PHYSICS OF TRAP GENERATION AND ELECTRICAL BREAKDOWN IN ULTRA-THIN SiO_2 AND SiON GATE DIELECTRIC MATERIALS

Paul E. Nicollian

Physics of Trap Generation and Electrical Breakdown in Ultra-thin SiO_2 and SiON Gate Dielectric Materials

Paul E. Nicollian

TITLE:
Physics of Trap Generation and Electrical Breakdown in Ultra-thin SiO_2 and SiON Gate Dielectric Materials

AUTHOR:
Nicollian, Paul E.

ISBN: 978-90-365-2563-3

KEYWORDS:
Reliability, breakdown, TDDB, trap generation, dielectrics, oxide, oxynitrides, SiO_2, SiON, PNO.

Copyright © 2007 by Paul E. Nicollian

PHYSICS OF TRAP GENERATION AND ELECTRICAL BREAKDOWN IN ULTRA-THIN SiO_2 AND SiON GATE DIELECTRIC MATERIALS

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the Rector Magnificus,
prof. dr. W. H. M. Zijm,
on account of the decision of the graduation committee,
to be publicly defended
on Friday 31 August 2007 at 16.45

by

Paul Edward Nicollian

born on 1 September 1961
in Summit, New Jersey USA

The dissertation has been approved by Promotor:
prof. dr. J. Schmitz

and Assistant Promotor:
prof. dr. ir. F. G. Kuper

To my Mother, Rose Nicollian

COMPOSITION OF THE DOCTORAL COMMITTEE

Chairman:	
prof. dr. W. H. M. Zijm	University of Twente
Secretary:	
prof. dr. K. Hoede	University of Twente
Promotor:	
prof. dr. J. Schmitz	University of Twente
Assistant Promotor:	
prof. dr. ir. F. G. Kuper	University of Twente/NXP
Internal Members:	
prof. dr. ir. A. J. Mouthaan	University of Twente
prof. dr. ir. B. Nauta	University of Twente
External Members:	
prof. dr. M. A. Alam	Purdue University
prof. dr. G. Groeseneken	IMEC/Katholieke University
prof. dr. J. Suñé	University of Barcelona Autonoma

ACKNOWLEDGEMENTS

In my professional journey, I have been fortunate in having the opportunity to spend the past 22 years in the employ of Texas Instruments Incorporated. In this environment, I have worked with, learned from, and been nurtured by many extraordinarily talented individuals. I have also benefited from interactions with researchers from other institutions.

I would like to thank the following individuals for the collaborations, contributions, discussions, and mentoring that have improved my understanding of the field and/or enhanced the quality of my work: William R. Hunter, Anand T. Krishnan, Mark Rodder, Douglas T. Grider, Vijay K. Reddy, Ajith Amerasekera, George Brown, Sunil V. Hattangady, Rajesh B. Khamankar, Jerry C. Hu, Chris Bowen, Srinivasan Chakravarthi, Cathy A. Chancellor, Jerry Seitchek, Thomas Anderson, John Kuehne, and Kwame N. Eason of Texas Instruments; Donelli J. DiMaria, Ernest Y. Wu, and James H. Stathis from IBM; Ashraful Alam from Purdue University; Robin Degraeve from IMEC; Kin P. Cheung from Rutgers University; Eric. M. Vogel from the University of Texas at Dallas; John Suehle from the National Institute of Standards; and David J. Dumin from Clemson University.

Management support provided by Srikanth Krishnan, Timothy Rost, James Ondrusek, Bharath Rajagopalan, Linton Salmon, David Spratt, Robert M. Wallace, and Robert Eklund is gratefully acknowledged.

I would like to thank my Promotor Jurriaan Schmitz for bringing me into the University of Twente PhD program and guiding me through the dissertation process. The author gratefully acknowledges the time and effort of the entire doctorate committee in evaluating this work, including Jurriaan Schmitz from the University of Twente; Fred Kuper from the University of Twente/NXP; Ton Mouthaan from the University of Twente; Bram Nauta from the University of Twente; Ashraful Alam from Purdue University; Guido Groeseneken from IMEC/Katholieke University; and Jordi Suñé from the University of Barcelona Autonoma.

This work was supported in its entirety by Texas Instruments Incorporated. All materials were processed at TI and all measurements were performed in TI laboratories.

Paul E. Nicollian

Dallas, Texas USA
August 2007

SUMMARY

This work spans nearly a decade of industrial research in the reliability physics of deeply scaled SiO_2 and SiON gate dielectrics. In this work, we will present our following original contributions to the field:

- Below 5V stress, the dominant mechanism for stressed induced leakage current in the off-state is tunneling via interface traps in films less than 35Å thick. This finding enhances the value of SILC measurements as a probe of trap generation. LV-SILC is a two-trap process and senses interface states at both top and bottom interfaces.

- A conclusive experimental proof of the IBM energy driven breakdown theory, showing that breakdown is indeed voltage rather than field driven in ultra-thin oxides. This work has been instrumental in ending the long running controversy in the industry on breakdown models.

- Experimental verification of the Bell Labs theory that anode hole injection through minority ionization remains a plausible breakdown mechanism down to 3.6V. This finding shows that holes continue to play a role in the degradation physics at low voltages. However, our experiments eliminate anode hole injection as the mechanism for breakdown below about 2.7V.

- Plasma nitridation of oxides significantly extends the reliability scaling limit of SiO_2 based films. Bulk trap generation rates are minimized and the film reliability is optimized when the nitrogen profile is uniform. Plasma nitrided SiON films are now widely used throughout the industry in high volume manufacturing.

- Reaction-diffusion theory applies for TDDB stress of ultra-thin NMOS SiON films. Measurable recovery effects are present, showing that quasi-equilibrium exists for NMOS under substrate injection conditions. This finding enables the determination of the mechanisms for trap generation and breakdown, showing that they are controlled by the release of two species of hydrogen (H^+ and H^0) from the anode in two separate anode reactions. H^+ and H^0 both create interface traps at the poly interface when they are released. After migrating into the dielectric, H^+ subsequently creates SiON bulk traps while H^0 subsequently creates interface traps at the pwell interface. The hydrogen species that controls breakdown is voltage dependent. The mechanism for breakdown transitions from hole induced H^+ desorption to electron induced H^0 desorption below the 2.7V threshold for vibrational excitation of Si-H bonds.

- Bulk traps control breakdown in SiO_2 dielectrics below 30Å. However, bulk traps are not always the defects that control breakdown in SiON films below 20Å. Below the 2.7V threshold energy for vibrational excitation of silicon-hydrogen bonds, the rate

limiting step is the generation of interface traps. However, a minimum of two traps is required to cause breakdown in SiON films down to 10Å EOT. At least one trap must be an interface state and at least one must be a bulk state.

- Our experimentally obtained trap generation power law exponent m being about 0.3, which is lower than the numbers reported by other researchers, is the only value that is consistent with the temperature and voltage dependence of trap generation and breakdown. This leads to the SiON bulk trap diameter being about 4Å, which is significantly lower than earlier estimates and results in the Weibull slope to remain greater than 1 down to the 12Å limit for physical oxide thickness.

These findings have been published or have been submitted for publication as follows:

1. P.E. Nicollian and A.T. Krishnan, “Two Trap Model for Low Voltage Stress Induced Leakage Current”, submitted to the *Journal of Applied Physics*, 2007
2. P.E. Nicollian, A.T. Krishnan, C.A. Chancellor, and R.B. Khamankar, “The Traps that cause Breakdown in Deeply Scaled SiON Dielectrics”, in *IEDM Technical Digest*, 2006, pp. 743-746
3. P.E. Nicollian, A.T. Krishnan, C. Bowen, S. Chakravarthi, C. Chancellor, R. Khamankar, “The Roles of Hydrogen and Holes in Trap Generation and Breakdown in Ultra-thin SiON Dielectrics”, in *IEDM Technical Digest*, 2005, pp. 403-406
4. P.E. Nicollian, G.C. Baldwin, K.N. Eason, D.T. Grider, S.V. Hattangady, J.C. Hu, W.R. Hunter, M. Rodder, A.L.P. Rotondaro, “Extending the Reliability Scaling Limit of SiO_2 through Plasma Nitridation”, in *IEDM Technical Digest*, 2000, pp. 545-548
5. P.E. Nicollian, W.R. Hunter, J.C. Hu, “Experimental Evidence for Voltage Driven Breakdown Models in Ultra-thin Gate Oxides”, in *Proceedings of the IRPS*, 2000, pp. 7-15
6. P.E. Nicollian, M. Rodder. D.T. Grider, P. Chen, R.W. Wallace, S.V. Hattangady, “Low Voltage Stress-Induced-Leakage-Current in Ultra-thin Gate Oxides”, in *Proceedings of the IRPS*, 1999, pp. 400-404

SAMENVATTING

Dit werk omvat bijna een decennium van industrieel onderzoek naar de betrouwbaarheidsfysica van extreem dunne SiO_2 en SiON gate-diëlektrica. In dit proefschrift worden de volgende originele bijdragen aan het veld gepresenteerd:

- Het is aangetoond dat in lagen dunner dan 35 Å, belast met minder dan 5 V, het dominante mechanisme voor stress-veroorzaakte lekstroom (SILC) in de uittoestand het tunnelen via grensvlaktoestanden is. Door deze bevinding zijn SILC-metingen zeer waardevol als een methode om het ontstaan van traps te bestuderen. SILC bij lage spanning treedt op via twee traps, en is gevoelig voor grensvlaktoestanden aan beide elektrodes.

- Het energie-gedreven doorslagmodel van IBM is definitief experimenteel bevestigd. Hiermee is aangetoond dat de doorslag wordt veroorzaakt door de aangelegde spanning, en niet het elektrisch veld. Deze bevindingen hebben geholpen om een langdurende controverse in de halfgeleiderindustrie ten aanzien van doorslagmodellen te beeindigen.

- De theorie van Bell Labs dat anode-gatinjectie via minderheidsionisatie ook beneden 3.6 V een waarschijnlijke veroorzaker van doorslag is, is experimenteel bevestigd. Zo is gedemonstreerd dat ook bij lage spanningen gaten een rol blijven spelen in de degradatie. Onze experimenten sluiten daarentegen uit, dat anode-gatinjectie nog een rol speelt beneden 2.7 V.

- De grens aan het schalen van SiO_2-gebaseerde diëlektrica bepaald door betrouwbaarheidseisen, wordt significant verlegd met plasmanitridatie. Wanneer het stikstofprofiel uniform is, is de snelheid van trapgeneratie minimaal en de betrouwbaarheid van het diëlektricum maximaal. Plasma-genitrideerde SiON diëlektrica worden nu breed en massaal toegepast door IC-fabrikanten.

- De reactie-diffusietheorie beschrijft het degradatiegedrag onder TDDB (tijdsafhankelijke diëlektrische doorslag) belasting van ultra-dunne NMOS SiON dielektrica. Er zijn meetbare effecten van herstel, wat aangeeft dat in NMOS structuren quasi-evenwicht heerst onder substraat-injectiecondities. Op basis van deze bevindingen kunnen de mechanismes voor trapgeneratie en doorslag worden bepaald. Deze worden gecontroleerd door het vrijkomen van twee verschijningsvormen van waterstof, H^+ en H^0, aan de anode uit twee afzonderlijke reacties. H^+en H^0 vormen na vrijkomen grensvlak-traps aan het polysilicium-interface. Na migratie in het diëlektricum vormt H^+ SiON bulktraps, terwijl H^0 grensvlaktraps vormt aan het interface met het substraat. Het hangt van de stressspanning af welk van de verschijningsvormen van waterstof doorslag veroorzaakt. Het mechanisme voor doorslag verschuift van gaten-geïnduceerde H^+

desorptie naar elektronen-geïnduceerde H^0 desorptie beneden de 2.7 V drempel voor vibrationele excitatie van Si-H covalente bindingen.

- Bulk traps bepalen de doorslag in SiO_2 diëlektrica dunner dan 30 Å; maar dit is niet altijd het geval in SiON lagen dunner dan 20 Å. Beneden de bovengenoemde drempel van 2.7 V is de beperkende stap het genereren van interface traps. Minimaal twee traps zijn benodigd om een doorslag in SiON lagen te veroorzaken, tot 10 Å EOT. Hiervoor zijn tenminste een interface trap en een bulk trap nodig.

- Experimenteel hebben we een waarde van 0.3 gevonden voor de exponent *m* in de generatie van traps. Deze waarde is lager dan anderen rapporteren; maar tevens de enige waarde die consistent is met de temperatuur- en spanningsafhankelijkheid van trapcreatie en doorslag. Hieruit volgt een trapdiameter van 4 Å, aanmerkelijk kleiner dan eerdere schattingen. Deze waarde leidt tot een Weibull helling groter dan 1 tot aan de limiet van 12 Å (fysieke dikte).

Deze bevindingen zijn (of worden) gepubliceerd in:

1. P.E. Nicollian and A.T. Krishnan, “Two Trap Model for Low Voltage Stress Induced Leakage Current”, submitted to the *Journal of Applied Physics*, 2007
2. P.E. Nicollian, A.T. Krishnan, C.A. Chancellor, and R.B. Khamankar, “The Traps that cause Breakdown in Deeply Scaled SiON Dielectrics”, in *IEDM Technical Digest*, 2006, pp. 743-746
3. P.E. Nicollian, A.T. Krishnan, C. Bowen, S. Chakravarthi, C. Chancellor, R. Khamankar, “The Roles of Hydrogen and Holes in Trap Generation and Breakdown in Ultra-thin SiON Dielectrics”, in *IEDM Technical Digest*, 2005, pp. 403-406
4. P.E. Nicollian, G.C. Baldwin, K.N. Eason, D.T. Grider, S.V. Hattangady, J.C. Hu, W.R. Hunter, M. Rodder, A.L.P. Rotondaro, “Extending the Reliability Scaling Limit of SiO_2 through Plasma Nitridation”, in *IEDM Technical Digest*, 2000, pp. 545-548
5. P.E. Nicollian, W.R. Hunter, J.C. Hu, “Experimental Evidence for Voltage Driven Breakdown Models in Ultra-thin Gate Oxides”, in *Proceedings of the IRPS*, 2000, pp. 7-15
6. P.E. Nicollian, M. Rodder. D.T. Grider, P. Chen, R.W. Wallace, S.V. Hattangady, “Low Voltage Stress-Induced-Leakage-Current in Ultra-thin Gate Oxides”, in *Proceedings of the IRPS*, 1999, pp. 400-404

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION **1**

1.1 Scaling trends 2
1.2 Motivation for this work 3
1.3 Stages of degradation of a dielectric under stress 4
1.3.1 Trap generation regime 4
1.3.2 Soft breakdown 5
1.3.3 Post breakdown regime 6
1.3.4 Hard breakdown 7
1.4 Soft breakdown detection techniques 7
1.5 Statistical concepts for dielectric reliability 9
1.5.1 Basic definitions 10
1.5.2 The Weibull distribution 11
1.5.3 Experimental considerations for applying Weibull statistics 12
1.6 Reliability projection 13
1.6.1 Field and voltage dependence 13
1.6.2 Area scaling 14
1.6.3 Ramped voltage breakdown 15
1.6.4 Temperature dependence 16
1.7 Band bending and electric fields 17
1.7.1 NMOS energy band diagram 17
1.7.2 Relationship between gate voltage, oxide voltage, and band bending 19
1.7.3 Quantum effects 19
1.7.4 Some important properties of oxide electric fields 19
1.8 Equivalent oxide thickness 20
1.9 Capacitance-voltage characteristics 22
1.9.1 Components of the 1-dimensional capacitance 22
1.9.2 Effect of interface traps and oxide fixed charge 23
1.9.3 Techniques for measuring C-V characteristics 24
1.10 Chapter summary 27
1.11 References 29

CHAPTER 2 CHARGE TRAPPING AND STRESS INDUCED LEAKAGE CURRENT **35**

2.1 Introduction 35
2.2 Time-0 transport mechanisms 35
2.2.1 Fowler-Nordheim tunneling 35
2.2.2 Direct tunneling 37
2.2.3 Effects of nitrogen 40

2.3 Charge trapping and trap generation 42
2.3.1 The trap potential well 43
2.3.2 Determination of the effective trapped charge density and centroid 45
2.3.3 Stress and measurement techniques for characterizing charge trapping 46
2.3.4 Charge trapping kinetics 50
2.3.5 Detrapping effects 51
2.3.6 Electron and hole traps 56
2.4 Stressed induced leakage current due to bulk traps (SILC) 58
2.4.1 Post stress I_G-V_G characteristics of sub-60Å oxides 58
2.4.2 Elasticity of trap assisted tunneling through bulk traps 59
2.4.3 Transport model for stress induced leakage current 61
2.4.4 Kinetics of trap generation 64
2.5 Stressed induced leakage current due to interface traps (LV-SILC) 64
2.5.1 Energy dependence of trap generation 65
2.5.2 Post stress I_G-V_G characteristics of sub-40Å oxides 66
2.5.3 Model for LV-SILC 70
2.6 Chapter summary 84
2.7 References 88

CHAPTER 3 MODELS FOR DIELECTRIC BREAKDOWN 93

3.1 Introduction
3.2 Tunneling and carrier energy 93
3.2.1 Definition of tunneling processes 93
3.2.2 Direct tunneling 93
3.2.3 Ballistic Fowler-Nordheim tunneling 94
3.2.4 Steady state Fowler-Nordheim tunneling 96
3.3 Field based breakdown models 98
3.3.1 The 1/E Model 99
3.3.2 The E-Model 100
3.3.3 Applications and limitations of field driven breakdown models 102
3.4 Gate voltage driven breakdown 104
3.5.1 The polarity gap 104
3.5.2 Poly doping experiments 107
3.5.3 Substrate hot electron experiments 109
3.5.4 Formulations of the V_G Model 110
3.5 Chapter summary 112
3.6 References 113

CHAPTER 4 MECHANISMS FOR TRAP GENERATION AND BREAKDOWN 117

4.1 Introduction 117
4.2 Reaction – diffusion theory
4.2.1 Quasi-equilibrium 118

4.2.2 Trap generation kinetics when the the reaction product is a charged species 119
4.2.3 Trap generation kinetics when the reaction product is a neutral species 121
4.3 Impact ionization in the gate dielectric 123
4.3.1 Stabilization of the electron energy distribution at high fields 123
4.3.2 Experimental evidence for impact ionization 124
4.4 Features of anode hole injection 127
4.4.1 Majority and minority ionization 127
4.4.2 Experimental evidence for anode hole injection 129
4.5 Features of anode hydrogen release 133
4.5.1 Description of the mechanism 134
4.5.2 Experimental evidence for anode hydrogen release 134
4.6 Identification of the hydrogen species that cause trap generation 136
4.6.1 The stress method 137
4.6.2 Cathode interface reactions induced with a back bias 138
4.6.3 Bulk trap generation 141
4.6.4 Two reaction model 143
4.7 Comparison of breakdown data with scanning tunneling microscopy results 146
4.7.1 STM induced hydrogen desorption from silicon surfaces 146
4.7.2 Evidence of vibrational excitation of hydrogen from TDDB studies 148
4.8 Identification of the stress-induced defects that cause breakdown 150
4.8.1 Percolation models 150
4.8.2 SiO_2 gate dielectrics below 30Å 151
4.8.3 SiON films below 20Å 154
4.8.4 Implications for SiON thickness scaling 157
4.9 Anode hole injection below 3V 159
4.9.1 Minority ionization at low voltages 159
4.9.2 Majority ionization at low voltages 159
4.10 Chapter summary 160
4.11 References 163

CHAPTER 5 CLOSING REMARKS **169**

List of publications 171

Author biography 174

CHAPTER 1

Introduction

1.1 Scaling trends

The scope of this work spans nearly a decade of research as an industrial scientist at Texas Instruments Incorporated. During this timeframe, aggressive technology scaling has stimulated a tremendous amount of interest in the physics and modeling of gate dielectric reliability. At the beginning of this period, a long running controversy existed regarding whether time dependent dielectric breakdown (TDDB) followed an E-Model [1,2] or 1/E-Model [3] dependence. This issue was never definitely resolved. The E-Model was widely applied because it yielded more conservative lifetimes, albeit in an era where reliability requirements were often readily attainable.

To maintain scaling as defined by Moore's Law, the number of transistors doubles every 2 – 3 years. Since total product failure rates have remained constant at about 10 to 100 fits, the allowed failure rate per transistor has continued to decrease. From 1995 to 2020, the failure rate per device must drop by a factor of 300 to 6,000 (Figure 1.1).

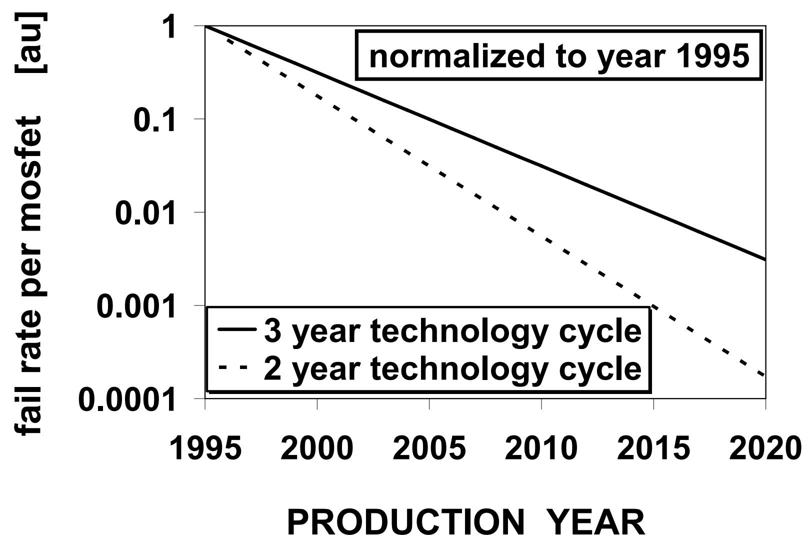

Figure 1.1. Failure rate per transistor vs. year in production, normalized to1995.

In addition, the industry has become increasingly concerned over the impact of high gate leakage due to direct tunneling resulting from thickness scaling [4]. Moreover, under stress, the gate current can increase by orders of magnitude due to the

generation of trap states (i.e. stress induced leakage current) [5]. In the midst of these concerns, it was reported that oxide thickness could not be reliably scaled below 24Å [6]. This triggered a flurry of activity in breakdown physics research which resulted in a myriad of discoveries, including a V_G Model for reliability projections [7-11] that is more optimistic than the E-Model. The V_G Model has been instrumental in continuing the scaling of reliable gate dielectrics below 24Å.

The introduction of plasma nitrided oxides (PNO) is another important development [12]. Earlier attempts to thermally incorporate nitrogen to block boron penetration from PMOS poly using NO, N_2O, and NH_3 ambients were plagued by mobility degradation [13]. This problem is minimized with PNO [14], and PNO has a further advantage of lower gate current [15,16]. Properly optimized, PNO improves the reliability of SiO_2 below 24Å, enabling deep scaling of oxide based films [16]. Plasma nitridation is now widely used throughout the industry.

Equivalent oxide thickness continues to be aggressively scaled to achieve higher performance. From 2001 to 2014, equivalent oxide thickness is projected to decrease from 14.5Å to 4.5Å [17] as shown in Figure 1.2. Because at least 2 mono layers are needed to attain the bulk insulating properties of an oxide film, the theoretical scaling limit is about 7Å [18,19]. To allow margin for manufacturability, the physical film thickness must be no less than about 12Å [19]. Accordingly, scaling equivalent oxide thickness below 12Å requires a dielectric with a higher dielectric constant than of oxide. Nitrogen incorporation achieves this goal. In theory, a pure nitride would enable scaling EOT down to about 6Å to 7Å, but it is not clear whether a functional (or reliable) device can be built under this severe a nitridation condition. Beyond this limit, high-k gate dielectrics are needed to continue EOT scaling [20].

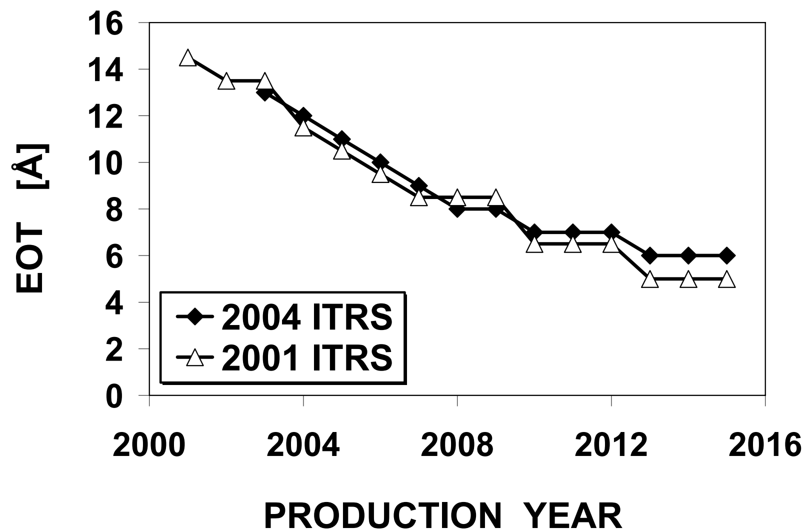

Figure 1.2. Equivalent oxide thickness vs. production year for high performance logic technologies [17]. © 2001, 2004, Semiconductor Industry Association.

As the manufacturability limit of 12Å physical film thickness is approached, an emerging trend to meet performance targets is a reduction in the rate at which voltage is scaled down with successive technology nodes. In Figure 1.3, the ITRS projections from the 2001 and 2004 roadmaps [17] are compared. The projected supply voltages (V_{CC}) have become starkly higher in the 3 year span between roadmaps. Moreover, it is common practice to overdrive V_{CC} beyond the ITRS roadmap to meet high performance requirements. Since TDDB is a strongly voltage driven phenomenon [11], this trend is becoming a significant new challenge for meeting gate dielectric reliability requirements.

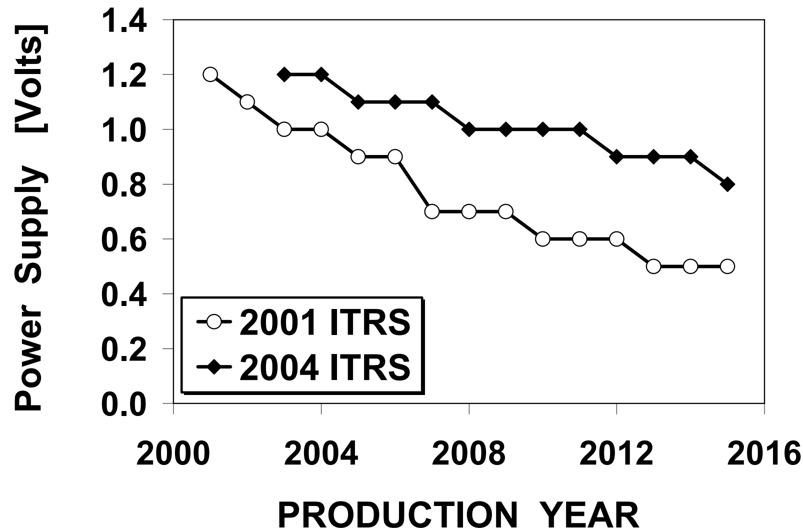

Figure 1.3. Power supply voltage vs. production year for high performance logic technologies, comparing 2001 and 2004 ITRS roadmaps [17]. © 2001, 2004, Semiconductor Industry Association.

1.2 Motivation for this work

In this dissertation, we will present our novel findings regarding the physical mechanisms of trap generation and breakdown in SiO_2 and SiON films with thickness ranging from about 10Å to 60Å EOT. This work has been driven by the need to attain world class reliability in high performance technologies, which are typically the most challenging to meet reliability requirements due to aggressively scaled oxide thickness, overdriven supply voltages, high operating temperatures, and large areas. All samples used for these studies were processed at Texas Instruments Incorporated. Except where noted, the data are from NMOS devices.

We begin with an overview of reliability and device physics principles in Chapter 1. In Chapter 2, we will discuss trap generation and its effects on transport, along with our discoveries regarding the role of interface traps on stress induced leakage currents. In Chapter 3, we will present our work on voltage driven dielectric breakdown models. In Chapter 4, we will show our findings on the physical mechanisms that result in trap

generation and breakdown, including the effects of electrons, holes, and hydrogen. The defects that are generated during stress and their impact on the reliability of the gate dielectric will be elucidated.

1.3 Stages of degradation of a dielectric under stress

Starting with a virgin 12Å NMOS PNO gate dielectric, we illustrate how a device degrades with time under a constant voltage stress in Figure 1.4. The normalized SILC increase $\Delta I/I(0)$ tracks the build up of traps during stress [5] and is defined as:

$$\Delta I/I(0) = [I(t) - I(0)] / I(0) \quad (1.1)$$

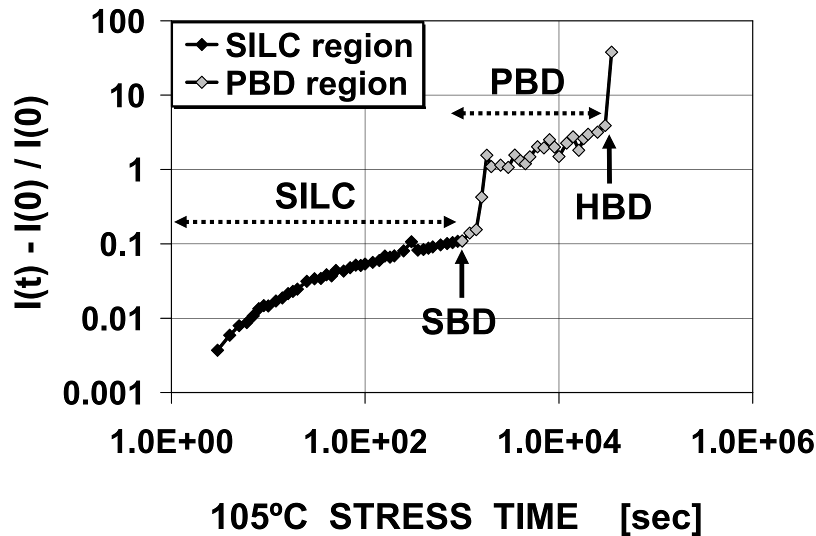

Figure 1.4. Normalized SILC increase vs. time for a 12Å NMOS PNO film stressed at +2.3V. The current was sensed at +1.0V. The gate area is 10^{-7} cm^2.

1.3.1 Trap generation (SILC regime)

Once the stress has begun, charge trapping and trap generation commence. When present, charge trapping through the filling of as fabricated hole traps follows an exponentially decaying trapping rate [21], and therefore becomes increasingly less important as the stress time increases. Under the stress conditions used for the 12Å production grade PNO film shown in Figure 1.4, positive charge trapping is insignificant over the entire time domain because the native trap density is low and the probability of a hole remaining trapped in such a thin dielectric is small (at low voltages, for inverted poly gate NMOS, the holes originate from impact ionization in the anode). Unless the trap is energetically shallow (with respect to the oxide conduction band), the trapped hole will tunnel into the cathode at high electric fields.

The generation of new trap states follows a power law in time (and fluence) [5],

$$N(t) = b_t t^m \quad (1.2)$$

$$N(Q) = b_Q Q^m \quad (1.3)$$

In dielectrics thinner than about 50Å, these states give rise to an increase in gate current (SILC) due to trap assisted tunneling of electrons through these stress generated defects [5]. In thicker films, electrons can become permanently trapped in these sites, but this is also a negligible effect in our 12Å PNO dielectrics because unless the trap is energetically deep (with respect to the oxide conduction band), the trapped electron will tunnel into the anode under high electric fields. Trapped charge has been observed below 30Å by injecting substrate hot carriers into the oxide at low oxide electric fields to minimize de-trapping probabilities [22]. SILC is predominantly seen in the direct tunneling regime because it can be masked by Fowler-Nordheim conduction at higher voltages [5]. As SILC is a tunneling process, its I-V relationship is exponential. Stress induced leakage current will be discussed in more detail in Chapter 2.

1.3.2 Soft Breakdown (SBD)

In thick oxides stressed at high voltages, the first breakdown event results in an abrupt and catastrophic increase in the gate current. Below about 50Å, a phenomenon known as soft breakdown appears [23,24]. SBD occurs when the density of traps generated during stress is sufficient to form a path that connects the anode to the cathode [25]. The post SBD resistance is significantly higher than catastrophic hard breakdown (HBD) and has been correlated to the spatial distribution of the traps that form the breakdown path inside the dielectric [26]. As the stress voltage is decreased, the probability that the 1ST breakdown is HBD continues to diminish due the reduced power and energy available to the device when breakdown occurs [26,27]. Consequently, the manifestation of the breakdown event is strongly affected by the specific electrical environment (e.g. current compliance) that the device experiences during stress [28].

Several mechanisms have been proposed to explain post SBD transport, including (a) the creation of a physically damaged region (PDR) at the anode [24], (b) variable range hopping transport (VRH) [29], (c) conduction through a quantum point contact (QPC) [30], (d) percolation through a nonlinear conductor network [31], (e) space charge limited current (SCLC) in the silicon electrode [32], and (f) electron co-tunneling through a coulomb blockade [33]. In the PDR model, the damaged region causes the direct tunneling current to increase through a reduction in barrier height and tunneling distance [24]. In the VRH model, transport is due to the hopping of a carrier from trap state i to trap state j, where average hopping distance is not constant but decreases with increasing temperature (hence the name "*variable range*" hopping) [34]. In the QPC model, ballistic transport occurs through a defect path that is sufficiently narrow (comparable to an electron wavelength) so that the energy is quantized perpendicular to the direction of propagation [30]. In the percolation model, conduction occurs through a network of traps that connect the anode and cathode [31]. SCLC is due to the current being limited by transport in the space charge region. In the electron co-tunneling

model, a coulomb blockade arises if the tunneling of one electron changes the voltage across the tunneling junctions, thus setting up a coulomb blockade that inhibits the tunneling of additional electrons. If there are N traps, there will be N+1 tunnel junctions in series. Transport can occur if multiple electrons tunnel at the same time (co-tunnel) through subsequent traps via virtual intermediate states, thus mitigating the effects of the coulomb blockade.

In the PDR model, the I-V relation is exponential. VRH and QPC both give rise to a sinh(V) dependence of the current, which becomes exponential at high voltage. For percolation, SCLC, and co-tunneling models, the I-V dependence is power law, $I \sim gV^{P}$. An interesting feature of the co-tunneling model is that it predicts that the slope of the power law I-V relationship is quantized, in agreement with experimental data. The value of the slope increases with oxide thickness and is determined by the number of traps in the percolation path [33].

