dependence of the damage should be considered carefully. Temporal degradation can be traced down to the microscopic mechanisms during HCI discussed in Section 6.1.1. Understanding these processes increases the accuracy of the model, however time-efficiency of such an approach for circuit-level tools may not be satisfactory. Some models based on compact analytical formulation may suffer from accuracy. In this section, various approaches to HCI modeling are discussed.

A. Modeling of microscopic mechanisms

In Hu's work, the time dependence of degradation was derived to be a power law, based on diffusion kinetics of hydrogen, with a time exponent of 0.5 (assumed 1-D hydrogen diffusion and infinite diffusion velocity into the poly). Although this is the widely accepted exponent, experimental observations generally show a spread of the exponent from 0.3 to 0.7 [132], [143]- [147]. Moreover, the degradation can saturate over time for LDD device structures due to series resistance and self-limiting (increasing V_T over time decreases the amount of hot carriers), further complicating the modeling of hot carrier time dependence. [136], [140], [141].

Based on the role of $Si-H$ bond breaking during hot carrier stress and the localization of the damage near the drain, the Reaction-Diffusion model provides an

explanation to the HCI time-exponents by extending the geometry dependence of the hydrogen diffusion [53], [54]. This approach gives a time-exponent of $\approx 0.3 - 0.5$ if the diffusing hydrogen species is atomic H and $\approx 1/3$ if H_2 is diffusing. However, as qualitatively suggested by Mahapatra through experiments, HCI behavior may not be alone attributed to the 2-D R-D model [120]. The reasoning behind this argument is the contribution of $Si - O$ bond breaking (and time-exponents ≥ 0.5) which are due to hot holes at high stress conditions.

Another signature of the $Si-H$ vs. $Si-O$ bond breaking is the recovery behavior of degradation. The MOSFETs stressed under hot carriers also show recovery, but the magnitude is much less than the case of NBTI. The complete absence of recovery at $V_{GS} \approx V_{DS}/2$ in Mahapatra's work (V_D=4 and 4.5V, channel length of 0.28 μm and oxide thickness of 4.8 nm [120]) is contrary to the deuterium experiments which show a significant isotope effect suggesting hydrogen (not oxygen) is released from the interface. This discrepancy could be due to the measurement method (Mahapatra used charge pumping technique which has an inherent delay) or excessive degradation (saturation observed in [120]). Reaction-Diffusion framework extended to 2-D can explain reduced recovery in the case $Si - H$ bond breaking; the localized damage reduces the recovery rate [53].

B. Compact MOSFET-level modeling

Impact of hot carrier degradation on circuits and its implications on circuit design have been a major focus for product reliability. Since degradation occurs in individual MOSFETs, it has to be translated into compact, time efficient forms so that they can be used in circuit simulators.

The main parameters in compact modeling are the threshold voltage, V_T, and mobility, μ, degradation, and time-dependence. These parameters are related to the magnitude and spatial extent of degradaion as well as the transistor itself. It is not easy to include microscopics of degradation at the device level and use it for circuits

in a time-efficient manner since it may requires the knowledge of doping, potential and field profiles and hot carrier distributions. A 1-D compact model that converts localized interface trap density into V_T and μ degradation was suggested in [148]. This model did not explain the time-dependence or why the degradation occurs over the specified dimensions.

C. Empirical HCI modeling

Another type of modeling is based on empirical device degradation and does not deal with microscopic mechanisms. This method encapsulates the possible V_T, μ, series resistance or charge trapping related degradation. The age of a MOSFET is extracted from I_D, I_{sub} and I_g . The AC stress degradation in circuits is estimated from the DC characteristics by calculating the time interval in which hot carriers are generated during the voltage transients [149]. The maximum hot carrier activity, I_{sub}, occurs at rising edges of the gate bias for NMOSFETs and falling edges for PMOSFETs. Degradation in one cycle (coined as quasi-static approach) is calculated for the transistors in terms of signal delay and then extrapolated to the circuit activity by summing the contribution from each cycle. The circuit delay or frequency degradation or timing errors (glitches) are thus linked to MOSFET I_D degradation. This model was applied to circuits such as ring oscillators, pass-gate transistors, NAND and NOR gates, multiplexors and showed good agreement with experimental results [150].

In general, the degradation depends on several parameters: i) load and parasitic capacitances (such as $C_{drain-gate}$ or coupling due to interconnect layers) and related voltage overshoots [151], [152], ii) number of fanouts [151], iii) transistor sizing (e.g., in buffers) [152], iv) activity of previous stages in circuit (rise and fall times can be affected) [152], v) the type of gates (e.g., NAND, NOR etc) [150], vi) the function of a MOSFET (such as in an inverter vs pass-gate transistor where bi-directional activity is seen) [150], [152] and vii) statistical inter-die or intra-die variations [153]. These individual parameters might also depend on degradation (such as in i, iv, vi).

Therefore, the overall picture is quite complicated and quasi-static modeling might be misleading. Moreover, most reliability tools focus on one type of mechanism. In the case of both NBTI and HCI degradation, the parametric shifts of a single gate or subsequent stages in a circuit requires a thorough analysis.

6.2 Summary

Although hot carrier degradation of MOSFETs has been studied for decades, there is still debate about the exact mechanisms and their impact on MOSFET characteristics. The role of hot electrons or hot holes (or both), whether injection into the gate oxide is needed or not, the types of bonds broken that generate interface traps, the evolution of spatial extent of damage region and the significance of bulk defects are not yet understood well. Most importantly, from a device and circuit lifetime angle, the time-dependence and the source of power-law behavior are not clear.

To identify microscopic mechanisms during hot carrier stress and relate them to device and circuit degradation under realistic operating conditions, a set of new experimental tecniques and theoretical investigations are listed in the following chapters. First, the theoretical tools will be described in the next chapter. Then, to answer the open questions in the hot carrier literature and provide a consistent analysis of the problem, experimental approach will be discussed in subsequent chapters.

7. HCI MODELING

7.1 Introduction

A thorough modeling of hot carrier degradation should consist of MOSFET-level analysis as well as macroscopic observations such as time-dependence and recovery of degradation. In fact, microscopic mechanisms and macroscopic trends are related and cannot be treated separately for robust modeling. In order to understand these aspects of HCI, a framework spanning device process, electrical characteristics and hot carrier analysis is implemented. The tools can be jointly used with experimental results.

The flow of hot carrier degradation modeling is shown in Fig. 7.1. From process and device simulations, hot carrier distributions in energy and space can be obtained and incorporated into the geometry of hydrogen diffusion and k_F physics in 2D R-D simulations to obtain the time dependence and to calibrate voltage and temperature dependeces. Finally, the resulting hot carrier degradation can be modeled compactly for circuit tools.

These set of tools provide a powerful framework to study hot carriers and degradation. Previously, other groups focused on a few aspects of the phenomena and treated others phenomenologically. The capabilities at hand potentially presents a complete approach.

In this chapter, the use of TSUPREM4, MEDICI and SMC simulations will be discussed and verification of the tools will be described for a case study of Drain-extended-MOSFET (DeMOS) transistors.

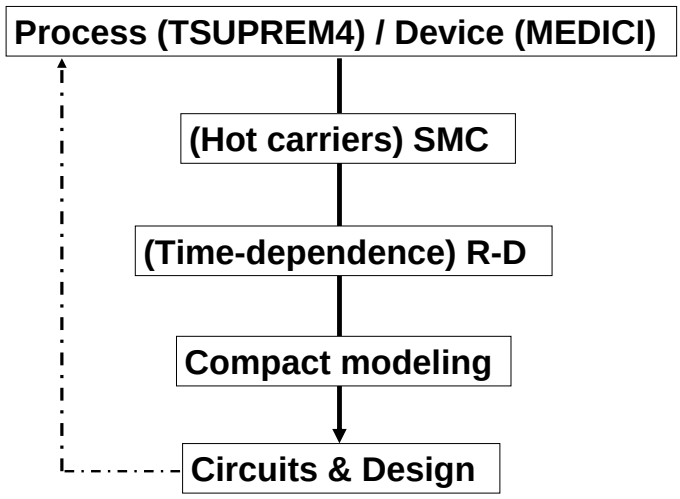


Fig. 7.1. Flow of the HCI degradation modeling. The device-level simulations are performed for obtaining the time, voltage and temperature dependence. The trends are then translated into compact models for circuits.

7.2 Process and Device Modeling

For accurate modeling, it is important the represent the Si device as realistic as possible, i.e., with correct device dimensions and doping profiles resulting from a process simulator. In this work, TSUPREM is used for process modeling, then the electrical simulations are performed on MEDICI. The device mesh and electrostatic results are then used for SMC (Monte Carlo code from Bell Labs, [154]) to get hot carrier distributions. Fig. 7.2 shows the device structure of DeNMOS which is taken as an example for the modeling framework. The DeNMOS works as an I/O transistor, and can handle higher voltages due to its asymmetric structure [155], [156]. The drain of DeNMOS is actually the n-well on p-substrate, compatible with CMOS technology. This transistor has the same oxide thickness as the logic MOSFETs. The extended drain region which is otherwise reserved for logic PMOSFETs provides a voltage drop path when a high drain bias is applied, making the DeNMOS more tolerant to breakdown. Drain-extended-PMOSFET (DePMOS) has also similar structure.

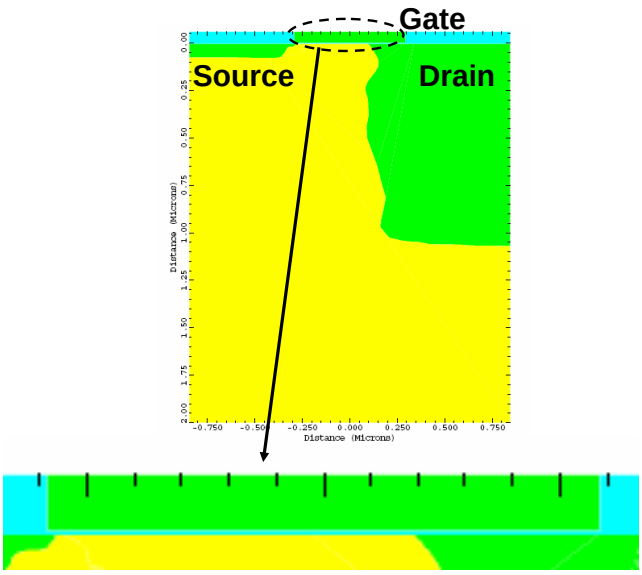


Fig. 7.2. DeNMOS device structure obtained from process simulations. The asymmetric drain is made of the n-well allocated for PMOS transistors of CMOS logic. This drain can withstand high voltages. A close-up view of the channel region is also given. The deck is used in Medici simulations.

For accurate results, MEDICI and SMC tools should first be calibrated with respect to experimental data for a range of voltages and temperatures. Fig. 7.3 and Fig. 7.4 show comparison of experimental I_{DS} vs V_{GS} characteristics with MEDICI simulation results. The high field mobility parameters and interface trap densities were adjusted to match experimental $I_{DS,lin}, (V_{DS} = 0.1V)$ and $I_{DS,sat}(V_{DS} = 1.8V)$ for both subthreshold and superthreshold region. For $V_{GS} < 0$, additional calibration of band-to-band tunneling (GIDL) and generation/recombination parameters are needed, but $V_{GS} \geq 0V$ is the main focus for this work. Fig. 7.5 and Fig. 7.6 also show the calibration results for two different temperatures. The calibrated simulations provide a good fit to experimental observations.

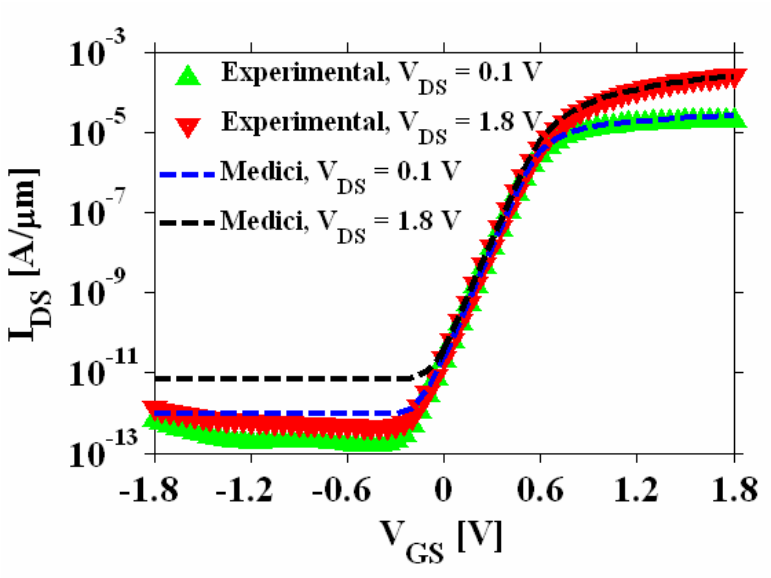


Fig. 7.3. Experimental data and MEDICI simulations for I_{DS} vs. V_{GS}. The simulations match the drain current for the subthreshold, linear and saturation regimes. I_{DS} is plotted in log-scale. MEDICI high field mobility parameters and interface trap density were calibrated to match the current.

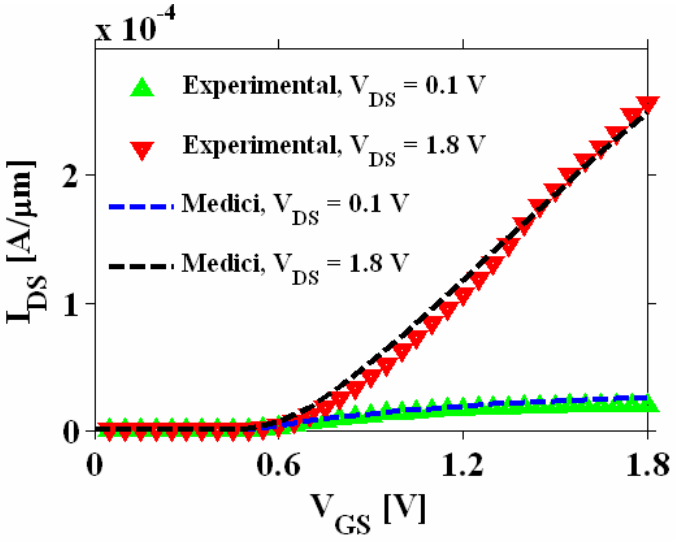


Fig. 7.4. Same as previous figure, I_{DS} in linear scale.

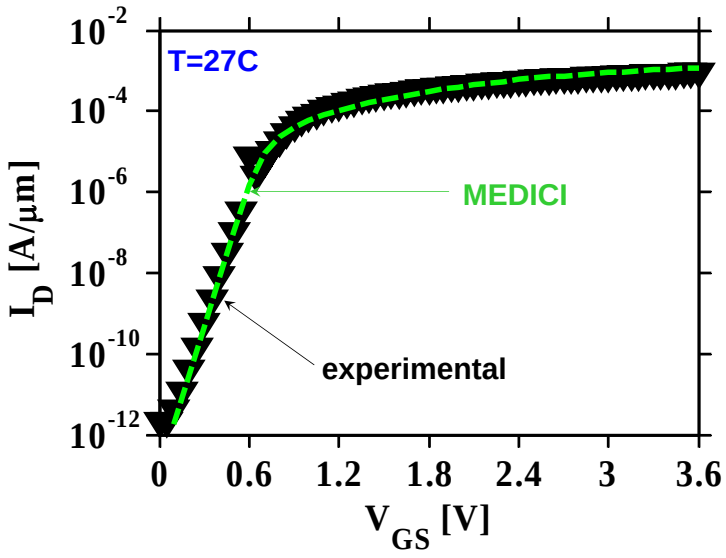


Fig. 7.5. Experimental data and MEDICI simulations for I_{DS} vs. V_{GS} at 27^oC. Temperature dependence is calibrated. $V_D = 0.1V$.

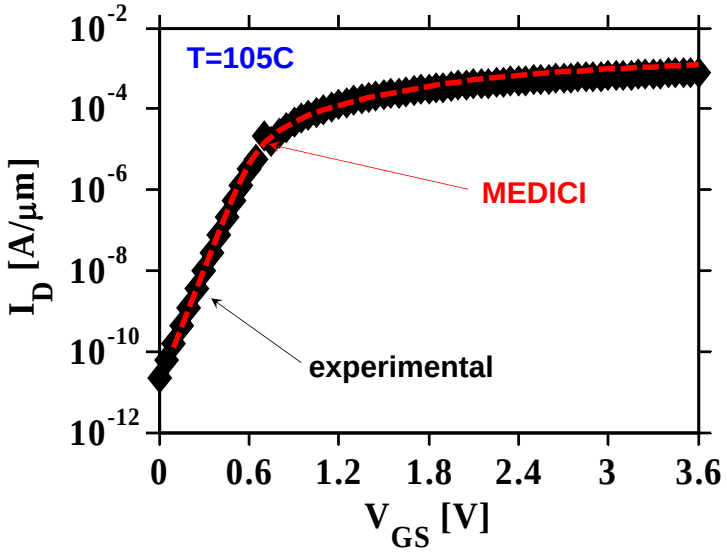


Fig. 7.6. Experimental data and Medici simulations for I_{DS} vs V_{GS} at 105^oC. $V_D = 0.1V$.

7.3 Hot Carrier Modeling by Simplex Monte Carlo (SMC)

Simplex Monte Carlo (SMC) is a full band simulator for carrier transport in Si devices. The advantage of SMC comes from its phase-space simplex algorithm

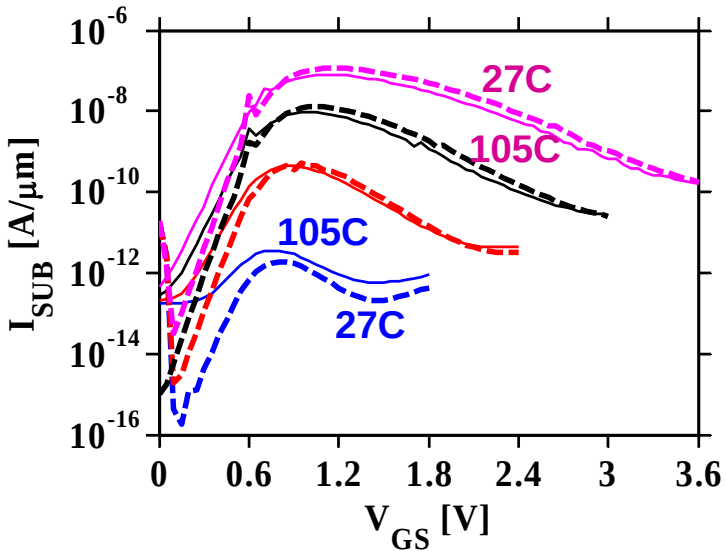


Fig. 7.7. Experimental I_{sub} vs V_{GS} for different V_{DS} at 27 (dashed) and 105^oC (solid lines). The temperature dependence shows a reversal from low to high drain voltage (=1.8, 2.4, 3.0 and 3.6V). V_{GS} is swept from 0 to V_{DS} for each drain voltage.

which can contain the accuracy of full band physics and efficiency in time through piecewise approximations to both energy and electrostatic potential. The k-space and real space are treated symmetrically, thus, exact integrations for the motion equations and selection of final state for a range of scattering mechanisms can be performed [154]. SMC was extensively calibrated for Si devices and used to explain several experimental phenomena [113], [157], [158], [159].

For inputs, SMC takes terminal currents and electrostatic potential of a MOSFET obtained by a device simulator and then computes hot carrier densities, distributions and currents based on those estimations. The calibrated MEDICI simulations were then used to study hot carrier characteristics and degradation. For robust predictions, SMC hot carrier dependence were first calibrated. Impact ionization and phonon scattering parameters were adjusted for to match experimental I_B/I_S ratio for different temperatures. Fig. 7.7 shows I_{sub} vs V_{GS} relationship for different drain voltages and temperatures. At the drain, V_{DS} = 1.8, 2.4, 3.0, and 3.6V were applied and

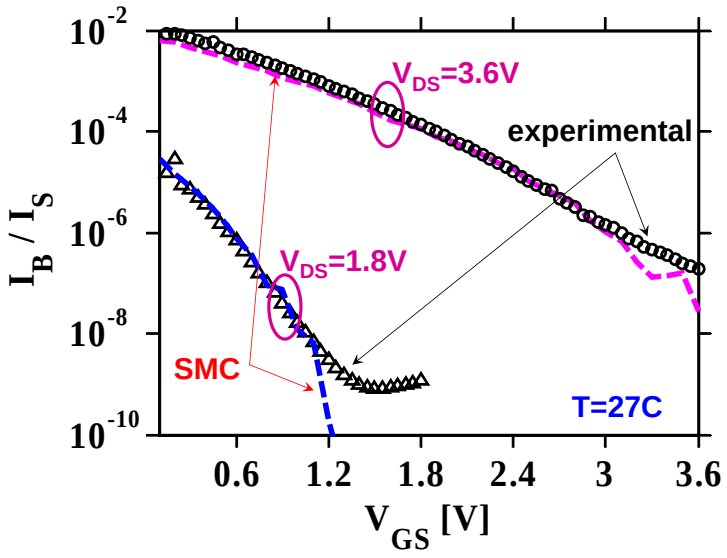


Fig. 7.8. The ratio I_B/I_S for experimental data (symbols) and SMC (lines) at $V_{DS} = 1.8V$ (triangles) and $3.6V$ (circles). Experiments were done at T=27^oC.

V_{GS} was swept from 0V to V_{DS} for each case. Substrate current which is a directly proportional to the hot carriers generated by impact ionization was measured for 27 and 105^oC. At low V_{DS}, high temperature I_{sub} is larger than that of 27^oC whereas for V_{DS} =3.6V, the trend is flipped; low temperature has higher substrate current. The reason for this reversal is attributed to the drain bias dependence of the distribution function and the spread of its tail due to temperature [113]. SMC ionization and phonon scattering parameters were adjusted to fit the experimental I_B/I_S ratio which reflects the impact ionization coefficent α_{II}:

$$I_B = \alpha_{II} \cdot I_{DS} \cdot \Delta L, \tag{7.1}$$

where ΔL is the pinch-off length in the channel. It is assumed that this length does not change significantly with bias and temperature. Fig. 7.8 compares experimental I_B/I_S ratio with SMC results. Fig. 7.9 presents the same for T=105^oC and the two temperature are compared in Fig. 7.10. These results imply satisfactory calibration of SMC for experimental trends. After calibrating MEDICI and SMC, hot carrier

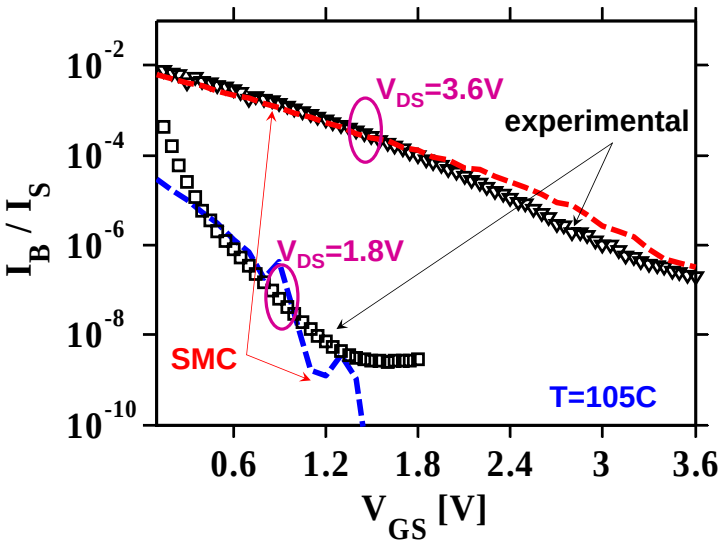


Fig. 7.9. The ratio I_B/I_S for experimental data (symbols) and SMC (lines)at $V_{DS} = 1.8V$ (squares) and $3.6V$ (triangles). Experiments were done at T=105^oC.

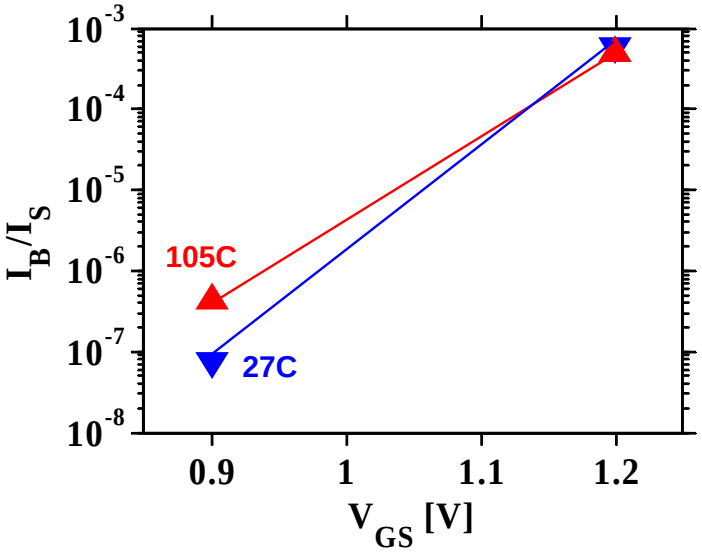


Fig. 7.10. I_B/I_S from SMC are compared for two temperatures. The ratio I_B/I_S for $V_G = 0.9V$, $V_D = 1.8V$ and $V_G = 1.2V$, $V_D = 3.6V$ are shown. The temperature dependence flips from low to high V_D.