In practice, the current appears to follow a power law below about 3.5V [35], but because a significant portion of this region is influenced by the build up of voltage drops in the semiconductor [30], a reliable fit of the data can only be obtained within a narrow voltage excursion. Above this voltage range, Fowler-Nordheim tunneling (an exponential process) eventually dominates the total current at a voltage that depends on oxide thickness. Moreover, for the 12Å films in Figure 1.4, the device breaks down below 3V. Accordingly, for deeply scaled dielectrics, there is some difficulty in the identification of the transport mechanism due to the small voltage excursion between the onset of strong inversion (or strong accumulation) and the breakdown of the oxide.

Noting that the identification of a transport mechanism must explain both voltage and temperature dependence, the strong temperature dependence of SBD [29, 31] rules out the PDR model, but VRH, percolation, and co-tunneling transport fit the temperature data in the power law I-V region.

1.3.3 Post breakdown (PBD regime)

After SBD, the device continues to degrade until catastrophic hard breakdown occurs [36]. In small area devices, after the 1ST breakdown event, the post breakdown (PBD) regime is typified by quantized jumps in the current following the occurrence of each successive breakdown event [37], as seen in Figure 1.4. The 1ST breakdown event (soft or hard) area scales according to Weibull statistics [38] and therefore the time to 1ST breakdown readily lends itself to conservative reliability modeling.

As SBD does not necessarily render transistors inoperative [39], reliability projection techniques have been proposed to increase the time to failure beyond the 1ST breakdown event [27,37,40]. One of these models is known as the prevalence method [27,40], where the hard breakdown distribution is shifted from the first breakdown time by a factor that depends on the stress conditions. Another technique is the successive breakdown method [37], which provides a methodology for determining the time at which a specified leakage criteria is exceeded following the occurrence of multiple soft breakdown events.

The specification of a time to fail beyond the occurrence of the 1ST soft breakdown would seem to be a natural extension for circuits with large drivers charging small capacitive loads, because the post breakdown resistance needs to be relatively small (hard) to cause failure [41]. However, there are other applications that are much more sensitive to dielectric breakdown. For example, analog receivers in wireless devices inherently have low signal to noise margin ratios, which are easily perturbed by a soft breakdown event. In dynamic logic circuits, pre-charged nodes are vulnerable to SBD if the resultant charge loss occurs prior to the read cycle [42]. Under certain conditions, a SBD resistance path as high as 500KΩ can increase the minimum operating voltage (V_{MIN}) of an SRAM cell to the extent where it is not functional [43]. Moreover, the prevalence and successive breakdown methods [27,37,40] apply to stable filaments (such as in Figure 1.4) that are spatially and temporally uncorrelated. However, a stable SBD filament is not always formed [36,44,45,46,47]. This failure mode is known as digital SBD [44], and the corresponding degradation behavior is called progressive breakdown [36-46]. When digital SBD occurs, the current continues to increase, and successive breakdown events are not spatially independent [47]. Additionally, the voltage and temperature acceleration factors have been reported to be different for stable vs. unstable filaments [47]. This adds significant complexity to reliability projections in the post breakdown regime, because oxide failures in a circuit are then an unknown mixture of two breakdown modes that have different kinetics. Accordingly, in the interest of developing a physical understanding that is useful for delivering generic reliability models that are safe to use for a wide range of semiconductor products, in this work, the time to failure is always taken as the time at which the 1ST breakdown event occurs.

1.3.4 Hard breakdown (HBD)

Except for the digital SBD mode, HBD and SBD are independent failures occurring at different spatial locations [48,49]. Hard breakdown (HBD) is characterized by an ohmic I-V relationship and a post breakdown resistance < 10KΩ. As the stress voltage increases, the time delay between SBD and HBD diminishes [45], and the 1ST breakdown becomes predominantly HBD above about 5V [26,27]. In thick oxides stressed at high voltages, HBD is catastrophic and results in a low resistance short between the two electrodes. If the stress is thermally limited with a resistance in the measurement path [49], below about 50Å, the HBD I-V curves exhibit the characteristics of a quantum point contact [50]. In ultra-thin dielectrics stressed at low voltages, the HBD and SBD regimes are differentiated by the magnitude of the post breakdown resistance.

1.4 Soft breakdown detection techniques

In thicker oxides stressed at higher voltages, a large and abrupt current jump (ΔI) clearly demarks the occurrence of breakdown. However, as can be seen in Figure 1.4, the exact time at which SBD occurs is not necessarily well defined at low voltages in ultra-thin dielectrics. The difficulty in detecting SBD increases with decreasing oxide thickness [39,51] and increasing gate area [51] due to the diminishing separation

between the SBD signal and background direct tunneling current. This problem can be somewhat mitigated by using a smaller area device [52].

Because degradation in the SILC regime follows a well behaved power law in time [53], one possible SBD detection technique is the change in the slope of the normalized SILC increase vs. time curve:

$$S(t) = d \log(\Delta I/I(0))/d \log t \tag{1.4}$$

A sensitive measure of SBD can be obtained by noting that one of the signatures of soft breakdown is the onset of complex time-dependent fluctuations in the gate current [23]. The physics of this phenomenon are not thoroughly understood. One theory for the origin of this random telegraph noise (RTS) is the reduction in the self-averaging of the SBD current resulting from it flowing through the small area SBD filament [54], whereas the current flows across the entire device area before SBD. It has also been proposed that RTS arises from the switching between 2 different SBD conduction paths [32]. An approach for the quantifying noise resulting from SBD is to determine the noise variance (NV) resulting from stress [55] which, for constant voltage stress is expressed as:

$$NV = \langle I_G^2 \rangle - \langle I_G \rangle^2 \tag{1.5}$$

The onset of SBD is identified by a large jump in the noise variance [55] and is determined from a running average of the last 5 current read points [56]. It has become a widely used technique for sensing the onset of SBD. Figure 1.5 shows the noise variance vs. time for the device stressed in Figure 1.4.

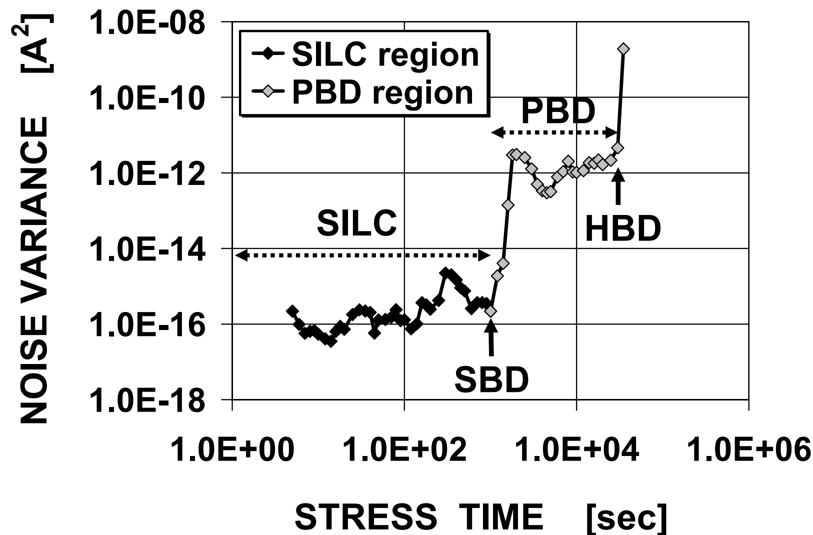

Figure 1.5. Noise variance vs. time for the same device shown in Figure 1.4.

Noise variance is less effective for sensing SBD in large area devices where the margin between noise signal and the direct tunneling background current is weak [51]. In this case, the algorithm recommended was:

$$D(t) = [I(t+\Delta t) - I(t)] / I(t) \quad (1.6)$$

The advantage of this technique is that D(t) is a decreasing function of time in the SILC regime. Accordingly, a sudden increase in D(t) signals the onset of SBD [36,51].

For extremely small area devices, the quantization of trap generation becomes apparent, because a single trap can have a measurable impact on the total gate current [57]. Therefore, excessive RTS may be observed even in the SILC regime for extremely small area devices, thereby reducing the effectiveness of noise based SBD triggers [57]. Accordingly, an abrupt current jump re-emerges as a viable monitor for the occurrence of SBD [57], provided that ΔI is chosen so that RTS peaks are not falsely identified as breakdown events, and the current has not yet evolved beyond SBD into the PBD regime when the breakdown trigger is reached. A figure of merit for determining the appropriate current jump criteria is the plateau in the Weibull slope vs. the value of the current jump used to reckon SBD [57], as illustrated in Figure 1.6.

In this work, depending on the dielectric thickness, gate area, and stress conditions, we use noise variance and current jump methods to determine the time to soft breakdown.

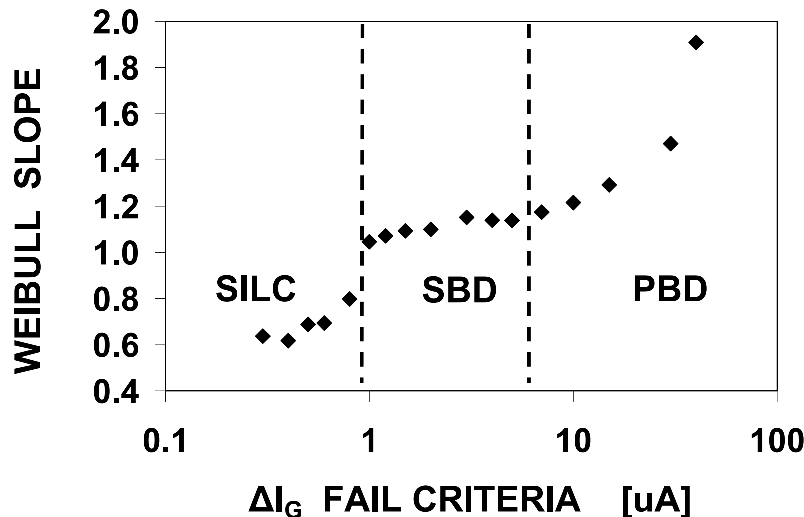

Figure 1.6. Weibull slope vs. delta I_G failure criteria. For the devices used for this figure, SBD is correctly identified when ΔI_G is between 1μA - 5 μA.

1.5 Statistical concepts for dielectric reliability

Because there are many excellent texts on reliability statistics, only the most salient features pertinent to our work will be presented. Unless otherwise noted, the material contained in this section follows [58-60].

1.5.1 Basic definitions

A function f(t) that describes the fraction of devices that fail within an interval of time Δt is called a probability density function (PDF). The cumulative distribution (CDF) is then the cumulative fraction of devices that have failed up to a time t. The CDF, denoted F(t), is related to the PDF via the following relationship

$$F(t) = \int_0^t f(y)\,dy \tag{1.7}$$

Accordingly, $F(t_2) - F(t_1)$ is the fraction of the population that fails in the interval $t_2 - t_1$. The fraction of the population that survives up to time t is known as the reliability function R(t) and is sometimes referred to as the survival function, where

$$R(t) = 1-F(t) \tag{1.8}$$

The failure rate h(t) in an interval Δt, often called the hazard rate or instantaneous failure rate, is the fraction of failures during the interval Δt relative to the surviving population:

$$h(t) = f(t)/R(t) \tag{1.9}$$

Inserting (1.8) into (1.9) yields:

$$h(t) = f(t)/(1-F(t)) \tag{1.10}$$

This leads us to the definition of the average failure rate (AFR), which is the integral of the instantaneous failure rate h(t) averaged over the interval $t_2 - t_1$:

$$AFR = \frac{\int_{t_1}^{t_2} h(y)\,dy}{t_2-t_1} \tag{1.11}$$

which after integration over the time interval t yields the following relationship:

$$AFR = -\ln(1-F(t))/t \tag{1.12}$$

The units of failure rate are FITS, where: 1 FIT = 1 failure per 10^9 power-on hours. In any real system, the failure rate can be partitioned into 3 distinct regions, which collectively, is known as the bath-tub curve due to its shape, as shown in Figure 1.7. They are: (1) Decreasing failure rate at short times (extrinsic failure period) where defective components are weeded out. It is customary to quantify the reliability in this portion of the bathtub curve by the PPM defective, where 1 PPM = 1 failure per 10^6 units. The extrinsic failure period is followed by: (2) A long, flat constant (intrinsic) failure

rate. Finally, (3) An increasing failure rate, or wear out period ensues. No single PDF with a fixed set of parameters can simultaneously fit all 3 regions of the bath tub curve.

1.5.2 The Weibull distribution

The Weibull distribution occupies a central role in the statistics of gate dielectric reliability. It is used extensively to scale reliability data from small test structure gate areas and relatively large cumulative fail fractions (due to finite sample sizes) to large product gate areas and low cumulative fail fractions corresponding to low product failure rates. It also carries important implications regarding the physics of dielectric breakdown. Time to breakdown (t_{BD}), charge to breakdown (Q_{BD}), and the critical trap density to breakdown (N_{BD}) all follow Weibull statistics.

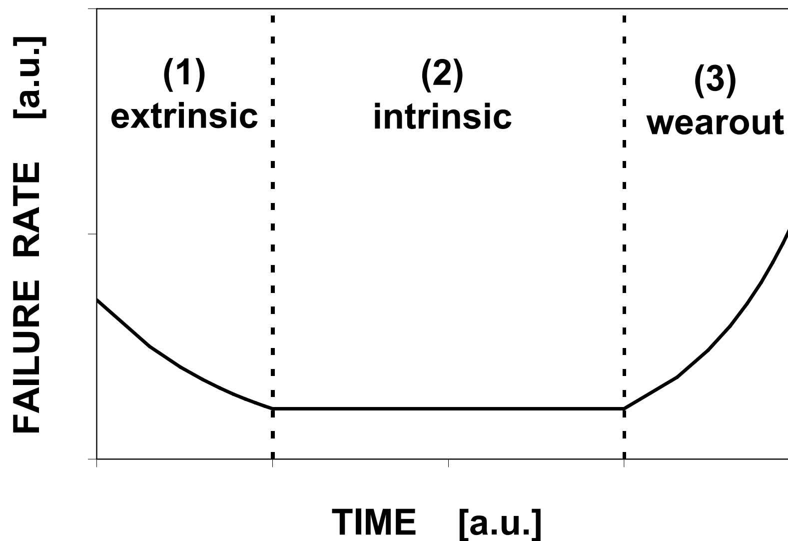

Figure 1.7. Failure rate vs. time bathtub curve.

When the weakest link in a system causes the entire system to fail, the distribution follows a smallest extreme value distribution. According to [61], the application of extreme value theory to reliability failures dates back at least to 1926, when its use was first reported in the textile industry. It was first demonstrated to apply to dielectric breakdown in 1986 [62]. The connection between Weibull and smallest extreme value distributions is that if a variable t has a Weibull distribution, then ln(t) has an extreme value distribution. The CDF for the Weibull distribution is:

$$F(t) = 1-\exp(-(t/\eta)^{\beta}) \quad (1.13)$$

β is the shape parameter or Weibull slope, and η is the scale parameter or characteristic life. For a unimodal distribution, the Weibull slope is independent of area, voltage, and temperature [38] and depends only on oxide thickness [25]. The Weibull slope decreases with decreasing thickness, reflecting the larger statistical spread in the smaller trap densities required to form a breakdown path across thinner oxides [25].

From (1.13), when $t = \eta$, then $F(t) \sim 0.63$. η is often referred to as $t_{63\%}$. The median failure time, when $F(t) = 0.5$, is called $t_{50\%}$. Inserting (1.13) into (1.12), the AFR for the Weibull distribution is:

$$AFR = \frac{(t_2/\eta)^{\beta} - (t_1/\eta)^{\beta}}{t_2 - t_1} \tag{1.14}$$

For $\beta < 1$, AFR decreases with time and describes region 1 (extrinsic) of the bath tub curve. For $\beta < 1$, without defect screening, an AFR reliability requirement cannot be met at all times less than the product warranty period [63]. For $\beta > 1$, AFR increases with time and models region 3 of the bath tub curve. For $\beta = 1$, AFR = $1/\eta$ (constant) and describes region 2 (intrinsic) of the bath tub curve. For $\beta = 1$, the Weibull function reduces to the exponential distribution. Setting $\lambda = 1/\eta$, the exponential PDF is:

$$f(t) = \lambda e^{-\lambda t} \tag{1.15}$$

An important property of the exponential distribution is that its constant failure rate implies a lack of memory property. This means that the occurrence of failure is not influenced by the amount of time that the device has previously been stressed.

In analyzing Weibull distributions, a quantity called the Weibit (W) is defined, so that plotting W vs. ln t_{BD} (or W vs. ln Q_{BD}) yields a straight line with slope β:

$$W(Q_{BD}) = \ln[-\ln(1-F)] = \beta \ln(Q_{BD}/\eta) \tag{1.16}$$

It can be shown that trap generation must follow a power law in fluence (1.3) to obtain the Weibull distribution (1.16) for the fluence to breakdown [64].

1.5.3 Experimental considerations for applying Weibull statistics

To use Weibull statistics for reliability assessments, a method is needed for determining the cumulative fail fraction $F(t_i)$. The effect of finite sample sizes on the estimation of the Weibull parameters should be also comprehended.

For a sample population N, a few possibilities for calculating $F(t_i)$ are (a) $i/(N+1)$, (b) $(i-0.3)/(N+0.4)$, and (c) $(i-0.5)/N$. All are cases of the Generalized Hazen Formula [60]

$$F(t_i) = (i - \alpha)/(N - 2\alpha + 1) \tag{1.17}$$

The latter two formulations (b) and (c) reduce statistical bias so that the CDF value better approximates the "true" median value for $F(t_i)$ that would result if the experiment were repeated an infinite number of times. While the validity of these formulas is debated in the literature, the differences between them are small. To illustrate this point, for the tenth failure out of 50 units, $F(t) = 0.196$ for (a), $F(t) = 0.192$ for (b), and $F(t) = 0.190$ for (c). Our Monte Carlo simulations of Weibull distributions for N ranging

from 50 to 5,000 units also show little difference between these approximations. To ensure consistency, we always use the same formula: $F(t_i) = i/(N+1)$.

We next address the impact of sample size on Weibull parameter estimates. In our typical TDDB experiments, we employ 4-5 stress voltages with about 50 devices each so that 200 to 250 devices are stressed in aggregate. $t_{50\%}$ and β can be determined for each stress leg using either the method of maximum likelihood or a regression analysis. It can be shown that averaging the 4-5 separately determined Weibull slopes is approximately the same as if 200 to 250 devices were stressed at a single stress condition [65, 66]. Accordingly, our experimental space is sufficient to ensure that there is a 90% probability that the Weibull slope is known within +/- 10% accuracy [66].

1.6 Reliability projection

To project reliability from test data to product circuits, voltage (or field), temperature, area, and thickness must all be appropriately derated. In this section, we will introduce some of the basic reliability modeling concepts.

1.6.1 Field and voltage dependence

Projecting time to breakdown (t_{BD}) from stress to operating voltage requires a function that scales with either oxide electric field or gate voltage. There are 2 field based extrapolation methods, widely known as the 1/E-Model [3] and E-Model [1,2]. The 1/E-Model and E-Model are, respectively

$$\text{1/E-Model:} \quad t_{BD} = t_{0C}(A_G, T)\exp(B_C/E_{OX}) \qquad (1.18)$$

$$\text{E-Model:} \quad t_{BD} = t_{0E}(A_G, T)\exp(-B_E{*}E_{OX}) \qquad (1.19)$$

where t_{0C} is the 1/E-Model pre-factor, A_G is the gate area, T is the temperature, E_{OX} is the oxide electric field, and t_{0E} is the E-Model pre-factor. The field acceleration factor is defined as:

$$AF_E = -\partial \ln(t_{BD})/\partial E_{OX} \qquad (1.20)$$

The field acceleration factors for the 1/E and E Models are, respectively:

$$AF_E(\text{1/E-Model}) = B_C/E_{OX}^2 \qquad (1.21)$$

$$AF_E(\text{E-Model}) = B_E = \text{constant} \qquad (1.22)$$

Because the 1/E-Model AF_E increases with decreasing E_{OX}, its lifetime projections are more optimistic than the E-Model. There are 2 formulations for voltage driven breakdown models [7-11]. One is exponential in V_G and the other is a power law in V_G:

$$\text{Exponential Law Model:} \quad t_{BD} = t_{0V}(A_G, T, EOT)\exp(-B_V{*}V_G) \qquad (1.23)$$

Power Law Model: $t_{BD} = a(A_G, T, EOT)V_G^{-N}$ (1.24)

Where t_{0V} is the exponential law pre-factor, B_V is the voltage acceleration factor, a is the power law pre-factor, and N is the power law exponent. In contrast to the E and 1/E models, the V_G-Model thickness (EOT) dependence is carried in the pre-factors. The voltage acceleration factor is defined as:

$$AF_V = -\partial \ln(t_{BD})/\partial V_G \quad (1.25)$$

Power law lifetime projections are more optimistic than the exponential model because the exponential law acceleration factor is constant whereas the power law acceleration factor increases with decreasing voltage:

$$AF(\text{exponential law}) = B_V = \text{constant} \quad (1.26)$$

$$AF(\text{power law}) = N/V_G \quad (1.27)$$

Breakdown models will be discussed in more detail in Chapter 3 and the mechanisms that cause breakdown will be elaborated in Chapter 4.

1.6.2 Area scaling

We will now discuss how the Weibull distribution handles the area scaling aspects of t_{BD} and AFR. Assuming that the defects that cause breakdown are randomly distributed over the area of the gate oxide, Poisson statistics will describe the area dependence:

$$F = 1-\exp(-D^*A_G) \quad (1.28)$$

Re-writing (1.28) to obtain the Weibit on the left hand side of the equation:

$$\ln(-\ln(1-F)) = \ln(D^*A_G) \quad (1.29)$$

Taking 2 capacitors of different areas A_{G1}, A_{G2} that have the same defect density, we use (1.29) for the corresponding CDF's F_1, F_2 to write the difference in the Weibits:

$$\ln(-\ln(1-F_2)) - \ln(-\ln(1-F_1)) = \ln(A_2/A_1) \quad (1.30)$$

Increasing the gate area by a factor A_2/A_1 shifts the Weibit by a factor $\ln(A_2/A_1)$. This relation can be derived using the binomial distribution [67]. As β is independent of area [67], combining (1.13) and (1.30) yields the area scaling relationship

$$\eta_2/\eta_1 = (A_1/A_2)^{1/\beta} \quad (1.31)$$

Using (1.12) and (1.28) the area scaling relationship for the AFR is:

$$AFR_2/AFR_1 = A_{G2}/A_{G1} \quad (1.32)$$

1.6.3 Ramped voltage breakdown

Ramped voltage breakdown (RVB) is widely employed to characterize dielectrics. A common misconception is that E_{BD} is a material constant. We will show that this is not true. For any wave form $\tau(E_{OX})$ where Eox is varied in time, for a device stressed to breakdown

$$\int_0^{t_{BD}} dt/\tau(E_{OX}) = 1 \tag{1.33}$$

(1.33) is called the Lifetime Integral [2]. If E_{OX} is ramped linearly with time, then

$$E_{OX}(t) = R^*t \tag{1.34}$$

$$E_{BD} = R^*t_{BD} \tag{1.35}$$

where R is the ramp rate and E_{BD} is the oxide electric field at breakdown. For the E- Model, inserting (1.19), (1.34), and (1.35) into (1.33):

$$\int_0^{E_{BD}/R} dt/t_{0E}^*\exp(-B_E Rt) = 1 \tag{1.36}$$

Solving (1.36) for E_{BD},

$$E_{BD} = B_E^{-1}\ln(B_E R t_{0E} + 1) \tag{1.37}$$

(1.37) shows that E_{BD} increases with increasing ramp rate. It is also a function of temperature and gate area through the t_{0E} term (see (1.19)). E_{BD} vs. ramp rate data are shown for a 60Å oxide in Figure 1.8. It can be seen that at a fixed temperature and gate area, E_{BD} can be significantly modulated and is therefore not a material constant. Accordingly, in this work, we will only use RVB for relative comparisons.

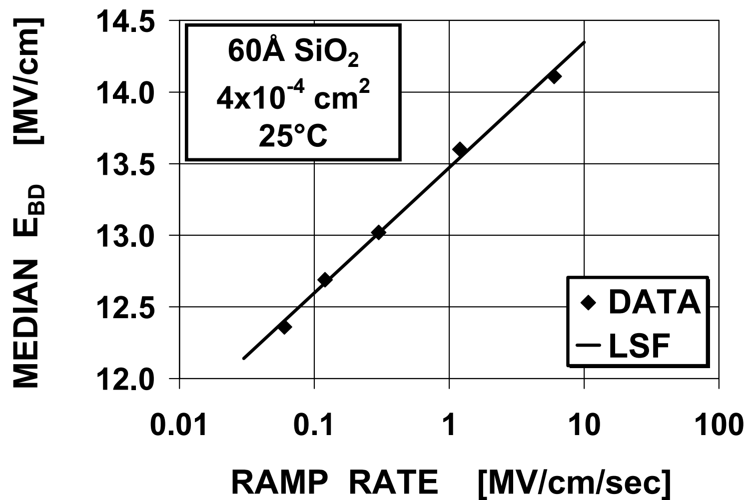

Figure 1.8. E_{BD} vs. ramp rate data and least squares fit (LSF) for a 60Å NMOS oxide.

1.6.4 Temperature dependence

When only thermal stresses are significant, the temperature dependence of oxide breakdown is commonly assumed to follow an Arrhenius Model. The Arrhenius Model predicts that the reaction rate increases exponentially with temperature [68]. The time to breakdown is inversely proportional to the reaction rate. Strictly speaking, the Arrhenius relationship applies to single step reactions in equilibrium [68]. Using the exponential V_G Model to illustrate Arrhenius temperature scaling of the time to breakdown:

$$t_{0V}(T) = t_0 \exp(E_A/k_B T) \quad (1.38)$$

where $t_{0V}(T)$ is the temperature dependent part of the V_G Model pre-factor, t_0 is a constant, E_A is the activation energy for the reaction to occur, and k_B is the Boltzmann constant. Also of relevance to our work, the Arrhenius equation is used to model the temperature dependence of solid state diffusion processes. For a system with an intermediate metastable state between reactants and products, a more complex temperature model described by the Eyring equation is used:

$$t_{0V}(T) = AT^z \exp(\Delta H/k_B T)\exp(-\Delta S/k_B) \quad (1.39)$$

Where A and z are constants, ΔH is the activation enthalpy, and ΔS is the activation entropy. For a system with thermal plus an additional stress, the Eyring equation becomes [58,69,70]:

$$t_{0V}(T, D) = AT^z \exp(\Delta H/k_B T)\exp(-\Delta S/k_B)\exp((B + C/T)D) \quad (1.40)$$

Where B and C are constants and D represents the intensive parameter corresponding to the 2ND stress. Additional stress terms may be added to (1.40). The problem with the application of (1.40) is that there are 5 unknown constants that must be determined and each additional stress type adds 2 additional parameters. In contrast, there are only 2 fitting parameters for the Arrhenius equation. In practice, the simpler Arrhenius relationship works well in many cases. The Eyring formulation has been applied to thick oxides to model temperature dependent field acceleration factors and field dependent activation energies [69,70]. Since voltage acceleration is temperature-independent in ultra-thin dielectrics [71], the Arrhenius model can be applied in this regime. However, a temperature dependence that fits neither Arrhenius nor Eyring formulations has been observed in ultra-thin oxides [72]. This is known as the "linear" model [73], where

$$t_{0V}(T) = t_0 \exp(-\alpha T) \quad (1.41)$$

Empirically, the temperature model that best fits appears to be process dependent [71]. Deviations from Arrhenius behavior have been reported below (approximately) room temperature [71]. Accordingly, temperature modeling of ultra-thin dielectric breakdown remains an open issue.

The Lifetime Integral (1.33), with t(E) replaced by t(T), can be used to calculate an effective operating temperature T_{EFF} of a product where the various chip functions

operate at different temperatures if the approximate operating time of each state is known within the warranty period. Since products typically contain a relatively small number of circuit blocks that are operating at different temperatures, the integral in (1.33) is replaced by a summation, so that for an Arrhenius Model:

$$T_{EFF} = -E_A/[k_B \ln(Y^{-1}\sum_i t_i \exp(-E_A/k_B T_i))] \qquad (1.42)$$

where Y is the warranty period and t_i is the time of the i^{TH} circuit block at temperature T_i.

1.7 Band bending and electric fields

Understanding dielectric reliability and carrier transport require knowledge of the electric fields, potentials, and energies that will be encountered in experiments and device simulations. We will assume that the reader is well versed in basic device physics, so we will briefly review the concepts that are most important to our work. Unless otherwise noted, the material contained in this section follows [74].

1.7.1 NMOS energy band diagram

A 1 dimensional energy band diagram for an NMOS capacitor with a p-type substrate (PWELL) and heavily doped n+ poly gate electrode with a positive gate voltage applied is shown in Figure 1.9.

The terminology employed is defined as follows:

E_{FN} = Fermi level in n+ poly
E_{FP} = Fermi level in p-type silicon
E_I = Intrinsic Fermi level
E_C = Silicon conduction band
E_V = Silicon valence band
E_G = Silicon band gap
E_{OX} = Oxide electric field
n_i = Intrinsic concentration
n_s = Electron concentration at pwell – oxide interface
N_A = Pwell acceptor doping concentration
N_D = N+ poly donor doping concentration
V_G = Gate Voltage
V_{MG} = Mid-gap Voltage
V_{OX} = Oxide voltage
X_{SI} = Silicon electron affinity
X_{OX} = Oxide electron affinity
Ψ_S = Surface potential in p-type silicon at oxide interface
Ψ_{Si} = Band bending in p-type substrate
Ψ_G = Band bending in n+ poly gate
Ψ_{BN} = Bulk potential in n+ poly gate
Ψ_{BP} = Bulk potential in p-type substrate

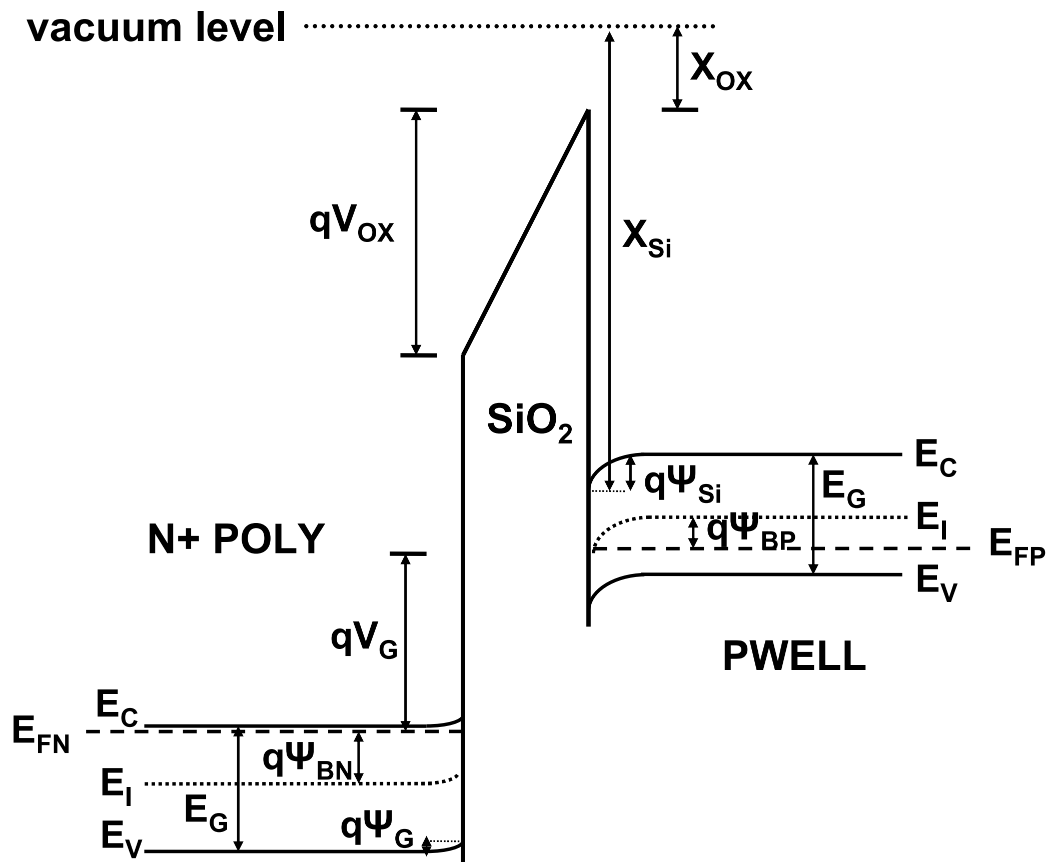

Figure 1.9. Energy band diagram for an n poly over pwell NMOS device with arbitrary oxide thickness. The pwell surface is weakly inverted. The bottom portion of the oxide energy gap has been omitted for clarity.

We use the pwell Fermi level (E_{FP}) as our energy reference. By applying a positive bias to the gate, the electrostatic potential of the n+ poly decreases with respect to the pwell. In the absence of carrier injection, the gate voltage is equal to the separation between the equilibrium Fermi levels E_{FP} and E_{FN}. A positive gate voltage creates a positive space charge at the n+ poly surface and induces an equal negative charge in the pwell. Surface charge results in an electric field and electrostatic potential (band bending) in the region in the semiconductor where the space charge exists, and goes to zero in the neutral bulk. Silicon band bending is positive when the bands are bent downward and corresponds to the accrual of negative space charge. The following conditions of surface potential in the pwell are

$\Psi_S < 0$	Accumulation of holes (majority carriers)
$\Psi_S = 0$	Flatband
$\Psi_B > \Psi_S > 0$	Depletion of majority carriers
$\Psi_S = \Psi_B$	Midgap, $n_S = n_I$ (intrinsic)

$\Psi_S > \Psi_B$ Inversion
$\Psi_S > 2\Psi_B$ Strong inversion

From Gauss's law, electric fields at the semiconductor surfaces give rise to an electric field in the oxide. The potentials and electric fields in the poly, pwell, and oxide are found by solving the 1 dimensional Poisson equation (classical solution). For thinner oxides at lower voltages, better accuracy is required and is obtained by simultaneously solving Poisson and Schrödinger equations (quantum solution). In our work, we utilize quantum solutions for dielectrics with thickness less than 40Å.

1.7.2 Relationship between gate voltage, oxide voltage, and band bending

Because the oxide electric field plays an important role in dielectric breakdown and transport, we will derive the 1 dimensional V_{OX} vs. V_G relationship to illustrate some of the basic features of electric fields in gate oxides ($E_{OX} = V_{OX}/t_{OX}$). Starting at the pwell Fermi level, the energy balance equation is

$$E_G - E_{FP} - \Psi_{Si} + X_{SI} = X_{SI} + V_{OX} + \Psi_G + E_{FN} - E_G - V_G \quad (1.43)$$

For uniform doping, at flat band, no charge, field, or voltage drop exists in any region of the device so that

$$V_G|_{Vox=0} = (E_{FP} - E_{FN})|_{Vox=0} = V_{FB} \quad (1.44)$$

For non-uniform doping, charge, field, and band bending cannot all be simultaneously zero so that strictly speaking, (1.44) is valid for uniform doping.

Re-arranging terms and inserting (1.44) into (1.43) yields:

$$V_{OX} = V_G - \Psi_{Si} - \Psi_G - V_{FB} \quad (1.45)$$

1.7.3 Quantum Effects

The application of quantum mechanics to the calculation of band bending and electric fields introduces corrections to the classical results when large numbers of free carriers are present. First, the peak of the free carrier distribution in an inversion or accumulation layer is expected to lie at the semiconductor surface in the classical world. Invoking quantum mechanics, the peak is displaced by about 10Å from the interface and is broader than classical predictions [75]. This results in an overestimate of the oxide thickness extracted from capacitance measurements. Another important aspect is that the density of states (DOS) are not a continuum, but are quantized and are therefore smaller than the classical result. Combined with the thickness effects, the band bending required to obtain a given free charge density is larger in the quantum solution [75,76].

1.7.4 Some important properties of oxide electric fields

- V_{FB} is typically on the order of -1V in an NMOS device. Accordingly, from (1.45), V_{FB} increases E_{OX} for inversion and reduces E_{OX} for accumulation.

- Since band bending is small in an electrode that is accumulated, E_{OX} is not sensitive to either n+ poly or PWELL doping levels in an accumulated NMOS device.

- For an inverted pwell, E_{OX} decreases as n+ poly doping decreases. E_{OX} also decreases with decreasing pwell doping, but this effect diminishes at high fields provided that N_D(poly) >> N_A(pwell).

- Once the pwell is strongly inverted or accumulated, band bending increases weakly with V_G. Under these conditions, E_{OX} is approximately a linear function of V_G [38]

  $$E_{OX} = k_1 V_G + k_0 \tag{1.46}$$

  where k_1 and k_0 are constants.

- The temperature dependence of the oxide field comes from Ψ_{Si} and Ψ_G through the temperature dependence of the respective majority carrier concentration. Accordingly, E_{OX} is insensitive to temperature between 25°C to 125°C [77].

- Even though t_{OX} is larger in a classical solution, at a given V_G, the quantum E_{OX} is lower compared to the classical result under strong accumulation or inversion, as the larger classical t_{OX} is offset by the higher pwell band bending in the quantum solution [75,76]. The classical counterpart is that E_{OX} is lower when Fermi-Dirac rather than Maxwell-Boltzmann statistics are utilized [77].

1.8 Equivalent oxide thickness

While the terminology equivalent oxide thickness (EOT) is widely used, its meaning is sometimes ambiguous. In this section, we will show from basic electrostatics that the equivalent oxide thickness of an arbitrary dielectric corresponds to the oxide thickness that has the same E_{OX} at a given V_G.

We will derive the equivalent oxide thickness for an NMOS device with an oxide-nitride stack as shown in Figure 1.10. We assume that there are no charges in the insulators, the pwell and poly are conductors in the neutral bulk, and pwell and n+ poly have the same permittivity. We define a 1-dimensional coordinate system with origin $x = 0$ at the neutral – depleted poly boundary and $x = t_T$ at the neutral – depleted pwell boundary. $V(0) = V_G$ and $V(t_T) = 0$.

The thicknesses of the poly space charge region, nitride, oxide, and pwell space charge region are t_P, t_N, t_{OX}, and t_{Si} respectively. We will solve the boundary value problem [78]

$$V_G = \int_0^{t_T} E\,dl \tag{1.47}$$

Using Gauss's Law

$$D = \varepsilon_i E_i = Q \tag{1.48}$$

where ε_i and E_i are the permittivity and field in the i^{TH} dielectric and D is the displacement field, we integrate (1.47) to obtain

$$V_G = Q(t_P/\varepsilon_{Si} + t_N/\varepsilon_N + t_{OX}/\varepsilon_{OX} + t_{Si}/\varepsilon_{Si}) \tag{1.49}$$

Re-writing (1.49) using C = Q/V:

$$V_G/Q = 1/C = 1/C_P + 1/C_N + 1/C_{OX} + 1/C_S \tag{1.50}$$

where C_P, C_N, C_{OX}, and C_S are the capacitance of the poly space charge, nitride, oxide, and pwell space charge respectively. If Q is known at a given V_G, then E_{OX} is known:

$$E_{OX}(V_G) = Q(V_G)/\varepsilon_{OX} = D(V_G)/\varepsilon_{OX} \tag{1.51}$$

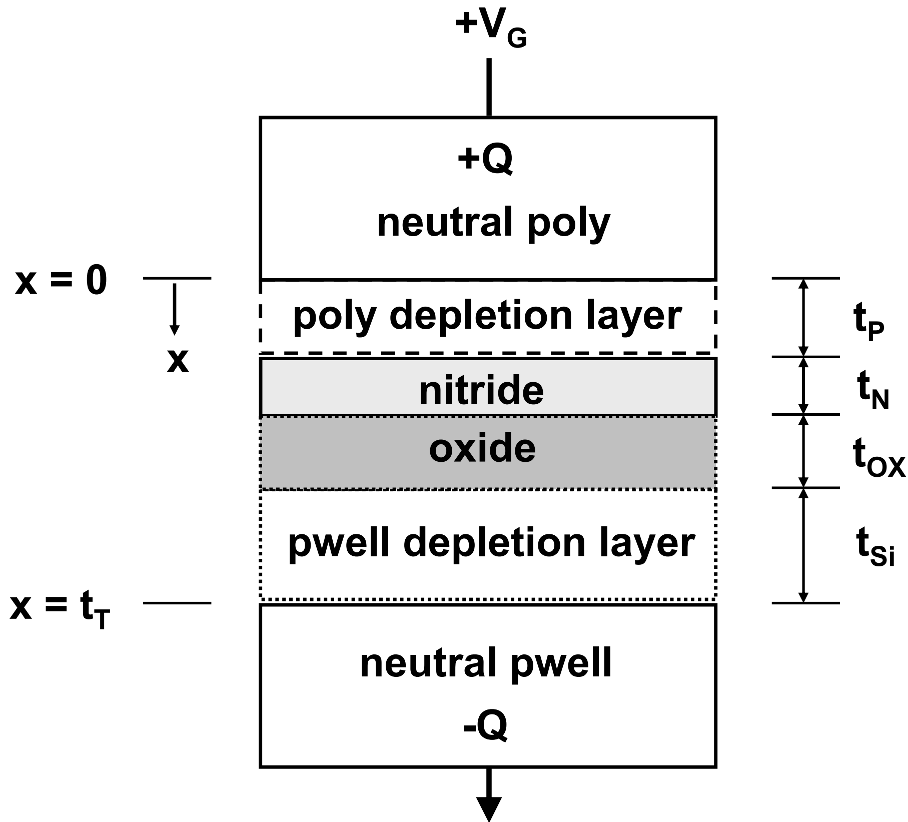

Figure 1.10. Stacked gate dielectric system for equivalent oxide thickness derivation.

Let us now treat the oxide – nitride stack as a single dielectric layer so that

$$1/C_{O'} = 1/C_N + 1/C_{OX} \tag{1.52}$$

Then (1.50) becomes

$$V_G/Q = 1/C = 1/C_P + 1/C_{O'} + 1/C_S \quad (1.53)$$

At a given V_G, if the value of $C_{O'}$ is picked using (1.52), then the charge Q in (1.53) will be the same as in (1.50). Therefore, the displacement field is unchanged. Accordingly, the electric field in the equivalent dielectric will be the same E_{OX} found in (1.51). Therefore, from (1.52), the electrical thickness of the equivalent dielectric is

$$EOT = \varepsilon_{OX}/C_{O'} = t_{OX} + (\varepsilon_{OX}/\varepsilon_N)t_N \quad (1.54)$$

For example, a 20Å stack with 10Å oxide and 10Å nitride will have an EOT of 15Å. In practice, the physical film thicknesses and dielectric constants of the stack components are not all known. By treating the stack as an equivalent oxide, the V_{OX} vs. V_G relation and EOT are found by matching an experimentally obtained C-V curve with a classical or quantum solution generated C-V curve. This simplification is a consequence of the uniqueness of electrostatic potentials theorem, which is that an electrostatic boundary value problem has exactly 1 solution [78]. However, the converse of this theorem is NOT true. There may be an infinite number of systems with the same solution. We have exploited this principle by replacing one system (the oxide – nitride stack) with a simpler system (an oxide) that has the same solution, but is easier to solve. The general form for (1.54) is

$$EOT = \sum_i (\varepsilon_{OX}/\varepsilon_i)t_i \quad (1.55)$$

For a pure SiON film with a physical thickness of t_{SiON}, (1.55) becomes

$$EOT = (\varepsilon_{OX}/\varepsilon_{SiON})t_{SiON} \quad (1.56)$$

In summary, the EOT of a dielectric stack corresponds to the oxide thickness that has the same V_{OX} vs. V_G relationship.

1.9 Capacitance-voltage characteristics

Because C-V characteristics are widely used in device simulators to determine electric fields, band bending, and EOT for the study of dielectric reliability and transport, we will briefly discuss the basic physics and measurement techniques. Unless otherwise specified, the review material follows [74].

1.9.1 Components of the 1-dimensional capacitance

From (1.45) the MOS device is a voltage divider with voltage drops across the poly space region, oxide, and pwell space charge region (3 capacitors in series). The voltage dependent capacitance of the poly and pwell space charge regions follows the free carrier response at the space charge – neutral region edge. In depletion, the doping

density of the pwell is found from the slope of the C-V curve. For metal gates or thick oxides, Ψ_G and t_P can be neglected so that

$$C = \frac{C_{OX}C_S}{C_{OX} + C_S} \quad \text{(depletion)} \quad (1.57)$$

$$C_S = \partial Q/\partial \Psi_{Si} \quad (1.58)$$

For a low frequency measurement (e.g. 1 Hz), $C \sim C_{OX}$ in strong accumulation or strong inversion. For strong accumulation, this is due to the small band bending resulting from the large majority density at the interface, i.e. the accumulation charge increases exponentially with band bending so that $C_S >> C_{OX}$. For strong inversion, the inversion charge also increasing exponentially with band bending and the inversion layer screens the space charge region from the surface field so that there is little majority carrier response to the gate voltage. Therefore, $C_S >> C_{OX}$ in inversion and for a metal gate device,

$$C = C_{OX} \quad \text{(inversion)} \quad (1.59)$$

$$C = C_{OX} \quad \text{(accumulation)} \quad (1.60)$$

For ultra thin gate dielectrics with poly gate electrodes, the poly space charge capacitance cannot be neglected so that

$$C = \frac{C_{OX}C_SC_P}{C_{OX}C_S + C_SC_P + C_PC_{OX}} \quad \text{(depletion)} \quad (1.61)$$

$$C = \frac{C_{OX}C_P}{C_{OX} + C_P} \quad \text{(inversion)} \quad (1.62)$$

$$C = C_{OX} \quad \text{(accumulation)} \quad (1.63)$$

$$C_P = \partial Q/\partial \Psi_G \quad (1.64)$$

For a poly gate device, C_S is still given by (1.58). C-V curves generated from classical or quantum solvers first match the experimental data in accumulation where $C = f(C_{OX})$. Once C_{OX} is known, the solution is matched to the data in inversion where $C = f(C_{OX}, C_P)$. Once C_{OX} and C_P are known the solution is matched to the depletion part of the C-V data where $C = f(C_{OX}, C_P, C_S)$.

1.9.2 Effect of interface traps and oxide fixed charge

Oxide fixed charges are charge states in the oxide that do not change occupancy during the C-V sweep. Accordingly, oxide fixed charge induces a parallel shift in the C-V curve

at any applied bias. In contrast, interface traps change occupancy with voltage when the trap state is swept through the Fermi level. A trap that is an acceptor state is neutral when empty and negatively charged when occupied by an electron (0/-). A trap that is a donor state is neutral when occupied by an electron and positively charged when empty (0/+). Accordingly, for an NMOS device with a heavily doped pwell ($E_F << E_I$), donor states will shift V_{FB} but will not significantly affect V_T, whereas acceptor states will shift V_T but will not significantly affect V_{FB}.

When free carriers are present in numbers that are not significantly greater than the interface trap density (i.e. weak accumulation or weak inversion), interface traps add a measurable capacitance that is in parallel with the space charge capacitance. Figures 1.11 and 1.12 show the effect of interface traps on the equivalent circuits for metal gate and poly gate electrodes respectively. While it is generally understood that traps at the pwell – oxide interface need to be accounted for in fitting C-V data, for poly gate devices, if the trap density is sufficiently high, interface traps at the poly - oxide interface cannot be ignored in C-V matched solutions without sacrificing some accuracy. While higher doping levels decrease the influence of interface traps, interface states can still be a factor even in degenerately doped poly. In the presence of interface traps, C'_S replaces C_S in (1.57) and C'_P replaces C_P in (1.61) and (1.62).

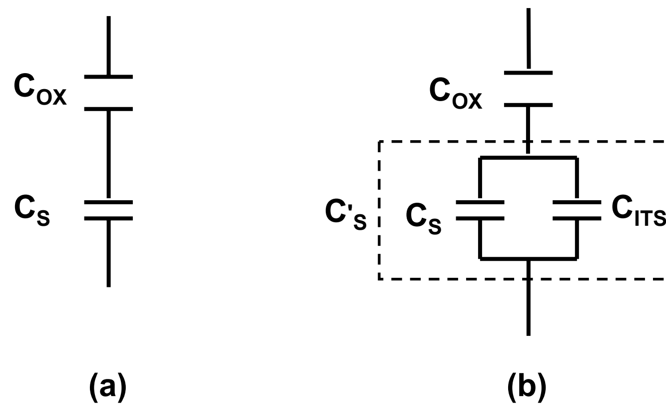

Figure 1.11. Equivalent circuit for a metal gate device (a) ideal (b) with interface traps at Si-SiO_2 interface.

1.9.3 Techniques for measuring C-V characteristics

If a small sinusoidal voltage (typically 20 mV – 50 mV) is superimposed on a C-V sweep, the capacitance will not rise back to C_{OX} in inversion, as the pwell space charge capacitance will saturate near its minimum because minority carrier generation cannot keep up with the ac signal (typically 100 KHz – 1 MHz). This is known as the small-signal high frequency method [79] and has been used extensively to determine oxide thickness and pwell doping densities of thick gate oxide devices. To determine the density of interface traps, a low frequency C-V measurement is utilized [80]. Unless the trap density is very large, interface states do not respond to a high frequency C-V measurement so that the doping density is obtained from the high frequency

measurement, and the interface trap density is determined by comparing the difference in C_S between high and low frequency curves. A quasi-static measurement [81] can be used in lieu of a low frequency sweep, where the capacitance is proportional to the displacement current resulting from a known voltage ramp rate. For poly gate devices where knowledge of the poly doping is desired, it is necessary to acquire a low frequency C-V curve to obtain C_P to solve (1.61) and (1.62).

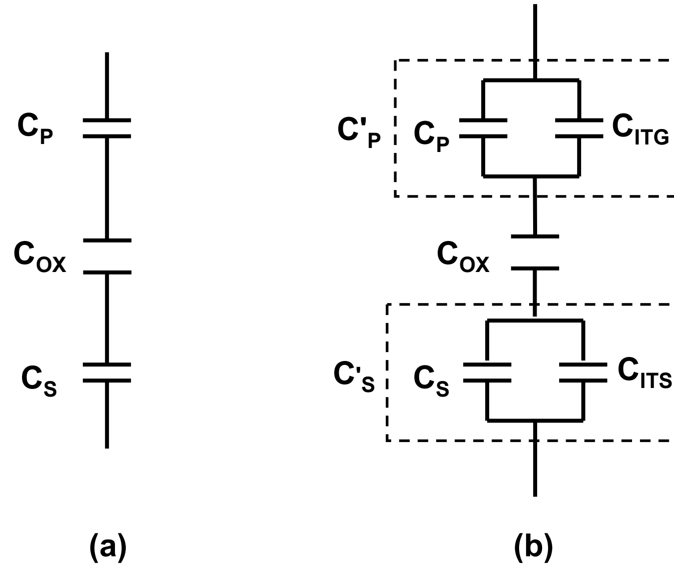

Figure 1.12. Equivalent circuit for a poly gate device (a) ideal (b) with interface traps at Si-SiO_2 and poly-SiO_2 interfaces.

The combination of high frequency and low frequency or quasi-static measurements (i.e. high-low methods) has several limitations: (a) If the doping is sufficiently high or the oxide is sufficiently thin, minority carrier generation rates cannot supply sufficient charge to maintain the equilibrium inversion charge density unless the ramp time is impractically long. (b) High gate leakage induces a shunt resistor in parallel with the oxide capacitance and causes skewing of the C-V curve [82]. For a quasi-static sweep from accumulation to inversion, the leakage opposes the displacement current and decreases the net capacitance in accumulation. When inverted, the leakage adds to the displacement current and increases the net capacitance, as shown in Figure 1.13. (c) the capacitor area must be sufficiently large to attain sufficient noise margin of the displacement current. (d) If the interface trap density is > 10^{11} $cm^{-2}eV^{-1}$, the high frequency curve will become distorted [83]. This effect is known as stretch-out, because the threshold voltage is shifted to higher values since charge conservation can be attained by capturing minority carriers in trap states rather than by depleting majority carriers. In practice, the high-low method produces acceptable results for hydrogen passivated interfaces when t_{OX} > 50Å.

The issue with minority carrier response (a) can be addressed by connecting the source and drain into the measurement circuit to supply minority carriers. This configuration will

yield low frequency C-V curves at moderate to high frequencies. This method is adequate down to about 25Å to 30Å (depending on the leakage of the dielectric). This approach can still be used to determine the change in capacitance resulting from stress between weak accumulation and weak inversion down to 12Å, but cannot be used for accurate EOT extractions at that small a thickness.

The charge-voltage (Q-V) [83] method can mitigate to some extent all the issues (a)-(d) outlined above. A known bias-independent capacitor is placed in series with the MOS device and the change in the internal node voltage (between capacitor and MOS gate) resulting from a bias applied to the series combination is measured. The change in charge at the internal node is the negative of the change in charge across the MOS device. This results in high noise immunity because large voltages are measured rather than small currents. This also enables C-V characterization of small area capacitors and slowly equilibrating devices. The technique provides a leakage correction by measuring the change in charge over a specific time interval for each step of the voltage ramp.

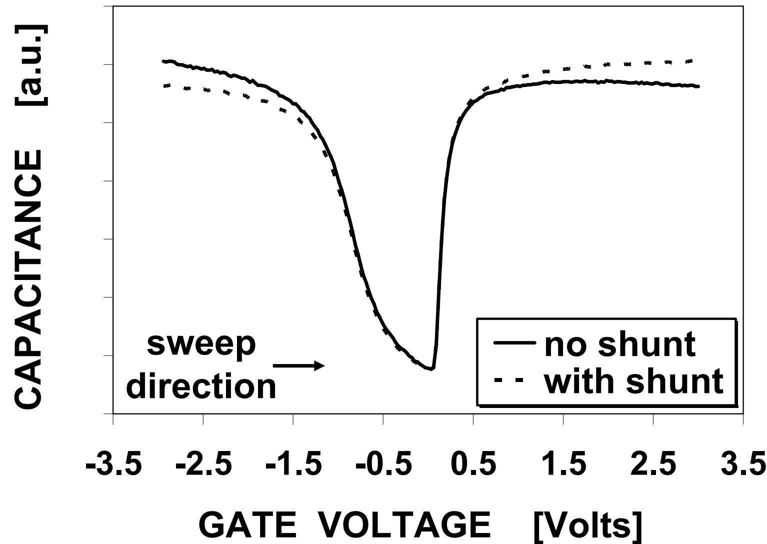

Figure 1.13. Effect of shunt resistance due to gate leakage on a quasi-static C-V measurement. With the sweep direction is from $-V_G$ to $+V_G$, the shunt path decreases the net capacitance in accumulation and increases in inversion.

The charge-capacitance (Q-C) method is an extension of the Q-V method [84]. In the Q-C method, additional stray and shunt resistances are factored in and a high frequency C-V is done simultaneously with the low frequency charge measurement. A method for accurately determining the Ψ_{Si} vs. V_G relationship without integration or differentiation naturally arises from the Q-V and Q-C methods. In the Q-C method, the device is driven into deep depletion with a large signal high frequency pulse to extract the true high frequency capacitance without the influence of interface traps. The doping profile is determined by taking the derivative of C_S with respect to Ψ_{Si} rather than V_G to eliminate the effects of stretch-out due to interface traps.

Between about 15Å to 30Å (depending on the leakage of the dielectric), multi-frequency measurements are employed [85]. In this method, the capacitance is extracted from the measurement of the impedance at two different frequencies.

Below 15Å to 20Å, the leakage may become sufficiently large that R-F C-V measurements at frequencies approaching 1 GHz must be made to obtain a useful C-V curve [86,87]. Employing this method requires specially designed test structures comprised of multiple poly fingers of short channel length with a large number of contacts to minimize series resistance [87]. R-F techniques have been shown to be valid at current densities as high as 1,000 $Amps/cm^2$ [87].

In summary, high gate leakage, presents significant challenges for acquiring accurate C-V data for device simulations. Extraction of interface trap densities is complicated by the presence of trap states at both top (poly) and bottom (pwell) interfaces. To accurately model electric fields and band bending, these effects must be factored in the device simulator.

1.10 Chapter summary

Despite years of investigation by many workers, significant gaps and challenges remain in the understanding of dielectric reliability physics. In the past decade or so, the primary economic driver for dielectric reliability studies has been deep thickness scaling and the consequences thereof such as higher electric fields, the onset of direct tunneling, the necessity to incorporate additional elements such as nitrogen in SiO_2 films, and discontinuities in the breakdown physics. As the thickness of SiON films approach the ultimate scaling limit, a new trend has emerged, which is a reduction in the rate of which the power supply voltage is scaled down with successive technology nodes to achieve higher performance. Since ultra thin dielectric breakdown is a voltage driven phenomenon, this direction has become a significant concern.

Except at very low fields, charge trapping under electron direct tunneling stress conditions is negligible in device grade SiO_2 and SiON films so that the generation of traps is the most important aspect of degradation prior to breakdown. When breakdown occurs, the breakdown can be soft or hard depending on the details of the stress conditions. In the literature, there are several opinions regarding the transport mechanism after soft breakdown. For deeply scaled dielectric films, there is some difficulty in the identification of the mechanism due to the small voltage excursion between the onset of strong inversion (or strong accumulation) and the breakdown of the dielectric. Accordingly, soft and hard breakdown are differentiated from the magnitude of the post breakdown resistance. Empirically, the resistance is typically less than 10KΩ after hard breakdown occurs.

To justify more aggressive operating conditions, several techniques have been proposed to extend the time to failure beyond the first breakdown event. The complexity of the degradation modes and circuit vulnerabilities make it a formidable challenge to

manage the risk of applying these methods to actual products. Accordingly, since the time to first breakdown is the safest metric to use in constructing reliability models, the detection of soft breakdown is a critical issue. Several techniques have been developed to address this problem such as a change in slope in the well-behaved trap generation regime, an increase in random telegraph noise, and the identification of a current jump that is consistent with the magnitude required to form a soft breakdown path. The appropriate technique to employ depends on the details of the sample and the stress. Because dielectric breakdown is a weakest link phenomenon, the Weibull distribution is of central importance in gate dielectric reliability. For an intrinsic film, the Weibull slope is related to the thickness dependent statistical spread in the trap density at breakdown. Weibull statistics are easy to apply and describe the area scaling of time to breakdown, charge to breakdown, trap density to breakdown, and failure rate. When using Weibull statistics, these aforementioned parameters are all reckoned from the occurrence of the 1ST breakdown event.

In the E and 1/E TDDB Models, the thickness dependence is carried in the exponent via the electric field, whereas the thickness dependence is contained in the pre-factor in the V_G Model. In all cases, the model pre-factor Weibull scales. While ramped voltage breakdown data are sometimes used for reliability assessments in lieu of TDDB modeling, the breakdown voltage depends on the ramp rate, so the breakdown distribution is arbitrary. Accordingly, absolute reliability projections must be obtained through an analysis of TDDB data. The temperature dependence of TDDB remains an open issue, although an Arrhenius relationship is commonly assumed due to its simplicity. There is some evidence that non-Arrhenius behavior may occur in ultra thin oxides but the conditions leading to this are not understood.

To analyze transport mechanisms and to perform reliability projections using the two field based TDDB models, the oxide electric field must be accurately known. This is accomplished by performing classical or quantum mechanical device simulations of the C-V characteristics. When large numbers of free charges are present, E_{OX} will be lower in the quantum case. Once a device is strongly inverted or accumulated, V_{OX} is linear in V_G, which greatly simplifies analysis. Since E_{OX} is weakly dependent on temperature, the V_{OX} vs. V_G relationship obtained from room temperature can also be applied to higher temperatures. Because band bending is small in an accumulated electrode, E_{OX} is insensitive to poly or pwell doping. When the pwell is inverted, the oxide field will be influenced by the poly doping over a wide range of voltage. For high N+ poly doping levels, E_{OX} will be insensitive to pwell doping in inversion at high voltages.

Although EOT is a loosely applied terminology, the uniqueness theorem for electrostatic potentials can be used to show that the equivalent oxide thickness of an arbitrary dielectric corresponds to the oxide thickness that has the same E_{OX} at a given V_G. Since EOT is obtained in C-V matched device simulations, accurate C-V measurements are essential. Oxide fixed charge causes a voltage independent parallel shift of the C-V curves but the effect of interface traps is influenced by the voltage dependent occupancy of the states. Traps at both Si-SiO_2 and poly-SiO_2 interfaces must be accounted for to accurately fit the C-V data. There are several measurement and device

level issues that complicate the ability to obtain valid capacitance data, such as minority carrier response, high densities of interface traps, signal to noise margin issues, and tunneling current through the dielectric. The later problem can be a significant concern and may require either multi-frequency or RF measurements.

1.11 References

[1] D.L. Crook, ”Method of Determining Reliability Screens for Time Dependent Dielectric Breakdown”, *Proceedings of the IRPS*, 1 (1979)
[2] A. Berman, “Time-0 Dielectric Reliability Test by a Ramp Method”, *Proceedings of the IRPS*, 204 (1981)
[3] I.C. Chen, S. Holland, and C. Hu, “A Quantitative Physical Model for Time Dependent Breakdown in SiO_2”, *Proceedings of the IRPS*, 24 (1985)
[4] K.F. Schuegraf, D. Park, and C. Hu, “Reliability of Thin SiO_2 at Direct Tunneling Voltages”, *IEDM Technical Digest*, 609 (1994)
[5] D.J. DiMaria and E. Cartier, “Mechanism for Stress Induced Leakage Currents in Thin Silicon Dioxide Films”, *Journal of Applied Physics*, **78**, 3883 (1995)
[6] J.H. Stathis and D.J. DiMaria, “Reliability Projection for Ultra Thin Oxides at Low Voltage”, *IEDM Technical Digest*, 167 (1998)
[7] D.J. DiMaria, “Explanation for the Polarity Dependence of Breakdown in Ultra Thin Silicon Dioxide Films”, *Applied Physics Letters*, **68**, 3004 (1996)
[8] J. Bude, B.E. Weir, and P.J. Silverman, “Explanation of Stress Induced Damage in Thin Oxides”, *IEDM Technical Digest*, 179 (1998)
[9] P.E. Nicollian, W.R. Hunter, and J. Hu, “Experimental Evidence for Voltage Driven Breakdown Models in Ultra Thin Gate Oxides”, *Proceedings of the IRPS*, 7 (2000)
[10] M.A. Alam, J. Bude, and A. Ghetti, “Field Acceleration for Oxide Breakdown – Can an Accurate Anode Hole Injection Model Resolve the E vs. 1/E Controversy?”, *Proceedings of the IRPS*, 21 (2000)
[11] E.W. Wu, et. al., “Voltage Dependent Voltage Acceleration of Oxide Breakdown for Ultra Thin Oxides”, *IEDM Technical Digest*, 541 (2000)
[12] S.V. Hattangady, et. al., ”Ultra Thin Nitrogen Profile Engineered Gate Dielectric Films”, *IEDM Technical Digest*, 495 (1996)
[13] C.G. Sodini, and K. S. Krisch, ”Silicon Oxynitride Gate Dielectrics for Scaled CMOS”, *IEDM Technical Digest*, 617 (1992)
[14] D.T. Grider, et. al., ”A 0.18µm CMOS Process Using Nitrogen Profile Engineered Gate Dielectrics”, *VLSI Technical Digest*, 47 (1997)
[15] M. Rodder, et. al., ”A 1.2V, 0.1µm Gate Length CMOS Technology: Design and Process Issues”, *IEDM Technical Digest*, 623 (1998)
[16] P.E. Nicollian, et. al., “Extending the Reliability Scaling Limit of SiO_2 through Plasma Nitridation”, *IEDM Technical Digest*, 545 (2000)
[17] Semiconductor Industry Association, *The International Technology Roadmap for Semiconductors*, 2001 edition, International SEMATECH: Austin, TX, 2001. Semiconductor Industry Association, *The International Technology Roadmap for Semiconductors*, 2004 Update, International SEMATECH: Austin, TX, 2004.