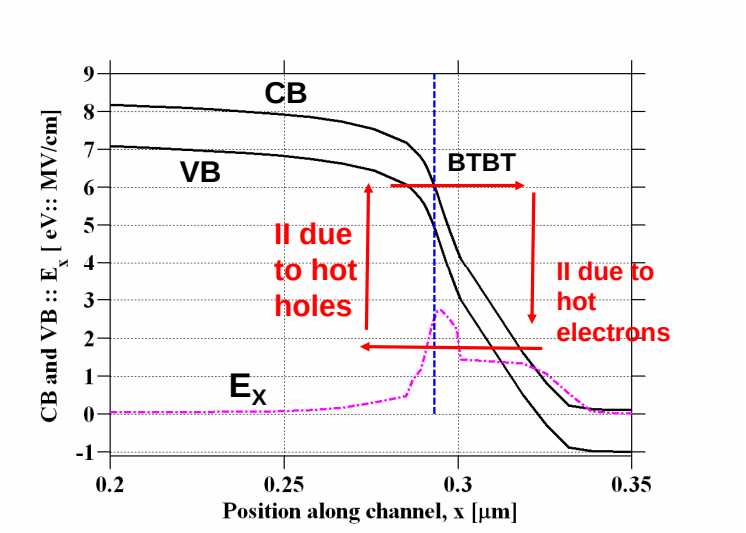


Fig. 7.11. Conduction and valence bands of DeNMOS and electric field near the drain. The vertical dashed line shows the gate edge. Carrier heating mechanisms are also depicted.

densities, corresponding spatial distributions and their impact on device reliability can be investigated. As a case study, DeNMOS off-state ($V_{GS} = 0V$, bias applied to drain) is considered. Fig. 7.11 shows lateral cross section of conduction and valence bands, and lateral electric field in the channel. The electric field peaks under the gate edge at the drain side and this leads to band-to-band tunneling of electrons towards the drain. At high drain bias, the drain current is dominated by the band-to-band tunneling component which also gives rise to the substrate current as depicted in Fig. 7.12. It can also be seen that the current component due to impact ionization (I_{SUB} calculated from SMC simulations) is orders of magnitude smaller than the band-to-band tunneling current. MEDICI simulations without impact ionization also agrees with the experimental data confirming the dominance of band-to-band tunneling in drain current [160]. As the electrons tunnel from the valence band to the conduction band, they experience an accelerating electric field inside the drain. The energetic carriers heated under this field can impact ionize in the drain and generate electron-hole pairs. Due to the field profile, the holes generated in the drain can be swept

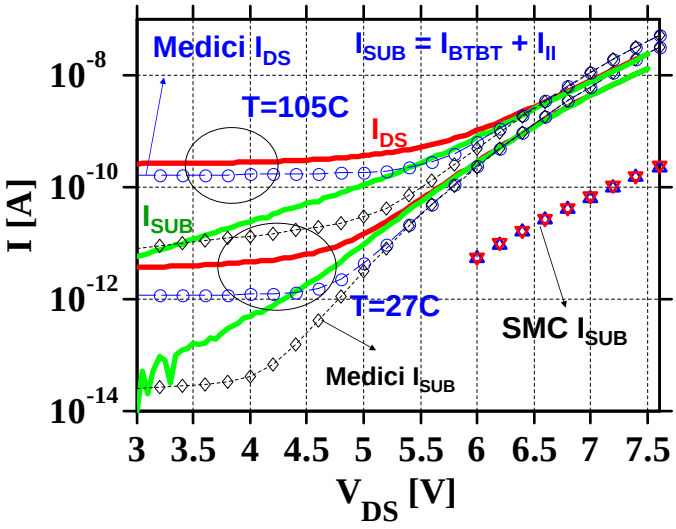


Fig. 7.12. Experimental drain and substrate current for two different temperatures. Medici currents are also compared. The drain and substrate currents are due to band-to-band tunneling at high drain bias. Impact ionization current, I_{SUB}, obtained from SMC simulations is much less than the drain current.

towards the gate and cause degradation at the *Si/oxide* interface or in the bulk gate oxide. SMC can also be used to obtain the hot carrier distributions in the device. Fig. 7.13 shows hot electron and hole distributions near the drain junction of the DeNMOS. Electrons having energies higher than 3.1 eV (conduction band offset between Si bulk and gate oxide) are considered hot, because it is assumed that only above that energy, they can be injected into the gate oxide and degrade the device. Similarly, hot holes are those with energies higher than 4.7 eV (valence band offset between Si and oxide). The hot electron density peaks underneath the spacer over the drain since they arise from band-to-band tunneling near the gate edge ($x = 0.2932\mu m$) and are accelerated in the drain electric field. Hot holes, on the other hand, peak over the gate-drain overlap. The hot carrier densities increase with higher drain bias as expected because the electric field which causes band-to-band tunneling and carrier heating increases with the drain voltage when gate bias is zero. The hot

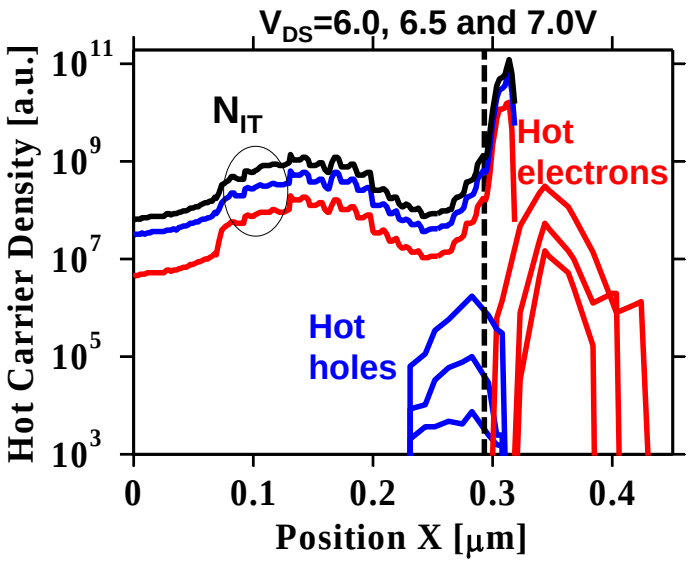


Fig. 7.13. Hot electron and hole densities inside the channel near the gate edge (vertical black line) for different drain voltages. Hot electrons are mainly outside the gate and generate holes via impact ionization. The holes are accelerated back towards the channel. Spatial interface trap density is also given. The density peaks where the hot hole and electron densities overlap.

carrier density also spreads in space with increasing drain bias. This is plausible since higher drain bias leads to higher electric field which then accelerates carriers to higher energies. So at a given energy, it is expected to find more carriers and more spatial spreading. Fig. 7.13 also compares spatial interface trap distribution with hot carrier densities. Spatial profiling of N_{IT} was obtained from charge pumping measurements discussed in the previous chapter. The interface trap density peaks right outside the gate edge over the spacer. This region corresponds to the intersection of the hot electron and hole densities. Fig. 7.14 shows integrated hot carrier densities over the whole device domain and integrated interface trap distribution in the channel for different drain bias. The individual integrated hot carrier quantities follow the N_{IT} versus V_{DS} trend and imply each of them separately or their intersection can be responsible for the interface trap generation. However, when only the densities

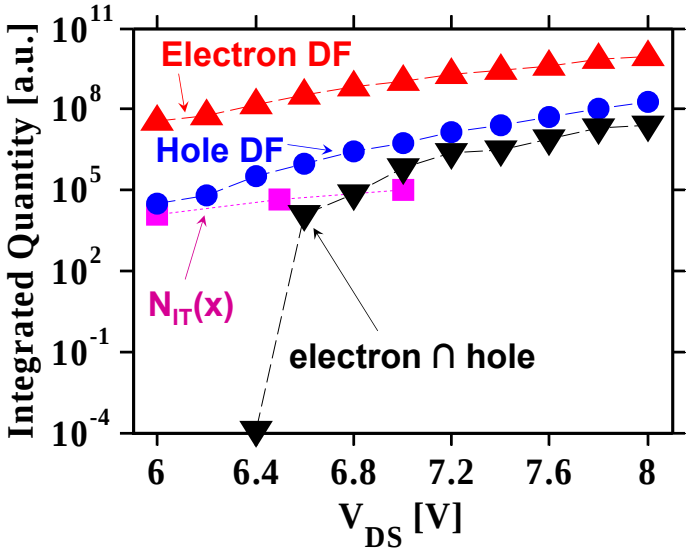


Fig. 7.14. Integrated hot electron, hole, their intersection and interface trap densities as a function of drain bias. The whole device domain is considered. The interface traps shows good correlation to the hot carrier densities.

underneath the gate oxide are integrated, only hot hole density shows good agreement with the experimental interface trap density trend as in Fig. 7.15.

7.4 Summary

A modeling framework that starts from process-level and extends to electrical and hot carrier simulations is presented. The framework was applied on a specific transistor, DeNMOS, to quantify the hot carrier degradation in off-state. The tools are applicable to other cases, such as on-state HCI for logic MOSFETs, as it will be discussed in the following chapters. For practical purposes, time-dependence of degradation is important so that lifetime extrapolations of MOSFETs and circuits can be achieved. The geometric localization of the hot carrier degradation can be incorporated into the Reaction-Diffusion framework for time-dependence [53]. The physics of bond-breaking, k_F, needs to be understood in terms of the contribution of

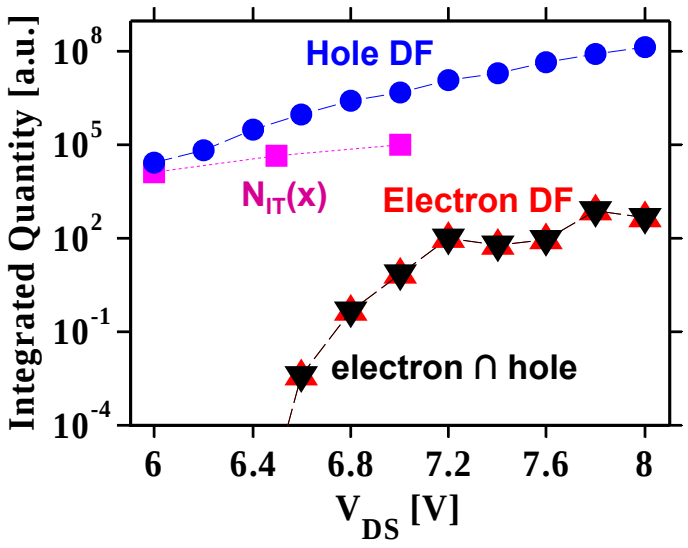


Fig. 7.15. Same as previous figure but only the region underneath the gate oxide is used in integration. Interface trap density compares well only with the hot hole density.

electrons or holes, the energies need for bond-breaking and as to whether injection into the gate oxide is required for creating interface traps. The spatial distribution of the defects are also important since it plays a role in time-dependence of degradation. These issues will be investigated with the help of experimental studies and calibrated framework presented in this chapter.

8. HOT CARRIER INJECTION CHARACTERIZATION

8.1 Introduction

The majority of the HCI work has been done during '80s and '90s when hot carrier degradation was very serious in CMOS technology. The large operating voltages ($V_{DD} = 5V$) resulted in significant damage, thus HCI reliability was a major concern as discussed in Chapter 6. It was speculated that if the operating voltages were lowered, the HCI would disappear; a prediction that was proven to be incorrect by recent technologies with lower V_{DD}s. In order to explain HCI better, one needs to understand carrier heating and damage generation phenomena. Particularly with technology scaling, the overall degradation which is a weighted combination of these mechanisms can change, affecting the impact on device and circuit operation.

The purpose of this chapter is to investigate HCI mechanisms for shorter channel MOSFETs stressed near operating voltages; conditions relevant to present technologies. First, the differences and similarities among the older and newer HCI observations will be discussed to identify the challenges of HCI characteristics. Then, degradation and recovery trends in technologically-relevant conditions are presented. In compliance with the previous chapter, the carriers responsible for damage, the type of degradation, time dynamics and the impact of degradation on MOSFET metrics are given.

8.1.1 HCI behavior of short and long channel MOSFETs

MOSFET dimensions and operating voltages are continually reduced according to the CMOS technology scaling. In older generations, it was observed that damage due

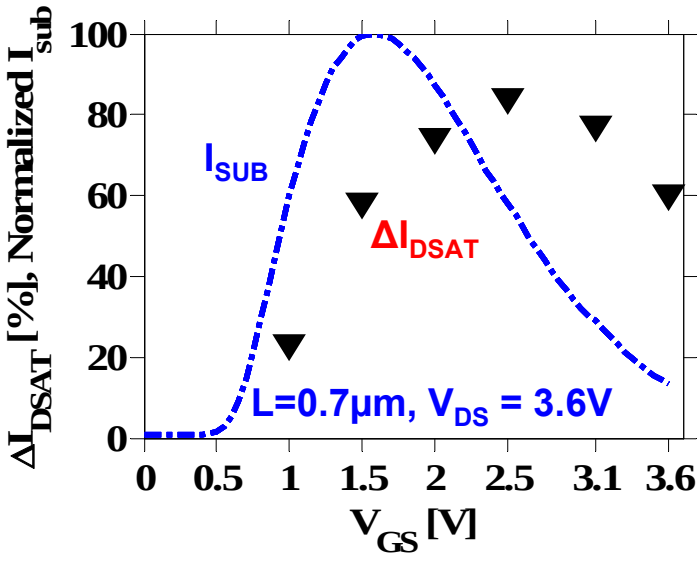


Fig. 8.1. I_{SUB} (dot-dashed line) and degradation (triangles) of an NMOSFET stressed at $V_D = 3.6$V and several different gate voltages. I_{SUB} shows a bell-shaped behavior with peak at $V_G = 1.6V$. The degradation, ΔI_{DSAT}, is also bell-shaped but peaks at a different V_{GS}, closer to $V_{GS} = V_D$.

to HCI was well correlated with the substrate current, I_{SUB}. However, with scaling, this trend is not consistent with the older generations anymore as shown in Fig. 8.1.

This behavior is not suprising since I_{SUB} reflects the amount of impact ionization which is a function of carrier energy in the channel. The degradation can involve additional mechanisms on top of energetic carriers: the carriers need to be directed towards the oxide and generate defects. The direction of carriers (following the electric field) changes with the gate bias, therefore I_{SUB} and degradation do not necessarily coincide.

As the channel lengths are shrunk, the degradation peaks at $V_G = V_D$ as representatively shown in Fig. 8.2. This is partly due to the direction of carriers near the drain as well as the gate voltage where energy of carriers are maximized. Fig. 8.3. shows that for short channel MOSFETs, the substrate current peaks near $V_G = V_D$ whereas in the long channel case, the I_{SUB} depicts a clear peak around $V_G = V_D/2$. The

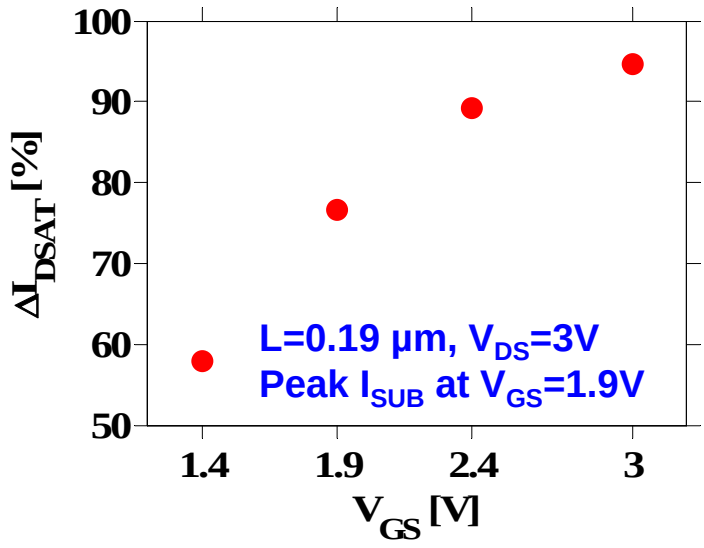


Fig. 8.2. Representative degradation of short channel MOSFETs. The degradation (circles) peaks at $V_G = V_D = 3V$ for a shorter channel MOSFET although the I_{SUB} has a bell-shape characteristics, peaking at $V_G = 1.9V$.

$V_{D,SAT}$ for a short channel device is smaller, therefore the average energy of carriers is higher (see discussion in Section 6.1). The V_G needs to be increased further than V_D to increase $V_{D,SAT}$ and thus reduce the average energy and I_{SUB}. As a result, the bell-shape for short channel devices does not appear unless V_G significantly exceeds V_D.

These results signify that hot carrier dynamics change with channel length and stress voltages, both playing a role on the degradation. These mechanisms need to be understood quantitatively in order to predict HCI for current generation technologies. In this thesis, the tools mentioned in Chapter 7 provide that essential insight.

8.1.2 Type of defects

The damage due to HCI impact MOSFET characteristics as exemplified in Fig. 8.4. The degradation cause reduction in the drain current, I_D, due to negatively

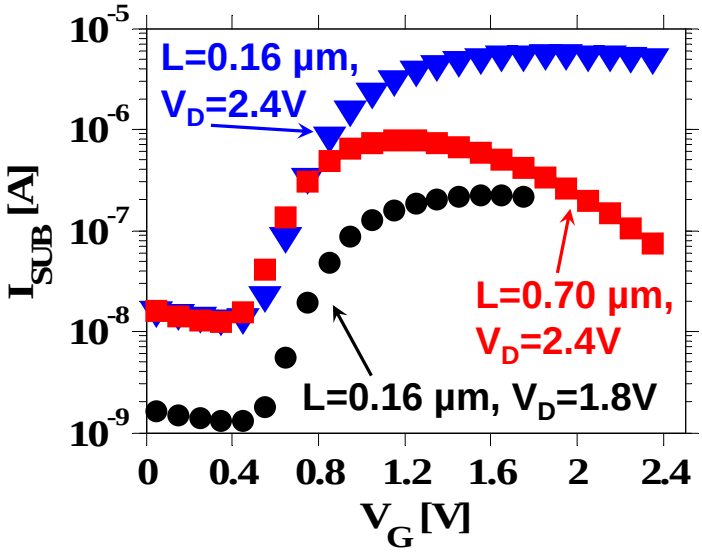


Fig. 8.3. As the channel length decreases (from $0.7\mu m$(squares) to $0.16\mu m$), the bell-shape of I_{SUB} disappears and the peak shifts to $V_G = V_D$ for both V_D (triangles and circles).

charged defects. These charges are believed to be originating from interface or bulk traps. The nature of the traps are attributed to broken chemical bonds; in the case of HCI, $Si - H$ (recovery in Fig. 8.4) and $Si - O$ are generally held responsible for the defects. The level of degradation at the $Si/oxide$ interface and bulk oxide depends on the stress condition. As discussed in the previous chapter, stress voltages directly affect the carrier energy in the channel; therefore lead to activation of different processes.

One such mechanism is bulk trap (N_{OT} in cm^{-3}) generation in the gate oxide due to carriers with high energies. This mechanism is prominent at higher stress voltages. Fig. 8.5 shows bulk trap density (measured by multi-frequency charge pumping (MFCP) discussed in Appendix A.6) for MOSFETs stressed at different voltages ($V_G = V_D$). As the stress voltages are scaled, the bulk defect creation reduces and becomes negligible at the lowest voltage. The voltage acceleration of N_{OT} is consistent with the values in [161]. Despite this fact, the MOSFET $I_D - V_G$ characteristics still degrade at this condition, as in Fig. 8.4. This can be attributed to

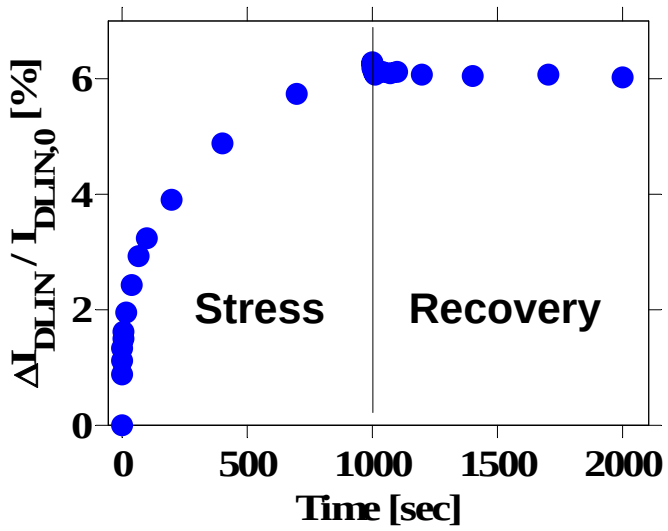


Fig. 8.4. HCI degrades the drain current of a MOSFET (L=0.16μm) over time. Stress at $V_G = V_D = 2V$ at T=30C. After 1000 seconds of stress, I_D recovers partially when held at $V_G = V_D = 0V$. Linear characteristics were measured at $V_G = 1.8V$, $V_D = 0.1V$ since measuring I_{DSAT} can degrade the MOSFET at $V_G = V_D = 1.8V$ condition.

interface damage created by the hot carriers. Fig. 8.6 compares interface (N_{IT}) and bulk traps (normalized, per unit area) for the aforementioned stress voltages. The lowest measurement frequency (5KHz) probes about x_{5KHz}=4.7$\dot{A}$ into the dielectric. The normalization of bulk traps is taken $N_{OT} \cdot x_{5KHz}$ which yields the areal density (in cm^{-2}) assuming trap density is constant over this depth. The interface traps are computed from the highest measurement frequency (2MHz). As the stress voltage increases, the bulk traps keep increasing, however the interface trap component tends to saturate. Therefore at high voltage, (N_{OT}) dominates, consistent with [16]. With voltage decreasing, both (N_{IT}) and (N_{OT}) decrease but since the latter does not increase with time (Fig. 8.5), interface traps dominate the overall degradation.

Fig. 8.7 gives the extracted time-exponents (n) of experimental degradation characteristics (Fig. 8.5). At low stress voltage, the N_{OT} does not change appreciably, as mentioned earlier. For the intermediate and high V_D, the exponents are about 0.35.

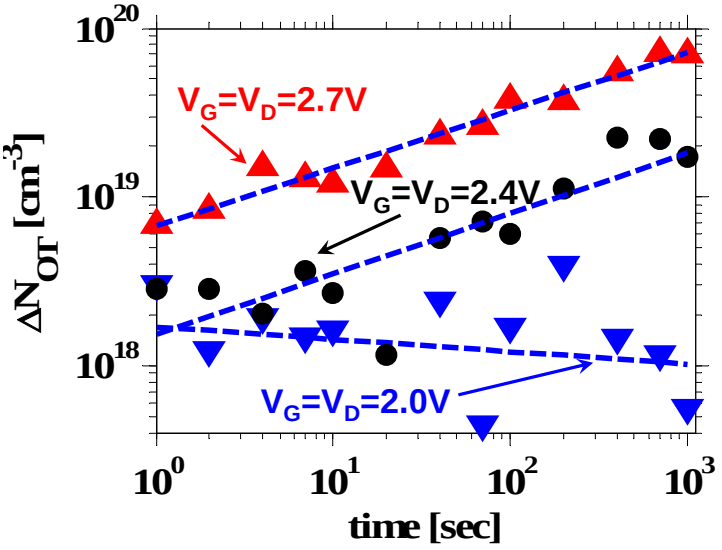


Fig. 8.5. Bulk traps, N_{OT}, extracted from multi-frequency charge pumping (MF-CP) technique (see Appendix A.6), scale with voltage. For $V_G = V_D = 2V$ (down-triangles), the generation is negligible; it does not increase with time unlike $V_D = 2.4V$ (circles) and $2.7V$ (up-triangles).

On the other hand, the interface trap time-exponent decreases with increasing V_D and becomes equal to that of N_{OT} at high voltage. The exponent $n \approx 0.35$ is in good agreement with [162] but lower than the values reported in [120], [163].

These results suggest that interface traps have two components that can appear as the stress conditions are changed. This is consistent with the role of $Si-H$ and $Si-O$ bond breaking. The total interface traps, N_{IT}, are a combination of broken $Si-H$ and $Si-O$ bonds. At low voltages where carriers have lower energies, $Si-H$ dominate, as in Fig. 8.8. The recovery is small unlike the NBTI case and the reason is the geometry effect as discussed in Section 4.2. As the voltage increase, $Si-O$ component overtakes the $Si-H$ and dominate the time dependence. The measured $Si-H$ time-exponent is around 0.5 whereas $Si-O$ exponent is 0.35. Therefore the total N_{IT} time-exponents reflect the two bond types: $Si-H$ and $Si-O$. At the intermediate stress conditions, the overall time-exponent somewhere in between.

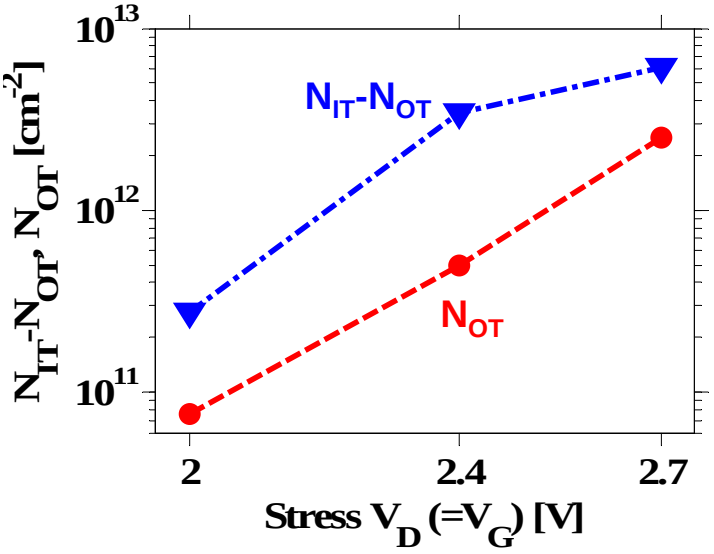


Fig. 8.6. Bulk traps extracted from MF-CP and interface traps measured by charge pumping increase with stress voltage; N_{IT} increases slower than N_{OT} at high voltage.

This is further verified analytically in Fig. 8.9. When the magnitude of the bulk trap component increases with voltage, it first becomes comparable to $Si-H$ (time-exponent n=0.44), then exceeds $Si-H$ (time-exponent n=0.37), with each case showing a near power-law trend. The $Si-H$ ($N_{IT}-N_{OT}$ in Fig. 8.6) component also increase with voltage however its rate is not as fast as $Si-O$ (N_{OT} in Fig. 8.6).

The type of defects also affect the recovery characteristics. It is known that broken $Si-H$ can be annealed when the stress is turned off. As for $Si-O$, passivation of dangling bonds by oxygen is not very effective, therefore the recovery is negligible. Experimental recovery trends for different stress conditions are shown in Fig. 8.10. As expected, relaxation of degradation is small for MOSFETs stressed at high voltages since the $Si-O$ component is more than $Si-H$. Broken $Si-O$ can be annealed by hydrogen released from $Si-H$, however this process is limited by the amount of available H; in other words, broken $Si-H$ bonds. Additionally, there is a competition

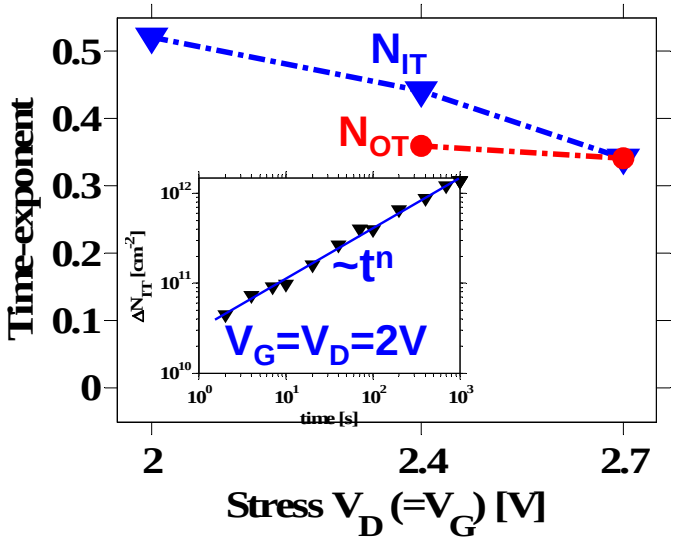


Fig. 8.7. Time-exponent of bulk traps are about 0.35 whereas that of interface traps change with stress voltage. When N_{OT} is negligible, N_{IT} exponent is around 0.5 and decreases with voltage, matching N_{OT} exponent at the highest stress condition. At $V_G = V_D = 2V$ (inset), the bulk trap density does not increase with time, therefore it is excluded from the plot.

between broken $Si-H$ and $Si-O$ to capture the free hydrogen, therefore recovery is limited.

Experimental results imply that as the stress voltages are scaled, the degradation is mainly due to interface traps originated from broken $Si-H$ bonds. At higher voltages, contribution from $Si-O$ increase. If accelerated tests with short stress time and high voltages are used to characterize HCI, the lifetime extrapolation can be misleading since the magnitude and time-dependence of $Si-O$ dominates at these conditions.

8.1.3 Carriers responsible for HCI

In the previous section, it was explained that for a MOSFET stressed near the operating voltages, the degradation is mostly related to interface traps. Voltage and

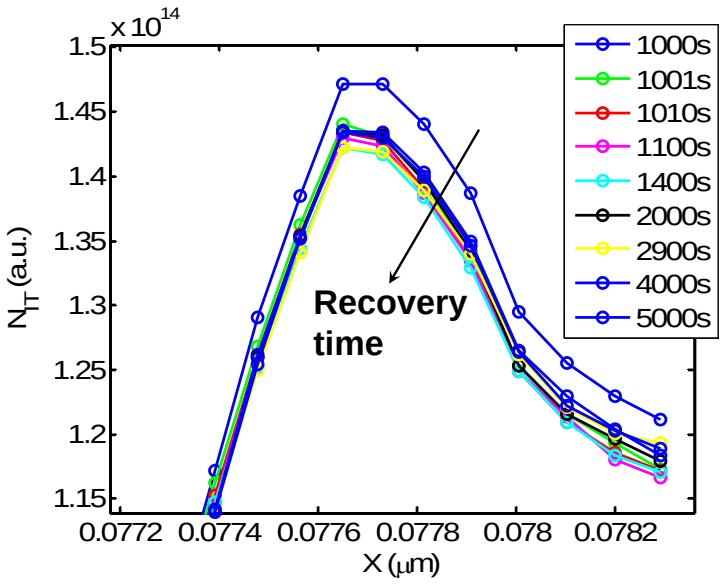


Fig. 8.8. Interface traps, measured by DCIV, recover, supporting the existence of $Si-H$ origin. Unlike NBTI, however, the fractional recovery is small. The gate extends to about x=0.079μm beyond which the spacers are present.

time-dependences were employed to isolate bulk trap component prominent at higher stress voltages. The next important question which has been debated in the literature for decades is what causes degradation, more specifically, which energetic carrier is generating the defects? Hot electrons, holes or both have been held responsible for HCI by different groups (Chapter 6). In this section, by utilizing experimental and theoretical tools, the carriers leading to defects will be identified for technologically-relevant NMOSFETs.

In order to determine which carriers generate interface defects, spatial profiling of the degraded region is performed by DCIV technique (Appendix A.7) during stress and recovery. DCIV is preferred over spatial charge pumping (Appendix A.5) because the doping profiles of the MOSFETs used in this study are not monotonic. It yields charge pumping not useful for this purpose. Fig. 8.11 shows build up of interface traps near the drain junction of the MOSFET . It is reasonable that the position of the degraded region and that of the carriers causing damage are correlated. The

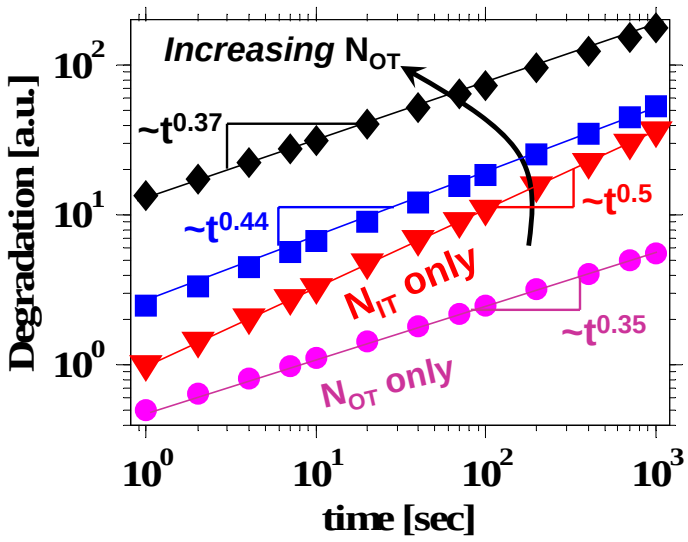


Fig. 8.9. Simulation of N_{IT} and N_{OT} components of degradation. As the stress voltage increases, the fraction of N_{OT} dominates the overall degradation. The time-exponent varies from that of N_{IT} to N_{OT} with voltage.

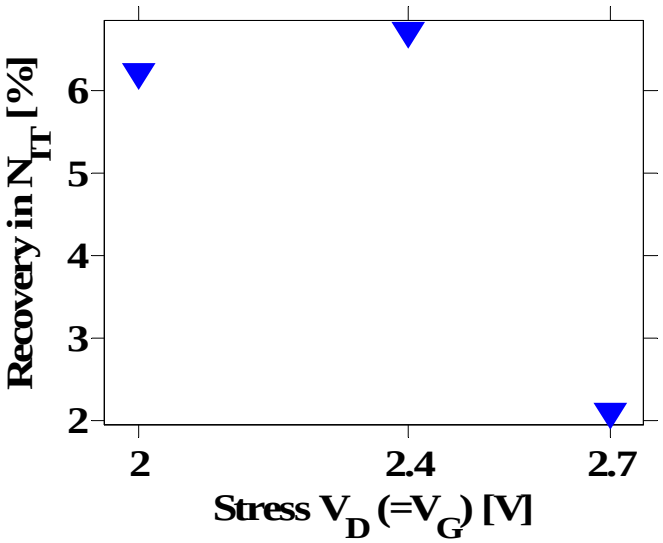


Fig. 8.10. Broken bonds partially recover when stress conditions are removed. The recovery is less for MOSFETs stressed at high voltages because the $Si-O$ breakage cannot be annealed by oxygen.

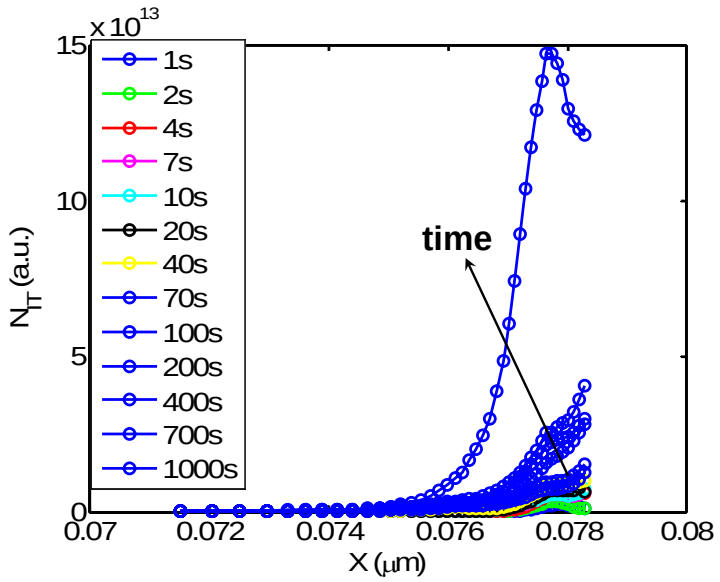


Fig. 8.11. N_{IT} increases over time during HCI. L=0.16μm NMOSFET stressed at $V_G = V_D = 2V$ for 1000 seconds. The degradation is localized near the drain end of the transistor. x=0 is the middle of the channel, gate extends to about x=0.079μm.

microscopic mechanism is not very clear on how hot carriers generate defects, however the magnitudes of the defects and hot carriers are generally related. For profiling of the carriers, the spatial distributions of hot electrons and holes are extracted from SMC. Fig. 8.12 compares the two near the drain of the NMOSFET. As it can be seen from Fig. 8.11, the position of the interface traps are in agreement with that of the hot electrons. Hot holes are also present, but their location is more towards the source. Since the holes, generated by impact ionization, gain energy in the potential near the drain, more hot holes are expected away from the drain. The connection between hot electrons and interface damage is further established in Fig. 8.13. Integrated hot electron distribution function (DF) agrees better with the trend of interface traps for MOSFETs having different channel lengths.

With higher stress voltages, the type of defects mostly consists of $Si-O$ as discussed in the previous section. Breaking the $Si-O$ bonds requires carriers with higher energies than those needed for breaking the $Si-H$. Particularly, hot holes are

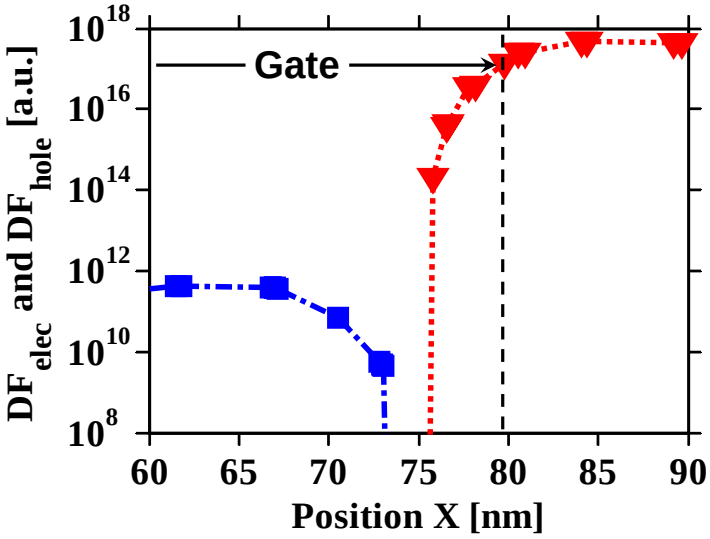


Fig. 8.12. Hot electrons and interface traps are spatially correlated. SMC is used to obtain the hot carrier distributions.

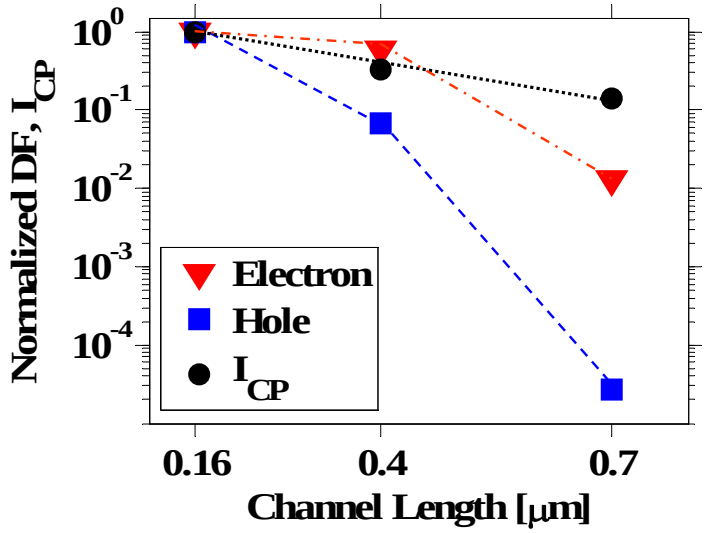


Fig. 8.13. Experimental charge pumping current ($\propto N_{IT}$) correlates better with hot electrons whereas hot hole density drops very rapidly as channel length increases. All MOSFETs were stressed at $V_G = V_D = 2.4V$ for 1000 seconds. Hot carrier distribution functions were integrated over space and energy for the entire device.

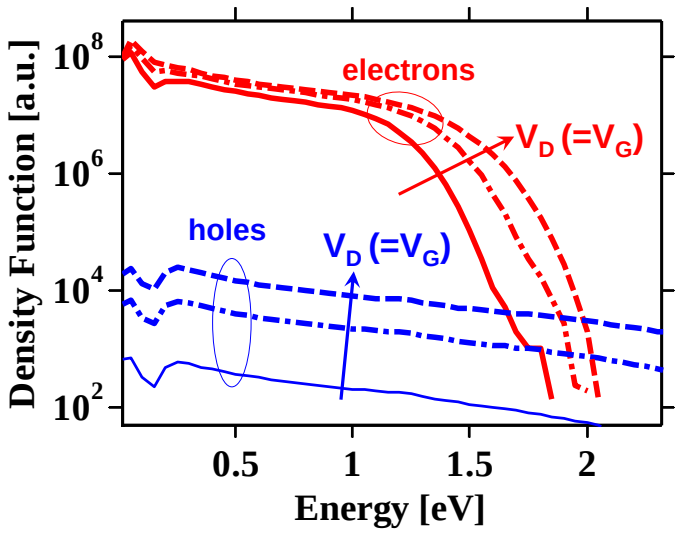


Fig. 8.14. Distribution functions of electrons and holes for L=0.16 μm NMOSFET stressed at (V_G =)V_D =2, 2.4 and 2.7V. The tail of the electron DF and hole distribution increases with V_D.

believed to take part in this mechanism. Fig. 8.14 shows the distribution functions (DF) of electrons and holes for MOSFETs stressed at different $V_G = V_D$ conditions. The tail of the electron distribution is populated significantly with increasing V_D whereas the hole distribution is raised over the entire energy range. The low energy portion of the electron DF is composed of channel electrons in the inversion layer and a fraction of these gain significant energy near the drain. Since holes come from impact ionization, they are proportional to the tail of electron DF. As a result, an increase in stress voltage directly impacts the hole DF, even at low carrier energies. From Fig. 8.14, it is not clear whether hot electrons (tail of DF) or holes break the $Si-O$ bonds. In order to clarify this, voltage dependences of electron and hole DFs, and experimental N_{IT} and N_{OT} are depicted in Fig. 8.15. It can be seen that the N_{OT} ($Si-O$) correlates better with the hole DF (energies $>$ 0.3eV), rather than the electron DF (energies $>$ 1.1eV). Moreover the N_{IT} has a trend similar to the N_{OT}. This is not suprising because as the stress voltage increases, the fraction of interface defects due to broken $Si-O$ follows, as in Fig. 8.9. Although the mechanism of defect

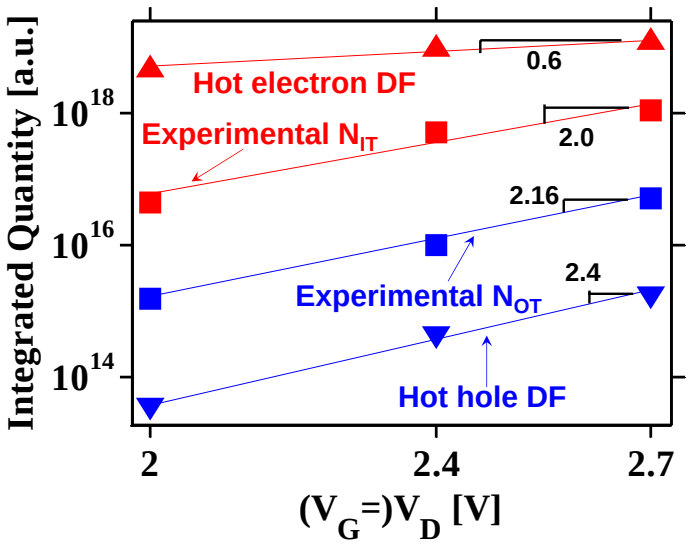


Fig. 8.15. Experimental $N_{IT}(= Si-H+Si-O)$ and $N_{OT}(= Si-O)$ are compared with integrated DF tail of electrons (E>1.1eV) and holes (E>0.3eV). Voltage accelerations are also given.

generation by hot carriers is not established, Fig. 8.14 presents clues for this process. With the stress voltages used in this study, the electron and hole DFs populate energies significantly less than the conduction and valence band offsets ($\approx$ 3.1 and 4.7 eV, respectively), even with the image-force barrier lowering. This signals that the hot carriers need not be injected into the oxide to generate interface or bulk traps. At $V_G = V_D$ stress, the local electric field near the drain favors electrons towards the interface. Therefore a carrier can convey energy to the bonds and accumulating energy transfer can dissociate atoms. This result is consistent with previous findings [125], [126]. Under hot carrier stress ($V_G = V_D$), the electron tranport in the channel is lateral: from source to drain. In order for electrons to interact with the chemical bonds at interface, the carriers need to be scattered so that their lateral movement is impeded. Since impact ionization, a scattering mechanism that can change the direction of electrons, is not required for HCI, alternative mechanisms are necessary. Such momentum change can come from electron-electron scattering or simply arise from roughness of the Si/oxide interface and dopant fluctuations in the channel and

drain. Even for advanced device geometries such as Ultra Thin Body (UTB) and nanowire FETs where transport is ideally 1-D, these non-idealities can scatter hot electrons towards the gate and result in degradation.

8.1.4 Summary

Spatial profiling (DCIV) of interface damage and theoretical (SMC) extraction of carrier distributions in energy and space yield that holes can break $Si-O$ bonds at higher stress voltages. For MOSFETs stressed near operating voltages, electrons correlate better with the interface damage emanating from broken $Si-H$ bonds. The carrier distributions in energy suggest that injection into the gate oxide is not necessary for creating defects. The results help clarifying these conflicted issues by means of powerful experimental and theoretical tools.

9. IMPACT OF DEGRADATION ON MOSFET CHARACTERISTICS

In the previous chapter, HCI induced interface and bulk oxide defects were discussed. Although given qualitatively, the effects of these defects on MOSFET characteristics were not presented in detail. Such a relationship is needed because the transistor and circuit lifetimes are generally assessed in terms of MOSFET metrics (i.e., I_{Drain}, g_m, V_T etc). As seen in Fig. 9.1, the defects and their impact on MOSFET characteristics may not appear to be correlated: time, temperature and voltage dependences of interface traps and therefore characterization of HCI should be performed carefully. In this chapter, interface trap (N_{IT}) generation and its relation to drain current (I_D) degradation will be investigated. First, the effect of measurement delay will be discussed because further investigation requires correction for any delay related artifacts. Then, the temperature dependence of I_D degradation and N_{IT} for worst-case HCI conditions will be presented. Finally, the correlation of time-dependences between N_{IT} and I_D, and their recovery characteristics will be given. Charge trapping/de-trapping effects will also be shown.

9.1 Measurement delay in HCI

In the previous chapter, it was shown that HCI, similar to NBTI, also shows recovery, although the fraction is much smaller than that of NBTI. This recovery is attributed to the annealing of dangling Si bonds by hydrogen. Recovery is more pronounced for lower stress voltages, where $Si-H$ breaking is the main mechanism behind interface trap generation.

In NBTI, it is known that recovery, i.e., measurement delay, causes artifacts: the magnitude and time-dependence of NBTI with recovery can be significantly different

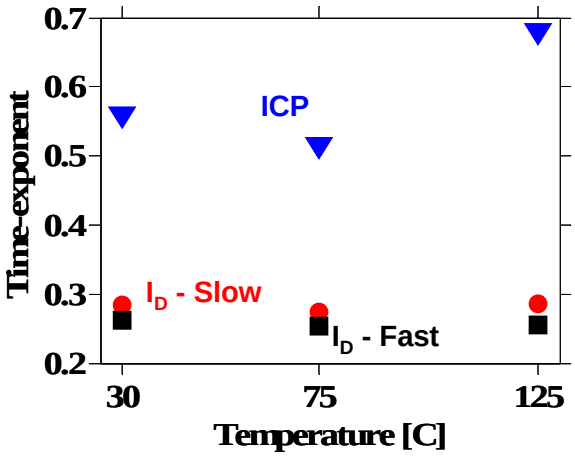


Fig. 9.1. Time-exponents of the power-law trends for I_{CP} (triangles) and ΔI_D (circles and squares) are markedly different under HCI stress. I_D measurements were done at $V_G = 1.8V$ and $V_D = 0.1V$. Fast (squares, 2 sec. delay) and slow (circles, 50 sec. delay) I_D measurements show a slight difference in time-exponents. MOSFETs were stresssed at $V_G = V_D = 2V$ for different temperatures.

than an ideal measurement. This can lead to erronous lifetime extrapolations for MOSFETs and circuits. Since HCI also recovers partially, the impact of delay should be evaluated first to avoid characterization discrepancies. In this section, the effects of measurement delay will be analyzed by conventional and ultra-fast techniques. The trends will be related to the theoretical behavior.

In conventional methods, the MOSFET is stressed under HCI, then the drain current is measured at predefined time intervals. Usually, between stressing and measuring the degradation, there is a time period due to equipment limitations. Moreover, if the measurement involves other techniques such as charge pumping, the delay can be larger, leading to larger recovery.

The degradation and recovery characteristics with different delay durations are shown in Fig. 9.2. With the fast measurement, the degradation and recovery are slightly higher than the slow measurement case because the recovery is smaller. The power-law exponents of stress are 0.255 and 0.286 (Fig. 9.1) for the fast and slow

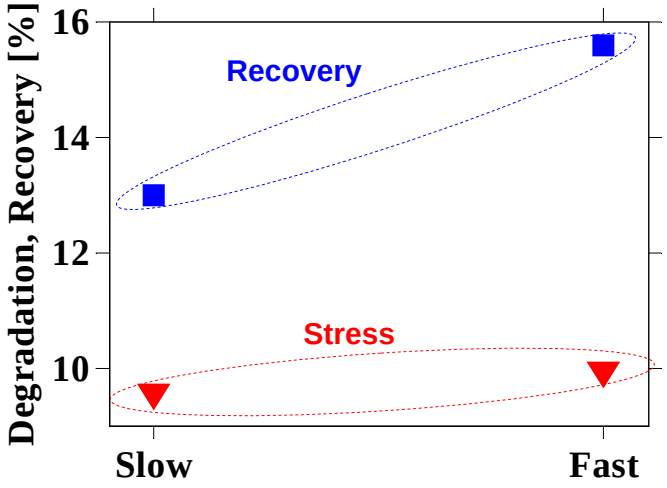


Fig. 9.2. Percentage degradation and recovery shows little change between slow and fast measurements. The delay in slow and fast measurements are 50 and 2 seconds, respectively. The HCI behavior captured by the fast measurement is closer to the intrinsic degradation. MOSFETs were stressed at $V_G = V_D = 2V$ for 10,000 seconds and recovery performed for 40,000 seconds.

measurements, respectively. The delay in conventional measurements is not very small, as evidenced by NBTI measurements. NBTI shows quick recovery even in ms time scale. In order to check the HCI recovery in short time periods, ultra-fast setup (Appendix A.8) was used. This setup is capable of measuring degradation and recovery in microsecond resolution. The recovery behavior is given in Fig. 9.3 which shows that the drain current does not change appreciably in short time scales. The recovery trend is within the noise window of the setup, therefore this result implies that the recovery is not significant and conventional techniques can measure HCI fairly accurately. Since the measurement delay does not affect the time-exponent of degradation noticeably, this suggests that intrinsic fractional recovery in HCI is indeed small as predicted by the 2-D R-D theory (discussed more in the next section) [53].

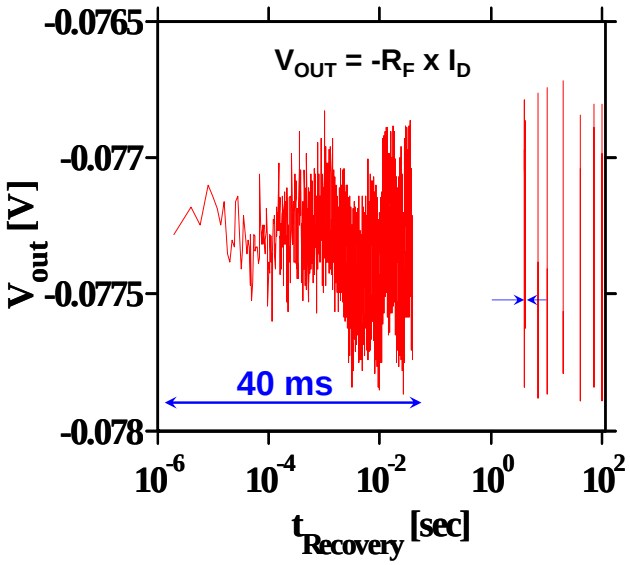


Fig. 9.3. Ultra-fast setup (Appendix A.8) results show that recovery is not significant at short time scales, down to microseconds. Due to the noise in the setup, the small amount of recovery cannot be observed clearly, but a fast-recovery component can be ruled out. Each measurement segment is 40 ms long. There is no interruption during the measurements or stress-recovery transition.

9.2 Temperature dependence

In addition to delay effect, another complication of HCI measurements arises from the temperature dependence. As discussed in Sec. 7.3 (Fig. 7.7), the substrate current (I_{SUB}) which is a measure of carrier energy, has a temperature dependence that reverses with stress voltage. For classical hot carrier studies with long-channel MOSFETs and larger stress voltages, it was shown that HCI reduces with temperature [164]. The reason is that the average energy of carriers in the channel of a MOSFET is reduced at high temperatures due to phonon scattering. As the operating voltages are scaled for present day technologies, this observation is reversed: higher temperatures generate more I_{SUB}. Since the MOSFETs can operate at higher temperatures in modern ICs, the implications on HCI are important.

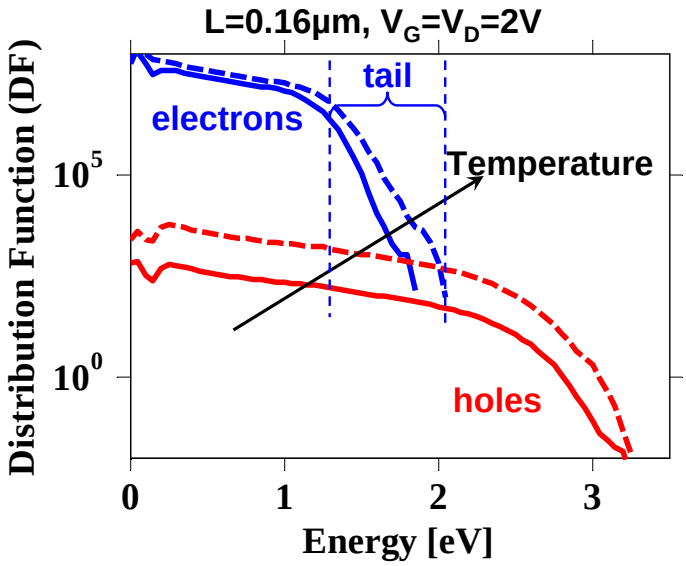


Fig. 9.4. With increasing temperature (T=25 (solid line) and $125C$ (dashed line) shown), the electron density increases over the entire energy region for a MOSFET (L=0.16μm) stressed at $V_G = V_D = 2V$. The increased electron population at thermal tail (E>1.1eV) of DF causes more impact ionization, thus hole density and I_{SUB} follows ($I_{SUB}(T = 125C) > I_{SUB}(T = 25C)$).

The reversed temperature dependence can be understood by looking at the carrier distribution functions (DF) in Fig. 9.4 and 9.5. The occupancy of electrons at high energy part of the DF (thermal tail) increases with temperature in both MOSFETs. When a MOSFET is stressed at lower voltages, the maximum energy a carrier can gain (without scattering) is roughly qV_D where q is the electronic charge. With lower V_D, the critical energies in HCI (impact ionization threshold, defect generation energy) lie within the range of this tail. As the stress temperatures increase, larger thermal tail can generate more holes via impact ionization and more damage due to electrons.

The increased damage at higher temperatures can be seen from Fig. 9.6 in which experimental charge pumping current (I_{CP}) and the interface trap density (per energy, D_{IT}) are shown. The I_{CP}, which is proportional to the interface trap density, decreases with stress temperature, however it is also intrinsically temperature-

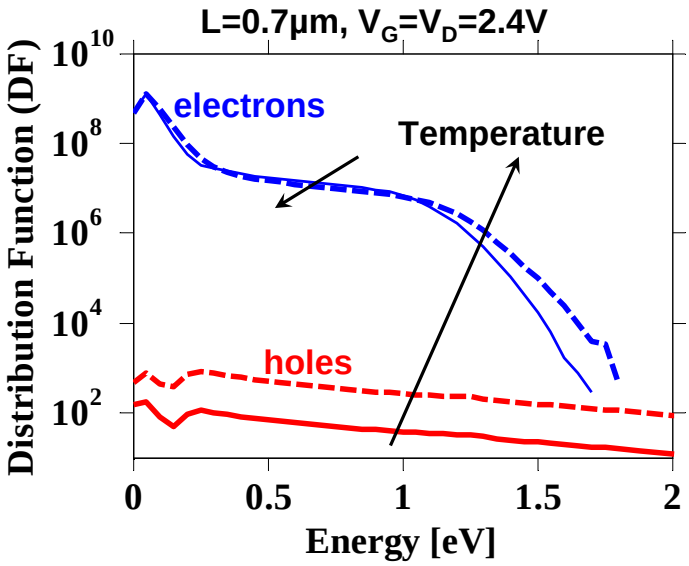


Fig. 9.5. Similar to the short channel MOSFET, higher temperature increases the electron density at the DF tail for a MOSFET (L=0.70μm) stressed at $V_G = V_D = 2.4V$. At the low energy part of the DF, the density of electrons at high temperature is lower than the low temperature DF, unlike the short channel (L=0.16μm). This could be due to suppressed DIBL.

dependent: the energy range (within the Si bandgap) probed by charge pumping varies with measurement temperature. The charge pumping current for HCI can be written as

$$I_{CP} = q{\cdot}A_{HCI}{\cdot}f{\cdot}D_{IT}{\cdot}\Delta E, \tag{9.1}$$

where A_{HCI} is the area degraded by HCI, f is the CP frequency, D_{IT} is the interface trap density (in units of $cm^{-2}eV^{-1}$) and ΔE is the energy range probed by charge pumping at a particular temperature, and given by (9.2)

$$\Delta E = 2k_BT[ln(\upsilon_{Th}n_i\sqrt{\sigma_n\sigma_p}\frac{|V_{FB}-V_T|\sqrt{t_Rt_F}}{\Delta V_G})], \tag{9.2}$$

where k_B is the Boltzmann's constant, T is the measurement temperature, υ_{Th} is the thermal velocity, n_i is the intrinsic carrier concentration, $\sigma_{n,p}$ are the capture cross sections for electrons and holes, V_{FB} is the flat band voltage, V_T is the threshold voltage, ΔV_G is the amplitude of the CP pulse, and $t_{R,F}$ are the rise and fall times of

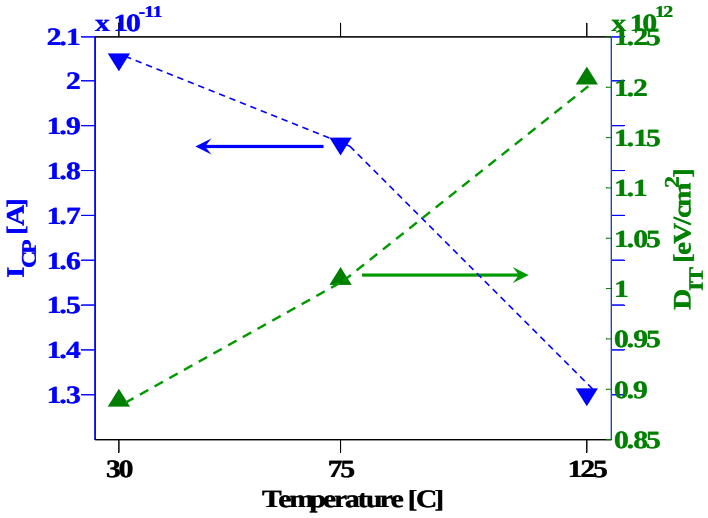


Fig. 9.6. Charge pumping current, I_{CP} (left axis), decreases with increasing stress temperature. After the correction for the energy range, D_{IT}(right axis) increases. The activation energy for D_{IT} is 0.014 eV (stress: L=0.16μm, $V_G = V_D = 2V$).

the CP pulse [165], [166]. The parameters υ_{Th} and n_i also depend on temperature. For the MOSFETs stressed at these conditions, the energy ranges are approximately 0.6, 0.48 and 0.28 eV at T =30, 75 and 125C, respectively. At higher temperatures, the energy range is smaller because the thermal emission of carriers from trap states are enhanced. Using this energy correction in (9.1) and (9.2), the corresponding D_{IT} values were extracted. Assuming uniform trap distribution in the bandgap, the interface trap density (D_{IT}) increases with temperature, consistent with expectation for short channel MOSFETs from SMC.