[18] S. Tang, R.M. Wallace, A. Seabaugh, and D. King-Smith, “Evaluating the Minimum Thickness of Gate Oxide on Silicon Using First Principles Method”, *Applied Surface Science*, **135**, 137 (1998)
[19] D.A. Muller, et. al., ”The Electronic Structure at the Atomic Scale of Ultra Thin Gate Oxides”, *Nature*, **399**, 758 (1999)
[20] See for example: G.D. Wilk, R.M. Wallace, and J.M. Anthony, “High-κ Gate Dielectrics: Current Status and Materials Properties Considerations”, *Journal of Applied Physics*, **89**, 5243 (2001)
[21] D.J. DiMaria, *The Physics of SiO_2 and its Interfaces*, edited by S.T. Pantelides, New York: Pergamon, 1978, pp. 160-178
[22] R. Degraeve, et. al., ”New Insights in the Relation Between Electron Trap Generation and the Statistical Properties of Oxide Breakdown”, *IEEE Transactions on Electron Devices*, **45**, 904 (1998)
[23] K.R. Farmer, R. Saletti, and R.A. Buhrman, “Current Fluctuations and Silicon Oxide Wear Out in Metal-Oxide-Semiconductor Tunnel Diodes”, *Applied Physics Letters,* **52**, 1749 (1988)
[24] S.H. Lee, B.J. Cho, J.C. Kim, and S.H. Choi, “Quasi Breakdown of Ultra Thin Gate Oxide Under High Field Stress”, *IEDM Technical Digest*, 605 (1994)
[25] R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, H.E. Maes, “A Consistent Model for the Thickness Dependence of Intrinsic Breakdown in Ultra Thin Oxides”, *IEDM Technical Digest*, 863 (1995)
[26] M.A. Alam, B.E. Weir, P.J. Silverman, Y Ma, and D. Hwang, “The Statistical Distribution of Percolation Resistance as a Probe into the Mechanics of Ultra Thin Oxide Breakdown”, *IEDM Technical Digest*, 529 (2000)
[27] J. Suñé, E.Y. Wu, D. Jimenez, R.P. Vollertsen, and E. Miranda, “Understanding Soft and Hard Breakdown Statistics, Prevalence Ratios, and Energy Dissipation During Breakdown Runaway”, *IEDM Technical Digest*, 117 (2001)
[28] B.P. Linder, et. al., ”Gate Oxide Breakdown Under Current Limited Constant Voltage Stress”, *VLSI Technical Digest*, 214 (2000)
[29] K. Okada and K. Taniguchi, “Electrical Stress Induced Variable Range Hopping Conduction in Ultra Thin Silicon Dioxides”, *Applied Physics Letters,* **70**, 351 (1987)
[30] J. Suñé and E. Miranda, “Post Soft Breakdown Conduction in SiO_2 Gate Oxides”, *IEDM Technical Digest*, 533 (2000)
[31] M. Houssa, T. Nigam, P.W. Mertens, and M.M. Heyns, “Model for the Current-Voltage Characteristics of Ultra-Thin Gate Oxides after Soft Breakdown”, *Journal of Applied Physics*, **84**, 4351 (1998)
[32] E. Miranda, J. Suñé, R. Rodriguez, M. Nafria, and X. Aymerich, “Soft Breakdown Fluctuation Events in Ultra Thin SiO_2 Layers”, *Applied Physics Letters,* **73**, 490 (1998)
[33] T. Nigam, S. Martin, and D. Abusch-Magder, “Temperature Dependence and Conduction Mechanism after Analog Soft Breakdown”, *Proceedings of the IRPS*, 417 (2003)
[34] N.F. Mott and E.A. Davis, *Electronic Processes in Non-Crystalline Materials*, 2ND edition, Oxford: Oxford University Press, 1979
[35] E. Miranda, J. Suñé, R. Rodriguez, M. Nafria, and X. Aymerich, “Detection and Fitting of the Soft Breakdown Failure Mode in MOS Structures”, *Solid-State Electronics*, **43**, 1801 (1999)

[36] F. Monsieur, et. al., "A Thorough Investigation of Progressive Breakdown in Ultra Thin Oxides: Physical Understanding and Application for Industrial Reliability Assessment", *Proceedings of the IRPS*, 45 (2002)
[37] M.A. Alam, R.K. Smith, B.E. Weir, and P.J. Silverman, "Statistically Independent Soft Breakdowns Redefine Oxide Reliability Specifications", *IEDM Technical Digest*, 151 (2002)
[38] E.Y. Wu, W.W. Abadeer, L.K. Han, S.H. Lo, and G.R. Hueckel, "Challenges for Accurate Reliability Projections in the Ultra Thin Oxide Regime", *Proceedings of the IRPS*, 57 (1999)
[39] B.E. Weir, et. al., "Ultra Thin Gate Dielectrics: They Break Down, but do they Fail?", *IEDM Technical Digest*, 73 (1997)
[40] J. Suñé and E.Y. Wu, "Statistics of Successive Breakdown Events for Ultra Thin Gate Oxides", *IEDM Technical Digest*, 147 (2002)
[41] B. Kaczer, et. al., "Impact of MOS Oxide Breakdown on Digital Circuit Operation and Reliability", *IEDM Technical Digest*, 553 (2000)
[42] B. Kaczer and G. Groeseneken, "Potential Vulnerability of Dynamic CMOS Logic to Soft Gate Oxide Breakdown", *IEEE Electron Device Letters,* **24**, 742 (2003)
[43] V. Ramadurai, N. Rohrer, and C. Gonzalez, "SRAM Operational Voltage Shifts in the Presence of Gate Oxide Defects in 90nm SOI", *Proceedings of the IRPS*, 270 (2006)
[44] T. Sakura, H. Utsunomiya, Y. Kamakura, and K. Taniguchi, "A Detailed Study of Soft and Pre-Soft Breakdowns in Small Geometry MOS Structures", *IEDM Technical Digest*, 183 (1998)
[45] B.P. Linder, S. Lombardo, J.H. Stathis, A. Vayshenker, and D.J. Frank, "Voltage Dependence of Hard Breakdown Growth and the Reliability Implication in Thin Dielectrics", *IEEE Electron Device Letters,* **23**, 661 (2002)
[46] T. Hosoi, P.L. Re, Y. Kamakura, and K. Taniguchi, "A New Model of Time Evolution of Gate Leakage Current after Soft Breakdown in Ultra Thin Gate Oxides", *IEDM Technical Digest*, 155 (2002)
[47] J.S. Suehle, B. Zhu, Y. Chen, and J.B. Bernstein, "Acceleration Factors and Mechanistic Study of Progressive Breakdown in Small Area Ultra Thin Gate Oxides", *Proceedings of the IRPS*, 95 (2004)
[48] J. Suñé, G. Mura, and E. Miranda, "Are Soft Breakdown and Hard Breakdown of Ultra Thin Gate Oxides Actually Different Failure Mechanisms?", *IEEE Electron Device Letters,* **21**, 167 (2000)
[49] J.C. Jackson, T. Robinson, O. Oralkan, and D.J. Dumin, "Non-Uniqueness of Time Dependent Dielectric Breakdown Distributions", *Applied Physics Letters,* **71**, 3682 (1997)
[50] J. Suñé, E. Miranda M. Nafria, and X. Aymerich, "Point Contact Conduction at the Oxide Breakdown of MOS Devices", *IEDM Technical Digest*, 191 (1998)
[51] J. Schmitz, H.J. Kretschmann, H.P. Tuinhout, and P.H. Woerlee, "Soft Breakdown Triggers for Large Area Capacitors Under Constant Voltage Stress", *Proceedings of the IRPS*, 393 (2001)
[52] F. Crupi, R. Degraeve, G Groeseneken, T. Nigam, and H.E. Maes, "On the Properties of the Gate and Substrate Current after Soft Breakdown in Ultra Thin Oxide Layers", *IEEE Transactions on Electron Devices*, **45**, 2329 (1998)

[53] K. Okada, H. Kubo, A. Ishinaga, and K. Yoneda, "A New Prediction Method for Oxide Lifetime and its Application to Study Dielectric Breakdown Mechanism", *VLSI Technical Digest*, 158 (1998)
[54] G.B. Alers, B.E. Weir, M.A. Alam, G.L. Timp, and T. Sorch, "Trap Assisted Tunneling as a Mechanism of Degradation and Noise in 2 – 5nm Oxides", *Proceedings of the IRPS*, 76 (1998)
[55] G.B. Alers, B.E. Weir, M.R. Frei, and D. Monroe, "J-Ramp on Sub-3nm Dielectrics: Noise as a Breakdown Criteria", *Proceedings of the IRPS*, 410 (1999)
[56] *JEDEC Standard No. 92*, Arlington: Jedec Solid State Technology Association, 2003
[57] B.E. Kaczer, R. Degraeve, R. O'Conner, P. Roussel, and G. Groeseneken, "Implications of Progressive Wear Out for Lifetime Extrapolation of Ultra Thin (EOT ~ 1nm) SiON Films", *IEDM Technical Digest*, 714 (2004)
[58] P.A. Tobias and D.C. Trindade, *Applied Reliability*, 2ND edition, Boca Raton: Chapman & Hall/CRC, 1995
[59] W. Nelson, Applied Life Data Analysis, New York: John Wiley & Sons, 1982
[60] L.C. Wolstenholme, *Reliability Modeling : A Statistical Approach*, Boca Raton: Chapman & Hall/CRC, 1999
[61] B. Epstein and H. Brooks, "The Theory of Extreme Values and its Implications in the Study of the Dielectric Strength of Paper Capacitors", *Journal of Applied Physics*, **19**, 544 (1948), and the references contained therein
[62] D. Wolters and J.F. Verwey, *Instabilities in Silicon Devices*, Volume 1, edited by G. Barbottin and A. Vapaille, North Holland: Elsevier Science Publishers B.V., 1986, pp. 315-362
[63] W.R. Hunter, "A Failure Rate Based Methodology for Determining the Maximum Operating Gate Electric Field, Comprehending Defect Density and Burn-In", *Proceedings of the IRPS*, 37 (1996)
[64] J. Suñé, "New Physics Based Analytical Approach to the Thin Oxide Breakdown Statistics", *IEEE Electron Device Letters,* **22**, 296 (2001)
[65] W.R. Hunter, private communication
[66] T. Anderson, private communication
[67] T. Nigam, R. Degraeve, G. Groeseneken, M.M. Heyns, and H.E. Maes, "Constant Current Charge to Breakdown: Still a Valid Tool to Study the Reliability of MOS Structures?", *Proceedings of the IRPS*, 62 (1998)
[68] R.E. Dickerson, H.B. Gray, and G.P. Haight Jr., *Chemical Principles*, 3RD edition, Menlo Park: Benjamin/Cummings, 1979
[69] J.W. McPherson and D.A. Baglee, "Acceleration Factors for Thin Gate Oxide Stressing", *Proceedings of the IRPS*, 1 (1985)
[70] M. Kimura, "Oxide Breakdown Mechanism and Quantum Physical Chemistry for Time Dependent Dielectric Breakdown", *Proceedings of the IRPS*, 190 (1997)
[71] E. Wu, J.M. McKenna, W. Lai, E. Nowak, and A. Vayshenker, "The Effect of Change of Voltage Acceleration on Temperature Activation of Oxide Breakdown for Ultra Thin Oxides", *IEEE Electron Device Letters,* **22**, 603 (2001)
[72] D.J. DiMaria and J.H. Stathis, "Non-Arrhenius Temperature Dependence of Reliability in Ultra Thin Silicon Dioxide Films", *Applied Physics Letters,* **74**, 1752 (1999)

[73] R. Degraeve, et. al., ”Temperature Acceleration of Oxide Breakdown and its Impact on Ultra Thin Gate Oxide Reliability”, *VLSI Technical Digest*, 59 (1999)
[74] S.M. Sze, Physics of Semiconductor Devices, 2ND edition, New York: Wiley, 1981, pp. 362-402
[75] J. Suñé, P. Olivo, and B. Ricco, “Quantum Mechanical Modeling of Accumulation Layers in MOS Structure”, *IEEE Transactions on Electron Devices,* **39**, 1732 (1992)
[76] C. Bowen, et. al., ’Physical Oxide Thickness Extraction and Verification Using Quantum Mechanical Simulation”, *IEDM Technical Digest*, 869 (1997)
[77] W.R. Hunter, “Gate Oxide Reliability: The Use of Simulation to Quantify Important Aspects of Lifetime Projection from TDDB Data”, *IRW Final Report*, 95 (1994)
[78] J.D. Jackson, *Classical Electrodynamics*, 2ND edition, New York: Wiley, 1975, pp. 33-45
[79] L.M. Terman, “An Investigation of Surface States at a Silicon/Silicon Oxide Interface Employing Metal-Oxide Silicon Diodes”, *Solid-State Electronics*, **5**, 285 (1962)
[80] C.N. Berglund, “Surface States at Steam Grown Silicon – Silicon Dioxide Interfaces”, *IEEE Transactions on Electron Devices*, **ED-13**, 701 (1966)
[81] M. Kuhn, “A Quasi-Static Technique for MOS C-V and Surface State Measurements”, *Solid-State Electronics*, **13**, 873 (1970)
[82] E.H. Nicollian and J.R. Brews, *MOS Physics and Technology*, New York: Wiley, 1982, pp. 592-601
[83] K. Ziegler and E. Klausmann, “Static Technique for Precise Measurements of Surface Potential and Interface State Density in MOS Structures”, *Applied Physics Letters,* **26**, 400 (1975)
[84] J.R. Brews and E.H. Nicollian, “Improved MOS Capacitor Measurements Using the Q-C Method”, *Solid-State Electronics*, **27**, 963 (1984)
[85] K.J. Yang and C. Hu, “MOS Capacitance Measurements for High Leakage Thin Dielectrics”, *IEEE Transactions on Electron Devices*, **46**, 1500 (1999)
[86] D.W. Barlage, J.T. O’Keeffe, J.T. Kavalieros, M.M. Nguyen, and R.S. Chau, “Inversion MOS Capacitance Extraction for High Leakage Dielectrics Using a Transmission Line Equivalent Circuit”, *IEEE Electron Device Letters,* **21**, 454 (2000)
[87] J. Schmitz, et. al., “Test Structure Design Considerations for RF C-V Measurements on Leaky Dielectrics”, *IEEE Transactions on Semiconductor Manufacturing*, **17**, 150 (2004)

CHAPTER 2

Charge trapping and stress induced leakage current

2.1 Introduction

In this chapter, we will present our work on the role of interface traps on the increase in leakage current through sub 35Å oxides following electrical stress. Before evaluating the effects of trap generation on post stress I-V characteristics, it is necessary to consider the transport processes in a virgin device in order to comprehend the changes that traps introduce. The effects from defects introduced by stress on the I-V behavior must then be accounted for to develop a model for post stress transport.

We begin this chapter with a brief review of the time-0 tunneling processes that are encountered in ultra thin device grade dielectrics. We then discuss charge trapping in as-fabricated and in stress generated sites. After subsequently reviewing what is known regarding the effects of stress generated bulk traps on gate leakage, we then show our findings on how interface traps influence stress induced leakage current in ultra thin gate dielectrics.

2.2 Time-0 transport mechanisms

In this section, we will discuss the electrode limited thermal tunneling mechanisms that are encountered in virgin device grade SiO_2 and SiON films. The operative tunneling mechanism is determined by the oxide thickness and V_{OX}. The two most important transport processes are Fowler Nordheim Tunneling (F-N) and Direct Tunneling (D-T). In this chapter, we will focus on the basic transport properties. We will expand this discussion to include the effects of tunneling on carrier energies in Chapter 3.

2.2.1 Fowler-Nordheim tunneling

Although tunneling through insulating films is not a new concept [1], the first observation of Fowler-Nordheim tunneling [2] in SiO_2 was reported in 1967 [3]. F-N tunneling occurs when the voltage drop across the oxide (V_{OX}) exceeds the barrier height (Φ_B) between the injecting electrode and the dielectric material so that the barrier shape becomes triangular, as illustrated in Figure 2.1 for two different oxide voltages. Using the WKB approximation with the classical turning point (the tunneling length) equal to the distance from the injecting interface to the point where the electron emerges into the oxide conduction band (X_{T1} and X_{T2} in Figure 2.1, where $X_T = \Phi_B/qE_{OX}$), the J_G-E_{OX} relationship for F-N tunneling through a triangular barrier is [2-4]

$$J_{FN} = AE_{OX}^2\exp(-B/E_{OX}) \quad (2.1)$$

$$A = q^3m/16\pi^2\hbar m_{OX}\Phi_B \quad (2.2)$$

$$B = 4(2m_{OX})^{1/2}\ \Phi_B^{3/2}/3q\hbar \quad (2.3)$$

q is the electron charge, m and m_{OX} are the electron effective masses in free space and in the oxide respectively, and ħ is the reduced Planck's constant. The parameters A and B depend on the model used for the energy-momentum relationship within the insulator band gap. The simplest is the parabolic dispersion relationship [4], and the other is the Franz dispersion relationship [5]. They are, respectively:

$$k(E) = (2m_{OX}/\hbar^2)^{1/2}E^{1/2} \quad (2.4)$$

$$k(E) = (2m_{OX}/\hbar^2)^{1/2}E^{1/2}(1-E/E_G)^{1/2} \quad (2.5)$$

E is the electron kinetic energy with respect to the oxide conduction band minima at a distance x from the injecting interface. The Franz formulation reduces to the parabolic relationship as the electron kinetic energy approaches zero (i.e.) $E \sim E_C$, where E_C is the SiO_2 conduction band edge. This implies that the effective mass is a function of the kinetic energy of the electrons that have tunneled into the oxide conduction band. There does not seem to be a consensus in the literature regarding which model is valid, as some groups report better fits to the parabolic model [4,6,7,8], whereas others find the non-parabolic model to be more suitable [5,9,10,11]. The lack of consensus may in part be due to issues with sample preparation or to the presence of charge trapping. It has been argued that the effective band gap in the context of the Franz formulation may be the energy separation between conduction band and one of the deeper lying valence bands, which could be as much as 18eV below the SiO_2 conduction band edge [6]. If this is indeed the case, the parabolic formulation would provide a satisfactory model for the E vs. k dispersion relationship over a wide range of bias (i.e. $E/E_G << 1$).

Equation (2.1) can be derived as follows: If k(E) and the energy difference between the tunneling energy and the oxide conduction band E(x) are known, the quantum mechanical transmission probability P(E) can be calculated using the WKB approximation [12] as:

$$P(E) = \exp[-2\int_0^{x_T} k_{OX}(x)\, dx] \quad (2.6)$$

Where k_{OX} is the imaginary part of the complex wave vector of the tunneling electron. The gate current can then be obtained from the integral [5]:

$$J = (4\pi q m_{OX} h^3)\int_0^{E} P(E')E'dE' \quad (2.7)$$

Equation (2.1) is obtained by integration of (2.6) and (2.7) with the parabolic dispersion relation given in (2.4). Inspection of (2.6) shows that the gate current is inversely proportional to the area under the barrier as shown in Figure 2.1. This is due to the reduction in the tunneling length with increasing oxide voltage. From (2.1), for a given E_{OX}, the F-N tunneling current J_G is independent of oxide thickness [4] and a plot of $\ln(J_G/E_{OX}^2)$ vs. $1/E_{OX}$ yields a straight line. This is often referred to as an F-N plot. The barrier height is determined by photoemission experiments or from the F-N plot. The electron barrier heights between silicon and SiO_2 conduction bands range from 2.90eV to 3.30eV [4,8,13-15], with the conduction band electron effective mass m_{OX}/m typically from 0.30 to 0.50 [6,7,10,13,15].

Mechanisms that can potentially modify the barrier height include quantum effects and image forces. Quantum effects introduce a voltage dependent barrier height due to occupancy of higher sub-bands [8]. In theory, the barrier height should also include a contribution from image force barrier lowering [2,4,16], but photon assisted tunneling experiments have shown that this effect is negligible for F-N transport [15]. This is not surprising, as the image force correction is expected to be insignificant for large barrier height interfaces because the image potential only slightly rounds the top of the barrier and does not affect the tunneling distance [6].

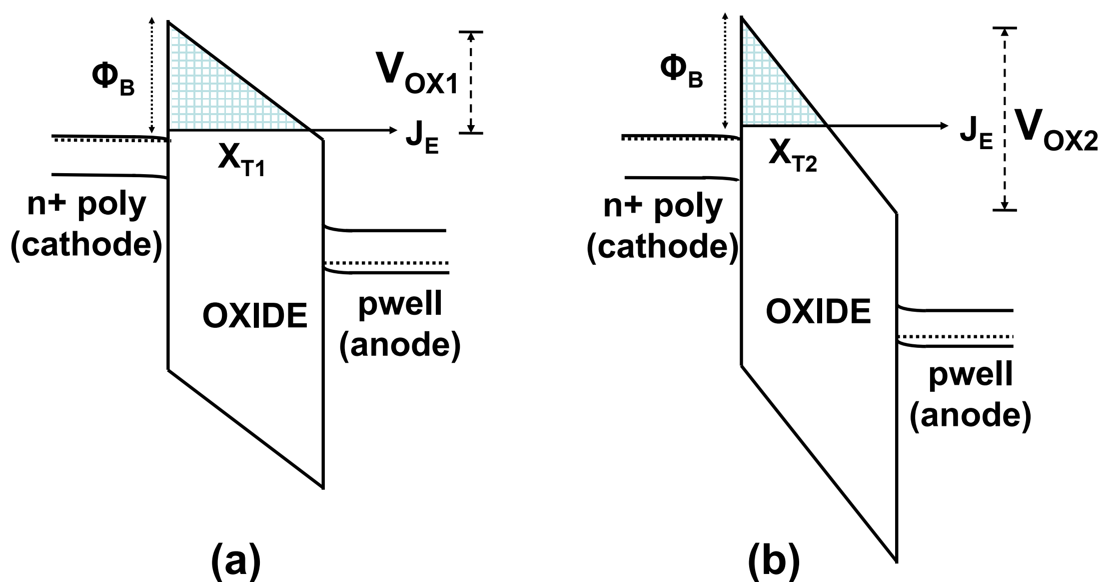

Figure 2.1. Energy band diagrams for NMOS F-N tunneling. Injection is from n+ poly. $V_{OX1} < V_{OX2}$, resulting in $X_{T1} > X_{T2}$. The gate current is inversely proportional to the barrier area (cross-hatched area in figure) so that $J_{G1} < J_{G2}$.

2.2.2 Direct tunneling

Direct tunneling occurs when the voltage drop across the oxide is less than the barrier height between the injecting electrode and the dielectric film, and the oxide thickness is less than one tunneling length [5,7,17]. For $V_{OX} < \Phi_B$, the barrier shape is trapezoidal

and the carrier tunnels from the cathode directly into the anode, without appearing in the oxide conduction band. Unlike F-N tunneling, the classical turning point (the location of the allowed states that the carrier tunnels into) for D-T is always t_{OX} for all $V_{OX} < \Phi_B$. Accordingly, D-T is not important, even at high V_{OX}, when the oxide thickness is significantly larger than a tunneling length (roughly 30Å - 40Å). In addition to tunneling of conduction band electrons, tunneling of valence band holes and electrons becomes significant in the direct tunneling regime. For NMOS in inversion, the gate current is only due to pwell conduction band electrons (CBE) for V_G < 1V. For V_G > 1V, tunneling of pwell valence band electrons (VBE) becomes an allowed transition because the n+ poly conduction band (anode) is now at a lower energy than the pwell valence band (cathode), as shown in Figure 2.2. Similarly, for inverted PMOS, gate current is solely due to tunneling of nwell valence band holes (VBH) for V_G > -1V, with tunneling of p+ poly valence band electrons appearing for V_G < -1V, as illustrated in Figure 2.3.

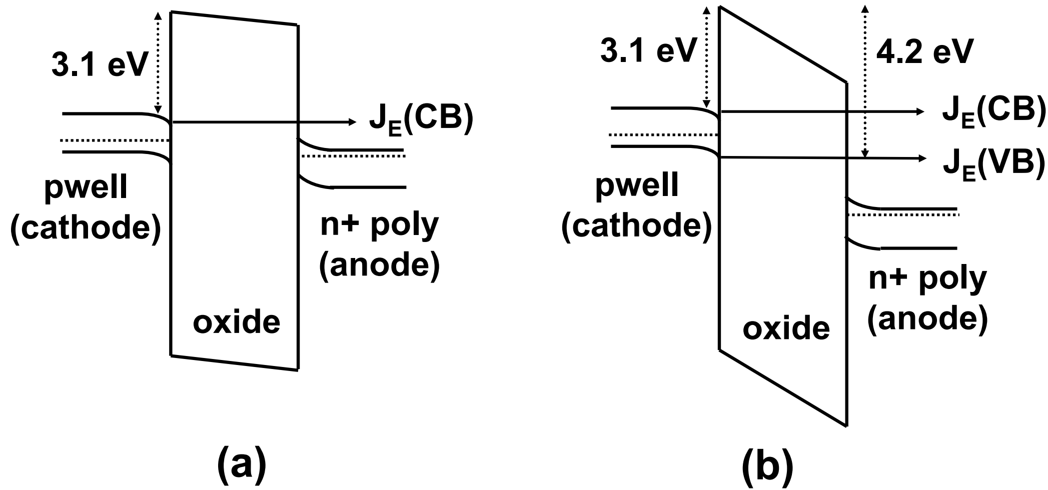

Figure 2.2. Energy band diagrams for NMOS direct tunneling in inversion (a) V_G < 1V, only conduction band electron tunneling allowed. (b) V_G > 1V, both conduction and valence band electron tunneling occurs.

Carrier separation is used to separate the CBE, VBE, and VBH tunneling components [18]. In this method, the current through the gate, drain, and substrate terminals are measured as shown in Figure 2.4. Since $|I_G| = |I_D| + |I_B|$, the magnitude and polarity of the currents allows the determination of the carrier flowing out of each terminal. I-V characteristics for an NMOS device are shown in Figure 2.5. In the on-state, the gate current is dominated by the tunneling of inversion layer electrons supplied by the drain. For V_G > 1V, the pwell VB is at a higher energy than the n+ poly CB so that pwell VBE can tunnel into the gate as shown in Figure 2.2-b. Holes are created in the VB following the tunneling of VBE into the gate, resulting in a hole current flowing out of the substrate contact. In the off-state, the gate current is dominated by the injection of conduction band electrons from the gate into the pwell, which subsequently diffuse into the drain region.

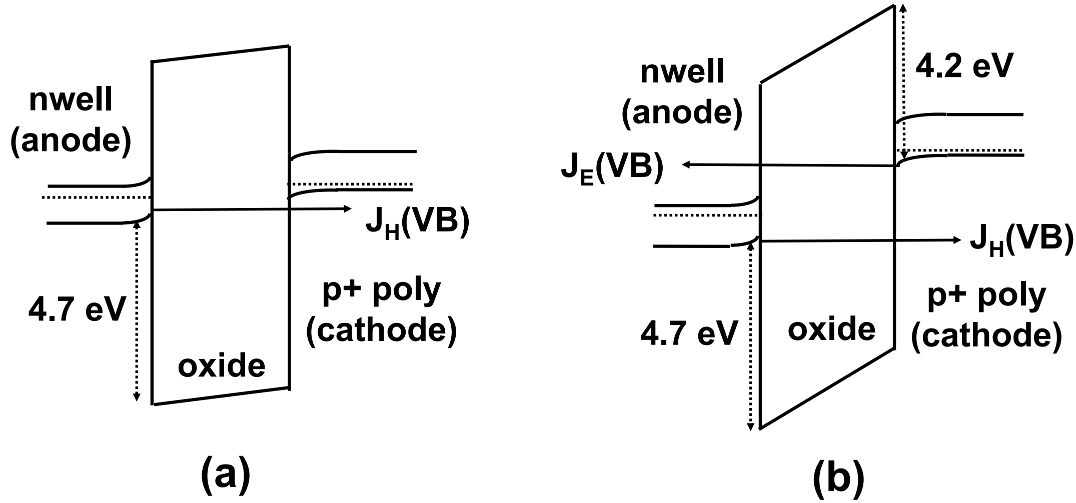

Figure 2.3. Energy band diagrams for PMOS direct tunneling in inversion (a) V_G > -1V, only valence band hole tunneling allowed. (b) V_G < -1V, both valence band electron and hole tunneling occurs.

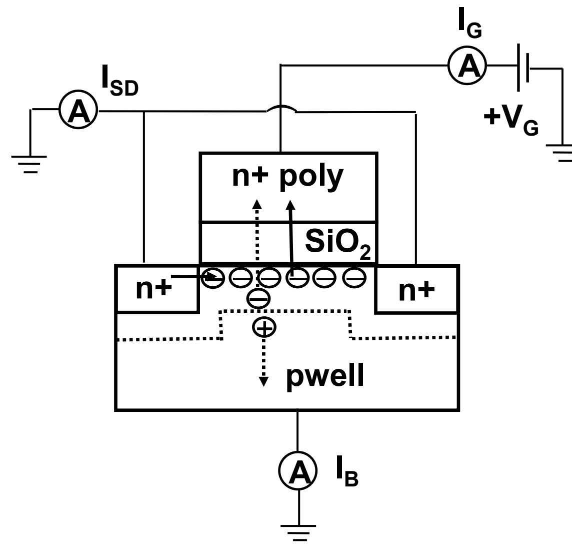

Figure 2.4. Carrier separation technique for inverted nMOS. The gate current is comprised of inversion layer holes supplied by the drain and valence band electrons supplied by the substrate.

Using the WKB approximation, the J_G-E_{OX} relationship for D-T tunneling through a trapezoidal barrier is [7,17]

$$J_{DT} = AE_{OX}^2/[1-((\Phi_B+qV_{OX})/\Phi_B)^{1/2}]^2\exp[(B/E_{OX}\Phi_B^{3/2})(\Phi_B^{3/2}-(\Phi_B-qV_{OX})^{3/2})] \quad (2.8)$$

Unlike F-N tunneling, D-T is thickness dependent at a given field. J_G vs. E_{OX} is calculated using (2.8) for 5 different oxide thicknesses and is plotted in Figure 2.6.

Because the Fowler-Nordheim tunneling length decreases with oxide field whereas the direct tunneling length is independent of field, the slope of the F-N tunneling current is larger that that of direct tunneling as shown in Figure 2.6. Accordingly, F-N tunneling dominates the total gate current at high fields. However, as t_{OX} decreases, ultimately, only direct tunneling will be observed when the oxide breaks down at $V_{OX} < \Phi_B$.

The band diagrams in Figures 2.2 and 2.3 show the interfaces to be abrupt. However, the transition from silicon to SiO_2 actually occurs over one monolayer [19]. Accordingly, as the physical thickness of the oxide is scaled down, the transition region will eventually cause the barrier shape to have sufficiently deviated from trapezoidal so that the potential is no longer a slowly varying function of distance. In this regime, the WKB approximation breaks down and (2.6) loses its validity.

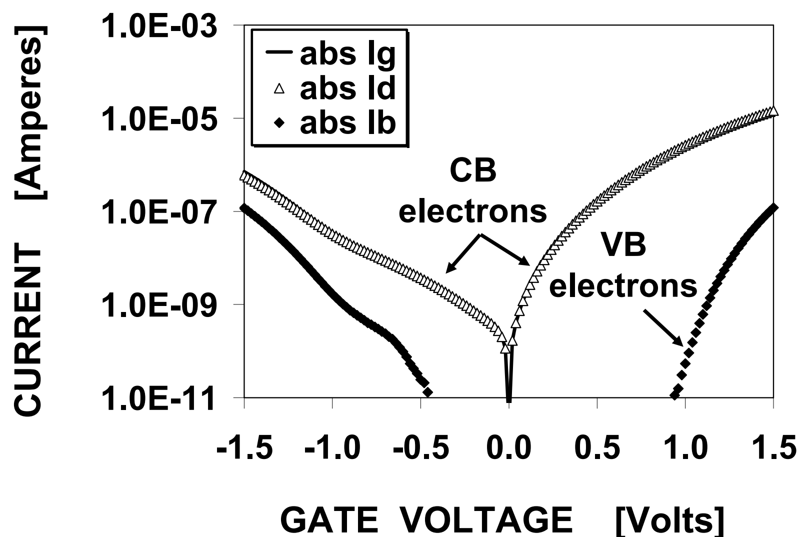

Figure 2.5. NMOS terminal currents for 12Å SiON films. In the on-state, $I_G \sim I_D$ because the gate current is dominated by tunneling of channel electrons supplied by the drain. The onset of VB electron tunneling occurs when the pwell valence band edge moves above the poly conduction band edge (V_G > 1V) where there are available states to tunnel into. For $V_G < 0$, $I_G \sim I_D$ because most of the electrons injected from the gate into the pwell diffuse into the drain regions and result in a drain current.

2.2.3 Effects of nitrogen

Because the gate current increases by about 0.4 decade/Å (depending on how $t_{PHYSICAL}$ and [N] are modified to obtain a given EOT), scaling the dielectric thickness from 35Å to 10Å can result in a 10 orders of magnitude increase in J_G [20]. Since the time to breakdown is inversely proportional to J_G, reducing gate leakage is a concern for both TDDB as well as power consumption. Accordingly, replacing SiO_2 with PNO provides a reduction in gate leakage from the higher dielectric constant due to the addition of nitrogen [21,22]. Incorporating nitrogen into an SiO_2 matrix occurs through an exchange

of an oxygen atom with a nitrogen atom. This results in material properties such as the dielectric constant, band gap, barrier height, and effective mass all being linear functions of the amount of nitrogen incorporated (up to 57 at% nitrogen) [23-25], as shown in Figures 2.7 though 2.10.

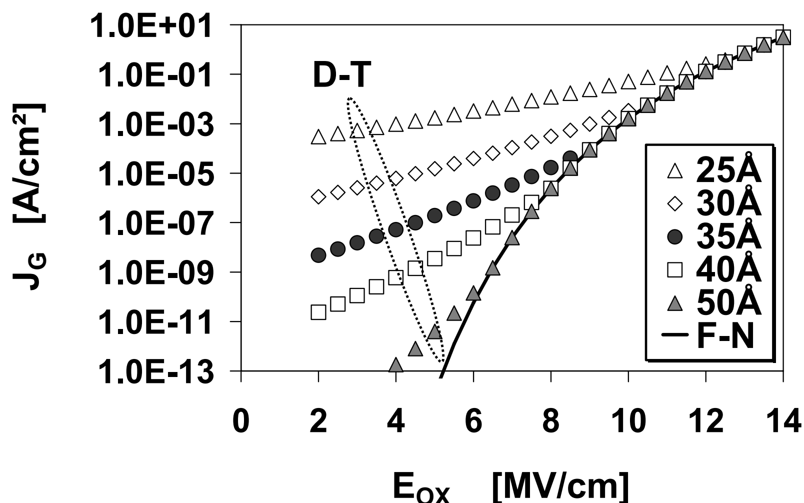

Figure 2.6. Calculated gate current densities for 5 different oxide thicknesses.

The increase in dielectric constant shown in Figure 2.7 realized through nitrogen incorporation reduces leakage by enabling a thicker physical oxide thickness at a given EOT. However, Figure 2.8 shows a reduction in barrier height with increasing nitrogen. Accordingly, reducing J_G through the addition of nitrogen requires a trade-off between ε_{SiON} against Φ_B and m_{OX}. The optimal nitrogen level has been reported to be about 30%, or approximately 50% equivalent Si_3N_4 composition [24].

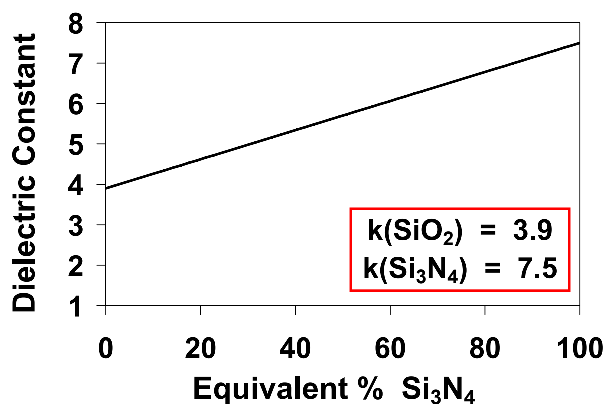

Figure 2.7. Dielectric constant vs. level of nitrogen incorporation.

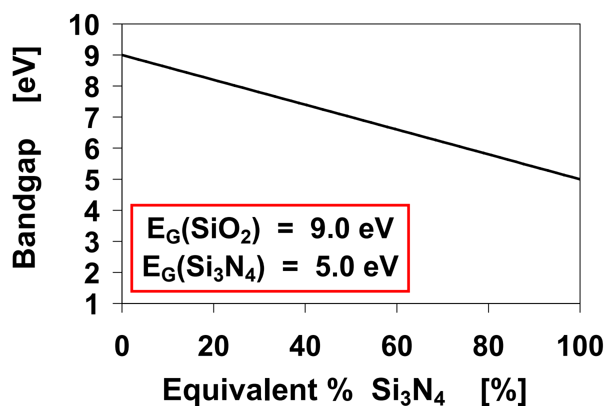

Figure 2.8. Band gap vs. level of nitrogen incorporation.

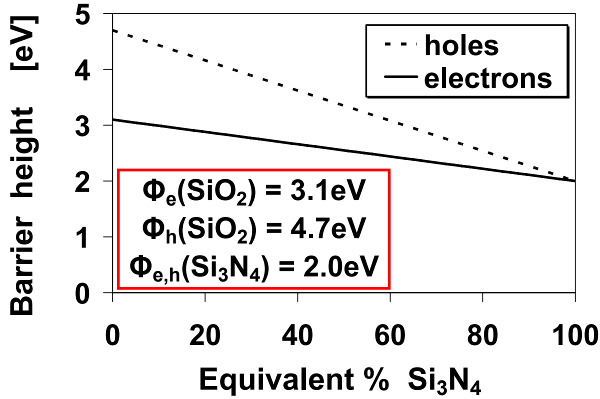

Figure 2.9. Barrier heights vs. level of nitrogen incorporation.

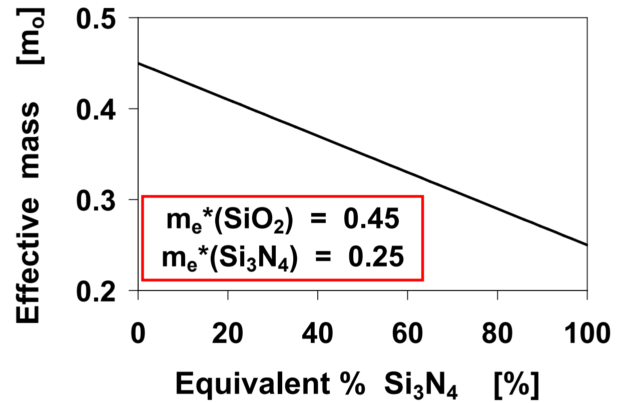

Figure 2.10. Effective mass vs. level of nitrogen incorporation.

There have been some reports regarding deviations from the linear effect of nitrogen levels on SiON material properties. Plasma enhanced CVD oxynitride films were shown to have a non-linear dielectric constant vs. nitrogen relationship, with ε_{SiON} being higher at lower nitrogen levels than predicted by a linear model [26]. This resulted in a reduction in gate current, which was optimal when the nitrogen profile was uniform. A recent work reported a quadratic dependence of SiON material parameters on nitrogen content over a limited range of nitrogen levels [27]. Another paper reported that a monolayer of interfacial nitrogen, properly engineered, can increase the effective thickness of the interfacial layer, resulting in a tenfold reduction in J_G relative to SiO_2 [28]. Accordingly, the strategy for the optimization of gate leakage in nitrided oxides is dependent on the approach used for nitridation.

2.3 Charge trapping and trap generation

A large number of papers on charge trapping and trap generation phenomena have appeared in the published literature over the past 50 years. Because of the enormity of findings reported over this long period of time, in this section, we condense this body of knowledge into a simple picture from results, which sometimes appear to be contradictory, that have been reported over many years. While the influence of trapped charge on device characteristics diminishes in importance as the oxide thickness is reduced (except in some high-k films), some of the defects that are generated during stress that result in charge trapping also give rise to stress induced leakage current in the ultra-thin dielectric regime. Accordingly, a holistic understanding of stress induced leakage current must comprehend charge trapping phenomena. Although much of the early investigations focused on charge trapping and trap generation resulting from exposure to high energy photons or from the introduction of impurities (which are well controlled in modern device technology), we will not discuss these effects in detail.

2.3.1 The trap potential well

Trap states can be classified as one of three different types of potential wells: Coulombic attractive (herein called Coulombic), neutral, and Coulombic repulsive (herein called repulsive). The 3 potential wells are illustrated in Figure 2.11 for electron traps [29].

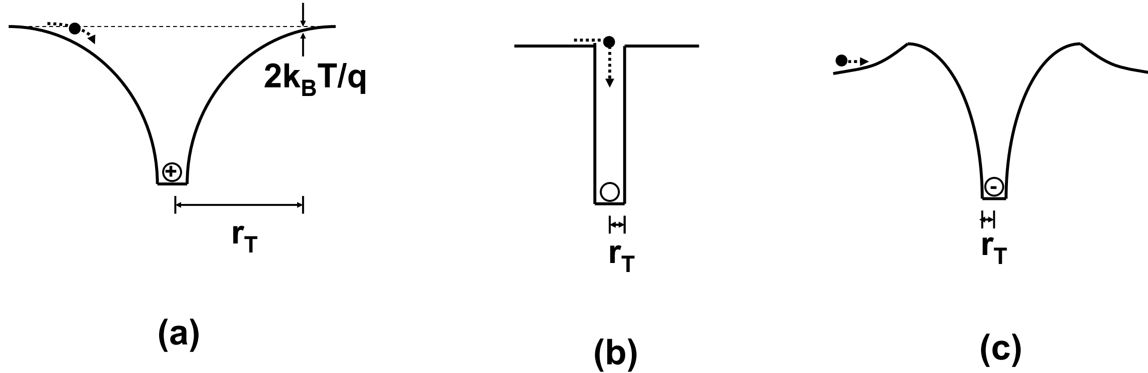

Figure 2.11. (a) coulombic, (b) neutral, and (c) repulsive electron trap potential wells. In (b), a rectangular well is used to illustrate capture in highly localized neutral traps, although the actual potential may have the form $V(r) = A/r^m$.

Coulombic electron traps are positively charged prior to capturing an electron and have large capture cross sections on the order of 10^{-12} to 10^{-15} cm^2 [29,30] (typically > 10^{-14} cm^2). The radius of these traps is the distance from the center of the well to where the coulomb potential $V(r) = A/r$ is about $2k_BT/q$ below conduction band edge, as shown in Figure 2.11-a. The large capture cross sections of coulombic traps arise from electron capture via energy loss through a series of excited states with large radii that are closely enough spaced in energy for one-phonon transitions to be possible between levels [30] (although the final transition to the lowest excited state to the ground state may be a multi-phonon emission process). Since the captured electron can be re-emitted through phonon absorption, at low applied fields, the effective capture cross section of coulombic centers increases with decreasing temperature due to decreasing number of phonons [30]. Another aspect leading to higher coulombic trap capture cross sections at lower temperatures is that as k_BT is reduced, the radius of capture increases because excited states with lower binding energies can participate in the trapping process [30]. The effective capture cross section of coulombic traps is also field dependent [30-32]. Below a field of 1 MV/cm, the capture cross-section shows an $E_{OX}^{-3/2}$ dependence [31,32], which can be accounted for by Frenkel-Poole barrier lowering, which increases the re-emission probability as shown in Figure 2.12. Between 1 MV/cm to 3 MV/cm, the capture cross section shows a stronger E_{OX}^{-3} dependence [32]. At these higher fields, electron heating is likely, which would further reduce the probability of capture beyond what is expected from Frenkel-Poole barrier lowering if the energy gained from the electric field is larger than that lost to phonon scattering [29,32]. Accordingly, the temperature dependence of the capture cross section is weak at high electric fields [32].

Most of the trap states that are electrically active in SiO_2 are neutral traps. They are uncharged prior to capturing an electron or hole and have capture cross sections on the order of 10^{-13} to 10^{-18} cm^2 [29,30,33,34] (typically < 10^{-14} cm^2). These capture cross are on the order of atomic dimensions and the neutral trap is therefore a highly localized potential well as shown in Figure 2.11-b. Unlike coulombic traps, neutral centers do not have excited states [35]. Consequently, ONLY multi-phonon transitions can occur during the capture process. Since the relaxation energies for capture through neutral centers have been calculated to be >1.5 eV in α-quartz [36], a large number of phonons must be emitted to capture a hole or electron in a neutral trap. Accordingly, the capture cross sections for neutral traps tend to be smaller than those of coulombic centers as observed. Capture into a neutral electron trap occurs when sufficiently large phonon vibrations move the energy level of the trap above the conduction band edge where a free electron can be captured [35]. Consequently, the capture cross sections of neutral traps decrease with decreasing temperature due to the reduction in lattice vibrations. Note that the temperature dependence of the neutral trap capture cross section is the opposite of that observed for coulombic traps.

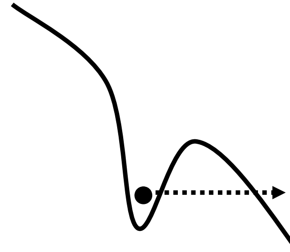

Figure 2.12. Schematic representation of electron de-trapping via tunneling arising from Frenkel-Poole barrier lowering of the coulombic potential in an applied electric field.

The electric field dependence is weaker for neutral traps compared to coulombic traps because there is no Frenkel-Poole barrier lowering for the neutral trap. However, for any trap potential $V(r) = A/r^m$, some amount of barrier lowering will be present. The expression for the barrier lowering is [33]

$$\Delta\Phi = (m+1)A(E_{OX}/mA)^{m/(m+1)} \quad (2.9)$$

The electric field dependence of the capture cross section σ is then given by [33]

$$\sigma = \sigma_o E_{OX}^{-3/(m+1)} \quad (2.10)$$

For a coulombic potential, (2.10) results in the expected $E_{OX}^{-3/2}$ dependence of the capture cross section due to Frenkel-Poole barrier lowering. From (2.10), the more localized the potential, the smaller the barrier lowering effect on the capture cross section. A localized neutral trap can be modeled as a dipole potential [30], where $V(r) = A/r^4$. Using (2.10), the field dependence of the neutral trap capture cross section is predicted to vary as $E_{OX}^{-0.6}$ for a dipole potential and has been verified in experiments

[33]. Electron heating at high fields would be expected to further increase the field dependence of the neutral trap capture cross section if the energy gained from the electric field is larger than that lost to phonon scattering [29,32]. From (2.10), once the field dependence of the capture cross section of a given trap has been determined experimentally, the trap potential $V(r) = A/r^m$ can be extracted. A neutral trap with a $E_{OX}^{-0.4}$ field dependence of the capture cross section has also been measured [33], which corresponds to an extremely localized trap potential $V(r) = A/r^{6.5}$.

The third classification of traps that we will discuss is the repulsive center (Figure 2.11-c). Although this is not an important defect in this work, we will introduce it to further elucidate the effects of the potential well on trapping behavior. A trap can be repulsive to electron capture if a net electron charge already resides on this center. This results in a potential profile that has a barrier to capture. Accordingly, the capture cross section will be extremely small for a repulsive trap and will be on the order of 10^{-18} to 10^{-21} cm^2 [29]. Repulsive traps do not have excited states [30] so that only multi-phonon transitions can occur during the capture process. Similar to a neutral center, the capture cross section of a repulsive trap should increase with increasing temperature. Accordingly, it is possible that repulsive traps may also be dipoles, albeit with the positive and negative charges having directionality that makes trapping unfavorable. Another example of a center that may be repulsive to electron capture might be when a neutral trap is in very close proximity to a metal gate cathode [29]. Unlike coulombic and neutral traps, the capture probability of repulsive centers increases with increasing field, particularly if significant electron heating occurs, as hot carriers would have a higher probability than thermal carriers for surmounting the repulsive energy barrier.

In practice, the determination of whether a trap center is coulombic, neutral, or repulsive is typically based on the experimentally derived capture cross section, field dependence, and temperature dependence.

2.3.2 Determination of the effective trapped charge density and centroid

Charges trapped in the gate dielectric modify the electric field. Positive charges increase the cathode field whereas negative charges reduce it. Accordingly, charge trapping can modify the gate current from the time-0 trap-free case [37]. After stressing a device, the change in the gate voltage at a given current level resulting from an arbitrary charge distribution in the oxide $\rho(x)$ is for +/- V_G respectively [38]:

$$\Delta V_G^+ = -(1/C_{OX})\int(x/t_{OX})\rho(x)dx \tag{2.11}$$

$$\Delta V_G^- = -(1/C_{OX})\int(1-x/t_{OX})\rho(x)dx \tag{2.12}$$

The limits of the integrals in (2.11) and (2.12) are 0 to t_{OX}, where $x = t_{OX}$ corresponds to the Si-SiO_2 interface. The centroid of the charge distribution (integrated from 0 to t_{OX}) is:

$$\bar{x} = \frac{\int x\rho(x)dx}{\int \rho(x)dx} \tag{2.13}$$

Placing (2.11) into (2.13) yields:

$$\Delta V_G^+ = -\bar{x}Q/\varepsilon_{OX} \qquad (2.14)$$

Q is the trapped charge in the oxide. To find the centroid of the trapped charge, we use (2.11) and (2.12) to evaluate the ratio $\Delta V_G^+/(\Delta V_G^+ + \Delta V_G^-)$:

$$\frac{\Delta V_G^+}{\Delta V_G^+ + \Delta V_G^-} = \frac{\int(x/t_{OX})\rho(x)dx}{\int(x/t_{OX})\rho(x)dx - \int(1-x/t_{OX})\rho(x)dx} \qquad (2.15)$$

Inserting (2.13) into (2.15), we obtain [38]

$$\bar{x} = t_{OX}(1+ \Delta V_G^-/\Delta V_G^+)^{-1} \qquad (2.16)$$

Equations (2.14) and (2.16) apply when the dielectric is thick enough so that there is no significant steady state current due to trap assisted tunneling.

2.3.3 Stress and measurement techniques for characterizing charge trapping

In order to study charge trapping phenomena, pre-existing traps must be charged by injecting carriers from either the cathode or anode. To create new traps, carriers with sufficient energy must be injected from one of the contacts. Traps can be generated (a) directly from these injected carriers, (b) through the creation of electrons and holes resulting from impact ionization in the oxide, (c) from anode hole injection, or (d) from anode hydrogen release.

Fowler-Nordheim injection is commonly used for these purposes. For F-N tunneling in SiO_2 and SION films, the injected carriers are predominantly electrons due to the electron barrier heights typically being much smaller than hole barrier heights. The kinetic energy of tunneling carriers is primarily controlled by the gate voltage. High electric fields can be applied across the oxide in this manner. and injection can be achieved for both polarities of gate voltage. However, because F-N tunneling occurs for oxide electric fields greater than about 6 MV/cm as shown in Figure 2.6, trapped charge studies can only performed at relatively high fields. Because de-trapping rates increase with increasing E_{OX}, the use of F-N tunneling as a probe of trapping phenomena has limitations regarding the fraction of traps that are occupied during the measurements [39,40]. Moreover, F-N tunneling is not an efficient source for hole injection due to the high barrier for hole tunneling.

Substrate hot electron injection (SHE) and substrate hot hole injection (SHH) are also widely used methods for studying trapping phenomena [41,42]. NMOS is utilized for SHE and PMOS is utilized for SHH. We will use SHE to illustrate the basic concepts. The technique requires a minimum of 5 device terminals: Gate, drain, source, substrate, and injector. If the device is a gate controlled diode where an n+ diffusion ring surrounds the device, then only 4 terminals are needed (the source contact is eliminated). The injector can simply be an n-type substrate under the pwell [41] or may be an n+p

junction in close proximity (within a diffusion length) of the device [42]. In conventional CMOS logic technologies where p-type substrates are utilized, the latter approach is used to define the injector as illustrated in Figure 2.13-a for an NMOS device [42].

An oxide electric field sufficient to invert the channel is applied, with a bias applied to the substrate. Electrons are supplied from the forward biased n+p junction and subsequently diffuse into the reverse biased pwell region. These carriers gain kinetic energy in the pwell space charge region and are heated when they arrive at the $Si\text{-}SiO_2$ interface. The electrons that reach the interface with sufficient energy to overcome the barrier are injected into the oxide and give rise to a dc gate current as illustrated in Figure 2.13-b. The advantage of SHE/SHH techniques is that the carrier energy is controlled by the substrate voltage rather than the gate voltage so that it is possible to inject carriers at low or high oxide fields. While the substrate bias controls the carrier energy, the forward biased pn junction injector provides control over the fluence of the carriers that will be injected into the gate oxide. Moreover, it is possible to inject holes into the gate dielectric with SHH if a PMOS device is used. In short, SHE and SHH provide the flexibility that allows separate control of oxide field, carrier energy, and carrier fluence. Accordingly, unlike F-N tunneling, SHE and SHH can be used to fill (charge) trap states at low oxide electric fields where de-trapping rates are lower.

Another technique designed to inject carriers at low oxide fields is avalanche injection [43]. In the method, a voltage pulse is applied to the gate to drive the substrate into deep depletion until avalanche breakdown is attained. The minority carriers created are accelerated towards the interface in the space charge region and those that have been heated to sufficiently high energies are injected over the barrier to give rise to an ac gate current. The advantage of this technique is that only a simple capacitor structure is needed, whereas SHE and SHH require devices fabricated with a minimum of 4 terminals. NMOS is used for avalanche electron injection (AEI) and PMOS is used for avalanche hole injection (AHI).

The most common measurement techniques used to sense the build-up or charging of trap states involve C-V based methods, which were discussed in Section 1.9. The change in flat band voltage ΔV_{FB} is used to determine build-up of charge in the dielectric resulting from stress. Since both ΔV_{FB} and ΔV_G^+ sense the change in the oxide field, ΔV_{FB} can be substituted for ΔV_G^+ in (2.14) and provides trapping information at lower oxide electric fields than are possible with F-N tunneling. However, as discussed in Section 1.9, the generation of a high density of interface traps can influence ΔV_{FB} and provide erroneous information on the trapped charge distribution inside the dielectric. To overcome this problem, the change in mid gap voltage ΔV_{MG} is often used. Strictly speaking, if the interface traps in the top half of the band gap are only acceptor-like (-/0) and the interface traps in the bottom of the band gap are only donor like (0/+), then all interface traps will be uncharged at V_{MG} and will not contribute to the MOS capacitance. In this case, the change in mid gap voltage is an accurate measure of trapped charge. However, the condition that all interface traps are neutral at mid gap voltage cannot be expected to be generally true for every arbitrary processing and stress condition.

Indeed, for ultra thin PNO dielectrics stressed at low voltages, only acceptors traps are observed for NMOS PBTI [44] and only donors are observed for PMOS NBTI [45,46].

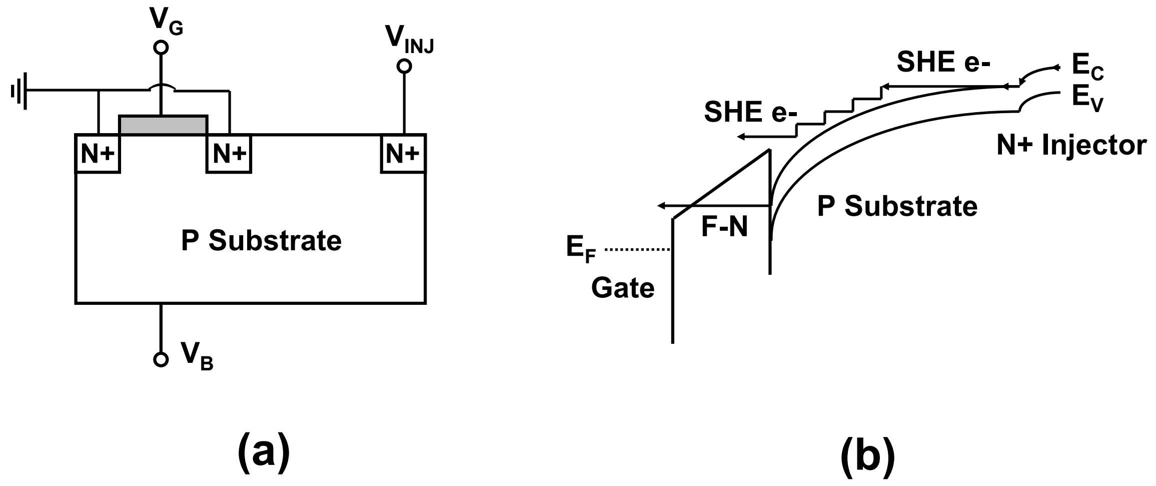

Figure 2.13. (a) Bias configuration for SHE. (b) Band diagram for SHE. Electrons are injected from the forward biased n+p junction into the reverse biased pwell. The electrons are heated in the pwell space charge region prior to being injected into the oxide after they arrive at the Si-SiO$_2$ interface. F-N tunneling is shown for comparison.

The change in threshold voltage ΔV_T of a MOSFET is another metric for monitoring the build up of charges resulting from stress. However, ΔV_T senses all of the charges in the MOS system, including interface traps. The approach used to circumvent this problem is to use the sub-threshold slope to separate the contribution from interface traps [47]. The sub-threshold regime corresponds to weak inversion, where the drain current is dominated by carrier diffusion and increases exponentially with gate voltage [47]. The sub-threshold slope is a figure of merit for the change in the sub-threshold drain current resulting from a change in gate voltage. The units of the sub-threshold slope are mV/decade so that the lower the sub-threshold slope the better. In the absence of interface traps, the sub-threshold slope (S) is given by [47]

$$S = \ln(10)\partial V_G/\partial(\ln I_D) = (k_B T/q)\ln(10)(1+C_S/C_{OX}) \quad (2.17)$$

In (2.17), the contribution from the poly silicon space charge capacitance is omitted for simplicity. From (2.17), it can be seen the lower the silicon space charge capacitance, the better the sub-threshold slope. Accordingly, interface traps, which introduce a capacitance in parallel to the space charge capacitance, will increase S. In the presence of interface traps, the expression for the sub-threshold slope will be [47]:

$$S(t)/S(0) = (1+ (C_S + C_{IT})/C_{IT}) / (1 + C_S/C_{IT}) \quad (2.18)$$

S(t) and S(0) are the post-stress and time-0 sub-threshold slopes respectively. If the band bending as a function of gate voltage is known, the contribution to the shift in threshold voltage component due to interface traps can be determined:

$$C_{IT} = \partial Q_{IT}/\partial \Psi_{Si} \tag{2.19}$$

$$\Delta V_T(\text{interface traps}) = t_{OX} Q_{IT}/\varepsilon_{OX} \tag{2.20}$$

Inspection of (2.17) and (2.18) show that for decreasing oxide thickness and increasing substrate doping, an increasingly larger density of interface states is required to observe changes in sub-threshold slope. Accordingly, for state of the art technologies, the sub-threshold slope may not provide an adequate level of sensitivity.

Charge pumping (C-P) is also a method used to characterize interface traps [48,49]. In the context of charge trapping, the utility of this technique is that while C-V methods probe all of the charges in the MOS system, charge pumping is only sensitive to the response of fast interface states. Accordingly, comparison of C-V and C-P data provide another method for separating the effects of trapped charge from interface traps [48,50]. In the C-P technique, an ac pulse is applied to the gate of a 4 terminal device (only 3 terminals are needed with a gate controlled diode where the diffusion completely surrounds the device) and the substrate current (I_{CP}) is monitored. If the channel length is sufficiently small so that all of the minority carriers that have populated the inversion layer in the on-state portion of the cycle flow back into the source-drain contacts or recombine with majority carriers via interface traps when the pulse is removed, then the substrate current will only be due to interface traps. For a square wave, the substrate current measured during each cycle is linearly proportional to the frequency and the total interface trap density, indicating that a constant recombination charge is measured in every cycle. The total interface trap density is given by [48]:

$$N_{IT} = I_{CP}/(qA_G f) \tag{2.21}$$

Eventually, as the channel length increases, some of the inversion charge will not have returned to the source-drain regions before the field changes sign, so that an additional substrate current will be measured due recombination in the bulk. This part of the total pumped current is called the geometric component and introduces difficulties in extracting trap densities. If the geometric component is absent, the I_{CP} vs. V_G characteristic will saturate after the device has been biased into strong inversion. If the geometric component is significant, then I_{CP} will increase with the magnitude of the gate pulse since a larger amount of inversion charge must return to the source-drain junctions or recombine in the bulk. The geometric component can be minimized by applying a saw tooth pulse, since the rate of change of gate voltage is slower for this waveform. However, I_{CP} will not be a linear function of frequency for a saw tooth signal. For a saw tooth pulse, the mean interface trap density will be proportional to the slope of I_{CP}/f vs. log f and is given by [49]:

$$N_{IT} = \log(e)\ (d(I_{CP}/f)/d\log f)\ /\ (2qk_B T A_G) \tag{2.22}$$

2.3.4 Charge trapping kinetics

Charge trapping in as-fabricated and stress generated sites, detrapping, and trap creation can occur simultaneously. In this section, we will review the kinetics of charge trapping when the same carrier type that is injected is captured. We will defer our discussion of detrapping effects and of hole trapping resulting from impact ionization of injected electrons to subsection 2.3.5. We will cover the resolution of charge trapping from trap generation effects in Section 2.4

1ST order kinetics are commonly applied to study the evolution of charge trapping when the reaction rate is linearly proportional to a single reactant [34]. The rate equation for trapping in N_0 trapping sites with a capture cross section σ is

$$dN_T/dQ = \sigma[N_0 - N_T(Q)] \tag{2.23}$$

$$\int dN_T/[N_0 - N_T(Q)] = \sigma dQ \tag{2.24}$$

The integral in (2.24) is of the form

$$\int dx/(a+bx) = (1/b)\log(a+bx) \tag{2.25}$$

with $a = N_0$ and $b = -1$. Inserting (2.24) into (2.25) and integrating with the time-0 trap density equal to zero, i.e. $N_T(0) = 0$ yields

$$N_T(Q) = N_0(1 - e^{-\sigma Q}) \tag{2.26}$$

Equation (2.26) is the mathematical description of the charge trapping kinetics of traps with a single capture cross section. Using $Q = Jt/q$, and defining the time constant $1/\tau = \sigma J/q$, (2.26) can be written in the time domain as

$$N_T(t) = N_0(1 - e^{-t/\tau}) \tag{2.27}$$

We define the charge trapping efficiency as

$$\eta(Q) = dN_T(Q)/dQ = (N_0\sigma)e^{-\sigma Q} \tag{2.28}$$

From (2.28), it can be see that the charge trapping efficiency is a decreasing function of fluence (or time). The maximum trapping efficiency occurs at $Q = 0$ (or $t = 0$):

$$\eta_{MAX}(Q) = \eta(0) = (N_0\sigma) \tag{2.29}$$

N_0 is the saturation value of $N_T(t)$. Once N_0 is known, σ, can be determined from (2.29).

We will now discuss the range of validity of first order kinetics [51]. Charge that is trapped will give rise to a displacement current

$$J_D = q\,dN_T/dt = (N_0\sigma J)e^{-\sigma Jt/q} \tag{2.30}$$

The ratio of the displacement current to the total current is

$$J_D/J = (N_0\sigma)e^{-\sigma Jt/q} = (N_0\sigma)e^{-\sigma Q} \quad (2.31)$$

From inspection of (2.31) and (2.28),

$$J_D/J = \eta(Q) \quad (2.32)$$

For first order kinetics (2.27) to be valid, $J_D/J << 1$ or equivalently, $\eta(Q) << 1$ [51]. Accordingly, this condition may be violated for short stress times, large trap densities, or large capture cross sections. Additionally, a capture event must not influence other capture events so that the capture cross section is time independent. This is mostly a concern for coulombic traps, which can have capture diameters on the order of 100Å. They can influence capture in other trap centers if they are present in sufficiently high density. Care must be taken in applying 1ST order kinetics to systems with large trap densities, especially if they are coulombic centers. It should also be noted that equation (2.26) implicitly assumes that the only field dependent term is the capture cross section. If the trap density (at a given fluence) is found to be field dependent, then 1ST order kinetics will not provide an adequate model for charge trapping. We will discuss this further in Section 2.3.5.

If there are traps with different capture cross sections present, then (2.27) is written as a summation

$$N_T(t) = \sum_{i}^{M} N_{0i}(1 - e^{-t/\tau_i}) \quad (2.33)$$

Equation (2.33) is evaluated by performing a regression analysis of the experimental data. When a result is obtained that is independent of the initial values chosen, then it is assumed that traps with M different capture cross sections are present. Since the contribution of a given trap to the total trapping rate (but not to the total trap density) vanishes when the time exceeds several time constants, the resolution of multiple capture cross sections is possible with this approach. The time constants of the different traps must be adequately separated temporally to provide ample resolution. An example of this technique is illustrated in Figure 2.14.

2.3.5 Detrapping effects

Once a carrier is trapped, it can subsequently be discharged by photons, phonons, impact ionization, or tunnel emission. Since we are concerned only with electrical trapping studies in this work performed in the dark, detrapping via photon excitation will not be considered further. Since phonons are involved in the capture of free carriers [29], they will also be involved in thermal detrapping processes, particularly at higher temperatures. In this subsection, we will review three field dependent detrapping mechanisms: (1) Detrapping of captured electrons via impact ionization from tunneling electrons. (2) Detrapping of trapped holes via annihilation from tunneling electrons. Modeling of the impact ionization detrapping mechanisms (1) and (2), which occur at high fields in thick oxides, requires only a simple modification to 1ST order kinetics. We

will then introduce (3) detrapping via tunnel emission, which is a significant effect over a wide range of experimental conditions. Tunnel emission is an important effect in studies of stress induced leakage current that will be covered in Section 2.4.

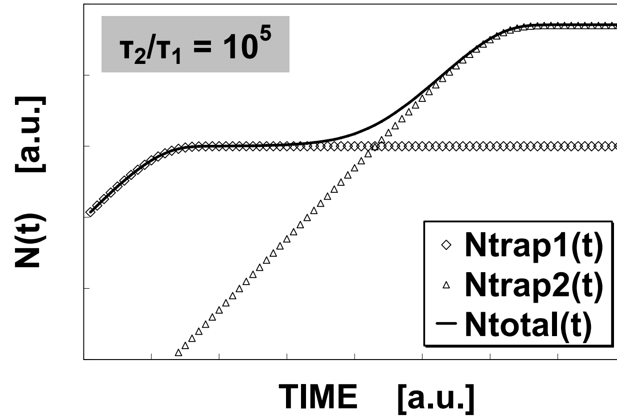

Figure 2.14. N(t) vs. t curves. A good fit is obtained with two time constants.

The application of first order kinetics requires that the only field dependent parameter in (2.26) is the capture cross section. If this is not the case, then the saturation trap density will be a function of electric field. Detrapping rates for both impact ionization and tunnel emission increase with increasing oxide field. A signature of detrapping via impact ionization between electrons in the oxide conduction band and trapped electrons is that the detrapping rate increases as the fluence of injected electrons increases [52]. The electric field and oxide thickness must be sufficiently large for impact ionization to be significant. The approach to modeling detrapping effects due to impact ionization is to assume a dynamic balance between trapping and detrapping processes [52].

To determine the trap density as a function of time, we will define the rate of the trapping process as $(\partial N_T/\partial t)_+$ and the rate of the detrapping process as $(\partial N_T/\partial t)_-$ respectively.

$$(\partial N_T/\partial t)_+ = (J\sigma/q)[N_0 - N_T(t)] \quad (2.34)$$

$$(\partial N_T/\partial t)_- = -(J/q)\beta N_T(t) \quad (2.35)$$

Where β is the detrapping efficiency. Equation (2.34) is the 1ST order kinetics rate used in the previous section. The steady state solution to (2.34) and (2.35) occurs when the trapping rate equals the detrapping rate:

$$(J\sigma/q)[N_0 - N_T(t)] = (J/q)\beta N_T(t) \quad (2.36)$$

$$N_T(ss) = N_0/(1+ \beta/\sigma) \quad (2.37)$$

The complete rate equation and solution are (2.38) and (2.39) respectively:

$$dN_T/dt = (J/q)[\sigma(N_0 - N_T(t)) - \beta N_T(t)] \quad (2.38)$$

$$N_T(t) = N_0(1-e^{-\sigma Jt(1+\beta/\sigma)/q}) \quad (2.39)$$

Equation (2.39) reduces to (2.27) when the detrapping rate is negligible.

We will now consider the case for hole trapping, where holes are created from the impact ionization of tunneling electrons. This process is important for oxide electric fields > 7 MV/cm in films that are thicker than about 300Å [40]. Subsequent detrapping of holes occurs through annihilation by tunneling electrons. The rate equation is [40]

$$dN_{TP}/dt = (J\sigma\alpha/q)[N_0 - N_{TP}(t)] - (J\sigma_{ep}/q)N_{TP}(t) \quad (2.40)$$

Where α is the band gap ionization probability and σ_{ep} is the annihilation cross section between trapped holes and free electrons. Note the similar forms of (2.40) and (2.38). The 1ST term on the right hand side of (2.40) is the trapping rate of holes $(\partial N_{TP}/\partial t)_+$, and the 2ND term on the right hand side of (2.40) is the detrapping rate of holes resulting from recombining with free electrons $(\partial N_{TP}/\partial t)_-$. The steady-state and general solutions to (2.40) are respectively:

$$N_{TP}(ss) = N_0/(1+ \sigma_{ep}/\sigma\alpha) \quad (2.41)$$

$$N_{TP}(t) = N_0(1-e^{-\sigma J\alpha t(1+ \sigma ep/\sigma\alpha)/q)}) \quad (2.42)$$

We will now consider detrapping via field emission. Using trapped electrons as an example, this process occurs when (1) trapped electrons tunnel into the oxide conduction band, (2) trapped electrons direct tunnel from the trap state to the anode conduction band, (3) trapped electrons are annihilated by holes that tunnel from the anode to the trapped electron. These mechanisms are illustrated in Figure 2.15. It can be seen that detrapping via direct tunneling of trapped electrons into the anode conduction band (process 2) can occur even at low fields if the trap is located within one tunneling length of the anode interface and the ground state energy of the trap is higher than the anode conduction band minimum.

Inspection of Figure 2.15 indicates that for oxide thickness greater than 2 tunneling lengths (about 60Å - 70Å), the detrapping process will only be a transient effect. After all centers that can be discharged have been detrapped, no dc gate current flows (when the field is too low for F-N tunneling) unless new traps are created during the detrapping process [53-55]. After stressing 100Å oxides at high fields, followed by recharging the created traps at low fields, subsequent measurement with no bias applied results in the gate current decreasing with time until the current has dropped below the noise floor [54,55] as shown in Figure 2.16. The current decay is proportional to 1/t [54,55]. A 1/t law for detrapping has also been observed at non-zero bias [56]. The current during the recharge cycle also follows a 1/t dependence [54,55].

The 1/t dependence of the discharge can be explained by assuming that the time constant in (2.27) is a function of the distance between the electron trap and the anode interface [57]. The farther the tunneling distance, the longer the time required to depopulate the trap. To model this effect, we replace τ with $\tau(x)$ in (2.27), where $\tau(x)$ is the tunneling time constant for a trap at the interface. Equation (2.27) becomes

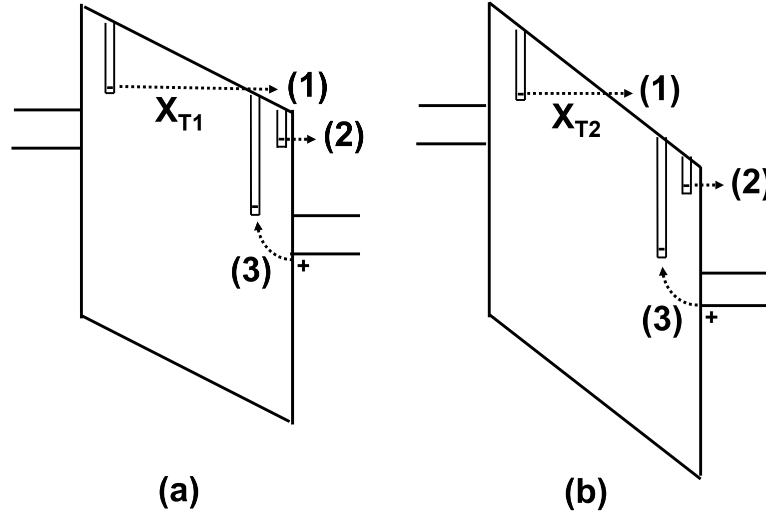

Figure 2.15. Electron detrapping via field emission. The electron detrapping processes illustrated are (1) tunneling of trapped electrons into the oxide conduction band, (2) direct tunneling of trapped electrons into the anode conduction band, and (3) recombination of trapped electrons with holes injected from the anode valence band. $E_{OX}(b) > E_{OX}(a)$ so that $X_{T2} < X_{T1}$, resulting in a higher detrapping probability for process (1) in band diagram (b).

$$N_T(x,t) = N_0(1 - e^{-t/\tau(x)}) \quad (2.43)$$

The time constant $\tau(x)$ can be expressed as [57]

$$\tau(x) = \tau(0)e^{\zeta x} \quad (2.44)$$

where $\tau(0)$ is the time constant at x = 0 and ζ is a tunneling constant. Inserting equation (2.44) into (2.43) yields:

$$N_T(x,t) = N_0[1 - \exp(-t/\tau(0)e^{\zeta x})] \quad (2.45)$$

Solving (2.45) for the value of x where the amount of states is 1/e time the total number of traps [57].

$$x(t) = \zeta^{-1}\ln[t/\tau(0)] \quad (2.46)$$

x(t) is known as the tunneling front [54], which is the boundary between filled and empty states at a given time described by (2.46). The tunneling front model is illustrated in Figure 2.17 and can be used to derive the 1/t dependence of detrapping [54]. The gate current I(t) due to discharging traps is given by

$$I(t)= qAN_T(x,t)v \qquad (2.47)$$

Substituting v = dx/dt, differentiating (2.47), and inserting it into (2.46) yields the 1/t law

$$I(t)= qAN_T(x,t)/(\zeta t) \qquad (2.48)$$

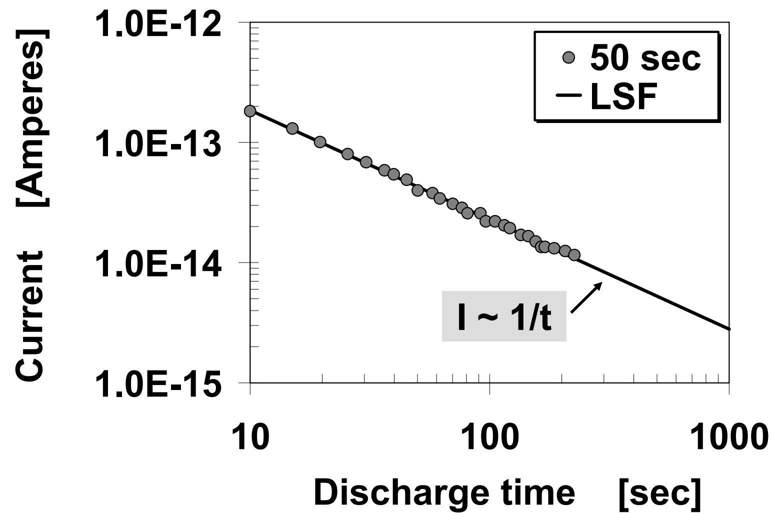

Figure 2.16. Gate current at V_G = 0V vs. discharge time for a 100Å oxide after the device was stressed at -11MV/cm, then recharged at -4V for 50 sec. The current decay is proportional to 1/t. After Scott, ref. [55]. © 1995 Electrochemical Society. Reprinted with permission.