The temperature also affects the recovery. Fig. 9.7 shows percentage recovery for CP and I_D and degradation in I_D. The recovery in I_{CP} increases with temperature. This behavior of I_{CP} can be interpreted as that of D_{IT} since percentage recovery is shown and energy range corrections on I_{CP} cancel out. The drain current degradation and recovery are also shown in Fig. 9.7. The I_D recovery is similar to I_{CP} although it depends less on the temperature. This can be explained by looking at I_D degradation

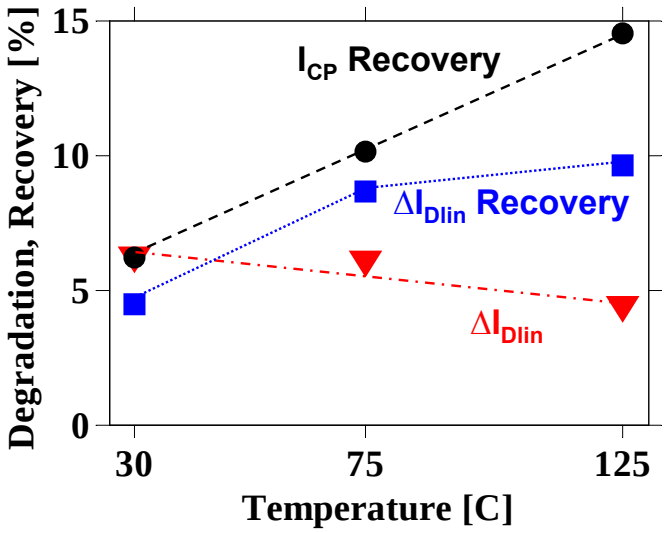


Fig. 9.7. $I_{CP}(= D_{IT})$ recovery increases with temperature. The drain current also recovers partially, however at higher temperatures, it becomes less dependent on D_{IT}. Mobility reduction with temperature dominates interface trap induced degradation and recovery. L=0.16μm MOSFETs were stressed at $V_G = V_D = 2V$ for 1000 seconds. $I_{D,lin}$ was measured at $V_G = 1.8V$ and $V_D = 0.1V$. Measurement delay is 50 seconds.

which decreases with temperature although the D_{IT} increases as in Fig. 9.6. When the temperature increases, the I_D of a MOSFET degrades (t=0, before HCI stress) due to mobility reduction with phonon scattering. Any impact of HCI degradation is therefore smaller at higher temperature since scattering is more effective than the defects. Thus the recovery in I_D due to reduction in D_{IT} is masked by mobility at high temperature. Fig. 9.8 compares percentage $N_{IT}(= D_{IT})$ recovery for low and high temperatures from 2-D R-D theory. Temperature dependence of stress (Fig. 9.6) and measurement delay (Fig. 9.9) are included in the simulations. It can be seen that recovery decreases with increasing temperature. This is because the hydrogen diffusion away from the dangling bonds is faster at high temperature and with 2-D diffusion, it is more difficult for them to anneal the bonds during recovery. Since

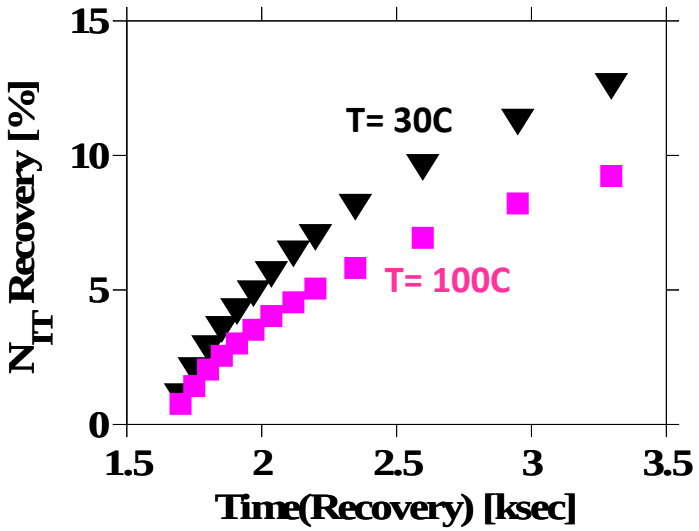


Fig. 9.8. R-D simulations show that recovery increases with decreasing temperature. This trend does not agree with the experimental observations.

diffusion, not annealing rate, is the limiting mechanism, high temperature reduces the recovery.

The recovery obtained from 2-D R-D simulations (Fig. 9.8) contradict the experimental observations (Fig. 9.7). This could be due to the assumption of uniform trap densities in the Si bandgap. The traps might concentrate on a narrow energy range [167]. Since the CP scans a limited energy range in the bandgap at different temperatures, the energy where D_{IT} peaks might not be probed.

Another possibility underlying the discrepancy between R-D simulations and I_{CP} recovery is that the simulations only consider interface traps. Experimental I_{CP}, however, can include both interface and bulk traps, as discussed in the previous chapter. Detrapping from bulk oxide traps is enhanced at higher temperatures. In order to check bulk oxide detrapping, ultra-fast setup was used for recovery. The representative waveforms for V_G and V_D are shown in Fig. 9.10. During recovery, the V_G is kept at negative voltage to induce detrapping. Linear drain current, which reflects both interface and bulk traps, is measured intermittedly. The resulting recovery in

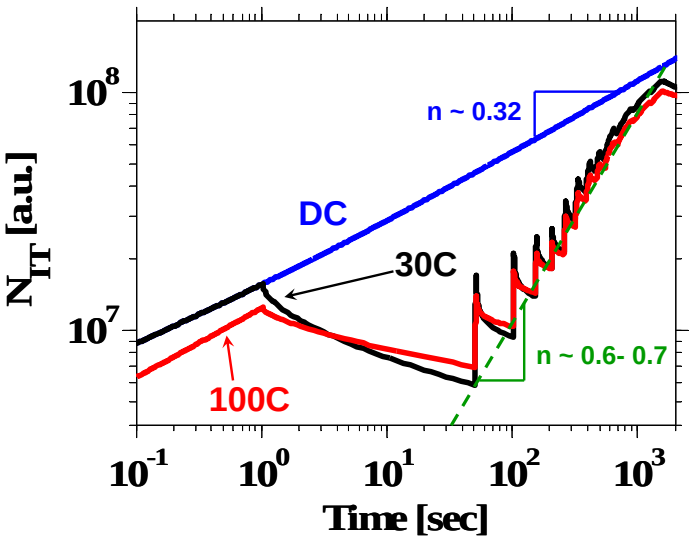


Fig. 9.9. 2-D R-D simulations for DC stress, with and without delay. The delay was assumed to be 30 seconds. The temperature dependence of fractional recovery is independent of delay. H_2 diffusion in R-D is implemented. The time-exponent for DC is around 0.32, with delay n=0.6-0.70.

I_D is given in Fig. 9.11. With a negative V_G, the I_D recovery increases to twice of $V_G = 0V$ for both temperatures. This result suggests that electrons are trapped into the bulk traps during stress. The applied V_G during recovery is less than the flat-band voltage $|V_{FB}|$ of channel region, however near the drain where trapping could take place, the $|V_{FB}|$ is smaller, therefore sufficient for detrapping.

The experimental temperature dependence shows higher interface trap generation and recovery for higher temperatures compared to the conventional HCI with larger stress voltages. The degradation characteristics are consistent with the hot electron distribution functions obtained from SMC and with R-D model with higher k_F. Also the weak temperature dependence suggests kinetic electron-induced dissociation. The experimental recovery trends are not well understood, the R-D theory predicts the opposite temperature dependence. The results could be explained by electron detrap-

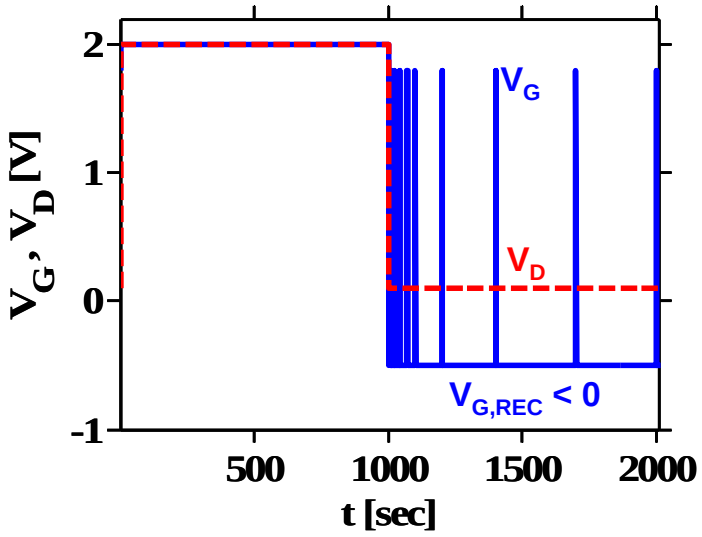


Fig. 9.10. Ultra-fast setup was used to check bulk trapping/detrapping. The NMOSFETs (L=0.16μm) were stressed at $V_G = V_D = 2V$ for 1000 sec and recovered for 1000 sec during which negative V_G was applied. $I_{D,Lin}$ was measured at pre-defined intervals.

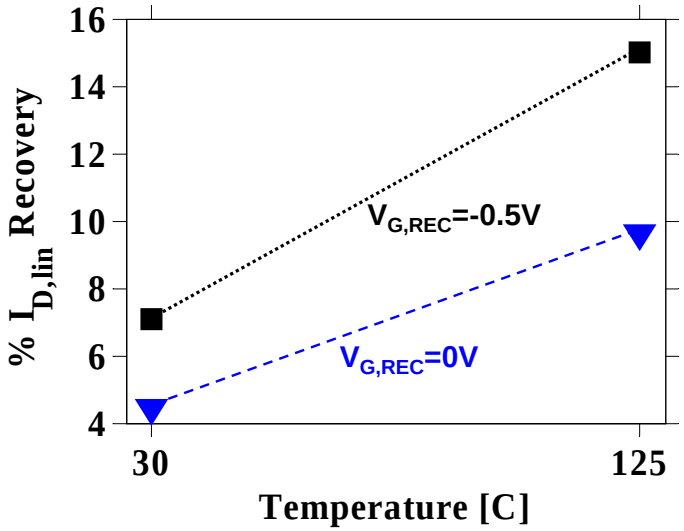


Fig. 9.11. The recovery with negative V_G is higher, implying electron detrapping. The fractional recovery also increases with temperature.

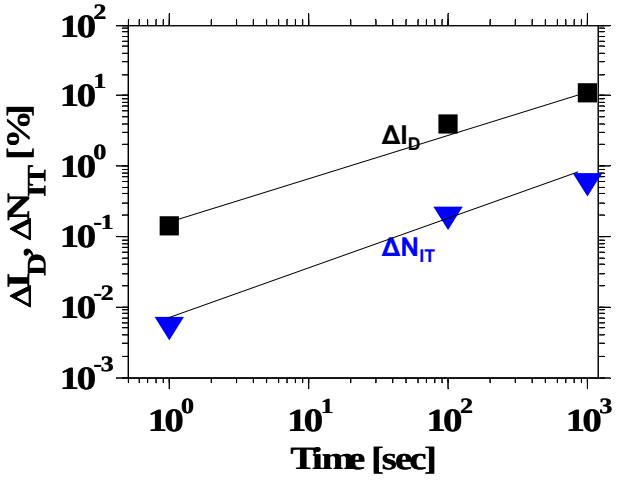


Fig. 9.12. When N_{IT} extracted from experimental data is used in MEDICI, the time-dependences of N_{IT} and I_D are the same. The spatial distribution of N_{IT} is included in the MEDICI simulations.

ping from bulk traps, a process activated by higher temperatures or with non-uniform interface trap density in the Si bandgap.

9.3 Location of HCI damage and screening effect

In the previous sections, the interface trap generation, the resulting charge pumping current and its qualitative impact on drain current degradation were discussed. I_{CP} and I_D show power-law time-dependence, however time-exponents are significantly different as shown in Fig. 9.1. This situation complicates correlating interface damage with MOSFET degradation for lifetime extrapolation. Fig. 9.12 compares the time-dependences of N_{IT} and I_D from MEDICI simulations. The N_{IT}, with its magnitude and spatial distribution, is extracted from experimental data and employed in the simulations. As it can be seen, MEDICI predicts the same time-dependence for both, contradicting the experimental results. This observation implies that the picture is not simple and N_{IT} measurements should be analyzed carefully. The conflicting results can be resolved by looking at the interface trap profiles and understanding

the effects on MOSFET characteristics. First, the experimental techniques such as DCIV and CP that measure spatial N_{IT} profiles are limited because probing deep into the drain region requires higher $|V_G|$ and the doping profiles of MOSFETs may not be convenient for this purpose. From Fig. 8.8 and 8.11, it can be seen that the N_{IT} can extend well into the spacer and drain because the maximum position probed by the measurement still has substantial N_{IT}. This is supported by the hot electron distribution function, as in Fig. 8.12, which peaks underneath the spacer region of the MOSFET. The average energy of an electron traveling towards the drain is $E_{ave}(x) = q \cdot \int_0^x E(x')dx'$, so the energy increases as the electron travels further into the drain and thus it can generate considerable damage in that region.

Second, the interface traps can be screened by the inversion layer carriers. The screening can reduce the effects of mobility degradation and electrostatic repulsion of charges. Therefore, depending on the gate bias, the MOSFET can feel only a fraction of the interface defects.

In this section, a detailed analytical degradation model that includes these two effects will be discussed. The model explains the experimental trends when a dynamic evolution of N_{IT} is present: the degradation starts from the spacers, then shifts into the channel of the MOSFET. The model predicts the voltage and time-dependence of the I_{CP} and I_D under given stress and measurement conditions. The doping, mobility, and N_{IT} profiles were approximated for a compact form that captures the overall degradation behavior.

9.3.1 Inversion layer screening

Trapped charges at interface states can reduce surface mobility due to Coulombic scattering and electrostatically repel channel carriers (i.e., shift V_T), overall reducing the I_D. When the inversion layer is present, screening of carriers can partially mask the effects of interface traps. In the case of HCI, mobility degradation is more

important because the interface defects are localized near the drain and their impact on V_T is less.

In this analytical model, the total channel mobility, μ_T, is given by Matheissen's rule:

$$\frac{1}{\mu_T} = \frac{1}{\mu_{Eff}} + \frac{1}{\mu_C}, \tag{9.3}$$

where μ_{Eff} is the effective channel mobility and μ_C is the Coulombic scattering mobility due to interface traps. The μ_C reduces as the N_{IT} increases and this leads to a decrease in μ_T. The μ_C is written as ([168])

$$\frac{1}{\mu_C(z, z_i, T)} = \frac{m^* q^3 N_{2D}(z_i)}{16\pi\bar{\epsilon}^2\hbar k_B T} \cdot F(z, z_i, T), \tag{9.4}$$

where z is the distance between the $Si/oxide$ interface and mobile inversion layer charges, z_i is the position of traps from the interface, T is the temperature, m^* is the effective electron mass, N_{2D} is the interface trap density, $\bar{\epsilon} = \frac{\epsilon_{Si}+\epsilon_{ox}}{2}$ is the average of Si and oxide dielectric constants, $\hbar$ is Plank's constant and $F(z, z_i, T)$ is the scattering rate [168]. The 2-D scattering rate is obtained from Fermi's Golden Rule and yields (9.5)

$$F(z, z_i, T) = \int_{\alpha=0}^{\pi/2} \left(1 - \frac{q_{sc}^2}{q_{sc}^2 + \frac{8m^* k_B T}{\hbar^2} sin^2\alpha}\right) \cdot exp\left[-2(z - z_i)\sqrt{q_{sc}^2 + \frac{8m^* k_B T}{\hbar^2} sin^2\alpha}\right] d\alpha, \tag{9.5}$$

where α is the scattering angle and q_{sc} is the screening factor obtained from the inversion layer density, N_{inv}, and average depth of the inversion layer, Z_{av}, as in

$$q_{sc} = \sqrt{\frac{q^2 N_{inv}}{\epsilon_{Si} Z_{av} k_B T}}. \tag{9.6}$$

The amount of screening and thus scattering depends on the gate bias since the inversion layer density and Z_{av} changes with voltage. The N_{inv} is approximated as (9.7)

$$N_{inv} = \frac{C_{OX}}{q} \cdot \left[V_G - V_T - 2k_B T \cdot ln(1 + \frac{V_G - V_T}{2k_B T}) - \frac{C_{OX}^2 \cdot (V_G - V_T)^2}{2\epsilon_{Si} \cdot q N_{poly}}\right], \tag{9.7}$$

with C_{OX} as the oxide capacitance, and N_{poly} is the poly-gate doping. The second term in (9.7) is due to the inversion layer capacitance and the third term represents the effect of poly-depletion [169].

The average inversion layer depth, considering the quantum effects, is calculated from

$$Z_{av} = \left[\frac{9\epsilon_{Si}(2\pi\hbar)^2}{16\pi^2 m^* q Q_{tot}}\right]^{1/3}, \tag{9.8}$$

where Q_{tot} is a sum of depletion (Q_D) and inversion layer charges ($Q_{inv} = qN_{inv}$)

$$Q_{tot} = \left(Q_D + \frac{11}{32} qN_{inv}\right), \tag{9.9}$$

and $Q_D = C_{OX} \cdot \left(V_T - V_{FB} - 2\frac{k_B T}{q} ln(\frac{N_A}{n_i})\right)$ [169].

The effective mobility is taken as $\mu_{Eff} = 32500 \cdot E_{eff}^{-1/3}$ where E_{eff} is the effective vertical electric field in the channel, given by $E_{eff} = \frac{Q_D + Q_{inv}/3}{\epsilon_{Si}}$ [169]. The drain current, $I_{D,Lin}$ can be found from (9.10)

$$I_{D,Lin} = \frac{V_D}{R_{Source} + R_{Drain}(time) + R_{Channel}(V_G, time)}, \tag{9.10}$$

The $R_{Channel}$ depends on the V_G and N_{IT} with

$$R_{Channel}(V_G, time) = \left(\frac{W}{L} \cdot \mu_T \cdot Q_{inv}\right)^{-1}. \tag{9.11}$$

The average channel and source/drain doping densities and oxide thickness are obtained from process files of the MOSFETs. The $R_{Source,Drain}$ are extracted from experimental $I_D - V_G$ after calibrating the model without any degradation. The R_{Source} is assumed to be constant over time whereas R_{Drain} increases due to N_{IT}. The interface traps deplete the region underneath the spacer and thus increase the path the electrons flow. The increase in path leads to higher drain resistance. The time-dependence of ΔR_D is taken to be the same as N_{IT} ($=t^{0.5}$).

The μ_T contains the voltage and N_{IT} (from (9.3)) dependences. The Q_{inv} also has a voltage dependence from (9.7). When the measurement gate voltage, $V_{G,m}$, is low, but $> V_T$, the inversion layer density and its screening of interface traps are smaller [170]. Therefore, the μ_C is also low as in Fig. 9.13. The total mobility, from (9.3), is reduced due to μ_C ($\mu_T < min(\mu_{Eff}, \mu_C)$). At low V_G, total degradation is less sensitive to R_{Drain} because the $R_{Channel}$ dominates. As the V_G is increased, the screening increases. Although the μ_{Eff} decreases due to surface scattering, the overall

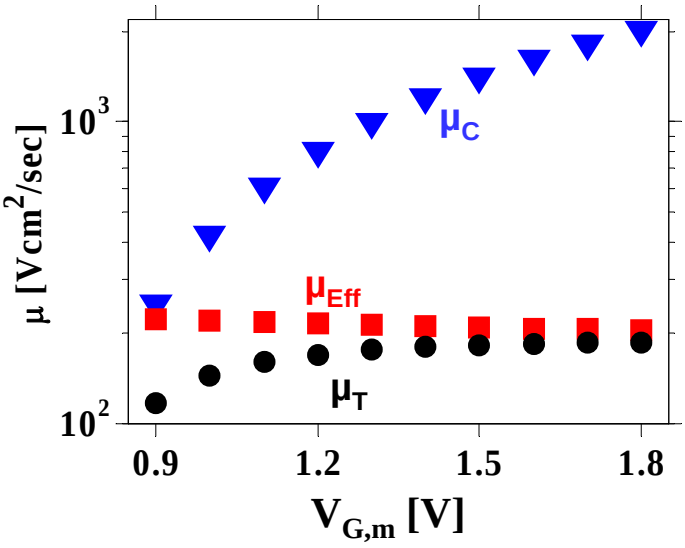


Fig. 9.13. At low measurement voltage, $V_{G,m}$, the reduction in total mobility is higher since the screening of inversion layer is less. As $V_{G,m}$ increases, so does screening and drain resistance degradation becomes more important compared to channel resistance degradation.

impact of interface traps on the total mobility is less at higher gate voltages. Since the channel resistance at this condition is lower, the MOSFET is affected more due to the degradation in R_{Drain}. Fig. 9.14 shows the degradation in transconductance ($=\frac{dI_D}{dV_G}$) extracted from experimental data. In accordance with the theory, the degradation is higher at lower V_G because of smaller screening. For higher V_G, the trend in μ_{Eff} dominates with a reduced contribution from the interface traps.

When I_D degradation measured at a lower $V_{G,m}$ is considered as in Fig. 9.15, the time-exponent of power-law is higher than the one measured at $V_{G,m} = 1.8V$. This is in good agreement with the $I_{CP}(N_{IT})$ exponent (n=0.56 in Fig. 9.1). This result also confirms that the impact of interface traps are felt more at lower $V_{G,m}$.

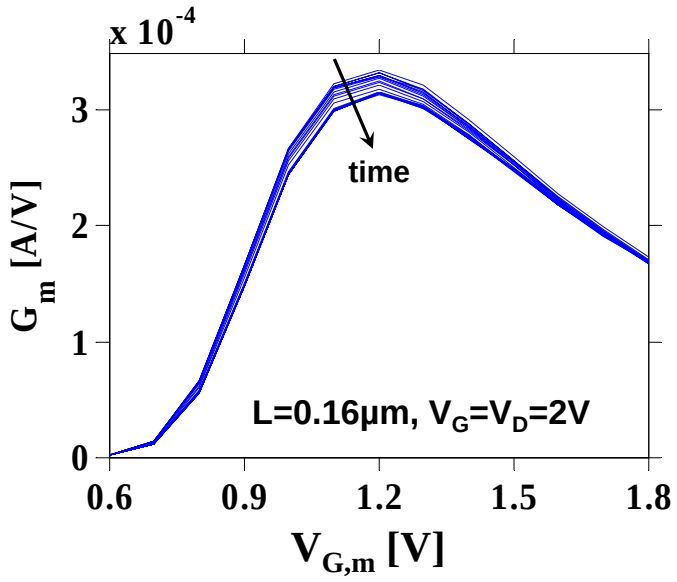


Fig. 9.14. Experimental transconductance degradation is higher at lower measurement $V_{G,m}$ ($\approx$ 1.2V) since inversion layer screening is less.

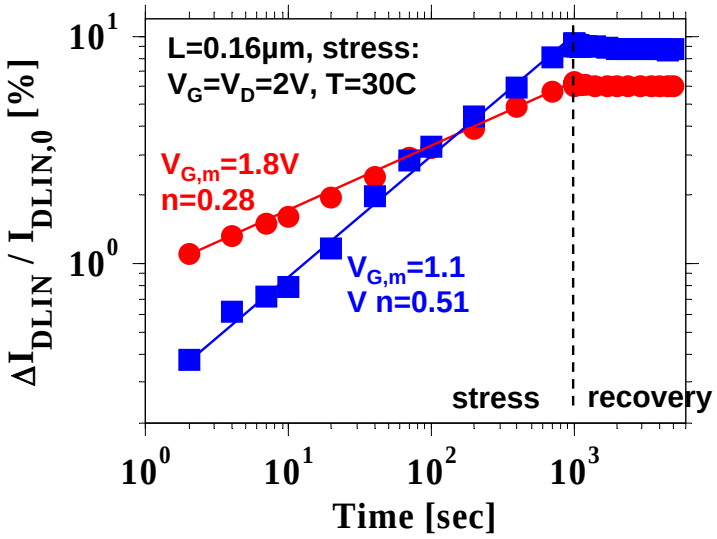


Fig. 9.15. Time-exponent of I_D degradation is higher when measured at lower measurement voltage, $V_{G,m} = 1.1V$. At this measurement voltage the I_D degradation can be correlated to N_{IT} which shows an time-exponent of 0.56.

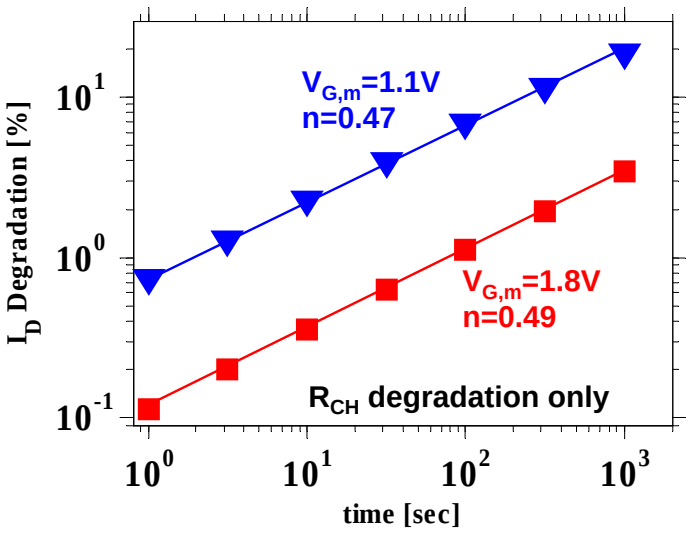


Fig. 9.16. Time-exponents of I_D degradation, obtained from analytical model, are the same for both measurement voltages. At lower $V_{G,m}$, the magnitude of I_D degradation is higher since screening is less.