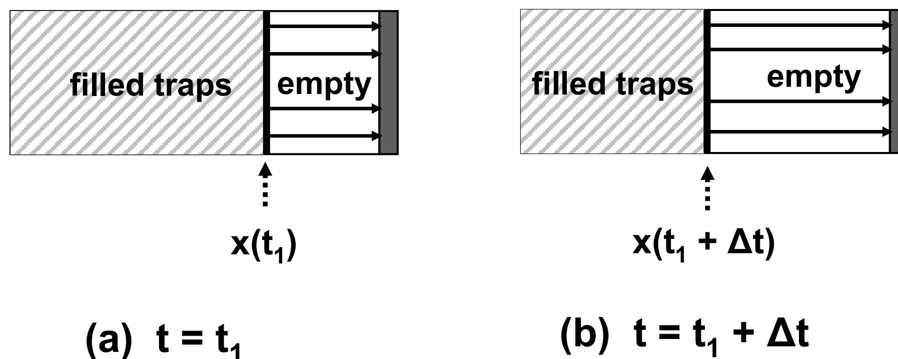

Figure 2.17. Illustration of the tunneling front model. The boundary of empty states propagates farther into the oxide as the discharge time increases.

2.3.6 Electron and hole traps

There is much confusion in the literature regarding the number and nature of traps that can exist in thermally grown oxides. This is largely due to the differences in processing and stress conditions that have been used for trapping studies over the years. To illustrate this point, as we discussed in Section 2.3.1, since the capture cross section is a function of temperature and electric field, failure to take this into consideration can introduce ambiguities in the analysis. Additionally, the gate electrode material [58,59] and oxidation process [34,58,60] can influence the results. It has also been reported that the capture cross sections of some electron traps change upon charging, discharging, and recharging [61]. Moreover, in some investigations, the effects of de-trapping may not have been properly accounted for, nor the effects of compensation from traps of the opposite polarity than are being evaluated.

After having reviewed many findings in the literature, in this section, we present a picture of charge trapping phenomena that is relevant to modern day technologies. We focus on the electrical aspects of oxide traps rather than the atomic nature of these centers. We do not consider the trapping studies involving the use of radiation, nor do we consider trapping effects at liquid nitrogen temperatures.

Traps are classified as either as-grown or stress generated. These two types of traps are differentiated by the magnitude of the electric field and carrier fluence needed to populate them. As-grown traps can be charged at low applied fields and small fluence, and are often characterized using SHE, SHH, AEI, and AHI. Stress-generated traps are often studied using Fowler-Nordheim injection at high fields or with SHE/SHH at moderate to high oxide electric fields. Following high field stress, SHE, SHH, AEI, and AHI at low oxide fields are often used to populate stress generated centers. Conversely, SHH and AHI stress may be followed by an F-N filling step to occupy traps with electrons, since F-N tunneling is often the most efficient way to provide a source of electrons in a PMOS device that has been fabricated with p+ poly gates.

At least four different electron traps can be present in modern device grade SiO_2. One is as-grown [34,61] and the other three are stress generated [61,62]. All four traps are neutral centers. The as-grown traps are water-related with capture cross sections on the order of 10^{-17} cm^2 to 10^{-18} cm^2 [34]. The presence of water-related centers was observed in experiments where dry O_2 grown oxides exposed to tritiated H_2O showed a one-to-one correspondence between the amount of hydrogen lost after avalanche injection and the density of electron traps produced [34]. The gate electrode material utilized can impact the density of water-related (hydrogenous) centers [58,59]. Aluminum gates results in a higher concentration of electron traps compared to other electrode materials [58,59]. In modern day technologies, the use of poly silicon gates, dry oxidation, and the appropriate post oxidation anneal can significantly reduce the concentration of water-related centers [63].

Stress generated electron traps can be created by either hole or electron injection. They are of particular importance since the build-up of these centers during stress tracks the breakdown of ultra thin oxide films [64]. When the density of as-fabricated electron traps

is small (as can be the case in modern technology), using equation (2.16), stress generated neutral electron traps have been found to have a spatially uniform distribution in the oxide (centroid = $t_{OX}/2$) [63,65,66]. However, if the concentration of water-related centers is sufficiently high, the post-stress electron trap distribution will, in general, not be spatially uniform because the spatial distribution of water-related electron traps in the oxide follows the profile of H_2O in the film [34].

The effects of the oxide electric field during an SHE filling step on trap occupancy following high field F-N stress was recently evaluated [56]. In this experiment, two stress generated electron traps were also identified. One of these states remains populated at fields as high as 11 MV/cm (high field trap) and the occupancy of the other approaches zero above 8 MV/cm (low field trap). The authors postulated that the low field trap may be energetically shallower than the high field trap. The generation kinetics of these two states is not the same. A later paper by the same group resolved the low field trap as being a single center and the high field trap as being two different centers [62]. The capture cross sections of the stress generated traps were determined to be between 10^{-14} cm^2 to 10^{-16} cm^2 [62].

At least 3 different hole traps can be present in modern device grade SiO_2 films. One is as-grown and the other two are stress-generated [67,68]. Unlike the stress generated center, once discharged, as-grown traps cannot be recharged without injecting holes [68]. Stress generated hole traps can be created by either hole or electron injection. However, injected electrons generate hole traps indirectly by providing a source of holes through impact ionization or anode hole injection. One stress generated center is known as the anti-neutralization positive charge (ANPC). It can be recharged without hole injection, but is not readily discharged unless the device is subjected to high field F-N injection of electrons [68]. An explanation proposed for the charging behavior of ANPC is that these states move to an energy level above the conduction band edge of the silicon after charging (with $-V_G$ applied). This would make the discharge process, where electrons tunnel from the silicon CB into the trap state (with $+V_G$ applied) less probable. The second type of stress generated centers are known as slow states, also referred to as anomalous positive charge (APC) or cyclical positive charge (CPC) [60,67-69]. Once generated, APC exchanges charge reversibly with silicon and can be readily charged and discharged without further hole injection by applying -/+ V_G respectively. This property of APC has been ascribed to this center having an energy level in the oxide band gap that is unchanged during cyclical charging and discharging and is located near the silicon conduction band edge [67,68], as illustrated in Figure 2.18.

There are some ambiguities regarding whether hole traps are coulombic or neutral centers, as capture cross sections have been reported between 10^{-13} cm^2 to 10^{-16} cm^2 under a variety of electrical stress, analysis, and process conditions [33,67,70,71]. It is conceivable that the capture cross sections at the high end of this range are not accurate because if the density of such large traps is sufficiently high, they will give rise to large displacement currents that render standard 1ST order kinetics techniques inaccurate as discussed in section 2.3.4. Most estimates of both as-grown and stress-generated hole trap capture cross sections are between 10^{-13} cm^2 to 10^{-14} cm^2. Unlike

electron traps, which are spatially distributed uniformly throughout the oxide film, hole traps tend to be concentrated near the oxide interfaces [29,66,67,68,72].

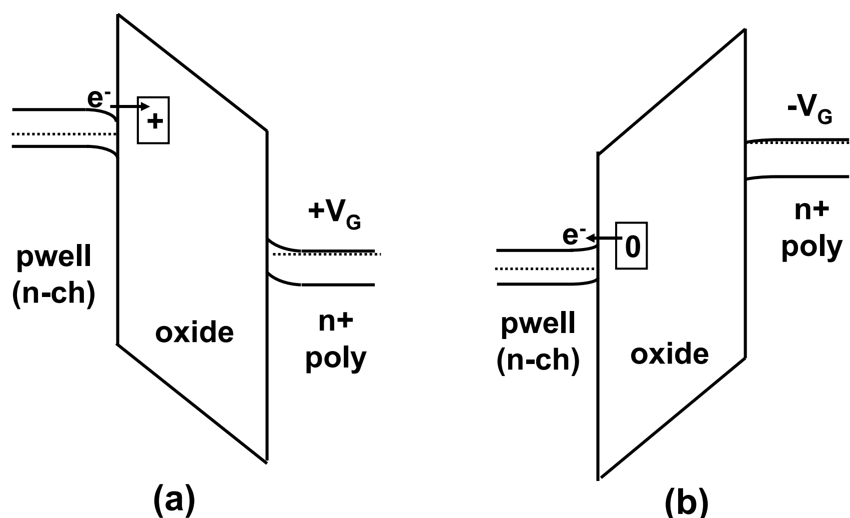

Figure 2.18. (a) Discharging of APC at +V_G. (b) Positive charging of APC under –V_G.

2.4 Stress induced leakage current due to bulk traps (SILC)

In this section, we will discuss the effects of bulk traps on the post stress electrical behavior of ultra thin dielectrics. While the discharge of bulk traps only results in transient currents in thick oxides, this is not always the case in thin films. When stress induced leakage current was first recognized as a device instability, the primary concerns were reliability issues such as DRAM refresh time, EEPROM data retention, and power dissipation. In this thesis, SILC will be used extensively as a tool for the analysis of the trap generation and breakdown physics of ultra thin dielectrics.

2.4.1 Post stress I_G-V_G characteristics of sub-60Å oxides

Figure 2.19 shows the post-stress current vs. time characteristics of 65Å, 85Å, and 130Å oxides [53]. As expected, the post-stress discharge current is a transient effect for the 85Å and 130Å films. However, the post-stress current through the 65Å oxides is a steady state signal. Figure 2.20 shows the fresh and post-stress I-V characteristics of a 45Å NMOS oxide [73]. The current increase is seen primarily in the direct tunneling regime. The post stress leakage current can be significant when the stress voltage exceeds the 5V threshold for trap generation [73]. As can be seen in Figure 2.20, following a ramped voltage stress to 5.85V, J_G increased by about a factor of 100 at low sense voltages [73]. Since the energy threshold for positive charge trapping via impact ionization is greater than 8eV, this is ruled out as the mechanism for the current increase [73]. Interface traps are also eliminated as the mechanism for SILC (at 45Å to 55Å thickness) because the current increase is similar for NMOS and PMOS, but the number of PMOS interface traps is much lower than NMOS [73] (although interface traps do play an important role in SILC when the oxide thickness is scaled below 35Å

[74]). The SILC increase is weakly dependent on the polarity of both stress and sense voltages, indicating that it is due to trap-assisted tunneling through bulk electron (neutral) traps [53,73], which are uniformly distributed spatially in the oxide [63,65,66]. In contrast, positive charges tend to be located closer to one of the interfaces [29,66,67,68,72]. Moreover, the temperature dependence of SILC and electron trap generation is similar [73].

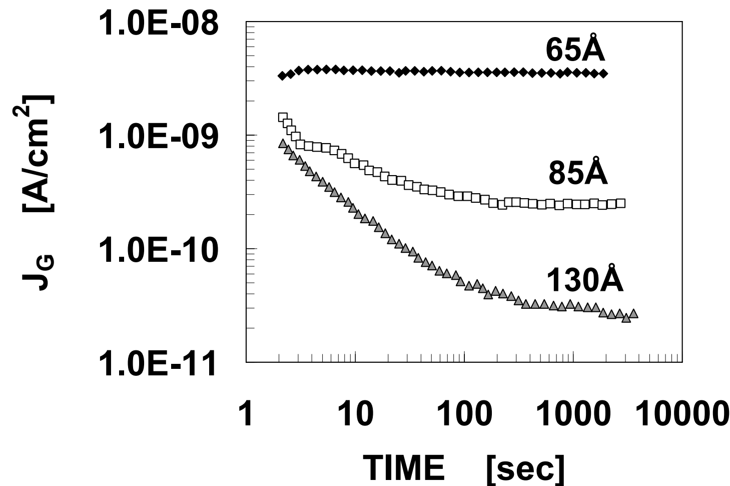

Figure 2.19. Gate current density at 5.4MV/cm vs. time after 9.5MV/cm stress for 20 C/cm^2 to 30 C/cm^2. After Moazzami, ref. [53]. © 1992 IEEE. Reprinted with permission.

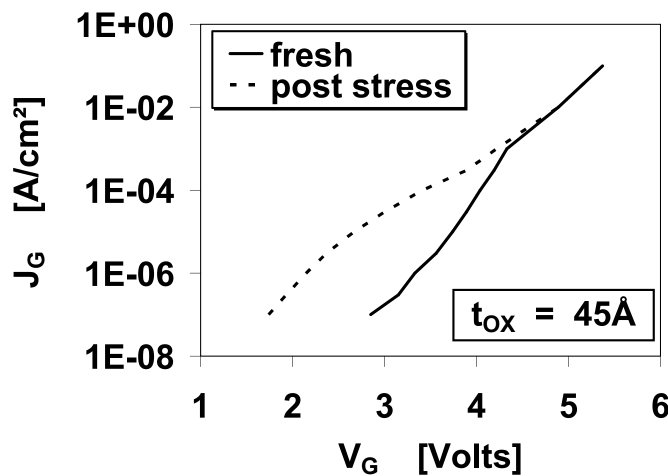

Figure 2.20. Fresh and post ramp-stress I-V characteristics for 45Å NMOS oxides. After DiMaria, ref. [73]. © 1995 American Institute of Physics. Reprinted with permission.

2.4.2 Elasticity of trap assisted tunneling through bulk traps

In this section, we will present the physical picture for stress induced leakage current through bulk traps. Figure 2.21 shows a band diagram for trap assisted tunneling. The

two paths shows are: (1) Elastic trap assisted tunneling, where there is no energy loss when the electron is captured by the trap. (2) Inelastic trap assisted tunneling, where there is an energy relaxation after the electron is captured by the trap.

To determine whether there is an energy loss in the trap assisted tunneling process, the quantum yield is measured [75]. In quantum yield experiments, carrier separation is performed as shown in Figure 2.4, but in a PMOSFET with a n+ poly gate electrode with a negative gate voltage applied [75,76]. Electrons injected from the gate electrode into the n-type substrate impact ionize in the space charge region to form electron hole pairs. Electrons are collected in the substrate and the created holes flow out the source/drain contacts. The quantum yield, which is the average number of electron hole pairs created per injected electron, is:

$$\gamma = I_{SD}/I_G \qquad (2.49)$$

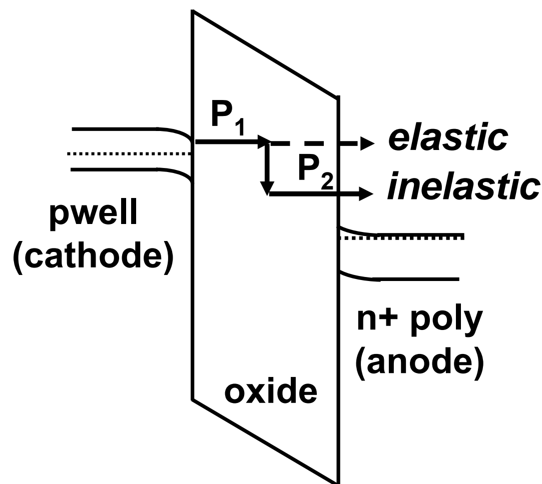

Figure 2.21. Band diagram for trap assisted tunneling through bulk traps (SILC). Both elastic and inelastic tunneling processes are shown. P_1 is the tunneling probability from the cathode to the trap and P_2 is the tunneling probability from the trap to the anode.

The quantum yield vs. electron energy follows a universal curve [77], independent of thickness provided that tunneling is ballistic [75,76], recombination is insignificant [76], and the kinetic energy gained by electrons that have tunneled into the substrate is negligible prior to the impact ionization event [78]. Universal curves for quantum yield are shown in Figure 2.22. It can be seen that the threshold energy for impact ionization is about 1.7 eV, or $(3/2)E_G$. The curves are universal (independent of thickness) below about 6eV, above which scattering broadens the electron energy distribution so that transport is no longer ballistic [76]. The electron energies in Figure 2.22 are calculated theoretically [76] and by the following equation [75]

$$\langle KE_{OX,ANODE} \rangle = qE_{OX}\lambda(1-\exp(X_T - t_{OX})/\lambda)) \qquad (2.50)$$

Where λ is the energy relaxation length and $<KE_{OX,ANODE}>$ is the average kinetic energy in the oxide at the anode interface. We will present a derivation of (2.50) in Section 3.2.4. One of the useful features of Figure 2.22 is that the electron energy can be determined from a simple quantum yield measurement. After stressing an oxide to create bulk traps, the quantum yield is given by:

$$\gamma^{STRESS} = \Delta I_{SD}/\Delta I_G \quad (2.51)$$

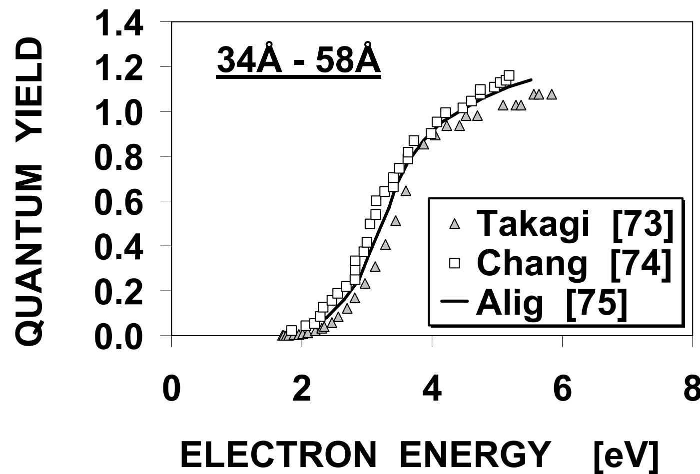

Figure 2.22. Universal curves for quantum yield vs. electron energy in unstressed devices. After Takagi, Ref. [75], Chang, Ref. [76], and Alig, Ref. [77]. Reprinted with permission.

After stress, a significant drop in quantum yield is observed [75], indicating that electrons undergoing trap assisted tunneling create fewer electron hole pairs. Consequently, electrons undergoing SILC enter the anode at a lower energy compared to a virgin device. The post stress electron energy in the substrate, determined from the measured quantum yield and the universal relationship in Figure 2.22, is plotted in Figure 2.23 vs. the electron energy for elastic tunneling [75]. It can be seen that after stress, electrons lose about 1.5eV of energy during the tunneling event [75]. Therefore, SILC is an inelastic transport process. Subsequent estimates of the energy loss resulting from trap assisted tunneling range from 1eV to 2eV [79]. These numbers are consistent with calculations of the relaxation energies for capture through neutral centers in α-quartz [36].

2.4.3 Transport model for stress induced leakage current

In this section, we will follow [80] in deriving phenomenological equations for SILC. The current component from the cathode to the trap (J_{CT}) is given by

$$J_{CT} = c\sigma N_T P_1[f_C(1-f_T) - f_T(1-f_C)] \quad (2.52)$$

Where c is a constant, σ is the capture cross section, N_T is the trap density, and f_C and f_T are the Fermi functions for the cathode and trap respectively. P_1 is the tunneling probability from the cathode to the trap (see Figure 2.21) obtained from the WKB approximation in equation (2.6). The term $f_C(1-f_T)$ represents the flux from the cathode to the trap and the term $f_T(1-f_C)$ represents the flux from the trap to the cathode. Equation (2.52) reduces to

$$J_{CT} = c\sigma N_T P_1(f_C-f_T) \tag{2.53}$$

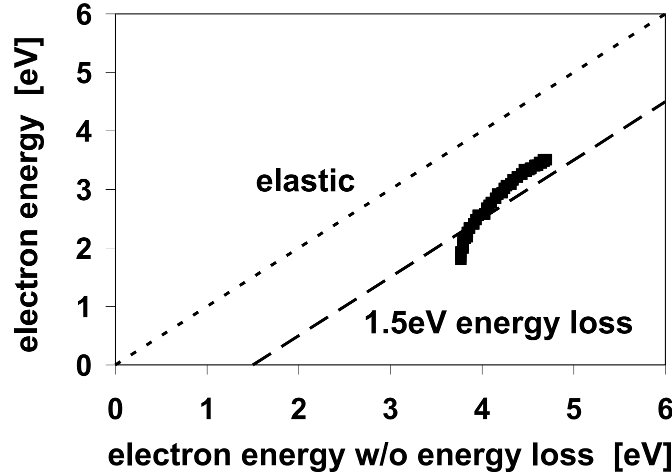

Figure 2.23. Post stress electron energy determined from quantum yield as a function of electron energy in an elastic tunneling process. After Takagi. Ref. [75]. © 1996 IEEE. Reprinted with permission.

Similarly, the current component from the trap to the anode is

$$J_{TA} = c\sigma N_T P_2(f_T-f_A) \tag{2.54}$$

Where P_2 is the tunneling probability from the trap to the anode (see Figure 2.21) obtained from the WKB approximation in equation (2.6), and f_A is the Fermi function at the anode. At steady state, $J_{CT} = J_{TA} = J_{SILC}$. Therefore,

$$J_{SILC} = c\sigma N_T(f_C-f_A)[P_1P_2/(P_1+P_2)] \tag{2.55}$$

From (2.55), J_{SILC} will be maximized for electrons that tunnel into trap locations where $P_1P_2/(P_1+P_2)$ is the highest. In Figure 2.24, it can be seen that this occurs at $x = t_{OX}/2$. Therefore, the traps with the largest contribution towards SILC are located midway between the anode and cathode. This is consistent with the conclusion that SILC is due to tunneling through neutral bulk traps [53,73] since they have a spatially uniform distribution in the oxide with a centroid of $t_{OX}/2$ [63,65,66].

Substituting $P = P_1 = P_2$ in (2.55):

$$J_{SILC} = \tfrac{1}{2}c\sigma N_T P(f_C - f_A) \quad (2.56)$$

The direct tunneling current at time-0 J(0), which is J_{DT} in (8) can be written as

$$J(0) = AP^2(f_C - f_A) \quad (2.57)$$

Eliminating the tunneling probability between (2.56) and (2.57),

$$J_{SILC} = \tfrac{1}{2}c\sigma N_T[(f_C - f_A)/A]^{1/2}J(0)^{1/2} \quad (2.58)$$

The only voltage dependent term in (2.58) is $J(0)^{1/2}$. Accordingly, SILC has an exponential dependence on oxide voltage and thickness. Setting the voltage independent terms on the left hand side of (2.58) equal to "K":

$$J_{SILC} = KJ(0)^{1/2} \quad (2.59)$$

It is common practice to express the experimentally measured SILC as the normalized SILC increase $\Delta I/I(0)$ or $\Delta J/J(0)$. The normalized SILC increase is a useful metric because it is proportional to the density of stress generated traps [73], i.e.

$$\Delta J/J(0) = N(Q) = bQ^m \quad (2.60)$$

A power law in time can be substituted for fluence. In the SILC framework,

$$\Delta J = J(t) - J(0) \quad (2.61)$$

$$J(t) = J_{SILC} + J(0) \quad (2.62)$$

Combining (2.59), (2.61), and (2.62) yields:

$$\Delta J/J(0) = J_{SILC}/J(0) = K/J(0)^{1/2} \quad (2.63)$$

The SILC increase is sometimes expressed as $J(t)/J(0)$ instead of $\Delta J/J(0)$. The sensitivity of the normalized SILC increase to the voltage that it is measured at is

$$d[\Delta J/J(0)]/dV_G = K^*d[1/J(0)^{1/2}]/dV_G \quad (2.64)$$

In practice, $\Delta J/J(0)$ corresponding to the build-up of bulk traps is a weak function of sense voltage [73]. Accordingly, K is << 1. Similarly, the only thickness dependent term in (2.58) is $J(0)^{1/2}$ so that

$$d[\Delta J/J(0)]/dt_{OX} = K'^*d[1/J(0)^{1/2}]/dt_{OX} \quad (2.65)$$

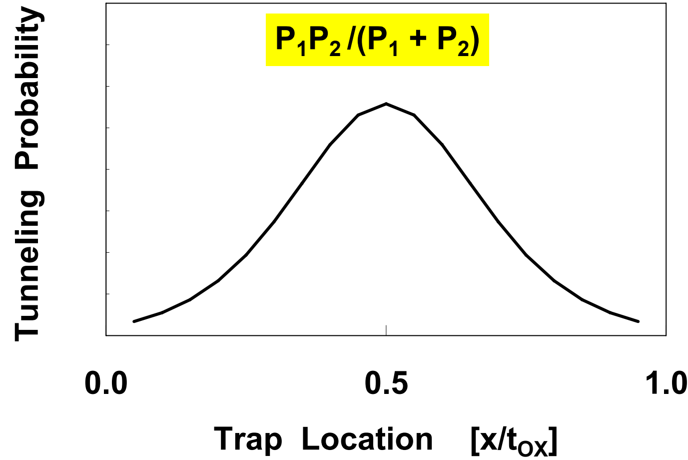

Figure 2.24. Tunneling probabilities $P_1P_2/(P_1+P_2)$ as a function of trap position in the oxide, showing that the traps with the largest contribution to SILC are located at $t_{OX}/2$.

2.4.4 Kinetics of trap generation

Both trap generation and charge trapping may occur simultaneously in an oxide under stress. To separate these effects in a SILC measurement [81], the charge trapping term (2.26) with the substitution $N_T(Q) = \Delta J/J(0)$ is added to (2.60):

$$\Delta J/J(0) = N(Q) = N_0(1 - e^{-\sigma Q}) + bQ^m \tag{2.66}$$

Equation (2.66) is evaluated by performing a regression analysis of the experimental data. When a result is obtained that is independent of the initial values chosen, then it is assumed that the correct solution has been found. In practice, this is possible when the trap generation term dominates at long stress fluence (or time), or if the effects of charge trapping are weak. For ultra-thin device grade oxides stressed at low voltages, this is often the case and (2.60) is adequate for characterizing trap generation.

$\Delta I/I(0)$ is plotted for a 28Å NMOS oxide stressed at +4.0V in Figure 2.25. When measured in the on-state, $\Delta I/I(0)$ is weakly dependent on sense voltage and tracks the build-up of bulk traps. However, $\Delta I/I(0)$ is significantly higher in the off-state when the sense voltage is near V_{FB} [74]. This will be discussed in detail in Section 2.5.

2.5 Stress induced leakage current due to interface traps (LV-SILC)

In this section, we will present our novel findings regarding the role of interface traps in stress-induced leakage. We will call the respective mechanism low voltage stress induced leakage current, (LV-SILC), which becomes important below the ~ 5eV to 6eV threshold energy for bulk trap generation at oxide thicknesses less than 35Å.

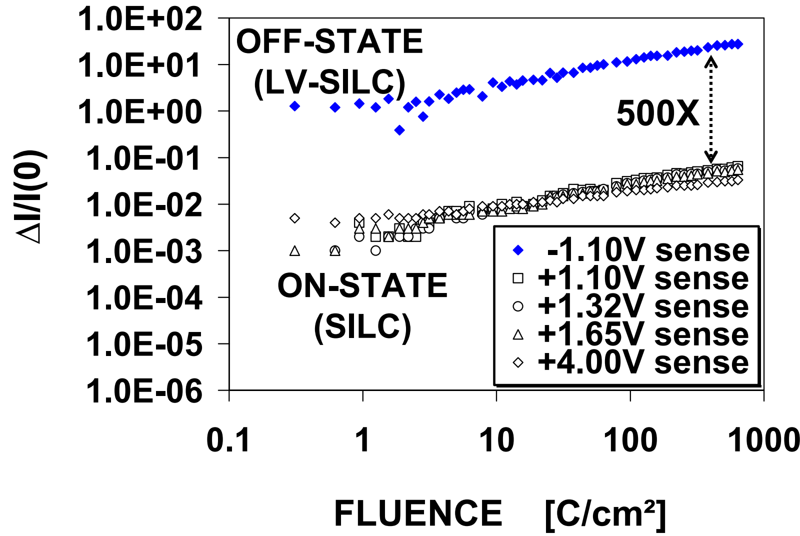

Figure 2.25. Normalized SILC increase for a 28Å NMOS oxide stressed at +4.0V. ΔI/I(0) is weakly dependent sense voltage only when measured in the on-state. After Nicollian, Ref. [74]. © 1999 IEEE.

2.5.1 Energy dependence of trap generation

To determine the energy dependence of trap creation, we must first determine the trap generation rate dN(Q)/dQ, which is found by differentiating (2.60)

$$dN(Q)/dQ = P_G = mbQ^{m-1} \tag{2.67}$$

P_G (or equivalently, dN(Q)/dQ) is equal to ΔI/I(0). Equation (2.60), solved for Q = Q_{BD} is

$$Q_{BD} = (N_{BD}/b)^{1/m} \tag{2.68}$$

N_{BD} is the trap density at breakdown. Inserting (2.68) into (2.67) yields

$$P_G = mN_{BD}/Q_{BD} \tag{2.69}$$

P_G for m = 1 is plotted vs. gate voltage for oxide thickness ranging from 22Å to 50Å in Figure 2.26 [82]. Strictly speaking, this plot is only valid if K <<1 in (2.64) and K' <<1 in (2.65). The slope of the curve begins to flatten out when the gate voltage exceeds about 5V-6V, corresponding to the threshold energy for trap generation through anode hydrogen release or anode hole injection. Accordingly, SILC measurements track the known energy dependence of trap generation. Note that trap creation still occurs below 5V, but the generation rate rapidly diminishes with decreasing voltage. From Figure 2.26, it can be inferred that while the trap generation efficiency is higher when a device is subjected to Fowler-Nordheim tunneling stress, traps are still created, albeit at a lower rate, when an oxide is stressed in the direct tunneling regime.

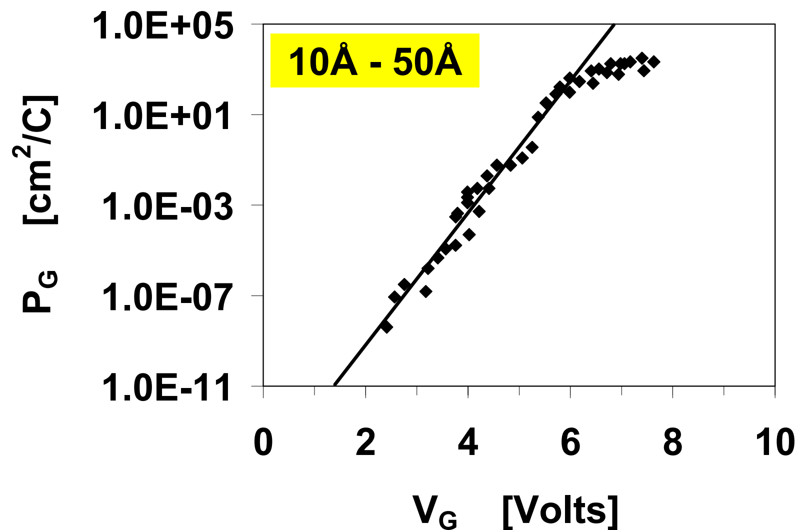

Figure 2.26. Trap generation rate from the normalized SILC increase ΔJ/J(0) vs. gate voltage for oxide thickness ranging from 10Å to 50Å. The solid line is an exponential fit to the data for V_G < 5.5V. After Stathis, Ref. [82]. © 2001 IEEE. Reprinted with permission.

2.5.2 Post stress I_G-V_G characteristics of sub-40Å oxides

Figure 2.27 shows the pre and post-stress I-V characteristics of 28Å NMOS oxides [74]. The devices were progressively stressed for 1 second intervals at incrementally higher voltage until soft breakdown occurred. The short stress time interval was chosen to maximize the range of stress voltages used to facilitate the observation of SILC mechanism threshold energies. A sense voltage dependent low voltage SILC increase (LV-SILC) is observed in the off-state until SBD occurs, indicating that the current increase is not due to either trap assisted tunneling through bulk traps or to charge trapped in the oxide. As the onset of F-N tunneling is around -5V in the off-state for this device, LV-SILC is only observed in the direct tunneling regime.

Pre and post-stress I-V characteristics are shown for 33Å and 37Å NMOS oxides in Figures 2.28 and 2.29 respectively [74]. Comparison of Figures 2.27 - 2.29 shows that LV-SILC diminishes as oxide thickness is increased, and is no longer apparent in the 37Å film (i.e. t_{OX} becomes greater than 1 tunneling length). The weakly sense voltage dependent SILC mechanism that is characteristic of trap assisted tunneling through bulk traps is clearly recognizable when the stress voltage exceeds 5V.

The effect of the measurement delay time on LV-SILC is shown in Figure 2.30 [74]. The device was stressed at +4.2V for 30 seconds. In Figure 2.30, the voltage is stepped in 100 mV increments, with the delay time per step ranging from 0.1 second to 50 seconds per step. The post-stress I-V characteristics are not significantly affected by step delay time, indicating that LV-SILC is not a transient effect.