9.3.2 Location of damage

Although Fig. 9.15 shows that I_D degradation is more sensitive to interface traps at low $V_{G,m}$, it is not clear why measurements at different V_G have different time-exponents, i.e., n=0.28 at $V_{G,m}$=1.1V and n=0.51 at $V_{G,m}$=1.8V. Fig. 9.16 compares two scenarios, simulated by the analytical model. The interface trap generation was assumed to take place only in the channel. In other words, the interface traps affect only the $R_{Channel}$ in (9.10); the $R_{Drain}(t) = R_{Drain}(0)$ is taken as a constant. The magnitude of the degradation for lower V_G is higher but the time-exponents, which are close to the N_{IT} exponent (n=0.5), are nearly the same for the two voltages. This problem of $R_{Channel}$-only degradation not explaining the experimental observation can be solved if the location of HCI damage is considered. As discussed earlier, there can be substantial interface trap generation underneath the spacer, outside the channel region. However, the degradation in drain resistance does not strongly depend on gate bias ($V_{G,m}$), so the time-exponents and magnitudes of I_D degradation

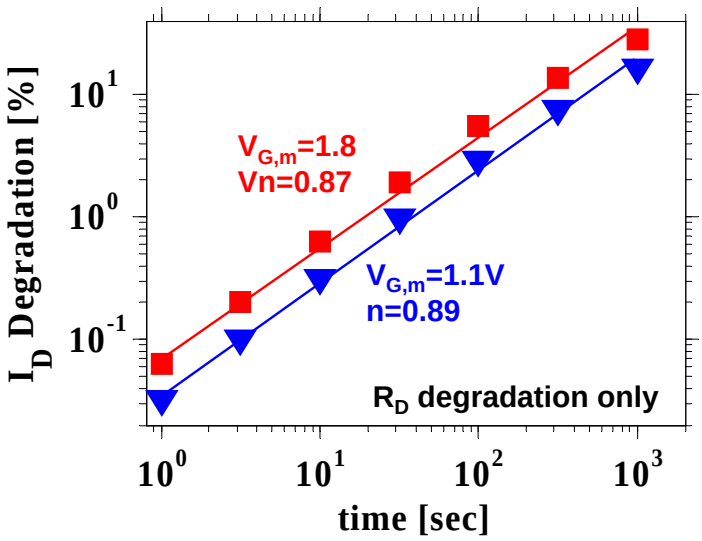
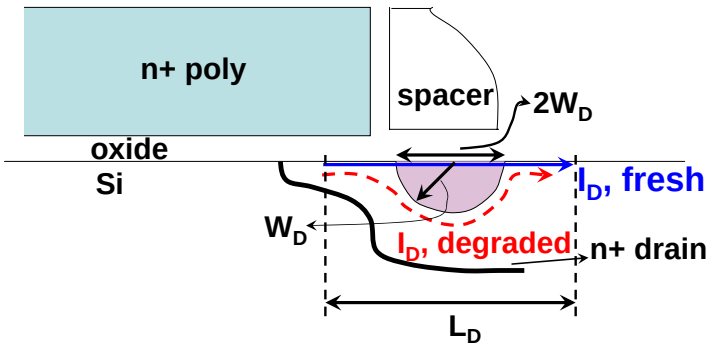


Fig. 9.17. Degradation in drain resistance is not sensitive to measurement $V_{G,m}$. Channel degradation is neglected in analytical model. Time-exponents of I_D degradation are also higher than that of N_{IT} (n=0.5).

at two measurement voltages are nearly the same as in Fig. 9.17 if only R_{Drain} is considered. The degradation was taken to be only underneath the spacers, the channel resistance in (9.10) is assumed to be constant. The generation of interface traps underneath the spacer region leads to depletion having width W_D. The depletion region is approximated to have a semi-circular shape as in Fig. 9.18. The W_D increases with time assuming $q\pi W_D^2(t)ZN_D = qN_{IT}(t)2W_DZ$ where Z is the width of the MOSFET and N_D is the drain doping. The channel electron have to follow a longer path around the depleted region, therefore the drain resistance increases. The time-exponents are higher than that of N_{IT} (n=0.5) since ΔR_{Drain} is significantly smaller than $R_{Drain}(t=0)$ and $R_{Drain} = R_{Drain}(t=0)+\Delta R_{Drain}$ changes appreciably only at later times in Fig. 9.19. Therefore only R_{Drain} is not satisfactory to explain the experimental trends and this suggests that both R_{Drain} and $R_{Channel}$ variations should be considered simultaneously.



$$\frac{\Delta R_D(t)}{R_D} = \frac{\pi W_D(t)}{L_D + (\pi - 2)\cdot W_D(t)}$$

Fig. 9.18. The current follows a straight path before stress. The depletion lengthen the path of the electrons, thus increases the drain resistance over time. Semi-circular depletion region is assumed and L_D is extracted from $R_{Drain}(t = 0)$.

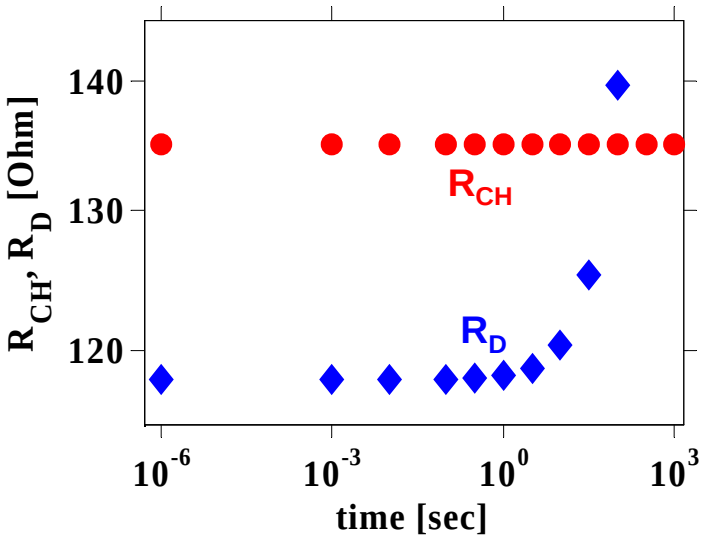


Fig. 9.19. Drain resistance increases more at later times since $\Delta R_{Drain} < R_{Drain}(t = 0)$. After considerable change in R_{Drain} $(t > 1sec.)$, the I_D degradation increases steeper.

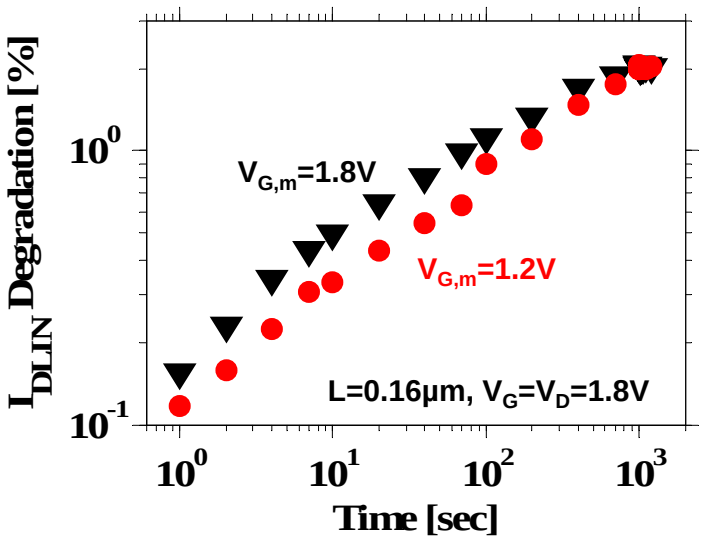


Fig. 9.20. Degradation measured at two voltages are nearly the same for stress under $V_G = V_D = 1.8V$ with L=0.16μm.

Fig. 9.20 and 9.21 show experimental degradation and I_{CP} characteristics of a MOSFET stressed at $V_G = V_D = 1.8V$. The CP measurement voltages were chosen such that only the channel region is probed. The MOSFET is stressed at lower voltages to slow down the dynamic evolution of degradation. It is observed that although the interface traps in the channel do not increase, the degradation in I_D is still present and the time-exponents measured at two voltages are nearly the same. This can be explained by the help of Fig. 9.17: the degradation (Fig. 9.20) is mainly underneath the drain at this stress condition. The CP (Fig. 9.21) begins to increase later in stress which is a sign of channel degradation.

9.3.3 Dynamic evolution of damage region

In the literature, it has been proposed that the degradation starts underneath the spacer, and slows down as schematically shown in Fig. 9.22 [140], [141]. During the saturation of R_{Drain} degradation, channel resistance begins to increase due to interface traps. Using the analytical model, this N_{IT} evolution is implemented. The

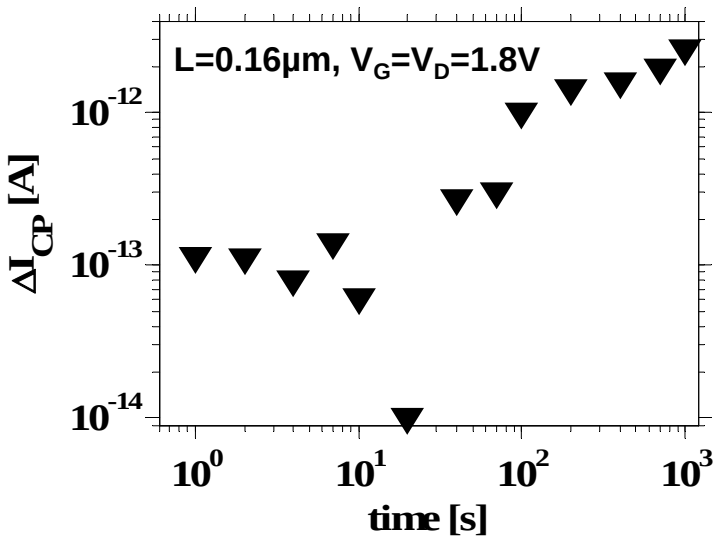


Fig. 9.21. The I_{CP} does not change significantly during stress. The CP voltages were adjusted to scan only the channel. The N_{IT} begins to increase at later stages ($t > 100sec$).

channel and drain resistances change as shown in Fig. 9.23: the drain resistance increases first, then saturates. The interface trap build-up in the channel starts at later stage. The I_D degradation as a function of measurement V_G is shown in Fig. 9.24. At earlier times where R_{Drain} is dominant, the degradation increases with V_G for a particular time. This is because the $R_{Channel}$ decreases with V_G and degradation in R_{Drain} becomes more effective on I_D. At later stages, the drain resistance saturates and channel resistance begins to increase. The $R_{Channel}$ strongly depends on the gate bias because of the screening. As a result, the I_D degradation is more pronounced at lower measurement V_G. Under this N_{IT} evolution, the overall time-dependence of I_D degradation is shown in Fig. 9.25. Although approximate, the analytical model captures the time-dependence and magnitude trends of experimental degradation. The measurement at lower $V_{G,m}$ predicts a higher time-exponent than $V_{G,m} = 1.8V$. The time-exponents of both cases are lower than that of N_{IT}, but this can be due to the initial extraction of drain and channel resistances, and the specific increases in these components.

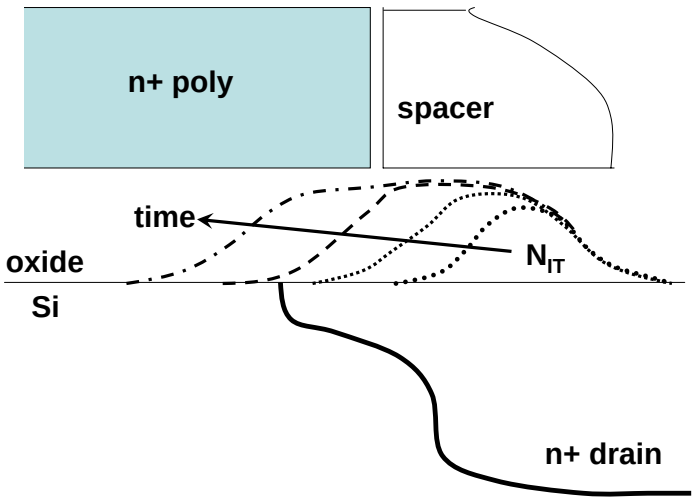


Fig. 9.22. Degradation initially is underneath the spacer, then moves into the channel over time. The degradation in the spacer region saturates at later times.

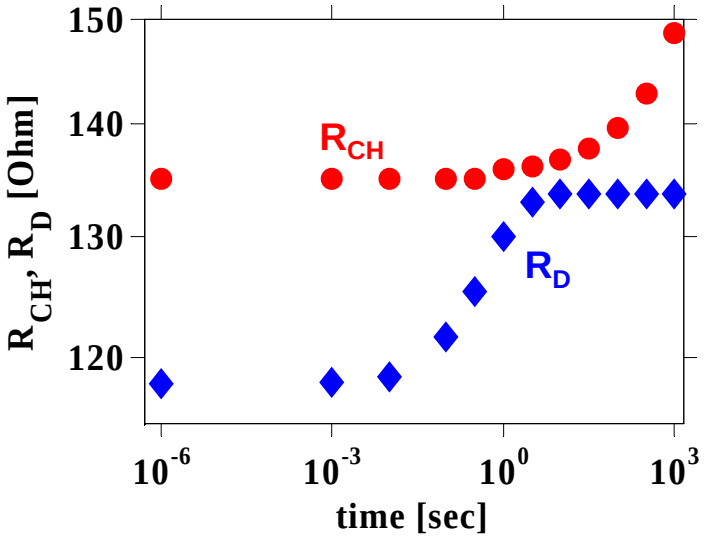


Fig. 9.23. Dynamic evolution of interface damage is modeled analytically. The degradation is initiated in the drain, then spreads into the channel region. The interface trap build-up in both regions is assumed to follow the experimental N_{IT} time-exponent.

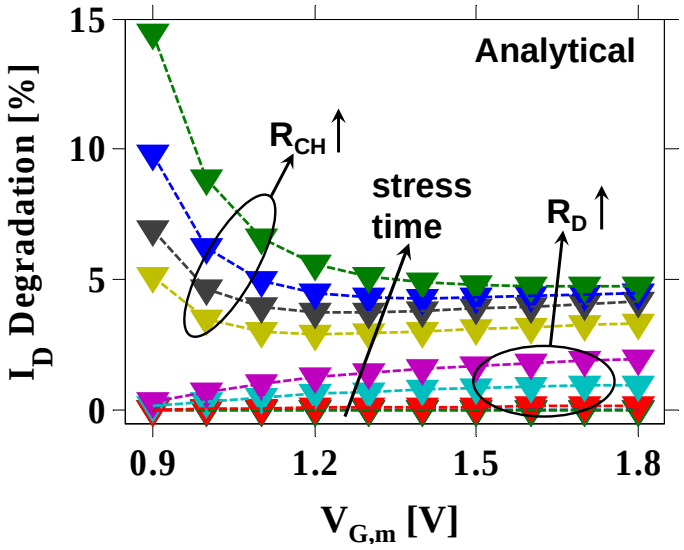


Fig. 9.24. The voltage dependence of I_D degradation switches over time. Initially, the drain resistance is dominant: increases with $V_{G,m}$. Later the channel resistance overtakes and measured degradation decreases with increasing $V_{G,m}$.

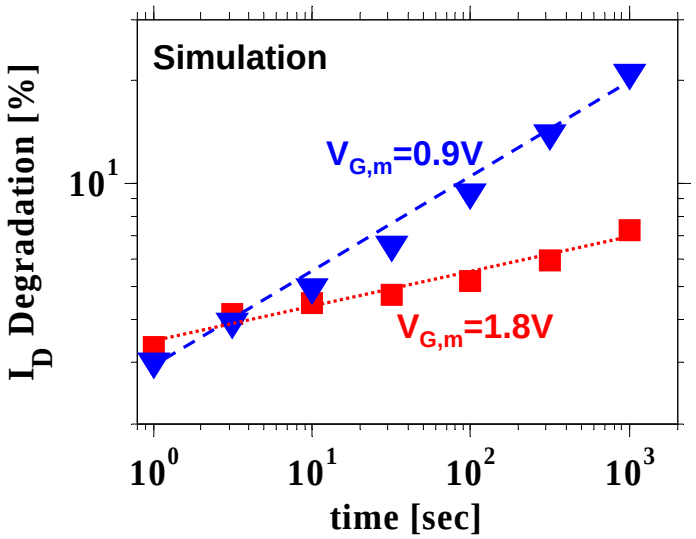


Fig. 9.25. Simulation results of I_D degradation measured at two voltages. The screening and dynamics of degradation location results in different time-dependences.

The I_D degradation is correlated to interface traps at lower measurement V_G and high V_G involves screening of inversion layer carriers. The analytical model captures experimental degradation trend only by means of the aforementioned spatial evolution of interface traps. The results confirm that the degradation first starts underneath the spacer, then evolves towards to channel. The dynamic behavior leads to unique time-dependence which cannot explained by considering drain or channel degradation individually.

9.4 Summary

The time-dependence of interface trap generation and its impact on MOSFET I_D are investigated by stress and recovery characteristics. In order to obtain the intrinsic time-dependence, a set of recovery, temperature dependence, and voltage dependence experiments were performed. It was seen that interface traps and I_D degradation can be correlated if the latter measurements are done at lower voltages because the screening of inversion layer carriers are supressed. The location of HCI damage, and it dynamic evolution, also affects the time-dependence. This dynamic behavior from spacer to channel degradation is captured in an analytical model.

10. RECOMMENDATIONS

In this thesis, the time-dependence of NBTI and HCI degradation mechanism were discussed. It is desired that this work contributes to the present knowledge and provide insights for future studies. From the context of this work, several topics of interest can be found. This section aims to suggest some of these directions for future research.

The time-dependence of degradation is a consequence of hydrogen diffusion in the R-D model. On the other hand, the voltage and temperature dependences are encapsulated in the phenomenological parameters, $k_{F,R}$ and $D_{hydrogen}$. The physical mechanism are not exactly known or characterized but rather represented as reaction rates. Particularly for HCI, the parameter k_F needs to be investigated, e.g., how the energetic carriers can break chemical bonds. The voltage and possible temperature dependences can be quantified better if a deeper understanding can be obtained.

Besides degradation, recovery is another puzzling phenomena. Experimental and theoretical recovery trends do not agree well for both NBTI and HCI. With the ultra-fast setup, it was observed that both mechanisms have $log(time)$ behavior. Although the fractional recovery can be captured by theory, the time-dependence is not well understood. Mechanisms on top of hydrogen passivation can be studied to resolve this issue. Furthermore, for HCI, the temperature dependence is the opposite of experimental results. The theory needs to account for this as well.

Due to scaling requirements, the technology is evolving faster into MOSFETs with strain and high-k dielectrics. The strain generally improves the MOSFET characteristics however it can also affect the damage mechanism through k_F, thus change the magnitude of degradation. Therefore an optimization of performance vs. reliability is possible. In terms of high-k dielectrics, due the presence of interfacial oxide layer, the NBTI and HCI mechanisms are likely to be dominated by interface traps.

Nevertheless, the trapping/detrapping can add up to the interface degradation, so modifications to the R-D framework might be needed.

Scaling also shrinks MOSFET dimensions, i.e., channel length and width. Given the active area of such ultra-scaled devices, there can be few interface traps in a MOSFET. The theoretical approach outlined in this thesis, i.e., characterized by rate and diffusion equations cannot be applicable for individual MOSFETs. Statistical behavior due to finite number of defects and their impact on device behavior can be studied by Monte Carlo approach. This method can be useful for both degradation and recovery.

The geometry-dependent R-D model discusses the implications of the standard theory. According to the geometry dependence, the MOSFET dimensions (particularly width) need to be scaled shorter than 20nm. Because the current technology MOSFETs are still larger than those, experimental evidence for this effect does not yet exist. As the scaling goes on, the geomety effect can be checked experimentally and correlated to the R-D model.

Finally, for circuit-level degradation, combined effects of NBTI and HCI can be considered. Because HCI is prominent during the transitions, the signal shape can directly affect the degradation. NBTI can delay the signal over time, thus shifting the maximum HCI condition. The transitions depend on path the signal follows in the circuit. Therefore accurate modeling needs to consider quasi-static activity, including the degradation in the previous stage. Similarly, recovery trends and its contribution to the signal behavior can be implemented.

11. SUMMARY

MOSFET degradation due to NBTI and HCI are studied for the purpose of developing compact reliability models of VLSI circuits. Starting from the microscopic mechanisms, transistor-level NBTI models are implemented by using the Reaction-Diffusion (R-D) framework. Stress, recovery, AC stress, saturation dynamics, the role of diffusing hydrogen species and implication for future generation FETs are discussed. Then, compact models for MOSFET and circuit degradation are presented.

The investigation of HCI degradation is performed using a similar approach to the NBTI case. Experimentally calibrated process, device, and hot carrier simulation tools work in tandem to obtain HCI-specific degradation metrics. The HCI for short channel MOSFETs stressed at lower voltages is characterized. The conflicted issues in the literature are summarized and resolved by joint use of powerful experimental and theoretical studies. After identifying degradation mechanisms, the HCI damage is correlated to interface traps at those operating conditions. The interface trap behavior for short channel MOSFETs stressed under realistic conditions can be explained by the expanding the geometry dependence of the R-D model.

Lifetime extrapolation, being strongly dependent on time-dependence of degradation, of MOSFETs and circuits can be facilitated by the Reaction-Diffusion Framework for both NBTI and HCI based on the role of interface trap generation. The time-dependence can be affected by a myriad of physical processes during stress and recovery. These additional mechanisms, explored through experiments and theory, are incorporated into the R-D model. Present day CMOS technologies with given operating conditions can possibly benefit from such an approach which is suitable for integrating reliability into the device and circuit development stages.

LIST OF REFERENCES

LIST OF REFERENCES

[1] Semiconductor Industry Association (SIA), The International Technology Roadmap for Semiconductors, 2005 Update.

[2] V. Reddy et al., "Impact of negative bias temperature instability on digital circuit reliability," IEEE IRPS Proceedings, pp. 248-254, 2002.

[3] A. T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, S. Krishnan, "NBTI impact on transistor and circuit: models, mechanisms and scaling effects [MOSFETs]," IEEE IEDM Technical Digest, pp. 349-352, 2003.

[4] W. Abadeer, W. Ellis, "Behavior of NBTI under AC dynamic circuit conditions," IEEE IRPS Proceedings, p. 17, 2003.

[5] Y. Chen, J. Zhou, S. Tedja, F. Hui, A. S. Oates, "Stress-induced MOSFET mismatch for analog circuits," IEEE Integrated Reliability Workshop Final Report, p. 41, 2001.

[6] M. Agostinelli, S. Lau, S. Pae, P. Marzolf, H. Muthali, S. Jacobs, "PMOS NBTI-induced circuit mismatch in advanced technologies," IEEE IRPS Proceedings, p. 171, 2004.

[7] M.L. Reed, J.D. Plummer, "Chemistry of Si-SiO2 interface trap annealing," JAP 63 p.5776, 1988.

[8] M. A. Alam, S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," Microelectronics Reliability, vol.45(1) p. 71, 2005.

[9] D. K. Schroder, J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," Journal of Applied Physics, 94 (1), p. 1, 2003.

[10] B. E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, J. Electrochem. Soc. vol.114, p.266, 1967.

[11] E. H. Nicollian and J. R. Brews, MOS Physics and Technology, Wiley- Interscience, New York, 1982.

[12] C. E. Blat, E.H. Nicollian, and E. H. Poindexter, Mechanism of negative bias temperature instability, J. Appl. Phys. 69, p. 1712, 1991.

[13] M.A. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETs," IEEE IEDM Technical Digest, p.345, 2003.

[14] Reisinger et al., "Analysis of NBTI degradation and recovery behavior based on ultra-fast V_T measurements," IRPS Proc., pp. 448-453, 2006.

[15] A K.O. Jeppson, C.M. Svensson, "Negative bias of MOS devices at high electric fields and degradation of MNOS devices," Journal of Applied Physics, 48 p.2004, 1977.

[16] S. Mahapatra, P. B. Kumar, M. A. Alam, "A new observation of enhanced bias temperature instability in thin gate oxide p-MOSFETs," IEEE IEDM Technical Digest, p. 337, 2003.

[17] A. Krishnan, S. Chakravarthi, P. Nicollian, V. Reddy, S. Krishnan, "Negative bias temperature instability mechanism: The role of molecular hydrogen," APL, vol. 88, pp.153518:1-3, 2006.

[18] J.P. Campbell, P.M. Lenahan, A.T. Krishnan, S. Krishnan, "NBTI: an atomic scale defect perspective," IEEE IRPS Proc., pp. 442-447, 2006.

[19] S. Chakravarthi, A. Krishnan, V. Reddy, C.F. Machala, S. Krishnan, "A comprehensive framework for predictive modeling of negative bias temperature instability," IEEE IRPS Proceedings, p.273, 2004.

[20] N. Kimizuka, K. Yamaguchi, K. Imai, T. Iizuka, C.T. Liu, R.C. Keller, T. Horiuchi, "NBTI enhancement by nitrogen incorporation into ultrathin gate oxide for 0.10 m gate CMOS generation," VLSI Symposium Technical Digest, p.92, 2000.

[21] S. Rangan, N. Mielke, E.C.C. Yeh, "Universal recovery behavior of negative bias temperature instability," IEEE IEDM Technical Digest, p. 341, 2003.

[22] A. Krishnan et al.,"Material dependence of Hydrogen diffusion:Implications for NBTI degradation," IEEE IEDM Proc., 2005.

[23] S. Mahapatra, B.P. Kumar,T.R. Dalei,D. Saha, M.A. Alam, "Mechanism of negative bias temperature instability in CMOS devices: degradation, recovery and impact of nitrogen," IEEE IEDM Proc. pp.105-108,2004.

[24] S. T. Pantelides, S. N. Rashkeev, R. Buczko, D. M. Fleetwood, R. D. Schrimpf, "Reactions of hydrogen with $Si-SiO_2$ interfaces," IEEE Trans. Nuclear Sci., vol. 47(6), pp. 2262-2268, 2000.

[25] A. Stesmans, "Dissociation kinetics of hydrogen-passivated Pb defects at the (111) Si/SiO_2 interface," Physical Review B, 61, pp. 8393-8403, 2000.

[26] L. Tsetseris, S.T. Pantelides, "Migration, incorporation, and passivation reactions of molecular hydrogen at the $Si-SiO_2$ interface," PRB B 70, pp. 245320-245325, 2004.

[27] N.K. Jha, V.R. Rao, "A new oxide trap-assisted NBTI degradation model," IEEE Elec. Dev. Lett., vol. 26(9), pp. 687-689, 2005.

[28] A.E. Islam, H. Kufluoglu, D. Varghese, M.A. Alam, "Critical analysis of short-term negative bias temperature instability measurements: Explaining the effect of time-zero delay for on-the-fly measurements," Appl. Phys. Lett. vol. 90, pp.90, 083505-(1-3), 2007.

[29] S. Zafar, B.H. Lee, J. Stathis, A. Callegari, T. Ning, "A model for negative bias temperature instability (NBTI) in oxide and high-κ pFETs," Symposium on VLSI Technology Dig., pp.208-209, 2004.

[30] O. Penzin, A. Haggag, W. McMahon, E. Lyumkis, K. Hess, "MOSFET degradation kinetics and its simulation," IEEE Trans. Elec. Devices, vol. 50(6), pp.1445-1450, 2003.

[31] B. Kaczer et al., Kinetics of Negative Bias Temperature Instability, Proc. of Int. Rel. Physics Symposium, p. 381, 2005.

[32] M. Ershov et al., "Dynamic recovery of negative bias temperature instability in p-type metaloxidesemiconductor field-effect transistors," App. Phys. Lett. vol. 83(8), pp.1647-1649, 2003.

[33] B. Kaczer, R. Degraeve, V. Arkhipov, N. Collaert, G. Goreseneken, M. Goodwin, The influence of recovery and temperature on the NBTI power-law exponents, Program Book. of SISC, p. 108, 2004.

[34] D. Varghese, D. Saha, S. Mahapatra, K. Ahmed, F. Nouri, M.A. Alam, "On the dispersive versus Arrhenius temperature activation of NBTI time evolution," To appear in IEEE IEDM Proc., 2005.

[35] C. Shen et al., "A fast measurement technique of MOSFET $I_d - V_g$ characteristics," IEEE Electron Device Lett., vol. 27(1), pp.5557, 2006.

[36] B. C. Paul, K.H. Kang, H. Kufluoglu, M. A. Alam, K. Roy, "Temporal Performance Degradation under NBTI: Estimation and Design for Improved Reliability of Nanoscale Circuits," Design, Automation and Test in Europe (DATE), pp.1-6, 2006.

[37] K. H. Kang, H. Kufluoglu, M. A. Alam, K. Roy, "Efficient Transistor-Level Sizing Technique under Temporal Performance Degradation due to NBTI," to appear in International Conference on Computer Design (ICCD) 2006.

[38] K. Kang, I. J. Chang, B. C. Paul, H. Kufluoglu, M. A. Alam, K. Roy, "Temporal Parametric Yield Model of Nano-scale SRAM Array under Negative Bias Temperature Instability," submitted to IEEE TCAD, in review.

[39] Wang, W. Yang, S., Bhardwaj, S., Vattikonda, R., Vrudhula, S., Liu, F., Cao, Y.,"The Impact of NBTI on the Performance of Combinational and Sequential Circuits," IEEE DAC Proc., pp. 364 - 369, 2007.

[40] K. Kang, K. Kim, A. E. Islam, M. A. Alam, K. Roy, "Characterization and Estimation of Circuit Reliability Degradation under NBTI using On-Line IDDQ Measurement," Proc. Design Automation Conference, 2007.

[41] M. A. Alam and H. Kufluoglu, "On Quasi-Saturation of Negative Bias Temperature Degradation," The 208th Meeting of The Electrochemical Society (ECS), October, 2005.

[42] S. Ogawa and N. Shiono, Generalized diffusion-reaction model for the low-field charge build up instability at the Si-SiO2 interface, Phys. Review B, vol.51, no.7, p. 4218, 1995.

[43] C. J. Nicklaw, Z. Y. Lu, D. M. Fleetwood, "The structure, properties and dynamics of oxygen vacancies in amorphous $Si0_2$," IEEE Trans. Nuc. Sci. 49 (6) pp. 2667-2673, 2002.

[44] L. Tsetseris and S. T. Pantelides, "Migration, incorporation and passivation reactions of molecular hydrogen at the Si-Si02 interface," Physical Review B 70, pp. 245320-1–6, 2004.

[45] McPherson et al., "Complementary model for intrinsic time-dependent dielectric breakdown in Si02 dielectrics," Journal of Applied Physics 88 (9), pp. 5351-5359, 2000.

[46] Tsetseris et al., "Physical mechanisms of negative-bias temperature instability," Applied Physics Letters 86, pp. 142103-1–3, 2005.

[47] H. Kufluoglu and M. Alam, "A Generalized Reaction-Diffusion Model With Explicit $H - H_2$ Dynamics for Negative Bias Temperature Instability (NBTI) Degradation," IEEE TED 54(5), pp.1101-1107, 2007.

[48] J. Godet, A. Pasquarello, "Ab initio study of charged states of H in amorphous SiO_2," Microelectronic Engineering, vol. 80, pp.288-291, 2005.

[49] K. Hess et al., "Simulation of Si-Si02 defect generation in CMOS chips: From atomistic structure to chip failure rates," IEEE IEDM Proceedings, pp. 93-96, 2000.

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

[51] J. Kakalios, R. A. Street, and W.B. Jackson, Stretched-exponential relaxation arising from dispersive diffusion of hydrogen in amorphous silicon, Phys. Rev. Lett., v.59, p. 1037, 1987.

[52] I. A. Shkrob and A.D. Trifunac, Time-resolved EPR of spin-polarized mobile H atoms in amorphous silica: The involvement of small polarons, Phys. Rev. B, v.54, p. 15073, 1996.

[53] H. Kufluoglu and M.A. Alam, "A geometrical unification of the theories of NBTI and HCI time-exponents and its implications for ultra-scaled planar and surround-gate MOSFETs," IEEE IEDM Technical Digest, p. 113, 2004.

[54] H. Kufluoglu and M.A. Alam, "Theory of interface-trap-induced NBTI degradation for reduced cross section MOSFETs," IEEE Trans. Elec. Devices, vol. 53(5), pp.1120-1130, 2006.

[55] S. Mahapatra, P. Bharath Kumar and M. A. Alam, Investigation and Modeling of Interface and Bulk Trap Generation during Negative Bias Temperature Instability of p-MOSFETs, IEEE Trans. Electron Devices, v.51, p.1371, 2004.

[56] S. Mahapatra, unpublished results.

[57] Kumar, S.V., Kim, C.H., Sapatnekar, S.S., "An Analytical Model for Negative Bias Temperature Instability," IEEE ICCAD 2006., pp.493-496, 2006.

[58] Fernandez, R., Kaczer, B., Nackaerts, A., Demuynck, S., Rodriguez, R., Nafria, M., Groeseneken, G., " AC NBTI studied in the 1Hz-2GHz range on dedicated on-chip CMOS circuits," IEDM Proc., pp.1-4 , 2006.

[59] H. Kufluoglu, C. Prasad, M. Agostinelli, "History dependent recovery of NBTI under alternating DC and AC stress," Proc. of International Reliability Physics Symposium (IRPS), pp.690-691, 2007.

[60] Grasser, T., Gos, W., Sverdlov, V., Kaczer, B., "The Universality of NBTI Relaxation and its Implications for Modeling and Characterization," Proc. of International Reliability Physics Symposium (IRPS), pp.268-280, 2007.

[61] V. Huard and M. Denais, Hole trapping effect on methodology for DC and AC negative bias temperature instability measurement in PMOS transistors, IRPS Proc., p.40, 2004.

[62] Parthasarathy, C.R., Denais, M., Huard, V., Ribes, G., Vincent, E., Bravaix, A., "New Insights into Recovery Characteristics Post NBTI Stress," IRPS Proc., p.471-477, 2006.

[63] Shen, C., Li, M.-F., Foo, C. E., Yang, T., Huang, D. M., Yap, A., Samudra, G. S., Yeo, Y.-C., "Characterization and Physical Origin of Fast Vth Transient in NBTI of pMOSFETs with SiON Dielectric," IEDM Proc., pp.1-4, 2006.

[64] A.E. Islam, E. N. Kumar, H. Das, S. Purawat, V. Maheta, H. Aono, E. Murakami, S. Mahapatra, M.A. Alam, "Theory and Practice of Ultra-fast Measurements for NBTI Degradation: Challenges and Opportunities," to appear in IEDM 2007.

[65] A. Plonka, "Time-Dependent Reactivity of Species in Condensed Media, Lecture Notes in Chemistry," Springer, Berlin, 1986.

[66] A. Stesmans, "Dissociation kinetics of hydrogen-passivated Pb defects at the (111) Si/Si02 interface," Physical Review B, 61, pp. 8393-8403, 2000.

[67] S. Bedard, L. J. Lewis, "Diffusion of hydrogen in crystalline silicon," Physical Review B, 61(15), pp. 9895-9898, 2000.

[68] T. Nigam, "Growth kinetics, electrical characterization and reliability study of sub-5nm gate dielectrics," PhD. Thesis, IMEC, Belgium, p. 30, 1999.

[69] M. Agostinelli et al., "Random charge effects for PMOS NBTI in ultra-small gate area devices," IEEE IRPS Proceedings, pp. 529-532, 2005.

[70] M. A. Alam, B. E. Weir, P. Silverman, "A future of function or failure?" IEEE Circuits and Devices- The Electronics and Photonics Magazine, 18(2), pp.42-48, 2002.

[71] M. Ieong, B. Doris, J. Kedzierski, K. Rim, M. Yang, "Silicon device scaling to the sub-10 nm regime," Science, vol. 306 p.2057, 2004.

[72] Hergenrother et al., "The Vertical Replacement-Gate (VRG) MOSFET: a 50 nm vertical MOSFET with lithography-independent gate length," IEEE IEDM Technical Digest, pp. 75-78, 1999.

[73] R. Li et al., "50 nm vertical surround-gate MOSFET with S-factor of 75 mV/dec," Device Resarch Conference Digest, p. 63, 2001.

[74] L.T. Su, J.E. Chung, D.A. Antoniadis, K.E. Goodson, M.I. Flik, "Measurement and modeling of self-heating in SOI nMOSFETs," IEEE Trans. Electron Devices, 41 (1) p. 69, 1975.

[75] R.J.T. Bunyan, M.J. Uren, J.C. Alderman, W. Eccleston, "Activation of NBTI degradation in SOI PMOS by self-heating," IEEE International SOI Conference, p. 130, 1992.

[76] I. Kizilyalli, J.W. Lyding, K. Hess, "Deuterium post-metal annealing of MOSFETs for improved hot carrier realibility," IEEE Elec. Dev. Lett. vol. 18(3), pp. 81-83, 1997.

[77] M. J. Deen, Z. P. Zuo, "Edge effects in narrow-width MOSFETs," IEEE Trans. Electron Devices, 38 (8) p.1815, 1991.

[78] M. Saitoh, E. Nagata, T. Hiramoto, "Effects of ultra-narrow channel on characteristics of MOSFET memory with silicon nanocrystal floating gates," IEEE IEDM Technical Digest, p.181, 2002.

[79] H. J. Cho et al., "Fin width scaling criteria of body-tied FinFET in sub 50 nm regime," Device Research Conference Digest, p. 209, 2004.

[80] Y. Liu et al., "A highly threshold voltage-controllable 4T FinFET with an 8.5 n m thick Si-fin channel," IEEE Electron Dev. Letters, 25 (7), p. 510, 2004.

[81] F. L. Yang et al., "5 nm gate nanowire FinFET," VLSI Technology Technical Digest, p.196, 2004.

[82] E. J. Nowak et al., "Scaling beyond the 65 nm node with FinFET-DGCMOS," Proceeding of the Custom Integrated Circuits Conference, p. 339, 2003.

[83] M. Lemme et al., "Subthreshold characteristics of p-type triple-gate MOSFETs," Proceedings of European Solid-Statte Device Research Conference, p. 123, 2003.

[84] S. E. Thompson et al., "In search of 'forever,' continued transistor scaling one new material at a time, " IEEE Transactions on Semiconductor Manufacturing , 18 (1) p. 26, 2005.

[85] M. Specht et al., "Novel dual bit tri-gate charge trapping memory devices," IEEE Electron Dev. Letters, 25 (12), p. 810, 2004.

[86] B. S. Doyle et al., "High performance fully-depleted tri-gate CMOS transistors," IEEE Electron Dev. Letters, 24 (4), p.263, 2003.

[87] Y. Bin et al., "FinFET scaling to 10 nm gate length," IEEE IEDM Technical Digest, p.251, 2002.

[88] S. Maeda et al.,"Negative Bias Temperature Instability in Triple-Gate transistors," IEEE IRPS Proceedings, p. 8, 2004.

[89] M. Je, S. Han, I. Kim, H. Shin, "A silicon quantum wire transistor with one-dimensional subband effects," Solid State Electronics 44(12) p.2207, 2000.

[90] H. Majima, Y. Saito, T. Hiramoto, "Impact of quantum mechanical effects on design of nano-scale narrow channel n- and p-type MOSFETs," IEEE IEDM Technical Digest, p 733, 2001.

[91] Y. Cui, Z. Zhong, D. Wang, U.W. Wang, C.M. Lieber, "High performance silicon nanowire field effect transistors," Nano Letters, 3 (2), p. 149, 2003.

[92] O. Sang-Hyun, D. Monroe, J.M. Hergenrother, "Analytic description of short-channel effects in fully-depleted double-gate and cylindrical surrounding-gate MOSFETs," IEEE Electron Dev. Letters, 21 (9), p. 445, 2000.

[93] C. P. Auth, J.D. Plummer, "Scaling theory for cylindrical, fully-depleted, surrounding-gate MOSFETs," IEEE Electron Dev. Letters, 18 (2), p. 74, 1997.

[94] J. Wang, E. Polizzi, M. Lundstrom, "A computational study of ballistic silicon nanowire transistors," IEEE IEDM Technical Digest, p 695. , 2003.

[95] H. Lee, C. Lee, D. Park, Y. Choi,"A study of negative-bias temperature instability of SOI and body-tied FinFETs," IEEE EDL, 26(5), pp.326-328, 2005.

[96] H. Kufluoglu and M.A. Alam, "A computational model of NBTI and hot carrier injection time-exponents for MOSFET reliability," Journal of Computational Electronics, vol. 3(3), pp.165-169, 2004.

[97] D. Lachenal, F. Monsieur, Y. Rey-Tauriac, A. Bravaix, "HCI degradation model based on the diffusion equation including the MVHR model," Microelectronic Engineering, vol.84(9-10), pp. 1921-1924, 2007.

[98] C.G. Van de Walle, W.B. Jackson, "Comment on "Reduction of hot electron degradation in metal oxide semiconductor transistors by deuterium processing" [Appl. Phys. Lett. 68, 2526 (1996)]," Appl. Phys. Lett. vol.69, pp.2441, 1996.

[99] D. Saha, D. Varghese, S. Mahapatra, "Role of Anode Hole Injection and Valence Band Hole Tunneling on Interface Trap Generation During Hot Carrier Injection Stress," IEEE Electron Dev. Lett., vol. 27(7), pp. 585-587, 2006.

[100] S.C. Sun, J.D. Plummer, "Electron mobility in inversion and accumulation layers on thermally oxidized silicon surfaces," IEEE Trans. Electron Devices, vol. 27(8), pp.1497-1508, 1980.

[101] Y.C. Cheng, E.A. Sullivan, "Relative importance of phonon scattering to carrier mobility in Si surface layer at room temperature," J. Appl. Phys. vol.44(8), pp.3619-3625, 1973.

[102] S.M. Sze, "Physics of Semiconductor Devices," J. Wiley and Sons, 2nd Ed., 1981.

[103] T. H. Ning et al., "1 μm MOSFET VLSI Technology: Part IV-Hot electron design constraints," IEEE Trans. Electron Devices, vol.26, pp.346-353, 1979.

[104] C. Hu, S.C. Tam, F. Hsu, P. Ko, T. Chan, K. W. Terrill, "Hot Electron Induced MOSFET Degradation-Model, Monitor and Improvement," J. Solid State Circuits, vol.20(1), pp.295-305, 1985.

[105] I. Chen, J. Y. Choi, T. Chan, C. Hu, "The Effect of Channel Hot-Carrier Stressing on Gate-oxide Integrity in MOSFETs," IEEE Trans Electron Devices, vol.35(12), pp.2253-2258, 1988.

[106] E. Li, E. Rosenbaum, L.F. Register, J. Tao, P. Fang, "Hot Carrier Induced Degradation in Deep Submicron MOSFETs at 100C," IRPS Proc. pp.103-107, 2001.

[107] E. Li, E. Rosenbaum, J. Tao, P. Fang, "Projecting Lifetime of Deep Submicron MOSFETs," IEEE Trans. Electron Devices, vol. 48(4), pp.671-678, 2001.

[108] M.V. Fischetti, S.E. Laux, "Monte Carlo study of sub-band-gap impact ionization in small silicon field-effect transistors," IEDM Proc., pp.305-308, 1995.

[109] A. Abramo, C. Fiegna, F. Venturi, "Hot carrier effects in short MOSFETs at low applied voltages," J. Appl. Phys., vol.80, pp.889-893, 1996.

[110] J.D. Bude, "Gate current by impact ionization feedback in sub-micron MOSFET technologies," VLSI Tech. Digest, pp.101-102, 1995.

[111] L. Selmi, D. Esseni, "A better understanding of substrate enhanced gate current in VLSI MOSFETs and flash cells. II. Physical analysis," IEEE Trans. Electron Devices, vol.46(2), pp.376-382, 1999.

[112] D.S. Ang, T.W.H. Phua, H. Liao, C.H. Ling, "High-energy tail electrons as the mechanism for the worst-case hot-carrier stress degradation of the deep submicrometer N-MOSFET," IEEE Electron Dev. Lett., vol.24(7), pp.469-471, 2003.

[113] M. Mastrapasqua, J. Bude, M. Pinto, L. Manchanda, K.F. Lee, "Temperature and channel-length dependence of impact ionization in p-channel MOSFETs," VLSI Tech. Proc., pp.125-126, 1994.

[114] J.D. Bude, M. Mastrapasqua, "Impact ionization and distribution functions in sub-micron nMOSFET technologies," IEEE Elect. Dev. Lett., vol.16(10), pp.439-441, 1995.

[115] A. Abramo, C. Fiegna, F. Venturi, "Hot carrier effects in short MOSFETs at low applied voltages," IEDM Proc., pp. 301-304, 1995.

[116] C. Hu, "Lucky electron model of channel hot electron emission," IEDM Proc., pp.22-25, 1979.

[117] E. Takeda, A. Shimizu, T. Hagiwara, "Role of hot-hole injection in hot-carrier effects and the small degraded channel region in MOSFETs," IEEE Electron Dev. Lett., vol.4(9), pp.329-331, 1983.

[118] P. Heremans, R. Bellens, G. Groeseneken, H.E. Maes, "Consistent model for the hot carrier degradation in n-channel and p-channel MOSFETs," IEEE Trans. Electron Devices, vol.35(12), pp.2194-2209, 1988.

[119] S.K. Lai, "Interface trap generation in silicon dioxide when electrons are captured by trapped holes," J. Appl. Phys., vol.54(5), pp.2540-2546, 1983.

[120] S. Mahapatra, D. Saha, D. Varghese, P.B. Kumar, "On the Generation and Recovery of Interface Traps in MOSFETs Subjected to NBTI, FN and HCI Stress," IEEE Trans. Electron Devices, vol.53(7), pp.1583-1591, 2006.

[121] B. Eitan, D. Frohman-Bentchowsky, J. Shappir, "Impact ionization at very low voltages in silicon," J. Appl. Phys., vol.53, pp.1244-1247, 1982.

[122] B. Ricco, E. Sangiorgi, D. Cantarelli, "Low voltage hot electron effects in short channel MOSFETs," IEDM Proc., pp.92-95, 1984.

[123] S. Tam, F.C. Hsu, C. Hu, R.S. Muller, "Hot electron currents in very short channel MOSFETs," IEEE Electron Dev. Lett., vol.4(7), pp.249-251, 1983.

[124] K. Hess et al., "An alternative Interpretation of Hot Electron Interface Degradation in NMOSFETs: Isotope Results Irreconcilable with Major Defect Generation by Holes?," IEEE Trans. Electron Devices, vol.46(9), pp.1914-1916, 1999.

[125] Z. Chen et al., "On the Mechanism for Interface Trap Generation in MOS Transistors Due to Channel Hot Carrier Stressing," IEEE Electron Dev. Lett., vol.21(1), pp.24-26, 2000.

[126] K. Cheng et al., "Separation of Hot Carrier Induced Interface Trap Creation and Oxide Charge Trapping In PMOSFETs Studied by Hydrogen-Deuterium Isotope Effects," IEEE Electron Dev. Lett., vol.22(4), pp.188-190, 2001.

[127] K. Hess, I.C. Kizilyalli, J. W. Lyding, "Giant Isotope Effect in Hot Electron Degradation of Metal Oxide Silicon Devices," IEEE Trans. Electron Devices, vol.45(2), pp.406-416, 1998.

[128] D.S. Ang, T.W.H. Phua, H. Liao, C.H. Ling, "High-energy tail electrons as the mechanism for the worst-case hot-carrier stress degradation of the deep submicrometer N-MOSFET," IEEE Electron Dev. Lett., vol.24(7), pp.469-471, 2003.

[129] D.S. Ang, T.W.H. Phua, H. Liao, C.H. Ling, "Evidence for a composite interface State generation mode in the CHE-stressed deep-submicrometer n-MOSFET," IEEE Trans. Electron Devices, vol.51(12), pp.2246-2248, 2004.

[130] P. Avouris et al., "STM induced H atom desorption: Isotope Effects and site selectivity," Chem. Phys. Lett., vol.257, p.18, 1996.

[131] J.W. Lyding, K. Hess, I.C. Kizilyalli, "Reduction of hot electron degradation in MOS transistors by Deuterium Processing," Appl. Phys. Lett., vol.68, pp.2526-2528, 1996.

[132] D. Esseni, J. D. Bude, L. Selmi, "On Interface and Oxide Degradation in VLSI MOSFETs-Part I: Deuterium Effect in CHE Stress Regime," IEEE Trans. Electron Devices, vol.49(2), pp.247-253, 2002.

[133] T.-C. Shen, C. Wang, G. C. Abeln, J. R. Tucker, J. W. Lyding, Ph. Avouris, R. E. Walkup, "Atomic scale desorption through electronic and vibrational excitation mechanisms", Science vol. 268, p.1590, 1995.

[134] Z. Chen, J. Guo, P. Ong, "Evidence for energy coupling from the SiD vibration mode to the SiSi and SiO vibration modes at the SiO2/Si interface," Appl. Phys. Lett., vol.83(11), pp.2151-2153, 2003.

[135] B. Sun, G. A. Shi, S. V. S. Nageswara Rao, M. Stavola, N. H. Tolk, S. K. Dixit, L. C. Feldman, G. Lupke, "Vibrational Lifetimes and Frequency-Gap Law of Hydrogen Bending Modes in Semiconductors," Phys. Rev. Lett. vol. 96, p.035501, 2006

[136] F.C. Hsu, K.Y. Chiu, "Evaluation of LDD MOSFET's based on hot-electron-induced degradation," IEEE Electron Dev. Lett., vol.5(5), pp.162-165, 1984.

[137] F. Driussi, D. Esseni, L. Selmi, F. Piazza, "Damage Generation and Location in n- and p-MOSFETs biased in the Substrate Enhanced Gate Current Regime," IEEE Trans. Electron Devices, vol.49(5), pp.787-794, 2002.

[138] S. Mahapatra, C. Parikh, J. Vasi, V. R. Rao, C. R. Viswanathan, "A Direct Charge Pumping Technique for Spatial Profiling of Hot Carrier Induced Interface and Oxide Traps in MOSFETs," Solid State Electronics, vol.43(5), pp.915-922, 1999.

[139] A. Neugroschel, C. T. Sah, K. M. Han, M. S. Carroll, T. Nishida, J. T. Kavalieros, and Y. Lu, "Direct-current measurements of oxide and interface traps on oxidized silicon," IEEE Trans. Electron Devices, vol. 42(9), pp. 16571662, 1995.

[140] A. Raychaudhuri, M.J. Deen, W.S. Kwan, M.I.H. King, "Features and mechanisms of the saturating hot-carrier degradation in LDD NMOSFETs," IEEE Trans. Electron Devices, vol.43(7), pp.1114-1122, 1996.

[141] D.S. Ang, C.H. Ling, "A unified model for the self-limiting hot carrier degradation in LDD nMOSFETs," IEEE Trans. Electron Devices, vol.45(1), pp.149-159, 1998.

[142] M. Bourcerie, B.S. Doyle, J.C. Marchetaux, J.C. Soret, A. Boudou, "Relaxable damage in hot carrier stressing of nMOS transistors-oxide traps in the near interfacial region of the gate oxide," IEEE Trans. Electron Devices, vol.37(3), pp.708-717, 1990.

[143] E. Takeda, N. Suzuki, "An empirical model for device degradation due to hot-carrier injection," IEEE Electron Dev. Lett., vol.4(4), pp.111-113, 1983.

[144] R. Bellens, G. Groeseneken, P. Heremans, H.E. Maes, "Hot-carrier degradation behavior of N- and P-channel MOSFET's under dynamic operation conditions," IEEE Trans. Electron Devices, vol.41(8), pp.1421-1428, 1994.

[145] T. Wang, L. Chiang, N. Zous, C. Hsu, L. Huang, T. Chao, "A comprehensive study of hot carrier stress-induced drain leakage current degradation in thin-oxide n-MOSFETs," IEEE Trans. Electron Devices, vol.46(9), pp.1877-1882, 1999.

[146] D.S. Ang, "A new insight into the degradation behavior of the LDD N-MOSFET during dynamic hot-carrier stressing," IEEE Electron Dev. Lett., vol.22(11), pp.553-555, 2001.

[147] D.S. Ang, "A reassessment of ac hot-carrier degradation in deep-submicrometer LDD N-MOSFET," IEEE Electron Dev. Lett., vol.24(9), pp.598-600, 2003.

[148] Y. Leblebici, S.M. Kang, " A one-dimensional MOSFET model for simulation of hot-carrier induced device and circuit degradation," Intl. Symp. Circuts and Systems Proc., vol.1 pp.109-112, 1990.

[149] R.H. Tu, E. Rosenbaum, W.Y. Chan, C.C. Li, E. Minami, K. Quader, P.K. Ko, C. Hu, "Berkeley reliability tools-BERT," IEEE TCAD, vol. 12(10), pp. 1524-1534, 1993.

[150] K.N. Quader, P. Fang, J.T. Yue, P.K. Ko, C. Hu, "Hot carrier reliability design rules for translating device degradation to CMOS digital circuit degradation," IEEE Trans. Elec. Devices, vol. 41(5), pp.681-691, 1994.

[151] P. Fang, J. Ta, J.F. Chen, C. Hu, "Design in hot carrier reliability for high performance logic applications," IEEE Custom Integrated Circ. Conf. Proc., pp.525-531, 1998.

[152] Y. Leblebici, "Design considerations for CMOS digital circuits with improved hot carrier reliability," IEEE. J. Solid State Circuits, vol.31(7), pp.1014-1024.

[153] J.F. Chen, B.W. McGaughy, C. Hu, "Statistical variation of NMOSFET hot carrier lifetime and its impact on digital circuit reliability," IEEE IEDM Proc., pp.29-32, 1995.

[154] J. Bude, R. K. Smith, "Phase-space simplex Monte Carlo for Semiconductor Transport," Semiconductor Sci. Tech., vol.9(5S), pp.840-843, 1994.

[155] J. C. Mitros et al., "High-voltage drain extended MOS transistors for 0.18-μm logic CMOS process," IEEE Trans. Electron Devices, vol.48(8), pp.1751-1755, 2001.

[156] A. Chatterjee et al., "Analog integration in a 0.35 μm Cu metal pitch, 0.1 μm gate length, low-power digital CMOS technology," IEDM Proc., pp.211-214, 2001.

[157] J. Bude, T. Iizuka, Y. Kamakura, "Determination of threshold energy for hot electron interface state generation," IEDM Proc., pp.865-868, 1996.

[158] A. Ghetti, J. Bude, C.T. Liu, " Monte Carlo simulation of hot-carrier degradation in scaled MOS transistors for VLSI technology," IEDM Proc., pp.893-896, 1998.

[159] S. Mahapatra, S. Shukuri, J. Bude, "CHISEL flash EEPROM: Performance and scaling," IEEE Trans. Electron Devices, vol.49(7), pp.1296-1301, 2002.

[160] Varghese, D., Kufluoglu, H., Reddy, V., Shichijo, H., Krishnan, S., Alam, M.A., "Universality of Off-State Degradation in Drain Extended NMOS Transistors," IEDM Proc., 2006.

[161] Alam, M.A., Bude, J., Ghetti, A.,"Field acceleration for oxide breakdown-can an accurate anode hole injection model resolve the E vs. 1/E controversy?," IEEE IRPS Proc., pp.21-26, 2000

[162] M. Houssa, P.W. Mertens, M.M. Heyns, "Relation between stress induced leakage current and time-dependent dielectric breakdown in ultra-thin gate oxides," Semiconductor Science and Tech., vol.14, pp.892-896, 1999.

[163] Degraeve, R., Groeseneken, G., Bellens, R., Ogier, J.L., Depas, M., Roussel, P.J., Maes, H.E., "New insights in the relation between electron trap generation and the statistical properties of oxide breakdown," IEEE Trans. Elec. Devices, vol. 45(4), pp.904-911, 1998.

[164] T. Ong, P.K. Ong, C. Hu, "50- gate-Oxide MOSFET's at 77 K," IEEE Trans. Elec. Devices, vol. 34(10), pp.2129-2135, 1987.

[165] G. Groeseneken, H. E. Maes, N. Beltran, R. F. De Keersmaecker, "A Reliable approach to Charge Pumping Measurements in MOS Transistors," IEEE Trans. Electron Devices, vol. 31(1), pp. 42-53, 1984.

[166] Heremans, P., Van den Bosch, G., Bellens, R., Groeseneken, G., Maes, H.E., "Temperature dependence of the channel hot-carrier degradation of n-channel MOSFET's," IEEE Trans. Elec. Devices, vol. 37(4), pp.980-993, 1990.

[167] G. Groeseneken, H. E. Maes, "Basics and Applications for Charge Pumping in Submicron MOSFETs," Proc. Intl. Conf. Microelectronics, vol. 2, pp.581-589, 1997.