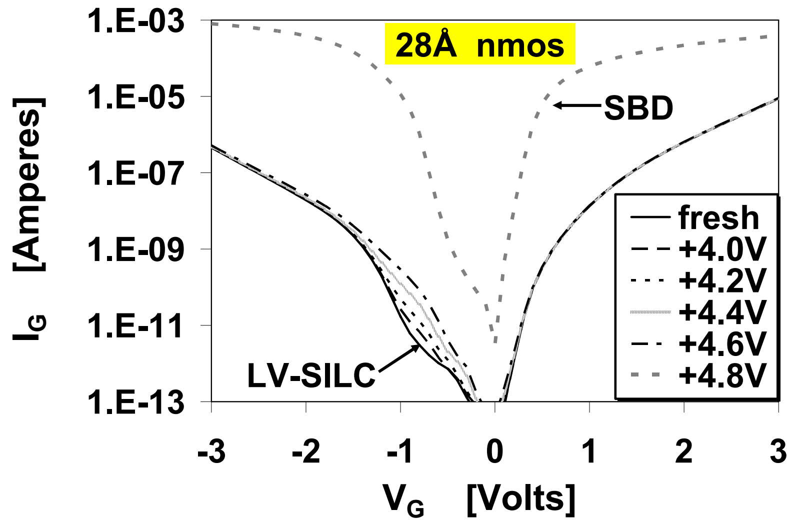

Figure 2.27. Fresh and post-stress I-V characteristics for a 28Å NMOS oxide. The device was incrementally stressed to higher voltages at 1 second intervals. After Nicollian, Ref. [74]. © 1999 IEEE.

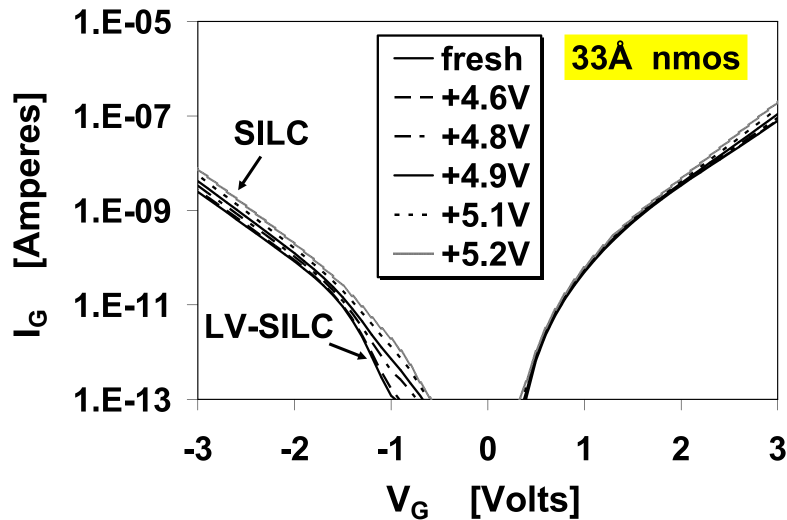

Figure 2.28. Fresh and post-stress I-V characteristics for a 33Å NMOS oxide. The device was incrementally stressed to higher voltages at 1 second intervals. After Nicollian, Ref. [74]. © 1999 IEEE.

LV-SILC is observed for both positive and negative gate polarity stress, as illustrated in Figure 2.31 for 28Å NMOS devices [74]. The devices were stressed at +/- 4.2V for 30 seconds. It can be seen that LV-SILC can occur whether the poly or pwell is the anode and occurs for both accumulation and inversion stress. Because the oxide

electric field is higher in the on-state compared to the off-state at a given $|V_G|$, it is possible to compare the effects of F-N vs. direct tunneling on LV-SILC. In Figure 2.31, the device is stressed in the F-N regime at +4.2V and is stressed in the direct tunneling regime at -4.2V. Accordingly, LV-SILC occurs for both F-N and direct tunneling stress.

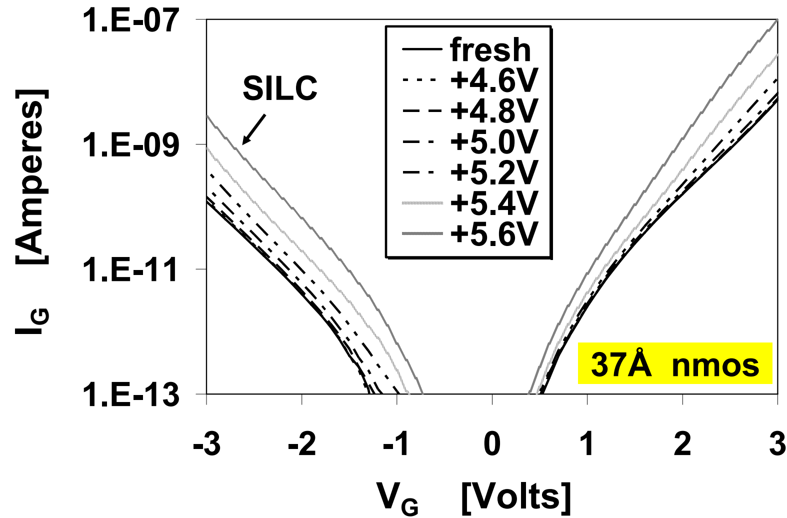

Figure 2.29. Fresh and post-stress I-V characteristics for a 37Å NMOS oxide. The device was incrementally stressed to higher voltages at 1 second intervals. After Nicollian, Ref. [74]. © 1999 IEEE.

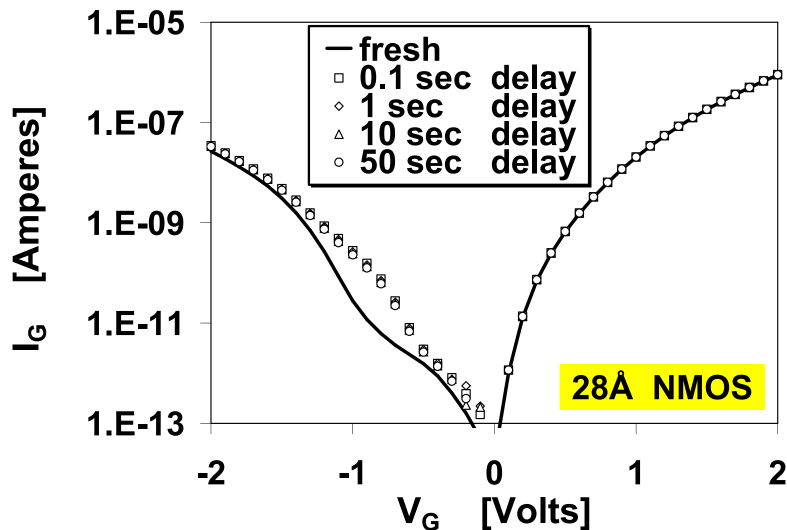

Figure 2.30. Fresh and post-stress I-V characteristics with varying sweep delay time for a 28Å NMOS oxide. After Nicollian, Ref. [74]. © 1999 IEEE.

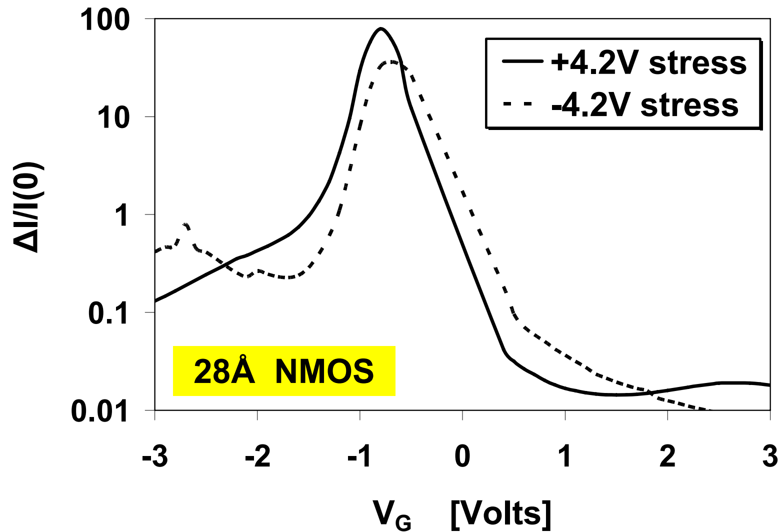

Figure 2.31. Normalized SILC increase ΔI/I(0) for 28Å NMOS devices after +/- 4.2V stress for 30 seconds. After Nicollian, Ref. [74]. © 1999 IEEE.

LV-SILC occurs for both NMOS and PMOS devices, as shown in Figure 2.32 [74].The degradation is highest near V_{FB} for both NMOS and PMOS, and is detected only when the sense voltage is +/- 1V from V_{FB}. This means that LV-SILC is observed only when the energy states within the anode and cathode band gaps are within the same range of electrostatic potential. This indicates that LV-SILC is due to tunneling via interface traps.

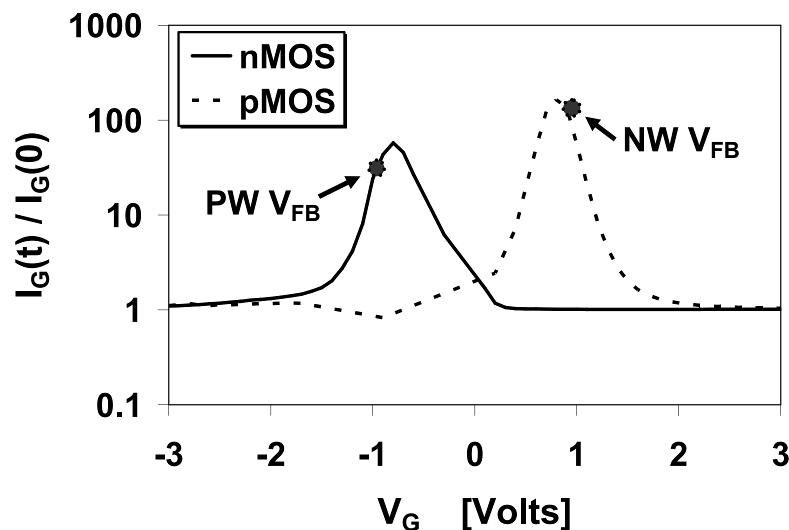

Figure 2.32. SILC increase vs. sense voltage for NMOS and PMOS oxides. In both cases, peak degradation occurs near V_{FB}. After Nicollian, Ref. [74]. © 1999 IEEE.

2.5.3 Model for LV-SILC

In this work, for the first time, interface traps have been identified as being a mechanism for SILC. However, tunneling via interface traps is not a new concept. In the late 1960's, tunneling from metal gates into as-grown Si-SiO_2 interface traps was inferred by noting that wafer processing that led to higher C-V extracted interface trap densities also had higher conductance when the device was biased so that the metal Fermi level aligned opposite the substrate band gap [83]. This effect was observed only when heavily doped p++ substrates were utilized. Under these conditions, the current through the dielectric could be due to either tunneling from the metal into unoccupied states in the silicon valence band, or into interface traps, as shown in Figure 2.33. For tunneling from the metal into the silicon valence band, the substrate must be degenerately doped (p++) so that the substrate Fermi level is at a lower energy than the valence band edge. This provides empty states in the valence band for electrons to tunnel into. For tunneling into interface traps to result in a steady-state current, the electrons that are captured by interface traps must recombine with holes in the accumulated p++ substrate [83].

We have fabricated our devices with poly gate electrodes, and the substrates are not degenerately doped. Accordingly, electrons cannot tunnel from the gate into the valence band as in Figure 2.33 because the pwell Fermi level is above the pwell valence band edge. Also, our LV-SILC peaks occur near V_{FB}, where the field across the oxide is zero. Under these conditions, the driving force for conduction is the separation between the cathode and anode Fermi Levels. In this operating mode, the current through the dielectric can only be due to tunneling into trapping centers, because no other states are available to tunnel into. Accordingly, at $V_G = V_{FB}$, no gate current would flow in our devices in the absence of trap states.

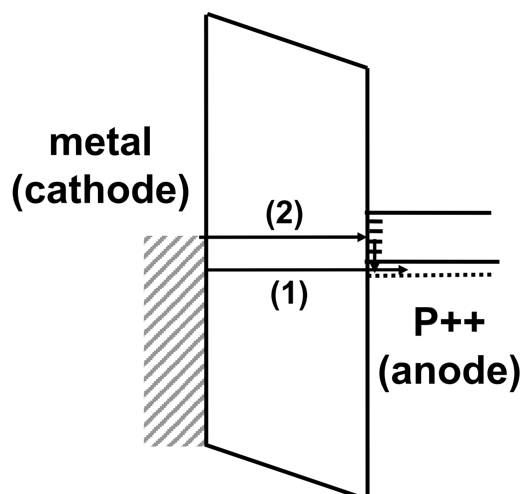

Figure 2.33. Band diagram for tunneling in a metal-oxide-p++ structure with $-V_G$ applied. The possible transport paths are (1) Tunneling of electrons from the metal into the p++ valence band. (2) Tunneling of electrons from the metal into as-grown interface traps at the p++ interface, followed by recombination with holes. After Dahlke, Ref. [83].

Another key difference between our experiment and the results in [83] is that the cathode is supply limited in the portion of the I-V sweep between V_{FB} and 0V in our devices. Figure 2.32 showed that LV-SILC remains significant under these bias conditions. In PMOS devices, the non-degenerately doped nwell Fermi Level is more than a few k_BT below the nwell conduction band edge. This adds complexity to the model for LV-SILC transport, as the possibility of tunneling from trap states in the cathode to trap states in the anode (i.e. a 2-trap tunneling process) becomes a viable transport mechanism. This process is illustrated for a PMOS (nwell) device biased near V_{FB} in Figure 2.34. Tunneling of valence band holes from the p+ poly into nwell interface traps is also a possibility, but the higher barrier height for this process makes it less likely.

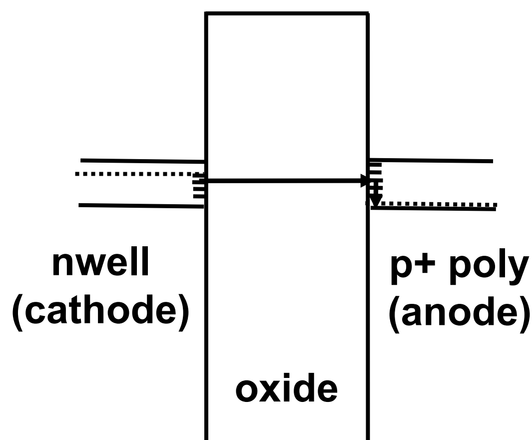

Figure 2.34. Band diagram for PMOS SILC at $V_G = V_{FB}$ illustrating a two-trap LV-SILC process. Electrons tunnel from interface traps below the nwell Fermi Level into interface traps above the p+ poly Fermi level. A steady-state current results when electrons captured in p+ poly-SiO_2 interface traps recombine with majority holes in the p+ poly. The driving force for tunneling is the separation in nwell and p+ poly Fermi Levels.

While papers have been subsequently published by other researchers after our discovery of the LV-SILC effect [84-86], there still some confusion regarding the details of the mechanism. In the remainder of this section, we will provide our interpretation of the phenomena to attempt to clarify LV-SILC effects in NMOS devices.

Band diagrams for LV-SILC are shown for two NMOS devices with differing poly and pwell doping densities at $V_G = V_{FB}$ in Figure 2.35. Only states below the n+ poly (cathode) Fermi Level can emit tunneling electrons and only states above the pwell (anode) Fermi Level can capture tunneling electrons. It can be seen that the energy range in which interface traps can participate in LV-SILC is diminished at the lower doping levels. For NMOS devices, LV-SILC is maximized when the poly is doped n++ and the pwell is doped p++. For the higher doping scenario in Figure 2.35-a, three possible LV-SILC processes are illustrated: (i) Tunneling from the n+ poly conduction

band into pwell interface traps (dashed horizontal arrow). (ii) Tunneling from n+ poly interface traps just below the poly Fermi level into pwell interface traps (solid horizontal arrow). (iii) Tunneling from n+ poly interface traps at energies that are significantly below the poly Fermi level into pwell interface traps (dash-dot horizontal arrow). Of these processes, the least likely is (iii) because the barrier height is the largest. Unless the poly is doped sufficiently high so that the Fermi Level is above the conduction band edge, tunneling from poly interface traps into pwell interface traps becomes a viable transport path. In the n+ poly (cathode), interface traps that are closest to the poly Fermi Level have the highest emission probability because the barrier height is the smallest.

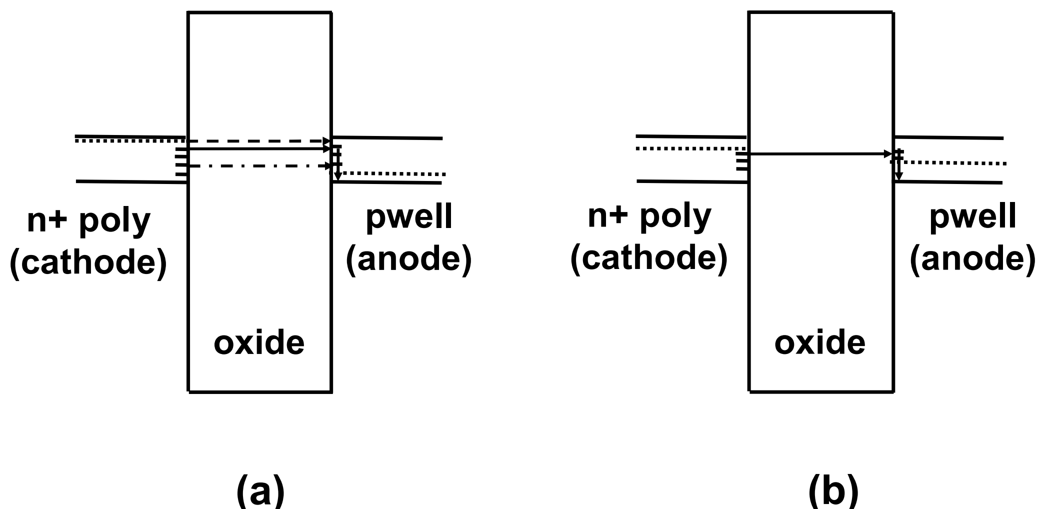

Figure 2.35. Band diagrams for NMOS LV-SILC at $V_G = V_{FB}$. The n+ poly and pwell doping densities are higher in (a). The dashed horizontal arrow represents tunneling from the n+ poly CB to pwell interface traps, while the solid and dash-dot horizontal arrows represent tunneling from n+ poly interface traps to pwell interface traps.

We utilize carrier separation to analyze the effects of interface trap generation on gate, drain, and substrate currents. The time-0 I-V characteristics for the 3-terminal device that will be stressed were shown in Figure 2.5. The pertinent features of Figure 2.5 are:

(1) For V_G > 0V, the gate current is primarily due to injection of conduction band electrons supplied by the drain terminal. Therefore, $I_G \sim I_D$.
(2) For V_G > 1V, electrons from the pwell VB are injected into the oxide. Since a hole will be injected into the pwell for every tunneling VB electron, a substrate current will arise.
(3) For V_G < 0V, most of the electrons that are injected from the gate into the pwell diffuse into the drain contact. Therefore, $I_G \sim I_D$. This is verified by comparing the substrate current with the drain floating versus all three terminals connected as shown in Figure 2.36 [87]. In the off-state, the substrate current is indeed higher with the drain floating since the electrons injected from the gate diffuse out of the pwell rather than drain contact. In the on-state, I_B is also higher with the drain floating (until pwell VB

tunneling becomes significant) due to increased generation rates due to the pwell going into deep depletion without the drain to supply inversion layer electrons.

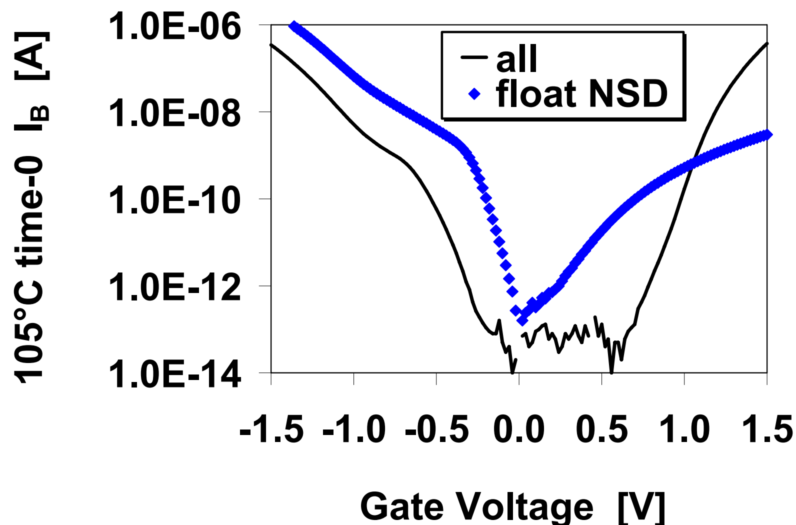

Figure 2.36. 105°C time-0 I_B vs. V_G characteristic of an NMOS device with 12Å EOT SiON gate dielectric with and without the drain floating during the I-V sweep.

The post-stress increase in gate, drain, and substrate currents with all terminals connected is shown for NMOS devices with 12Å device grade SiON dielectrics in Figures 2.37 – 2.39 [87]. All device terminals are connected during stress. To attain a relatively large interface trap density, the devices are stressed in inversion with a back-bias applied, as it has been shown that back bias increases SiON interface trap generation rates [44]. LV-SILC sweeps are performed on devices with gate oxide areas of 30 μm^2. Unlike the SiO_2 films in Figure 2.32, two LV-SILC peaks appear in the gate current for NMOS SiON films when all device terminals are connected during the sense operation [44]. Two LV-SILC peaks in PMOS SiON films have also been reported [46,87]. Three LV-SILC peaks are also present in the drain current (including the dip in the post-stress drain current that occurs at -0.64V) when all three terminals are connected during sense. Three LV-SILC peaks are observed for the substrate current when all terminals are connected during the sense operation. To interpret these data, the possible origins of a current increase resulting from the introduction of interface traps must be determined. We assume that interface traps can be generated at the poly-SiON interface, pwell-SiON interface, and at the NSD-SiON interface in the overlap region.

I. Mechanisms whereby <u>electron</u> tunneling between n+ poly and the pwell via interface traps can result in an increase in <u>substrate</u> current:

(1) For V_G < 0V, electrons tunnel from the n+ poly CB into pwell interface traps, followed by recombination with holes in the pwell (I_{B1}).

(2) For V_G < 0V, electrons tunnel from n+ poly interface traps into pwell interface traps, followed by recombination with holes in the pwell (I_{B2}).

(3) For $V_G < 0V$, electrons tunnel from the n+ poly VB into pwell interface traps, followed by recombination with holes in the pwell (I_{B3}).
(4) For $V_G > 0V$, electrons created in the pwell through BTBT tunnel from the pwell CB into n+ poly interface traps, followed by thermal emission into the n+ poly CB (for electrons trapped within a few k_BT of the n+ poly CB, or by recombination with holes in the n+ poly (I_{B4}).
(5) For $V_G > 0V$, electrons tunnel from the pwell VB into n+ poly interface traps, followed by thermal emission into the n+ poly CB (for electrons trapped within a few k_BT of the n+ poly CB, or by recombination with holes in the n+ poly (I_{B5}).

II. Mechanisms whereby electron tunneling between n+ poly and the pwell via interface traps can result in an increase in drain current:
(6) For $V_G < 0V$, electrons tunnel from n+ poly interface traps into the pwell CB. These electrons result in a drain current when they diffuse into the drain region (I_{D6}).
(7) For $V_G > 0V$, electrons tunnel from pwell interface traps into n+ poly interface traps. In this scenario, electrons that occupy pwell interface traps are supplied by the drain; giving rise to a drain current (I_{D7}).
(8) For $V_G > 0V$, electrons tunnel from pwell interface traps into the n+ poly CB. The electrons that occupy pwell interface traps are supplied by the drain, resulting in a drain current (I_{D8}).

III. Mechanisms whereby electron tunneling between n+ poly and the NSD overlap region via interface traps can result in an increase in drain current:
(9) For $V_G < 0V$, electrons tunnel from n+ poly interface traps into the NSD CB (I_{D9}).
(10) For $V_G < 0V$, electrons tunnel from n+ poly interface traps into NSD interface traps, followed by thermal emission into the NSD CB (for electrons trapped within a few k_BT of the NSD CB), or by recombination with holes in the NSD (I_{D10}).
(11) For $V_G < 0V$, electrons tunnel from the n+ poly CB into NSD interface traps, followed by thermal emission into the NSD CB (for electrons trapped within a few k_BT of the NSD CB), or by recombination with holes in the NSD (I_{D11}).
(12) For $V_G > 0V$, electrons tunnel from NSD interface traps into the n+ poly CB. The electrons that occupy NSD interface traps are supplied by the drain, resulting in a drain current (I_{D12}).

IV. Mechanisms whereby hole tunneling between n+ poly and the pwell via interface traps can result in an increase in substrate current:
(13) For $V_G > 0V$, electron hole pairs are created in the pwell via interface traps when the pwell is at mid-gap potential, followed by injection of holes from the pwell VB into the n+ poly VB (I_{B13}).

V. Mechanisms whereby interface traps result in a substrate current in the NSD/pwell junction.
(14) For $V_G > 0V$, VB electrons in the NSD surface space charge region tunnel into traps and are thermally emitted into the NSD CB through a thermal barrier (i.e. thermal emission tunneling). The resultant hole is injected into the pwell VB (I_{B14}).

We begin with the analysis of the I_G peak at -1.02V. Because the nearest I_D and I_B peaks are at -1.10V and -0.84V respectively, band diagrams are drawn for V_G ~ -1.1V and V_G ~ -0.8V and are shown in Figures 2.40 and 2.41. In this voltage range, the transport paths that can lead to an increase in substrate current are electron tunneling from the n+ poly CB into pwell interface traps followed by recombination with holes (Mechanism 1), electron tunneling from n+ poly interface traps to pwell interface traps followed by recombination with holes (Mechanism 2), or electrons tunneling from the n+ poly VB into pwell interface traps followed by recombination with holes (Mechanism 3). Therefore, the I_G peak at V_G = -1.02V requires traps at the pwell-SiON interface. The least likely is Mechanism 3 due to the high barrier height for this process. As the pwell is depleted at V_G = -0.80V, electron tunneling occurs against the direction of the applied field. The driving force is the energy separation of the cathode and anode Fermi Levels.

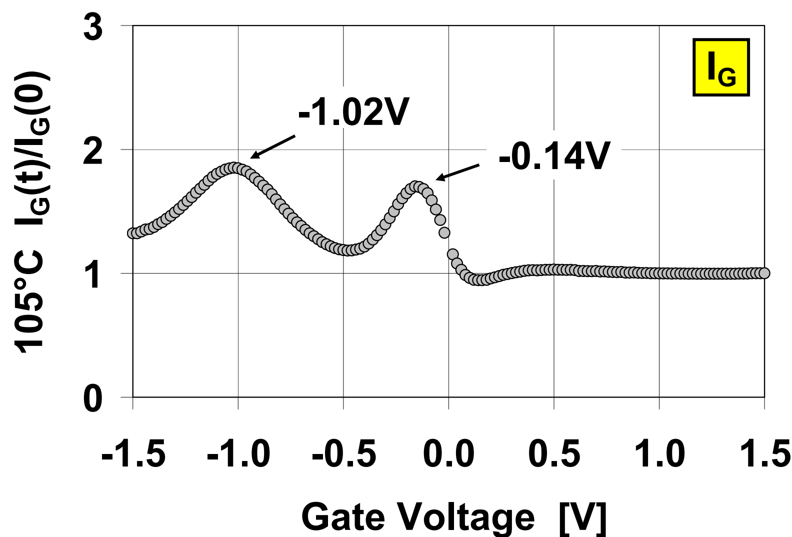

Figure 2.37. $I_G(t)/I_G(0)$ for a 12Å NMOS SiON gate dielectric stressed at V_G = +2.2V, V_B = -6V at 105°C for 300 sec. All terminals are connected during sense. Two LV-SILC peaks are observed in the gate current.

Figures 2.40 and 2.41 show that there are four transport paths that can give rise to a LV-SILC drain current in this voltage range. All involve emission from traps at the n+ poly interface. The least likely path is Mechanism 10 in Figures 2.40 and 2.41 because (i) the barrier height for tunneling is high for this process, as tunneling electrons must be emitted from interface traps near the n+ poly VB edge, (ii) the energy range of traps that can capture tunneling electrons in the NSD is narrow due to the proximity of the NSD Fermi Level to the NSD CB, (iii) the rate of recombination of tunneling electrons captured in NSD interface traps will be low due to the small number of VB holes in the heavily doped NSD. Therefore, the most likely explanations for drain current LV-SILC in this voltage range are tunneling from n+ poly interface traps into the pwell conduction band, followed by diffusion into the drain region (Mechanism 6), or tunneling from n+

poly interface traps into the NSD conduction band (Mechanism 9). We will show that the correct explanation is Mechanism 6.

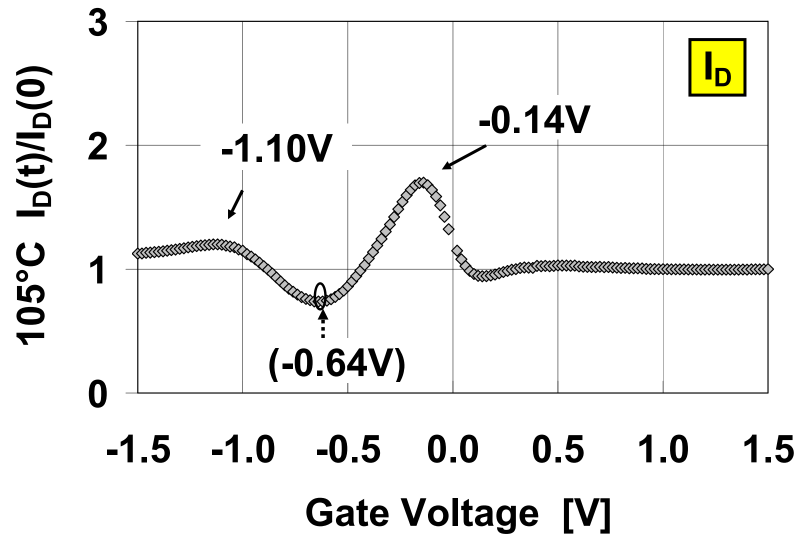

Figure 2.38. $I_D(t)/I_D(0)$ for a 12Å NMOS SiON gate dielectric stressed at V_G = +2.2V, V_B = -6V at 105°C for 300 sec. All terminals are connected during sense. Three LV-SILC peaks are observed in the drain current.

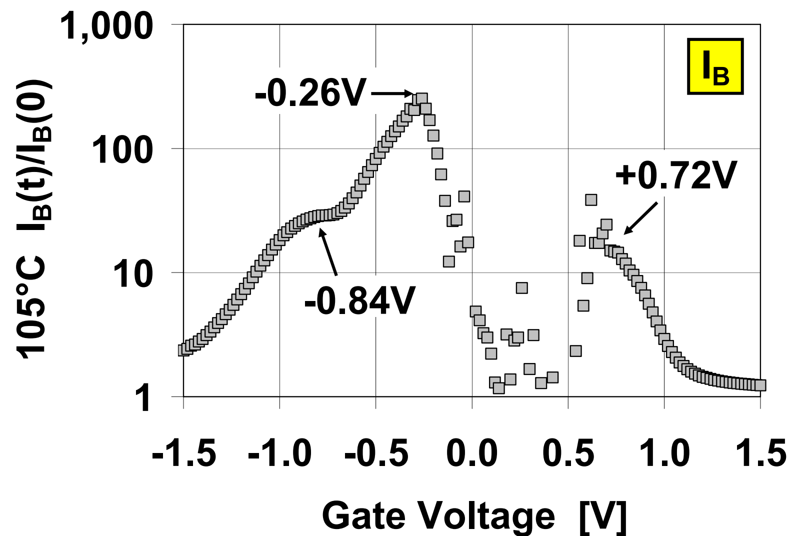

Figure 2.39. $I_B(t)/I_B(0)$ for a 12Å NMOS SiON gate dielectric stressed at V_G = +2.2V, V_B = -6V at 105°C for 300 sec. All terminals are connected during sense. Three LV-SILC peaks are observed in the substrate current.

We will now explain the dip in I_D centered at V_G = -0.64V and the peak at V_G = -0.14V in Figure 2.38. Figures 2.41 (V_G = -0.8V) and 2.42 (V_G = -0.2V) show that the only

LV-SILC paths that lead to an increase in I_D in the voltage range between V_{FB} and 0V involve tunneling into the small-area NSD overlap region. The post stress I-V characteristics are shown in Figure 2.43. In the off-state, the increase in substrate current due to tunneling into pwell interface traps reduces the number of electrons that are available to diffuse into the drain contact compared to the unstressed condition. However, an increase in I_D is observed when V_G is more negative than V_{FB} due to tunneling from n+ poly interface states into the pwell CB (over the entire gate area). This is not energetically feasible between V_{FB} and 0V, and while tunneling from poly interface states into the NSD is still possible, the area of the NSD overlap region is too small to compensate. Consequently, a drop in I_D occurs after stress until the magnitude of the gate voltage is reduced to a small enough value that the LV-SILC I_B peak has sufficiently dropped due to a diminished energy range of pwell interface traps that can participate in LV-SILC. Thereafter, LV-SILC in the gate terminal is dominated by tunneling into the NSD region, resulting in the coincident LV-SILC peaks in the drain and gate currents at V_G = -0.16V. The post-stress increase in I_D sensed with the pwell floating is shown in Figure 2.44. The dip at V_G = -0.64V vanishes, which supports the conclusion that it is due to the reduction in the number of electrons that are available to diffuse into the drain contact relative to the unstressed condition due to electrons injected from the gate recombining in pwell interface traps that were created during stress. The drain current peak at V_G = -1.10V has also disappeared, verifying that it was due to tunneling from n+ poly interface traps into the pwell CB, followed by diffusion of the electrons into the drain contact (Mechanism 6). Tunneling into pwell interface traps would not explain the disappearance of the drain current LV-SILC peak at V_G = -1.10V. Therefore, traps at both n+ poly-SiON and pwell-SiON interfaces participate in LV-SILC when the device is sensed near the V_{FB} for n+ poly over pwell. Only the peak at V_G = -0.16V remains with the pwell floating, so that this LV-SILC peak is indeed due to tunneling between the n+ poly and NSD overlap region. The possible transport paths are electrons tunneling from n+ poly interface traps into the NSD CB (Mechanism 9), or electrons tunneling from the n+ poly CB into NSD interface traps, followed by thermal emission into the NSD CB (for electrons trapped within a few k_BT of the NSD CB), or by recombination with holes in the NSD (Mechanism 11). The most likely is tunneling from n+ poly interface traps into the NSD CB (Mechanism 9) since recombination via NSD interface traps is a low probability process.