[168] Potbhare, S., Goldsman, N., Pennington, G., McGarrity, J.M., Lelis, A., "Characterization of $4H - SiC$ MOSFET Interface Trap Charge Density Using a First Principles Coulomb Scattering Mobility Model and Device Simulation," SISPAD Proc., pp.95-98, 2005.

[169] Taur, Y, Ning, T.H., "Fundamentals of Modern VLSI Devices" Cambridge, 2002.

[170] Chung, J.E., Ko, P.-K., Hu, C., "A model for hot-electron-induced MOSFET linear-current degradation based on mobility reduction due to interface-state generation," IEEE Trans. Elec. Devices, vol. 38(6), pp.1362-1370, 1991.

[171] J. S. Brugler, P. G. A. Jespers, "Charge Pumping in MOS Devices," IEEE Trans. Electron Devices, vol.16(3), pp.297-302, 1969.

[172] S. Mahapatra, C. Parikh, V. R. Rao, C. R. Viswanathan, J. Vasi, "A Comprehensive Study of Hot Carrier Induced Interface and Oxide Trap Distributions in MOSFETs Using a Novel Charge Pumping Technique," IEEE Trans. Electron Devices, vol. 47(1), pp. 171-177, 2000.

[173] P. Heremans, J. Witters, G. Groeseneken, H. E. Maes, "Analysis of the Charge Pumping Technique and Its Application for the Evaluation of MOSFET Degradation," IEEE Trans. Electron Devices, vol.36(7), pp. 1318-1335, 1989.

[174] Paulsen, R.E., White, M.H., "Theory and application of charge pumping for the characterization of Si-SiO2 interface and near-interface oxide traps," IEEE Trans. Electron Devices, vol. 41(7), pp. 1213 - 1216 1994

[175] Y. Maneglia and D. Bauza, "Extraction of slow oxide trap concentration profiles in metaloxidesemiconductor transistors using the charge pumping method," J. Appl. Phys. vol. 79, pp. 41874192, 1996.

[176] J. Maserjian, R. Kaw, J. Collier, "Tunnel injection into gate oxide traps," Annual. Reliability Phys. Symp., pp.26-33, 1975.

[177] A.M. Martirosian, T.P. Ma, "Lateral profiling of interface traps and oxide charge in MOSFET devices: charge pumping versus DCIV," IEEE Trans. Electron Devices, vol.48(10), pp. 2303-2309, 2001.

[178] C. T. Sah, A. Neugroschel, K. M. Han, and J. T. Kavalieros, "Profiling interface traps in MOS transistors by the DC current-voltage method," IEEE Electron Device Lett., vol. 17(2), pp. 7274, 1996.

APPENDIX

A. APPENDIX

A.1 Numerical Implementation of $H - H_2$ Model

To understand the implications of $H - H_2$ dynamics on NBTI degradation, the numerical model that incorporates both hydrogen species, their mutual coupling and the exchange with $Si-H$ bonds at the interface is developed. The diagram in Fig. 2.4 illustrates how the device domain is discretized and the simulations are implemented. At the interface, the N_{IT} term is coupled only to the $N_H^{(0)}$ term according to (2.1). The atomic hydrogen, once free from the interface, can diffuse away or gets converted into H_2 as in (2.8). Similarly, H_2 can also diffuse or dissociates into two Hs. Notice that H_2 cannot be generated directly from the $Si - H$ breaking. Time-dependent Newton-Raphson method is used to construct the Jacobian matrix which is given schematically as

$$\begin{bmatrix} \begin{bmatrix} n_{H_2}^{(M)} & \cdots & \cdot \\ \vdots & \ddots & \vdots \\ \cdot & \cdots & n_{H_2}^{(0)} \end{bmatrix} & \begin{bmatrix} c_H^{(M)} & \cdots & \cdot \\ \vdots & \ddots & \vdots \\ \cdot & \cdots & c_H^{(0)} \end{bmatrix} & \begin{matrix} \cdot \\ \vdots \\ \cdot \end{matrix} \\ \begin{bmatrix} c_{H_2}^{(M)} & \cdots & \cdot \\ \vdots & \ddots & \vdots \\ \cdot & \cdots & c_{H_2}^{(0)} \end{bmatrix} & \begin{bmatrix} n_H^{(M)} & \cdots & \cdot \\ \vdots & \ddots & \vdots \\ \cdot & \cdots & n_H^{(0)} \end{bmatrix} & \begin{matrix} \cdot \\ \vdots \\ a_{BC} \end{matrix} \\ \begin{matrix} \cdots & \cdots & \cdots \end{matrix} & \begin{matrix} \cdots & \cdots & a_{G-A} \end{matrix} & (n_{IT}) \end{bmatrix}$$

The molecular and atomic hydrogen blocks which are represented as n_{H_2} and n_H terms in the matrix constitute the diffusing species. The conversion reactions between the H_2 and H are shown by coupling blocks (c_{H_2} and c_H from (2.14)). The trap generation and annealing reaction in (2.1) is inserted into the a_{G-A} term along with the interface

trap density (n_{IT}) term. Finally, to conserve the hydrogen and trap densities in the simulation, a boundary condition is implemented through the term a_{BC}. The boundary condition is obtained from rate at which hydrogen changes at the $Si/oxide$ interface [96], [54]. Implementing box integration at the interface gives

$$\frac{\Delta}{2}\frac{dN_H}{dt} = J_{inward} - J_{away} \tag{A.1}$$

where Δ is the grid spacing of the discretization in the domain, J_{inward} is the flux of H towards the $Si/oxide$ interface, and J_{away} is the flux in the opposite direction. The J_{inward} is simply the interface trap generation term, dN_{IT}/dt, from breaking of $Si-H$ bonds and generating atomic H. The J_{away} encapsulates the diffusion of the hydrogen species, H and H_2, and their mutual conversion terms from (2.14). Therefore the overall boundary condition can be written as

$$\frac{\Delta}{2}\frac{dN_H}{dt} = \frac{dN_{IT}}{dt} + J_H + 2\cdot J_{H_2} + k_{H_2}N_{H_2} - k_{H_1}N_H^2. \tag{A.2}$$

The J represents the flux of the diffusing hydrogen species, i.e.,

$$J_X = D_X\frac{dX}{dy} \tag{A.3}$$

and the flux of H_2 is multiplied by 2 in (A.2) since a molecule contains two H atoms.

A.1.1 Simulation Parameters

Throughout the simulations, the following parameters were used: $k_F = 10^{-4}sec^{-1}$, $k_R = 8x10^{-9}cm^3/sec$, $k_{H_1} = 10^{-5}cm^3/sec$, $k_{H_2} = 10^{-1}sec^{-1}$, $D_{H_2} = 4.0x10^{-17}cm^2/sec$, $D_H = 2.8x10^{-17}cm^2/sec$, $T = 125^oC$. The time-dependence(time-exponents) of the degradation are independent of the parameters used above, they will only modify the magnitude of the degradation or shift in time. Since we are not fitting any particular experimental data in this work, these parameter values were taken to be estimates of corresponding physical mechanisms.

A.2 Diffusion length analysis

In deriving (4.5-4.7), the hydrogen concentration at the $Si/oxide$ interface, $N_H^{(0)}$, is assumed to be the same everywhere including the edges and corners. In reality, this may not be correct, however the results will only contain terms with minor scaling constants if the exact relations are to be obtained.

A.2.1 Planar MOSFET

Starting from the integral relation for the general case:

$$N_{IT}(t) = \frac{1}{W \cdot L} \int N_H^{(0)}(r) d^3 r. \tag{A.4}$$

The volume occupied by the hydrogen profile contains the rectangular, cylindrical and spherical sections as in (A.5)

$$\begin{aligned} N_{IT}(t) = \frac{1}{W \cdot L} \{ & \int_0^{\lambda_D} N_H^{(0)} (1 - \frac{y}{\sqrt{D_H t}}) W \cdot L \cdot dy + \\ & \frac{1}{2} \int_0^{\lambda_D} N_H^{(0)} (1 - \frac{r}{\sqrt{D_H t}}) 2\pi r \cdot L \cdot dr + \\ & \frac{1}{2} \int_0^{\lambda_D} N_H^{(0)} (1 - \frac{r}{\sqrt{D_H t}}) 2\pi r \cdot W \cdot dr + \\ & \frac{1}{2} \int_0^{\lambda_D} N_H^{(0)} (1 - \frac{r}{\sqrt{D_H t}}) 4\pi r^2 \cdot dr \} \\ = N_H^{(0)} \cdot \{ & \frac{W L \lambda_D}{2} + \frac{\pi \lambda_D^2 (L + W) + \pi \lambda_D^3}{6} \end{aligned} \tag{A.5}$$

Using

$$N_{IT} N_H^{(0)} = \frac{k_F N_0}{k_R}. \tag{A.6}$$

We obtain the following compact form for planar MOSFETs

$$N_{IT}^{PL}(t) = \sqrt{\frac{k_F N_0}{W \cdot L \cdot k_R}} [\frac{W L \cdot (D_H \cdot t)^{1/2}}{2} + \frac{\pi (L + W)(D_H \cdot t) + \pi (D_H \cdot t)^{3/2}}{6}]^{1/2}. \tag{A.7}$$

A.2.2 Triple-gate MOSFET

The diffusion contains 1-D portions for the top and side-gates and 2-D components at the corners of the Si-body

$$N_{IT}(t) = \frac{1}{3W}\{2\int_0^{\lambda_D} N_H^{(0)}(1-\frac{x}{\sqrt{D_H t}})W\cdot dx + \int_0^{\lambda_D} N_H^{(0)}(1-\frac{y}{\sqrt{D_H t}})W\cdot dy + \frac{1}{2}\int_0^{\lambda_D} N_H^{(0)}(1-\frac{r}{\sqrt{D_H t}})2\pi r\cdot dr\} + = N_H^{(0)}\cdot\{\frac{1}{2}\cdot(\lambda_D + \frac{\pi\lambda_D}{9W})\}. \quad (A.8)$$

From (2.6), the substitution gives

$$N_{IT}^{TRI}(t) = \sqrt{\frac{k_F N_0}{2\cdot k_R}}[(D_H\cdot t)^{1/2} + \frac{\pi(D_H\cdot t)}{9W}]^{1/2}. \quad (A.9)$$

A.2.3 Cylindrical MOSFET

The hydrogen profile is written in cylindrical coordinates and the integral becomes

$$N_{IT}(t) = \frac{1}{2\pi R\cdot L}\{\int_R^{R+\lambda_D} N_H^{(0)}(1-\frac{r-R}{\sqrt{D_H t}})2\pi\cdot r dr\} + = N_H^{(0)}\cdot\{\frac{(D_H\cdot t)^{1/2}}{2}\cdot(1+\frac{R}{(D_H\cdot t)^{1/2}})\cdot(2R+(D_H\cdot t)^{1/2}) - \frac{1}{3}[R^2 + R\cdot(R+(D_H\cdot t)^{1/2}) + (R+(D_H\cdot t)^{1/2})^2]\}. \quad (A.10)$$

Similar to the planar case, using (2.6)

$$N_{IT}^{CYL}(t) = \sqrt{\frac{k_F N_0}{R\cdot k_R}}\cdot\{\frac{(D_H\cdot t)^{1/2}}{2}\cdot(1+\frac{R}{(D_H\cdot t)^{1/2}})\cdot(2R+(D_H\cdot t)^{1/2}) - \frac{1}{3}[R^2 + R\cdot(R+(D_H\cdot t)^{1/2}) + (R+(D_H\cdot t)^{1/2})^2]\}^{1/2}. \quad (A.11)$$

A.3 Reaction-Diffusion Model in 2-D and 3-D

The interface trap density and the hydrogen density in the oxide are solved simultaneously through a time-dependent Newton-Raphson iteration scheme [96]. In 1-D,

the Jacobian matrix has only 2 off-diagonal entries. Additional off-diagonal pairs are introduced for 2-D and 3-D solutions when implemented in Cartesian coordinates. Cylindrical and spherical coordinates can also be incorporated with the appropriate modification of the matrix elements. At the gate/oxide interface, infinite diffusion velocity ($D_H = \infty$) is assumed. The boundary condition at the *Si/oxide* interface has to couple the trap density to the hydrogen density correctly and is very crucial for proper convergence [96]. The derivations of the boundary equations for linear, cylindrical and spherical coordinates are given below.

A.3.1 Boundary conditions for 1-D diffusion

The hydrogen diffusion satisfies

$$\frac{\partial N_H}{\partial t} = \frac{\partial F}{\partial x}, \tag{A.12}$$

where the hydrogen flux is given as $F = -D_H \frac{\partial N_H}{\partial x}$. Using box integration at the *Si/oxide* interface as depicted in Fig. A.1:

$$\frac{\partial}{\partial t}\int_0^{h/2} N_H \cdot dx = \int_0^{h/2} \frac{\partial F}{\partial x} \cdot dx \tag{A.13}$$

$$= F(atx = \frac{h}{2}) - F(atx = 0) = \hat{n_1} \cdot F_1 + \hat{n_0} \cdot F_0$$

$$= D_H \frac{\partial N_H}{\partial x} + \frac{dN_{IT}}{dt}. \tag{A.14}$$

Then the boundary equation for 1-D diffusion is obtained by (A.13)

$$\frac{h}{2}\frac{\partial N_H}{\partial t} - D_H \frac{\partial N_H}{\partial x} - \frac{dN_I T}{dt} = 0, \tag{A.15}$$

A.3.2 Boundary conditions for cylindrical coordinates

Writing (A.12) in cylindrical coordinates,

$$\frac{\partial N_H}{\partial t} = \nabla \cdot F = \frac{1}{r}\frac{\partial rF}{\partial r} \tag{A.16}$$

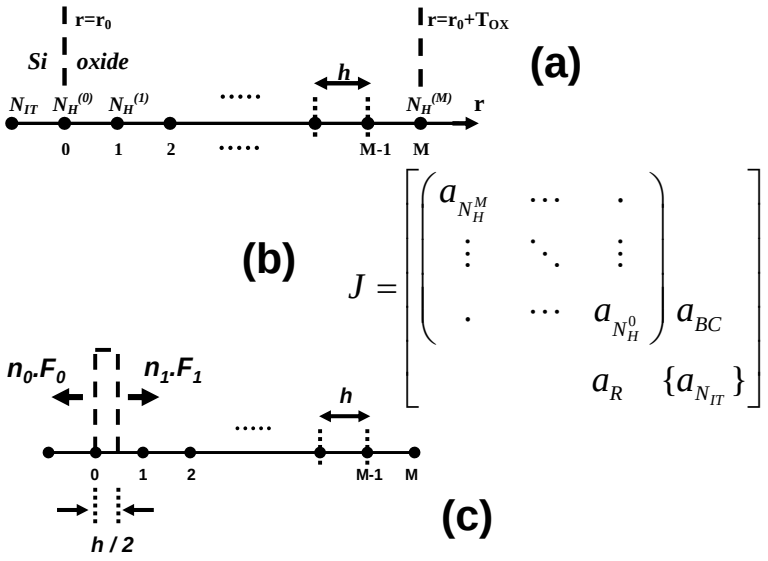


Fig. A.1. (a)The simulation domain of the R-D model is discretized uniformly from the $Si/oxide$ interface to the gate. The first node is the interface trap density term and the other nodes are the hydrogen density in the oxide. (b)The form of the Jacobian matrix used in the time-dependent Newton-Raphson iteration. The matrix contains the diffusion block for hydrogen and the interface trap block. The chemical reaction and the boundary condition terms provide the coupling between the two blocks for simultaneous solution. The hydrogen block is tri-diagonal in 1-D implementation. (c)Box integration is performed at the $Si/oxide$ interface to obtain the boundary equations. The nodes numbered 0-M represent the gate oxide. F_0 and F_1 are the hydrogen fluxes; n_0 and n_1 are the normal vectors at the interface.

and when the box integration is applied

$$\begin{aligned}\int_{r_0}^{r_0+h/2}\frac{\partial N_H}{\partial t}2\pi r\cdot dr &= \int_{r_0}^{r_0+h/2}\frac{1}{r}\frac{\partial(rF)}{\partial r}2\pi r\cdot dr \\ &= \pi(2r_0+\frac{h}{2})\cdot\frac{h}{2}\cdot\frac{\partial N_H}{\partial t} \\ = 2\pi[(r_0+\frac{h}{2})\cdot F(atr=r_0+\frac{h}{2}) - r_0\cdot F(atr=r_0)] & \\ = 2\pi[(r_0+\frac{h}{2})\cdot D_H\frac{\partial N_H}{\partial r} + r_0\cdot\frac{\partial N_{IT}}{\partial t}, & \qquad \text{(A.17)}\end{aligned}$$

and the boundary equation becomes

$$\frac{h}{2}\frac{\partial N_H}{\partial t} - \frac{2(r_0+h/2)}{2r_0+h/2}\cdot[D_H\frac{\partial N_H}{\partial r} - \frac{r_0}{r_0+h/2}\frac{dN_IT}{dt}] = 0 \qquad \text{(A.18)}$$

A.3.3 Boundary conditions for spherical coordinates

The box integration is performed in spherical domain, so

$$\frac{\partial N_H}{\partial t} = \nabla \cdot F = \frac{1}{r^2 sin\Theta}[sin\Theta \frac{\partial r^2 F}{\partial r}] = \frac{1}{r^2}\frac{\partial r^2 F}{\partial r} \tag{A.19}$$

$$\int_{r_0}^{r_0+h/2} \frac{\partial N_H}{\partial t} 4\pi r^2 \cdot dr = \int_{r_0}^{r_0+h/2} \frac{1}{r^2}\frac{\partial (r^2 F)}{\partial r} 4\pi r^2 \cdot dr \tag{A.20}$$

$$= [(r_0 + \frac{h}{2})^2 \cdot F(atr = r_0 + \frac{h}{2}) - r_0^2 \cdot F(atr = r_0)]$$

$$= [(r_0 + \frac{h}{2})^2 \cdot D_H \frac{\partial N_H}{\partial r} + r_0^2 \cdot \frac{\partial N_{IT}}{\partial t}, \tag{A.21}$$

$$\int_{r_0}^{r_0+h/2} \frac{\partial N_H}{\partial t} r^2 \cdot dr = (3r_0^2 + \frac{3hr_0}{2} + \frac{h^2}{4}) \cdot \frac{h}{2}\frac{\partial N_H}{\partial t} \tag{A.22}$$

Combining (A.19) and (A.22) results in

$$\frac{h}{2}\frac{\partial N_H}{\partial t} - \frac{1}{r_0^2 + \frac{hr_0}{2} + \frac{h^2}{12}} \cdot [(r_0^2 + h/2)^2 \cdot D_H \frac{\partial N_H}{\partial r} + r_0^2 \frac{dN_{IT}}{dt}] = 0 \tag{A.23}$$

A.4 Quantities of Bond-breaking and Annealing Reactions

In analytical derivation of the R-D time exponents, dN_{IT}/dt term is assumed to be negligible compared to dissociation and annealing reactions. Numerical simulations solves the coupled equations, there is no assumption. Fig. A.2 shows that the assumption is justifiable for the analytical calculations.

A.5 Charge Pumping

A. Basics of Charge Pumping Technique

Charge pumping technique presents a novel methodology to characterize interface damage after NBTI and HCI. This technique utilizes recombination of electron and holes at interface states and thus monitors the trap density at $Si/oxide$ interface.

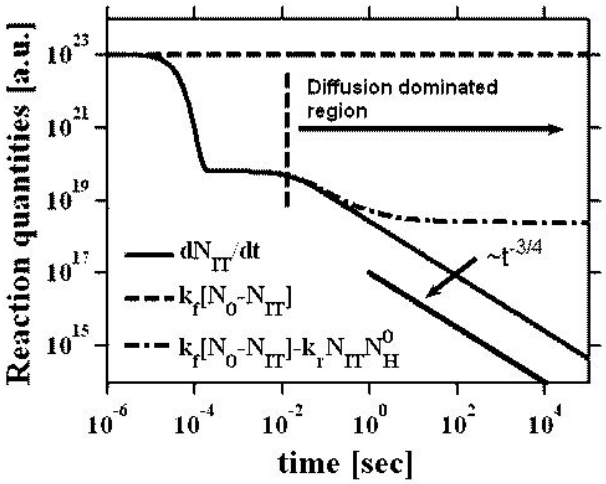


Fig. A.2. The components of (2.2), obtained from the numerical simulation. dN_{IT}/dt term is negligible compared to the bond-breaking and annealing terms in the diffusion-dominated regime.

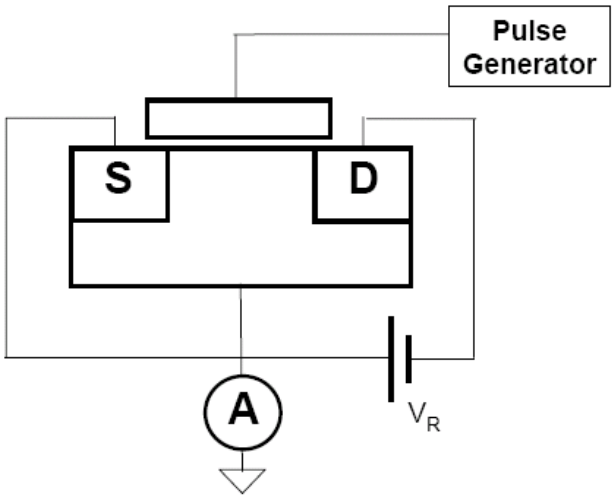


Fig. A.3. Charge pumping measurement setup. The MOSFET is biased in inversion and accumulation and the charge pumping current is monitored at the substrate.

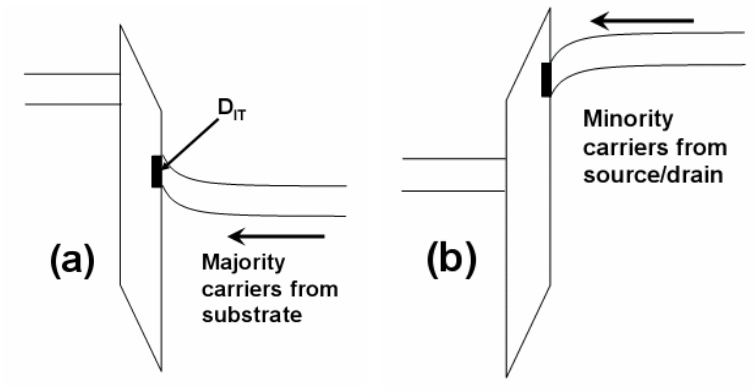


Fig. A.4. (a) Minority carriers from the source and drain fill the interface traps during inversion (NMOS). (b) During accumulation, majority carriers from the substrate recombine with the trapped carriers and contribute to the charge pumping current.

Fig. A.3 shows the experimental setup for the charge pumping measurements. The gate is pulsed so that the MOSFET goes through accumulation and inversion. The source and drain can be reversed biased slightly, with respect to the substrate. When the MOSFET is inverted, the minority carriers injected from the source and drain populates the channel as well as the interface states as in Fig. A.4(a). If the gate is quickly pulsed towards accumulation, these minority carriers can quickly leave the channel, flowing back to the source and drain. However, a fraction of the carriers remain trapped at the interface states if the transition from inversion to accumulation is fast. During accumulation, majority carriers from the substrate flow to the channel and recombine with the trapped minority carriers (Fig. A.4(b)). The charge pumped to the interface traps at each pulse is

$$Q_{cp} \propto q \cdot A_G \cdot N_{IT}, \tag{A.24}$$

where q is the electronic charge and A_G is the gate area. To be more accurate, the N_{IT} term is an integrated quantity; the distribution of the interface states in energy

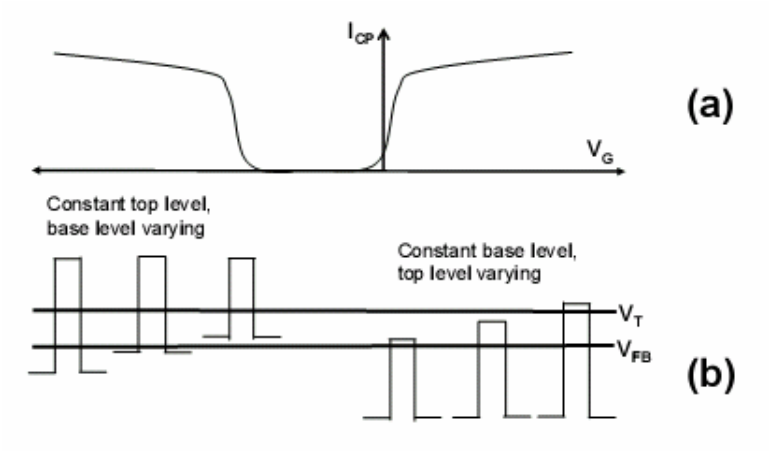


Fig. A.5. (a) Charge pumping current increases as the amplitude increases for a fixed-top-level pulsing. After the entire voltage range is probed, the current saturates. (b) Similar behavior is observed when the base-level fixed and the top level is increased.

need not be uniform, thus $N_{IT} = \int D_{IT}(E)dE$, where D_{IT} is expressed in cm^{-2}/eV. The charge pumping current is then proportional to the interface trap density N_{IT} as in (A.25),

$$I_{cp} \propto q \cdot A_G \cdot f \cdot N_{IT}, \tag{A.25}$$

with f being the pulse frequency. This current can be monitored at the substrate. Alternatively, the source or drain current can be measured instead of the substrate current [171], [172]. Ideally, the two currents are the same in magnitude and reflect the interface trap density. Several variations of the charge pumping measurements can be performed to obtain the interface trap density. The first type is keeping either the low or high level of the gate pulse and varying the other level. Fig. A.5(b) shows the qualitative picture: When the top level of the gate pulse, $V_{G,T}$, exceeds the local threshold voltage (V_T), the channel is inverted. Similarly, accumulation is reached when the low level of the gate pulse, $V_{G,L}$, is less than the flatband voltage (V_{FB}). By keeping the top level fixed and varying the amplitude, the pulse probes

the interface states distributed from V_T to V_{FB}. For low amplitudes, initially, the I_{CP} is small since only a small range of energies for $D_{IT}(E)$ in the Si bandgap is sensed. As the low-level of the pulse passes local V_{FB}, the maximum energy range is probed and the charge pumping current saturates as in Fig. A.5(a). Similarly, the other type of this technique keeps the base-level of the pulse fixed and sweeps the top-level. As the top-level scans the energies from V_{FB} to V_T, the I_{CP} increases and saturates. Theoretically, both fixed-level, variable amplitude measurements saturate at the same level. However, anomalies can be observed experimentally and slight increase in I_{CP} can exist during that saturation region [165]. A third method is to keep the amplitude of the pulse fixed and varying the base-level of the pulse. This method also scans the energies of the interface states, similar to the aforementioned methods. Fig. A.6. When both $V_{G,L}$ and $V_{G,H}$ are outside the range of V_{FB}-V_T, the charge pumping current is zero. As the base-level is increased, the top-level can exceed V_{FB}, thus begins to probe the interface states in the bandgap and I_{CP} increases. When $V_{G,L} < V_{FB}$ and $V_{G,H} > V_T$, all the energy levels in between are pumped so I_{CP} reaches its maximum and saturates. If the base voltage is further increased, then $V_T > V_{G,L} > V_{FB}$ and I_{CP} starts to decrease. Eventually, if $V_{G,L} > V_T$, the pulse is outside the $D_{IT}(E)$ range and I_{CP} again drops to zero.

B. Spatial Profiling of Interface Traps

The methods mentioned in the previous section are suitable for extracting N_{IT} for uniform degradation over the channel. In hot carrier damage, the degradation is often localized near the drain of the MOSFET due to the lateral electric field peaking. This localized damage alters the local V_{FB} and V_T near the drain. In this case, it is of great interest to determine the spatial variation of the interface degradation. However, during HCI, interface trap generation is usually accompanied by charge injection into the oxide and generation of bulk oxide traps (N_{OT}). These defects also change the local threshold and flat-band voltage, making the extraction of spatial

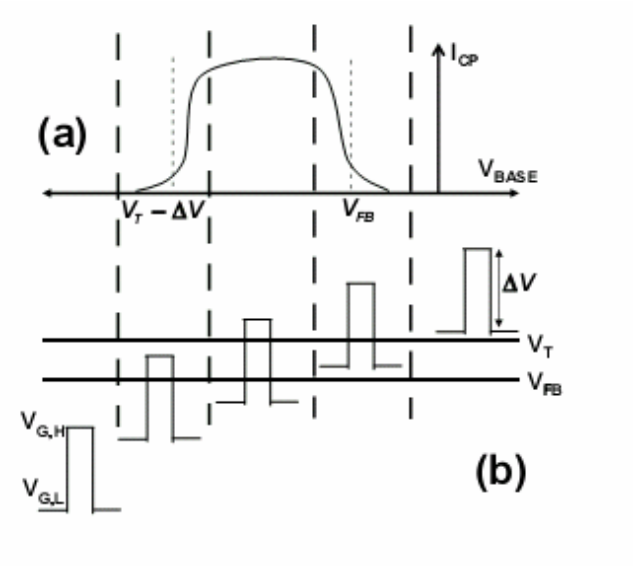


Fig. A.6. (a) Charge pumping current increases initially as more of the interface traps are sensed by increasing base-level, then current saturates when the top level exceeds the local threshold voltage. As the base level increases further and becomes higher than the flatband voltage, the current begins to decrease and after passing the threshold voltage, no current is measured. (b) Fixed amplitude, variable base-level measurement probes the available interface trap range similar to the variable amplitude methods.

interface trap distribution difficult [173]. One of the spatial profiling methods makes use of variable-amplitude technique for both fixed-base and fixed-top level cases in order to cancel the bulk oxide trap contribution [138]. Fig. A.7 schematically shows the experimental procedure in the presence of localized damage near the drain of an NMOSFET. Due to the source and drain diffusions, the local V_{FB} and V_T roll off near the junctions. The hot carrier damage from the interface and bulk traps cause changes in the V_{FB} and V_T near the drain.

For a virgin device, if pulses with fixed top-level and variable amplitude are applied to the gate, then the charge pumping current is proportional to the interface trap

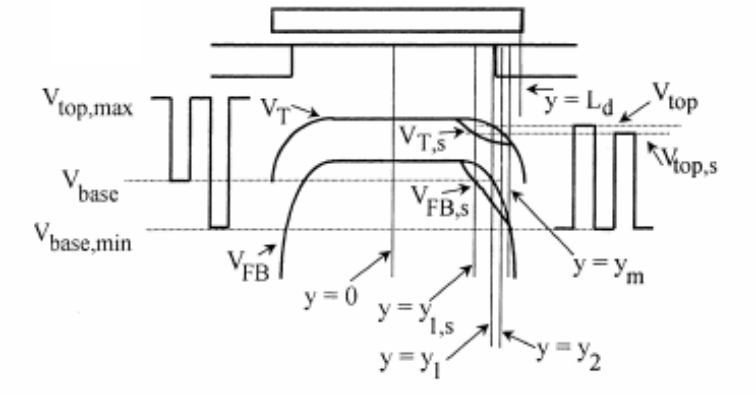


Fig. A.7. For a MOSFET, local threshold and flatband voltages change along the position. Variable-amplitude, fixed top and fixed base-level measurements scan different portions of the channel and spatial interface trap density can be extracted.

density in the channel outside the source and drain depletion regions since V_{FB} is less than $V_{G,L}$ towards the junctions.

$$I_{cp}(V_{base}) = q \cdot W_G \cdot f \cdot \int_{-y_1}^{y_1} N_{IT}(y)dy, \tag{A.26}$$

where W_G is the gate width, y is the distance along the channel and $V_{FB}(y_1) = V_{base}$. Similarly for fixed-base, variable amplitude pulses on an unstressed device, the regions near the source and drain junctions can be probed initially since $V_T > V_{G,H}$ towards the mid-channel; so

$$I_{cp}(V_{top}) = q \cdot W_G \cdot f \cdot [\int_{-y_m}^{-y_1} N_{IT}(y)dy + \int_{y_1}^{y_m} N_{IT}(y)dy], \tag{A.27}$$

with $V_T(y_1) = V_{top}$ and $V_{FB}(y_m) = V_{base,min}$.

After degradation, the local V_{FB} and V_T are changed because of the N_{IT} and N_{OT} near the drain. The charge pumping current of (A.26) becomes

$$I_{cp}^{stressed}(V_{base}) = q \cdot W_G \cdot f \cdot [\int_{-y_1}^{0} N_{IT}(y)dy + \int_{0}^{y_{1,s}} N_{IT,s}(y)dy], \tag{A.28}$$

and (A.29) can be rewritten as

$$I_{cp}^{stressed}(V_{top,s}) = q \cdot W_G \cdot f \cdot [\int_{-y_m}^{-y_2} N_{IT}(y)dy + \int_{y_{1,s}}^{y_m} N_{IT,s}(y)dy], \tag{A.29}$$

where $N_{IT,s}$ is the interface trap density after degradation, $V_{T,s}$ is the local threshold voltage after stress, $V_{FB,s}(y_{1,s}) = V_{base}$, $V_{T,s}(y_{1,s}) = V_{top,s}$, $V_T(y_2) = V_{top,s}$, $V_{FB,s}(y_m) = V_{FB}(y_m) = V_{base,min}$ since it is assumed that V_{FB} distribution is not affected by the interface and oxide traps at y_m deep inside the junction.

The change in base-level charge pumping current is

$$\Delta I_{cp}(V_{base}) = I_{cp}^{stressed}(V_{base}) - I_{cp}(V_{base}) \cong q \cdot W_G \cdot f \cdot \int_0^{y_{1,s}} N_{IT}(y)dy, \tag{A.30}$$

where $\Delta N_{IT}(y)$ is the interface trap density generated during stress. The assumption is that $\int_0^{y_{1,s}} \Delta N_{IT}(y)dy \gg \int_{y_1}^{y_{1,s}} N_{IT}(y)dy$ due to the fact that at a given V_{base}, $y_{1,s} \gg |y_{1,s} - y_1|$ and $\Delta N_{IT}(y) \gg N_{IT}(y)$ for significant damage.

For the top-level charge pumping, similarly,

$$\Delta I_{cp}(V_{top,s}) = I_{cp}^{stressed}(V_{top,s}) - I_{cp}(V_{top,s}) \cong q \cdot W_G \cdot f \cdot \int_{y_{1,s}}^{y_m} \Delta N_{IT}(y)dy. \tag{A.31}$$

Again the assumption is $\int_{y_{1,s}}^{y_m} \Delta N_{IT}(y)dy \gg \int_{y_2}^{y_{1,s}} N_{IT}(y)dy$. From (A.30) and (A.31),

$$\Delta I_{cp}(V_{top,s}) + \Delta I_{cp}(V_{base}) = \Delta I_{cp,max}(V_{top,max}, V_{base,min}) = q \cdot W_G \cdot f \cdot \int_0^{y_m} \Delta N_{IT}(y)dy. \tag{A.32}$$

From (A.32), if ΔI_{CP} and V_{base} are known, $V_{top,s}$ can be extracted after the degradation.

After the degradation at the interface and bulk oxide, the local threshold and flat-band voltages can be written as

$$V_{T,s}(y) = V_T(y) - \frac{q\Delta N_{OT}(y)}{C_{ox}} + \frac{q\Delta N_{IT}(y)}{2C_{ox}} \tag{A.33}$$

and

$$V_{FB,s}(y) = V_{FB}(y) - \frac{q\Delta N_{OT}(y)}{C_{ox}} - \frac{q\Delta N_{IT}(y)}{2C_{ox}}, \tag{A.34}$$

where C_{ox} is the oxide capacitance. Subtracting (A.34) from (A.33) for $y = y_{1,s}$

$$V_{top,s} - V_{base} = V_T(y_{1,s}) - V_{FB,s}(y_{1,s}) + \frac{q\Delta N_{IT}(y_{1,s})}{C_{ox}} \tag{A.35}$$

cancels the bulk oxide trap portion and yields a relationship for the interface trap density. $V_{top,s}$ can be extracted from (A.32), so the left side of (A.35) can be obtained. By measuring pre-stress I_{CP} for identical MOSFETs with different channel lengths (L_{ch}) and plotting I_{CP} vs. L_{ch} gives N_{IT} which then can be used to determine V_{FB} and V_T vs. y. V_{base} vs. $y_{1,s}$ relationship can be deducted from (A.30) and employing (A.35), spatial distribution of the interface traps can be profiled.

In order this method be useful, the V_{FB} and V_T profiles of the MOSFET must be monotonic, i.e., there should not be local peaks or dips in the profiles. Since this strongly depends on the doping profiles, features such as halo can limit this technique.

A.6 Multi-frequency Charge Pumping (MFCP)

This technique is a variation of the conventional charge pumping, however, it is utilized to measure the bulk oxide traps N_{OT} in addition to the interface traps N_{IT}, therefore discussed as a separate method. From (A.25), the charge pumping current, I_{CP} is a function of measurement frequency. The charge pumped per cycle is $Q_{CP} = I_{CP}/f$ is constant according to (A.25), however experimental I_{CP} actually increases with decreasing measurement frequencies [174], [175]. The reason for this is the bulk oxide traps which can contribute to the charge pumping current ($I_{CP,tot} = I_{CP,N_{IT}} + I_{CP,N_{OT}}$) at low frequencies. The bulk traps can communicate with the channel by means of tunneling in and out if sufficient time is allowed during charge pumping pulses.

Assuming first order dynamics, the trapped charge in the bulk oxide defects can be written as

$$\frac{\partial}{\partial t}\rho_{OT}(E,x,t) = -\frac{\rho_{OT}(E,x,t)}{\tau_T(E,x)} \tag{A.36}$$

where ρ_{OT} is the density of trapped charge per unit area (in units of $cm^{-2}eV^{-1}$), E, x, t are energy, distance into the oxide and time, respectively. τ_T is the tunneling time from a trap at energy E and position x. The tunneling time is given by

$$\tau_T(E,x) \approx \tau_0 e^{\alpha_i \cdot x} \tag{A.37}$$

where τ_0 is the tunneling time constant and α_i is attenuation constant for the oxide derived quantum mechanically [176]. The charge coming from bulk traps at a particular frequency is

$$Q_{cp,N_{OT}} = qA_G \int_{x_{min}}^{x_{ox}} N_{OT}(x) \left[1 - exp\left(\frac{-1}{f\tau_T(E,x,t)}\right)\right] dx \tag{A.38}$$

with A_G being the area of the gate and x_{ox} is the oxide thickness. Here, single energy level E_T for the traps is assumed so that $\rho_{OT} = N_{OT}(x)\cdot\delta(E - E_T)$ (N_{OT} in cm^{-3}). The term $(1 - exp\left(\frac{-1}{f\tau_T(E,x,t)}\right))$ is derived from (A.36) and represents the charge detrapped from oxide defects. It can be taken out of the integral by defining the maximum position at which trapped carriers can respond to the pulse frequency from (A.37), i.e.,

$$x_m(f) = \frac{1}{\alpha_i} ln(\frac{1}{f\tau_0}). \tag{A.39}$$

Then the integral becomes

$$Q_{cp,N_{OT}} = qA_G \left[1 - exp\left(\frac{-1}{f\tau_T(E,x,t)}\right)\right] \int_{x_{min}}^{x_m(f)} N_{OT}(x)dx \tag{A.40}$$

where the traps positioned at $x_{min} < x < x_m$ can follow the charge pumping pulses and the traps at $x > x_m$ are empty. Finally, the vertical profiling of the bulk traps can be extracted from

$$\frac{\partial Q_{cp,N_{OT}}}{\partial log(f)} \propto -\frac{2.3A_G}{\alpha_i} N_{OT}(x_m) \tag{A.41}$$

with the help of (A.41). Figures A.8 and A.9 show experimental Q_{CP}(per cycle) vs. $log_e(frequency)$ for two stress conditions, $V_G = V_D = 2V$ and $V_G = V_D = 2.7V$ (presented in Fig. 8.5). The bulk trap density is calculated from the slope of the line for each time step. In the higher stress voltage condition, the slope increases clearly over time. Also, the magnitude of Q_{CP} is significantly higher than the lower stress case.

A.7 Direct Current IV (DCIV) Technique

DCIV method can also be used to profile spatial interface traps distribution in the channel of a MOSFET [139]. Fig. A.10 shows the schematic of the technique on

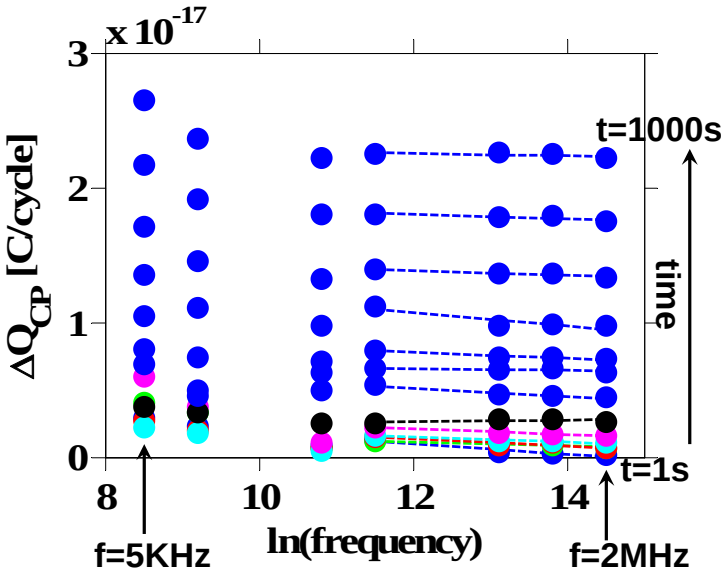


Fig. A.8. MFCP data for an NMOSFET, L=0.16μm, stressed at $V_G = V_D = 2V$. The magnitude of Q_{CP} increases over time however the slope of Q_{CP} vs. $log_e(frequency)$ does not have a clear trend. Therefore the extracted bulk trap generation is negligible. The measurement frequencies are 5, 10, 50, 100, 500 KHz, 1 and 2 MHz.

an NMOSFET. The gate is biased at flat band condition and the source-substrate junction is forward biased to inject minority electrons into the channel. The drain-substrate junction is reverse biased to form a depletion region near the drain. Holes from the substrate can recombine with the electrons provided by the source at the interface traps distributed over the channel. Due to the depletion, the region near the drain prevents electron-hole recombination. As the drain voltage is changed, so does the depletion width, so effectively, the recombination area is modified. As a result, the substrate current, which supplies the holes for recombination, changes with drain voltage. This principle can be used to profile the interface trap distribution in the channel region since the depletion region blocks the drain junction.

The substrate current can be obtained as

$$I_B = qW_G \int_{X_s}^{L-X} N_{IT}(X) \cdot R_S(X, V_{gb}, V_{sb}) dX \tag{A.42}$$

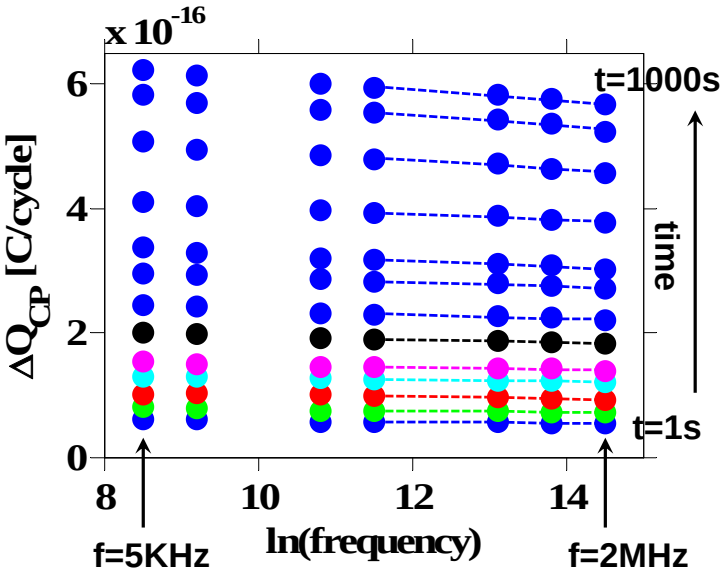


Fig. A.9. MFCP data for an NMOSFET, L=0.16μm, stressed at $V_G = V_D = 2.7V$. The slope of Q_{CP} vs. $log_e(frequency)$ increases over time, signaling N_{OT} generation.

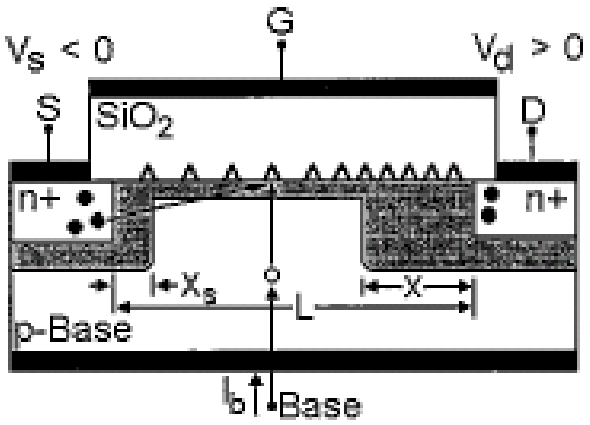


Fig. A.10. By changing the drain bias and thus depletion region width, the channel area over which the electrons from the source and holes from the substrate recombine can be modified and interface trap profiling can be performed. After [177].

where W_G is the channel width, X is the depletion width, L is the metallurgical channel length, R_S is the SRH recombination rate per unit interface trap density [139], V_{gb} and V_{sb} are the gate-substrate and source-substrate voltages [177].

The interface trap density is then obtained by

$$N_{IT}(X) = -(qW_G R_S(X, V_{gb}, V_{sb}))\frac{dV_{db}}{dX}\frac{dI_b}{dV_{db}} \tag{A.43}$$

with V_{db} being the drain-substrate voltage. $\frac{dV_{db}}{dX}$ can be extracted from device simulation or from analytical expressions [178].

A.8 Ultra-fast Measurement Technique

As it is discussed previously, the degradation due to NBTI and HCI can recover when the stress voltage is reduced because the dangling Si bonds can be annealed. The recovery can be captured via experimental techniques such as charge pumping or $I_D - V_G$, however these procedures have inherent delay, i.e., the measurement can take milliseconds to minutes for conventional characterization equipment. Therefore, the measurement results might not be accurate since the intrinsic stress and recovery behavior is masked by the measurement itself.

To circumvent this issue, ultra-fast measurement techniques have been developed [14], [35]. The time resolution of these setups can go down to microseconds and thus minimizes the measurement delay issue. Fig. A.11 shows the experimental setup. The op-amp acts as a current-to-voltage converter. The output voltage of the op-amp can be sampled with high resolution in time by fast acquisition systems, e.g, analog-to-digital converter or digital oscilloscope. Due to the virtual ground principle of the op-amp, the MOSFET current is forced through the amplifying resistor, R_F, and generates the output voltage, $V_{OUT} = -I_D \cdot R_F$. Fig. A.12 shows the behavior of the output voltage during stress and recovery schematically. When the MOSFET is stressed, because of the degradation, the I_D decreases over time. During recovery, the degraded current improves. These trends are repeated over subsequent cycles. The conventional (SMU) and ultrafast techniques are compared in Fig. A.13, Fig. A.14

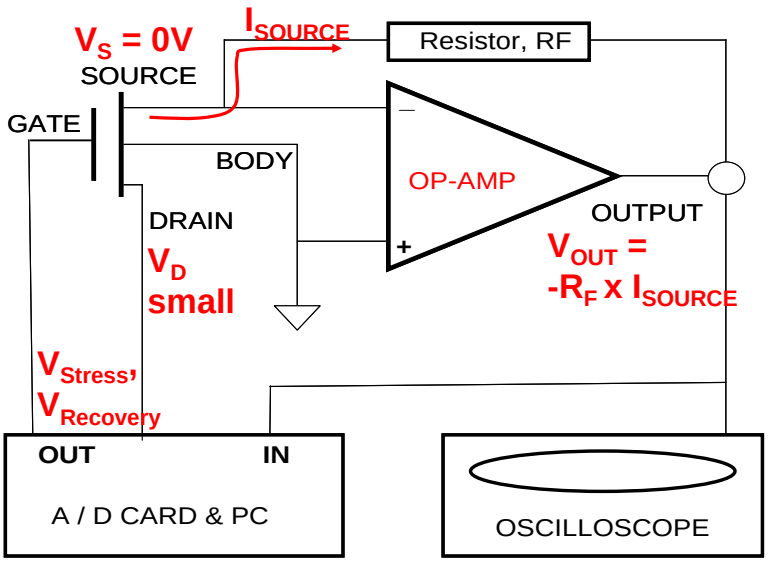


Fig. A.11. Degradation and recovery behavior of a MOSFET can be measured more precisely with the ultra-fast setup. The op-amp converts the MOSFET current to a voltage which can be read with high time-resolution.

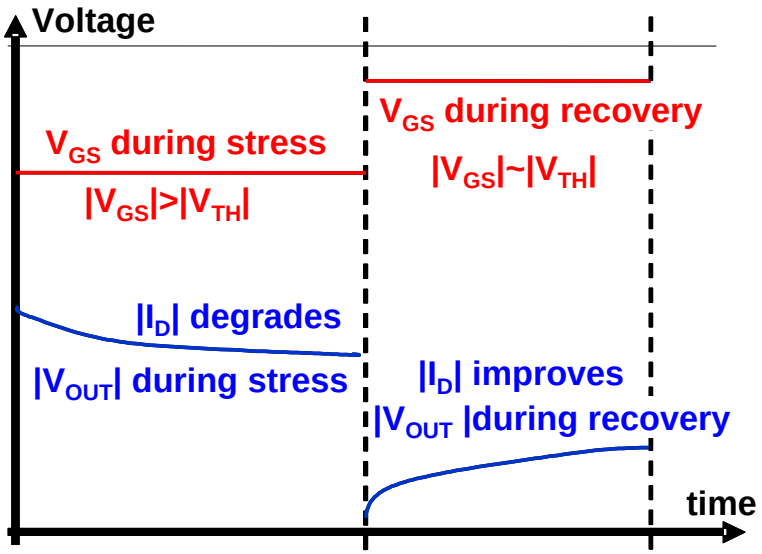


Fig. A.12. The schematic of the drain current, $|I_D|$, and thus $|V_{OUT}|$ degrades during stress and then improves with reduced stress conditions during which $|I_D|$ recovers.

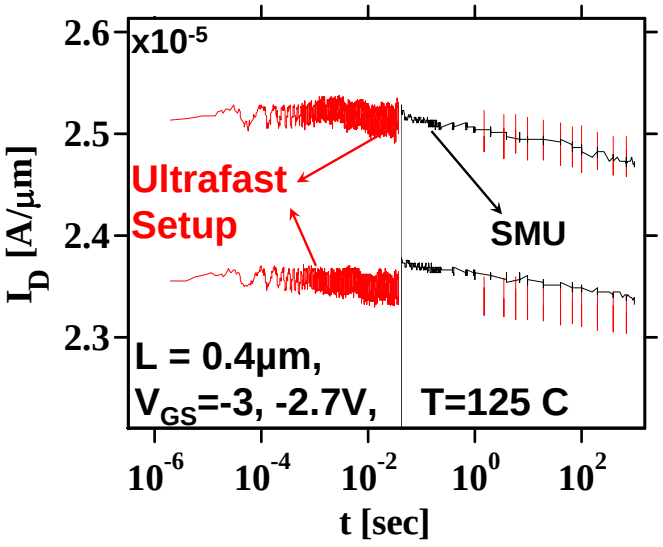


Fig. A.13. Data from ultrafast (red) and conventional (SMU, black) measurements overlap for NBTI stress. I_D decreases due to degradation. Ultrafast technique can measure down to microsecond regime whereas conventional SMU limited to millisecond resolution. The upper set is stressed at V_G=-3V and the lower for -2.7V.

and Fig. A.15. Both data overlap for two different stress voltages and temperatures, showing the accuracy of the ultrafast setup. The conventional method is limited to 1 ms while ultrafast can measure down to μs resolution. This comparison confirms that the ultrafast technique can measure accurately in shorter time scales.

Besides NBTI, the ultrafast technique can be useful for hot carrier degradation because of it also recovers, although its magnitude is not as significant as the NBTI case. The high resolution capture of the recovery can improve accuracy of the time-exponent during stress and the amount of recovery as well as providing a knob on $Si-H$ and $Si-O$ bond breaking conflict [132], [120].

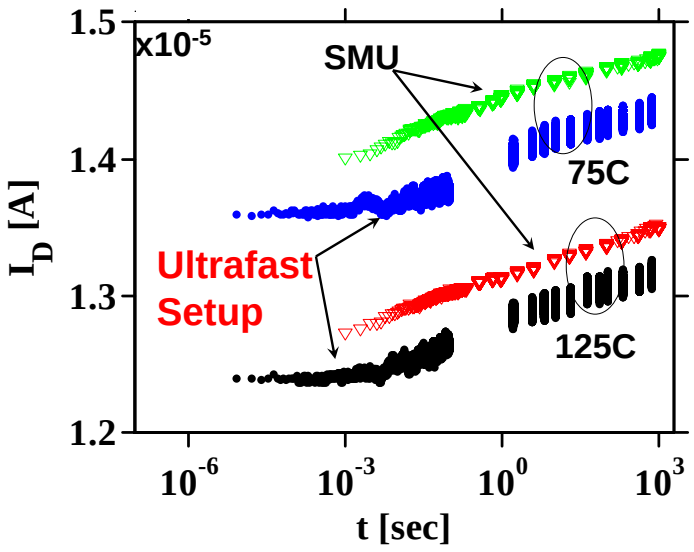


Fig. A.14. Ultrafast measurement also in good aggrement with the conventional method over different stress and recovery temperatures. Stress was done at $V_G = -3V$ and recovery was measured at $V_G = -1.1V$. Measurements were done at separate devices which show I_D (or V_T) variations.

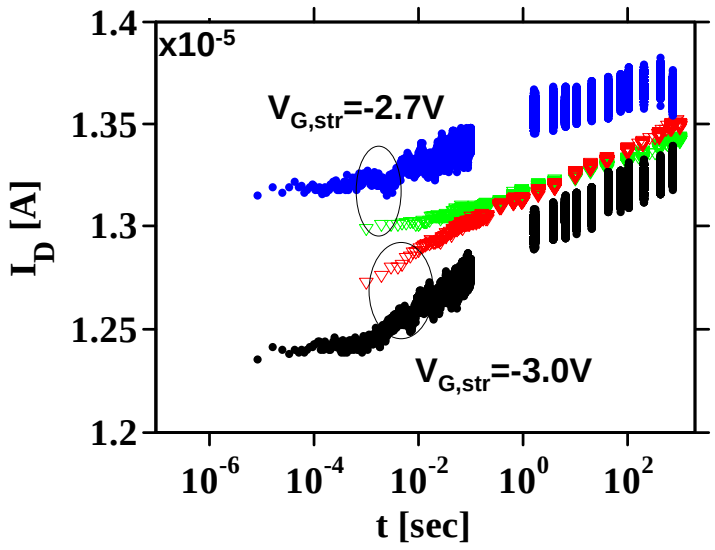


Fig. A.15. Comparison of ultrafast and conventional techniques for recovery where stress was done at two different voltages. V_G for recovery is -1.1V. The trends are similar: higher degradation (stressed at $V_G = -3V$) recovers faster. The variations on I_D is because separate MOSFETs were used in the experiment.

VITA

VITA

Haldun Kufluoglu obtained his B.S. and M.S. degrees in Electrical Engineering, in 2001 and 2003, respectively from Purdue University, West Lafayette, IN. Currently, he is pursuing his Ph.D. at the same institution.

His research interests include MOSFET reliability, experimental characterization and modeling of semiconductor devices. Particularly, his PhD research involves measurements and theoretical modeling of MOSFET degradation mechanisms such as NBTI, HCI and TDDB, and their implications on VLSI design. He also participates in off-state transistor reliability assessment. In 2006, he held a summer internship at Intel Corporation, LTD FE Q&R, Hillsboro, OR, working on experimental 65nm NBTI reliability and modeling.

Previously, he worked on MEMS that was interfaced with live neurons for biological sensor applications.