We now examine the LV-SILC peak in the substrate current at V_G = -0.26V seen in Figure 2.39. From Figure 2.42, The transport paths involving electron tunneling that can lead to an increase in I_B at V_G = -0.26V are electron tunneling from the n+ poly CB into pwell interface traps followed by recombination with holes (Mechanism 1), and electron tunneling from n+ poly interface traps to pwell interface traps followed by recombination with holes (Mechanism 2). The post stress activation energy for the substrate current is shown in Figure 2.45. The substrate current for the peak at V_G = -0.26V is thermally driven with an activation energy of 0.43eV, which is on the order of ½E_G(Si). Additionally, the surface potential at V_G = -0.26V is approximately mid gap for the pwell. Accordingly, this LV-SILC peak in the substrate current appears to be limited by electron hole pair generation through mid gap defects in the pwell created during stress. If the resulting hole is injected from the pwell into the poly, a substrate current peak

would appear (Mechanism 13). Therefore, in this voltage range, LV-SILC in the substrate terminal requires traps at the pwell interface. No corresponding peak is observed in I_G. This may be consequent of the orders of magnitude difference between the gate and substrate currents, as the increase in I_B at V_G = -0.26V is 4pA, while the gate current is 100x larger. In Figure 2.45, the peak in the activation energy versus gate voltage characteristic appears regardless of whether the drain is connected or floating, indicating that this is not a gate controlled diode surface state generation current [88] flowing through the pwell/NSD junction. The activation energy for the gate current after stress is nearly identical to the substrate current at V_G = -0.26V as shown in Figure 2.46, supporting the conclusion that the creation of electron hole pairs in the pwell through mid gap defects indeed results in carrier injection into the gate terminal.

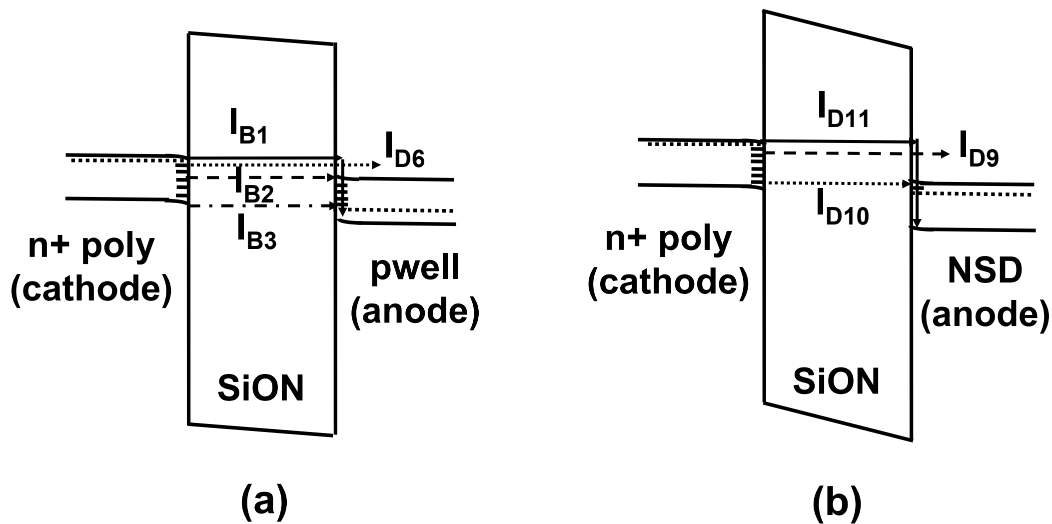

Figure 2.40. Band diagrams for LV-SILC at V_G ~ -1.1V for: (a) Tunneling processes between n+ poly and pwell. (b) Tunneling processes between n+ poly and NSD.

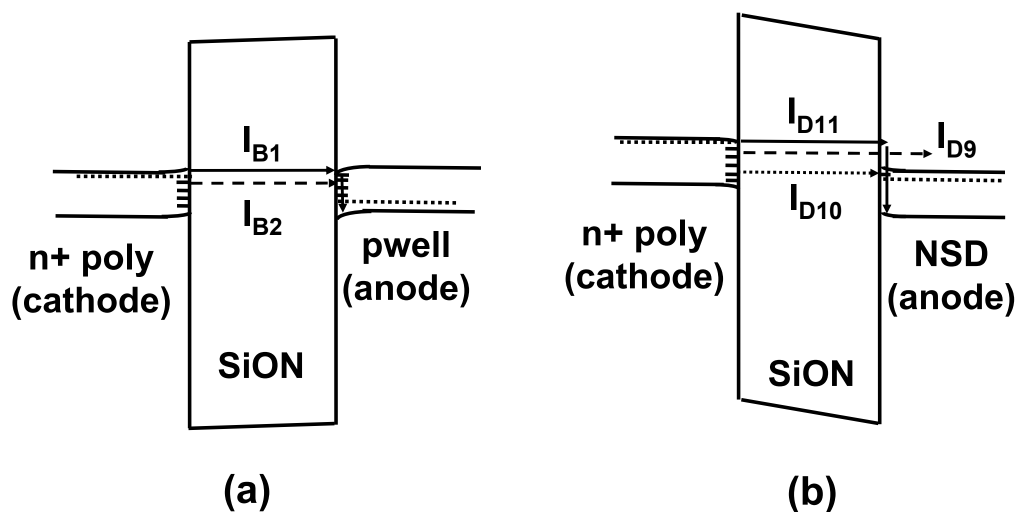

Figure 2.41. Band diagrams for LV-SILC at V_G ~ -0.8V for: (a) Tunneling processes between n+ poly and pwell. (b) Tunneling processes between n+ poly and NSD.

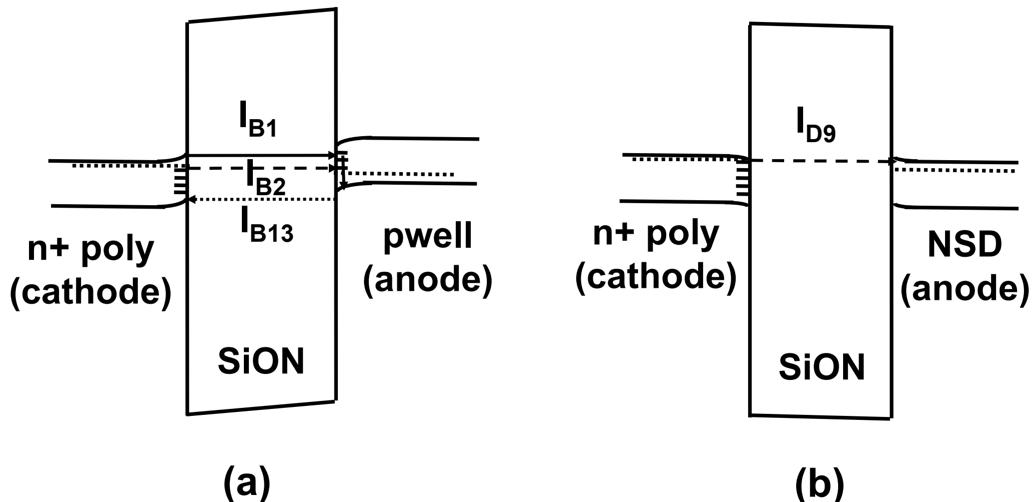

Figure 2.42. Band diagrams for LV-SILC at V_G ~ -0.2V for: (a) Tunneling processes between n+ poly and pwell. (b) Tunneling processes between n+ poly and NSD.

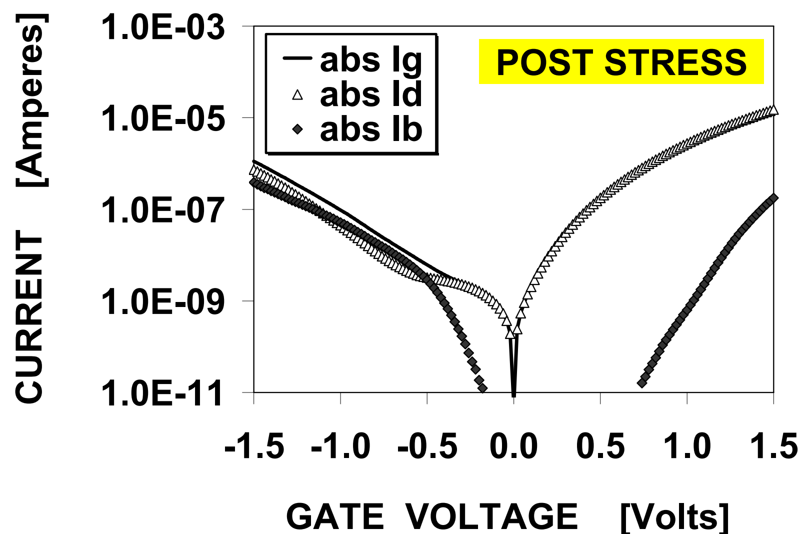

Figure 2.43. Post-stress NMOS terminal currents for 12Å SiON films. The time-0 I-V curves for this device are shown in Figure 2.5.

Our analysis of the LV-SILC peak at V_G = -0.26V shows that carrier separation increases the sensitivity for detecting interface traps and the effects on the device characteristics. This is further illustrated in Figure 2.47 where $I_B(t)/I_B(0)$ is plotted with the drain floating. A plot of $I_G(t)/I_G(0)$ has the same features as Figure 2.47 (not shown). The peak at V_G = -0.90V is still present, supporting the contention that interface traps at the pwell interface contribute to LV-SILC in this voltage range. It can be seen that a new peak appears at V_G = +0.16V when the drain is floating. In this bias condition, I_G has decreased from 16nA to 5pA since the drain cannot supply inversion layer electrons.

Concurrently, I_B has increased from 90fA to 5pA due to higher carrier generation resulting from the device being in deep depletion. Accordingly, this peak represents a small increase in the post-stress currents. From Figures 2.45 and 2.46, I_G and I_B are not strongly thermally activated for this LV-SILC peak. Accordingly, carrier creation in the pwell is not dominated by band gap or mid gap generation but must be due to some other process such as band-to-band tunneling (BTBT). Possible explanations for the LV-SILC peak in I_G and I_B at V_G = +0.16V include injection of electrons created in the pwell through BTBT (Mechanism 4), or injection of electrons from the pwell VB (Mechanism 5) which in either case, are captured in n+ poly interface traps, followed by thermal emission into the n+ poly CB, or by recombination with holes in the n+ poly. A band diagram for the n+ poly/SiON/pwell stack at V_G = +0.20V is shown in Figure 2.48. As there is no I_D LV-SILC peak, the n+ poly/SiON/NSD stack is not shown.

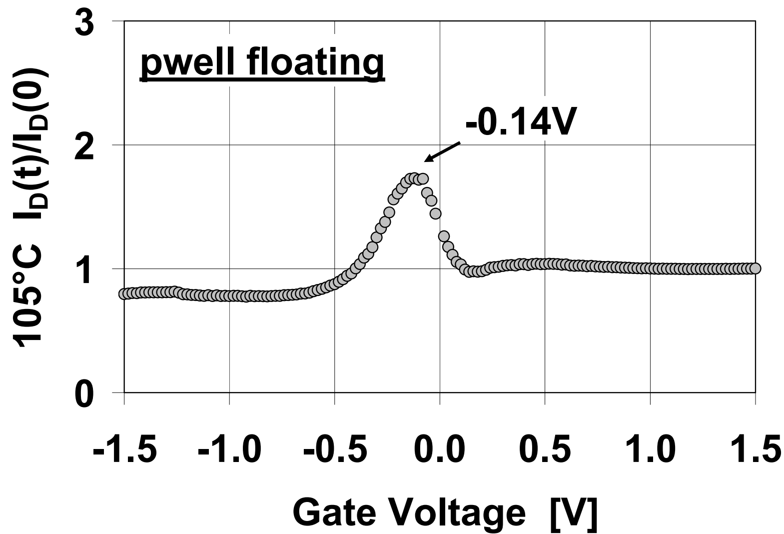

Figure 2.44. $I_D(t)/I_D(0)$ for a 12Å NMOS SiON gate dielectric stressed at V_G = +2.2V, V_B = -6V at 105°C for 300 sec. Only one LV-SILC peak is observed in the drain current when the pwell is floating during sense.

We will now analyze the LV-SILC substrate current peak at V_G = +0.72V that is detected when all device terminals are connected during the sense procedure as shown in Figure 2.39. There are no corresponding peaks observed in either I_D or I_G in Figures 2.37 and 2.38 respectively. Figure 2.45 shows that this peak has an activation energy of about 0.4eV, which is too high for tunneling of pwell VB electrons into n+ poly interface traps to be a viable explanation since the traps in the n+ poly that are unoccupied are too close to the edge of the n+ poly CB to result in a 0.4eV tunneling barrier. This activation energy is also too high for BTBT, and generation of electron hole pairs in the pwell through states within the silicon band gap is unlikely at this voltage since the pwell interface is fully inverted. Generation of electron pairs via mid gap states in the n+ poly is also unlikely since the surface potential is significantly less than mid gap. Tunneling of holes from the n+ poly VB to the pwell interface traps can also be ruled out since this would result in an I_D LV-SILC peak due to recombination of inversion

layer electrons (supplied by the drain) with the holes that are captured in pwell interface traps. Accordingly, there is no adequate explanation for this I_B peak that involves tunneling between n+ poly and pwell. The most likely explanation is that electrons in the NSD surface space charge region tunnel from the NSD VB into traps and are subsequently emitted into the NSD CB through a 0.4eV thermal barrier as shown in Figure 2.49. The resultant hole is injected into the pwell VB (Mechanism 14). For this peak, the sum of V_G + ΔH is approximately E_G(Si), which is consistent with a thermal emission tunneling process. Accordingly, the substrate current peak at V_G = +0.72V requires interface traps at the NSD-SiON interface and does not have a gate current component. However, a drain current component is present but cannot be detected.

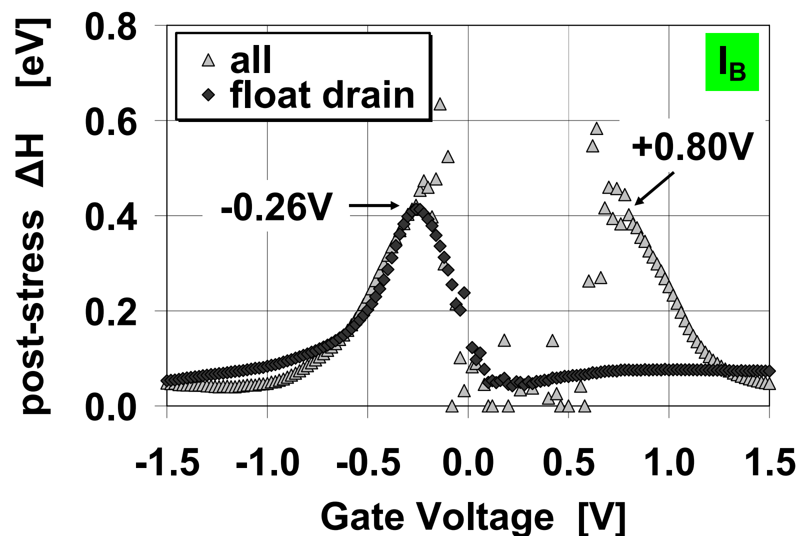

Figure 2.45. Post-stress activation energy for substrate current for a 12Å NMOS SiON gate dielectric stressed at V_G = +2.2V, V_B = -6V at 105°C for 300 sec. The LV-SILC peaks at V_G = -0.26V and V_G = +0.80V are thermally activated.

To summarize our analysis, the most likely physical explanations for the trap peaks observed in our data are as follows:

(1) V_G = -1.02V: This peak is comprised of two components; electrons tunneling from n+ poly interface traps into to pwell interface traps, and electrons tunneling from the n+ poly CB into pwell interface traps. A steady state current results when the electrons trapped in pwell interface states recombine with holes.
(2) V_G = -0.64V: This peak (actually a dip in the post stress drain current) is due to electrons tunneling from the n+ poly CB into pwell interface traps. The capture of these electrons in pwell interface states reduces in the number of electrons that are available to diffuse into the drain contact relative to the unstressed condition due to electrons injected from the gate into pwell interface traps recombine with holes.
(3) V_G = -0.26V: This peak is due to electron hole pair generation through mid gap traps at the pwell interface, with the resulting hole being injected from the pwell into the poly.

(4) V_G = -0.14V: This peak is due to electrons tunneling from n+ poly interface traps into the NSD CB.
(5) V_G = +0.16V: This peak is due to either (a) electron hole pair creation in pwell through GIDL or BTBT, followed by the injection of the electron into n+ poly interface traps, or (b) tunneling from pwell VB into n+ poly interface traps. In either case, the traps involved in the LV-SILC process are at the n+ poly interface.
(6) V_G = +0.72V: This peak is due to thermal tunneling emission of electrons in the NSD surface space charge region from the NSD VB into interface traps, and are subsequently emitted into the NSD CB through a 0.4eV thermal barrier. The resultant hole is injected into the pwell VB. This process occurs in the NSD surface space charge region and is a diode current rather than LV-SILC.

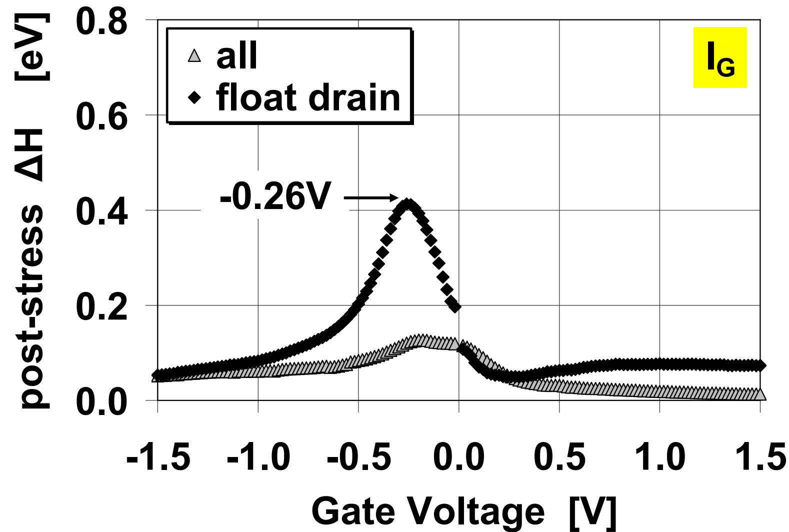

Figure 2.46. Post-stress activation energy for gate current for a 12Å NMOS SiON gate dielectric stressed at V_G = +2.2V, V_B = -6V at 105°C for 300 sec. The LV-SILC peak at V_G = -0.26V with the drain floating is thermally activated.

In this work, we have shown that using carrier separation in three terminal devices with all terminals connected as well when either the drain or pwell is floating, a total of six peaks are observed, where five of them contribute to LV-SILC. This technique significantly increases the information that can be extracted from LV-SILC measurements. We have shown that four LV-SILC peaks due to interface traps can be observed in I_B, three peaks are observed in I_D, and three peaks are observed in I_G. A fifth peak in the substrate current is also present, but it is due to thermal emission tunneling in the NSD surface space charge region and does result in LV-SILC. Our analysis shows that the three of the four I_B LV-SILC peaks involve traps at the pwell-SiON interface. Two trap peaks at the pwell-SiON interface have been observed in C-V measurements [44].

In this experiment, our most important finding is that when LV-SILC in the gate terminal is sensed near V_{FB}, the tunneling path is indeed a two trap process, where electrons

can tunnel from n+ poly interface traps into pwell interface traps, and can also tunnel from the n+ poly CB into pwell interface traps. In the off-state, the substrate current only senses traps at the pwell interface whereas depending on the bias, the drain current can sense traps at the poly or pwell interface. In the on-state, the LV-SILC in the drain current does not detect traps at any interface, while the substrate current senses traps at the poly interface. These findings are tabulated in Figure 2.50.

Note that LV-SILC is still observed as V_G approaches 0V. This corresponds to a diminishing range of energies over which interface traps can participate in the tunneling process. Therefore, LV-SILC is either an elastic tunneling process, or an inelastic tunneling process with a small energy loss [74].

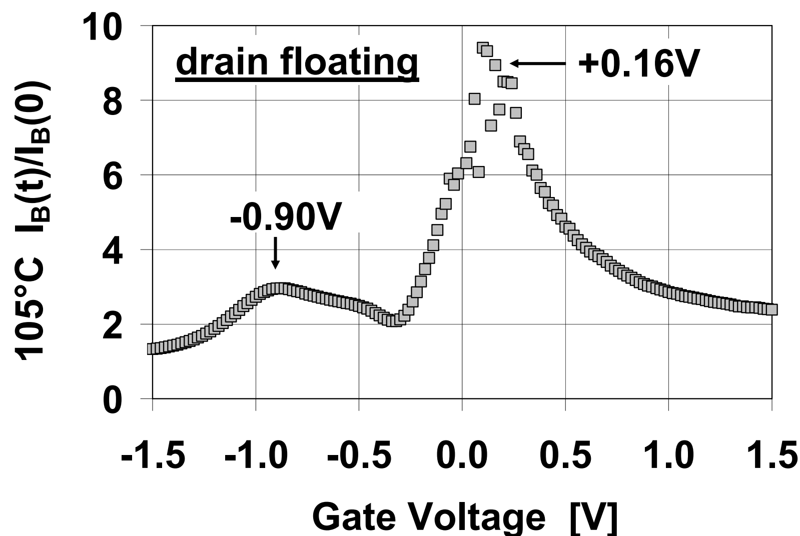

Figure 2.47. $I_B(t)/I_B(0)$ for a 12Å NMOS SiON gate dielectric stressed at V_G = +2.2V, V_B = -6V at 105°C for 300 sec. A new LV-SILC peak appears in the gate current when the drain is floating during sense.

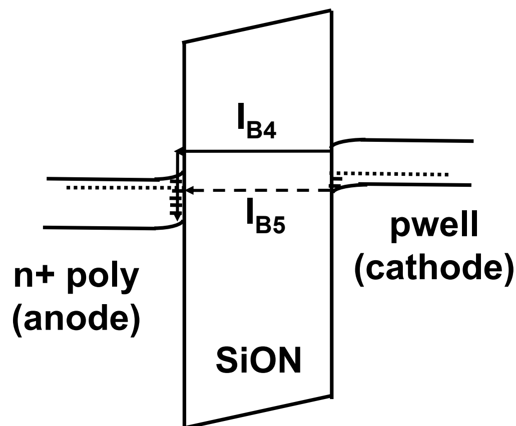

Figure 2.48. Band diagram for LV-SILC at V_G ~ +0.2V for tunneling processes between n+ poly and pwell.

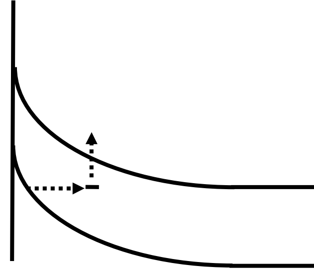

Figure 2.49. Band diagram at V_G ~ +0.2V for thermal emission tunneling in the NSD surface space charge region.

V_{PEAK} [V]	terminal sensing traps at poly interface	terminal sensing traps at pwell interface	terminal sensing traps at NSD interface
- 1.02	I_D, I_G	I_B, I_G	
- 0.64		I_B, I_D, I_G	
- 0.26		I_B, I_G†	
- 0.14			I_D, I_G
+ 0.16	I_B, I_G		
+ 0.72			I_B, I_D†

Figure 2.50. Tabular summary of the interface traps sensed in each device terminal. † Denotes that current due to interface states is present but too small to detect. All of the peaks represent LV-SILC current through the gate terminal except at V_G = + 0.72V.

2.6 Chapter summary

In this chapter, we reviewed time-0 transport, charge trapping, and stress induced leakage current through bulk traps (SILC) prior to presenting our original work regarding the role of interface traps on low voltage stress induced leakage current (LV-SILC).

In ultra thin, virgin device grade SiO_2 and SiON gate dielectrics, the important transport mechanisms are Fowler Nordheim tunneling (F-N) and direct tunneling (D-T). F-N tunneling occurs through a triangular barrier when the voltage across the oxide exceeds the barrier height. The slope of the I-V curve is steep for this mechanism because the tunneling distance decreases with increasing voltage. Direct tunneling occurs through a trapezoidal barrier when the oxide voltage is less than the barrier height and the oxide thickness is less than a tunneling length (about 40Å). The tunneling distance is always

t_{OX} for D-T. While the barrier height is too high for holes to participate in F-N tunneling in the poly gate SiON system, hole currents can be significant in the D-T regime. For p+ poly gate PMOS devices, holes can provide the dominant contribution to the gate current when the gate voltage is between 0V and -1V.

For SiON gate dielectrics, material properties such as the dielectric constant, band gap, barrier height, and effective mass are linear functions of the amount of nitrogen incorporated. The increase in dielectric constant realized through nitrogen incorporation reduces gate leakage by enabling a thicker physical oxide thickness at a given EOT. However, there is a concurrent reduction in barrier height and effective mass with increasing nitrogen. Accordingly, reducing J_G through the addition of nitrogen requires a trade-off between dielectric constant against the barrier height and effective mass.

To achieve a holistic understanding of SILC and LV-SILC phenomena, the effects of charge trapping on the device properties must also be considered. Because of the enormity of papers published in this area, we have compacted portions of this body of knowledge into a simple picture.

The two most important trap potentials are Coulombic and neutral centers. The determination of the type of center is typically based on the measured capture cross section and its dependence on field and temperature. Coulombic traps are the largest in size and have the strongest field dependence. At low fields, the capture cross section decreases with increasing field due to Frenkel-Poole barrier lowering, which increases the re-emission probability. The temperature dependence of the low field capture cross section increases with decreasing temperature due to the reduction in the number of phonons, and excited states with lower binding energies participating in the trapping process

Most of the traps in SiO_2 are neutral centers and their capture cross sections are smaller than Coulombic centers. Because the potentials for these states are highly localized, barrier lowering effects are weak, resulting in a weaker field dependence of the capture cross section compared to Coulombic traps. The temperature dependence of the capture cross section for neutral states is opposite that of Coulombic traps.
First order kinetics are widely applied to the analysis of charge trapping phenomenon. This model provides an adequate quantitative description of charge trapping when the reaction rate is proportional to a single reactant, no trapping parameter other than the capture cross section depends on the electric field, and the ratio of displacement current to total current is small. The latter condition may be violated for short stress times, large capture cross sections, or large trap densities. At high fields in thick oxides, detrapping can occur via impact ionization. Because detrapping due to impact ionization is a field dependent process, first order kinetic equations must be modified if these effects are present. Carriers can also be detrapped via field emission. This effect only gives rise to transient currents, with a decay rate that is inversely proportional to time.

There is much confusion in the literature regarding the number and nature of traps in SiO_2. This is partly due to processing differences and improper measurement

procedures. Traps are classified as being as-grown or stress generated. They are differentiated by the magnitude of electric field and carrier fluence required to fill them.

Early investigations on electron trapping focused on the role of water in the creation of as-grown electron traps. While these water-related centers are generally negligible in modern devices, three different stress generated electron traps have been identified. All are neutral centers that have spatially uniform distributions in the oxide. Their existence has been deduced from the field dependence of detrapping rates. Electron traps can be created by either hole or electron injection.

At least three different hole traps can be present in modern technologies. One is as-grown and the other two are stress generated. Hole traps can also be created by either electron or hole injection. One of the stress generated traps is the well known anomalous positive charge. These centers can be reversibly charged and discharged without further hole injection. The second type of stress generated hole trap can recharged without hole injection but can only be discharged at high fields. Unlike electron traps, hole traps tend to be concentrated near the oxide interfaces. There are some ambiguities regarding whether hole traps are Coulombic or neutral centers.

Below about 70Å thickness, a post stress steady-state current appears. It is primarily observed in the direct tunneling regime and can be significant when the stress voltage exceeds about 5V. This stress induced leakage current, or SILC, can result in a 100X increase in gate leakage. The mechanism for SILC transport is trap assisted tunneling through oxide bulk traps. The trap centers causing SILC are uniformly distributed neutral electron traps, and the traps with the maximum tunneling probability are located at $t_{OX}/2$. SILC is an inelastic transport process with an energy loss of about 1.5eV. In practice, the magnitude of the SILC increase is weakly dependent on sense voltage.

In this chapter, we presented our novel discovery regarding stress induced leakage currents in ultra thin gate dielectrics. Below about 35Å, another steady-state leakage instability appears. Unlike bulk trap SILC, the post stress current increase is sense voltage dependent for this new mechanism. It dominates the off-state post stress leakage current when the stress voltage is less than 5V. This effect, which we call low-voltage stress induced leakage current, or LV-SILC, occurs in both NMOS and PMOS, is observed after either negative or positive gate polarity stress, and can be induced from either F-N or D-T stress. For SiO_2, it is only observed in the direct tunneling regime when the sense voltage is +/- 1V from V_{FB}. This means that LV-SILC is a measurable effect only when the energy states within the anode and cathode band gaps are within the same range of electrostatic potential. This indicates that LV-SILC is due to tunneling via interface traps.

For SiO_2, the LV-SILC peaks occur at sense voltages near V_{FB}, where the electric field across the oxide is zero. Accordingly, the driving force for LV-SILC is the energy separation between anode and cathode Fermi Levels. This explains why the effect is only seen in oxides with thickness less than one tunneling length. Without trap states present in the silicon band gap, little gate current should be measurable at the flat band

condition. Only traps below the cathode Fermi Level can emit tunneling electrons and only traps above the anode Fermi Level can capture tunneling electrons. Accordingly, increasing the doping density of the silicon electrodes increases the energy range of traps that participate in LV-SILC. Because LV-SILC is still observed as the gate voltage approaches 0V, LV-SILC must be either an elastic tunneling process, or an inelastic tunneling process with a very small relaxation energy.

While we are the first to recognize the role of interface traps on post-stress leakage, tunneling via interface states is not a new concept. In the 1960's, tunneling from metal gates into as-grown Si-SiO_2 interface traps was inferred by noting that wafer processing that led to higher C-V extracted interface trap densities also had higher conductance when the device was biased so that the metal Fermi level aligned opposite the substrate band gap. It was proposed that the current through the dielectric was due to electrons tunneling from the metal into interface traps, followed by recombination with holes.

In our devices, because LV-SILC is observed in a bias regime where the cathode is supply limited (between 0V and V_{FB}), this raises the possibility that electrons can tunnel from cathode interface traps rather than from the cathode conduction band, making LV-SILC a 2-trap tunneling process. While this interpretation has been questioned by other researchers following our original discovery, we have presented new results in this dissertation showing that LV-SILC is a 2-trap tunneling mechanism in SiON gate dielectrics. This has been accomplished through careful analysis of carrier separation measurements in three terminal devices with all terminals connected as well when either the drain or pwell is floating. We factored in the generation of interface face traps at the n+ poly-SiON interface, pwell-SiON interface, and in the overlap region at the n+ poly-NSD interface.

These techniques significantly increase the information that can be extracted from LV-SILC measurements. A total of six trap peaks were observed, where five of them contribute to LV-SILC. We have shown that four LV-SILC peaks due to interface traps can be observed in I_B, three peaks are observed in I_D, and three peaks are observed in I_G. A fifth peak in the substrate current is also present, but it is due to thermal emission tunneling in the NSD surface space charge region and does result in LV-SILC in the gate terminal.

When LV-SILC in the gate terminal is sensed near V_{FB}, the tunneling path is indeed a two trap process, where electrons are emitted by both n+ poly interface traps and n+ poly CB and tunnel into pwell interface traps. In the off-state, the substrate current only senses traps at the pwell interface whereas depending on the bias, the drain current can sense traps at the poly or pwell interface. In the on-state, the LV-SILC in the drain current does not detect traps at any interface, while the substrate current senses traps at the poly interface.

2.7 References

[1] J. Frenkel, "On the Electrical Resistance of Contacts between Solid Conductors", *Physical Review*, **36**, 1604 (1930)
[2] R.H. Fowler and L. Nordheim, "Electron Emission in Intense Electric Fields", *Proceedings of the Royal Society A*, **119**, 173 (1928)
[3] E.H. Snow, "Fowler-Nordheim Tunneling in SiO_2 Films", *Solid State Communications*, **5**, 813 (1967)
[4] M. Lenzlinger and E.H. Snow, "Fowler-Nordheim Tunneling in Thermally Grown SiO_2", *Journal of Applied Physics*, **40**, 278 (1969)
[5] J. Maserjian, "Tunneling in Thin MOS Structures", *Journal of Vacuum Science and Technology*, **11**, 996 (1974)
[6] Z.A. Weinberg, "On Tunneling in Metal-Oxide-Silicon Structures", *Journal of Applied Physics*, **53**, 5052 (1982)
[7] M. Depas, B. Vermeire, P.W. Mertens, R.L. Van Meirhaeghe, and M.M. Heyns, "Determination of Tunneling Parameters in Ultra-Thin Oxide Layer Poly-Si/SiO_2/Si Structures", *Solid-State Electronics*, **38**, 1465 (1995)
[8] J. Sune, P. Olivo, and B. Ricco, "Quantum Mechanical Modeling of Accumulation Layers in MOS Structure", *IEEE Transactions on Electron Devices,* **39**, 1732 (1992)
[9] R. Stratton, G. Lewicki, and C.A. Mead, "The Effect of Non Parabolic Energy Bands on Tunneling through Thin Insulating Films", *Journal of the Physics and Chemistry of Solids*, **27**, 1599 (1966)
[10] B. Brar, G.D. Wilk, and A.C. Seabaugh, "Direct Extraction of the Electron Tunneling Effective Mass in Ultra thin SiO_2", *Applied Physics Letters*, **69**, 2728 (1996)
[11] R. Ludeke, E. Cartier, and A. Schenk, "Determination of the Energy Dependent Conduction Band Mass in SiO_2", *Applied Physics Letters*, **75**, 1407 (1999)
[12] E. Merzbacher, *Quantum Mechanics*, New York: Wiley, 1970, Chapter 7
[13] S. Horiguchi and H. Yoshino, "Evaluation of Interface Potential Barrier Heights between Ultra Thin Silicon Oxides and Silicon", *Journal of Applied Physics,* **58**, 1597 (1985)
[14] A.H. Goodman, "Photoemission of Electrons from n-Type Degenerate Silicon into Silicon Dioxide", *Physical Review*, **152**, 785 (1966)
[15] Z.A. Weinberg and A. Harstein, "Photon Assisted Tunneling from Aluminum into Silicon Dioxide", *Solid State Communications*, **20**, 179 (1976)
[16] J.G. Simmons, "Electric Tunneling Effect between Dissimilar Electrodes Separated by a Thin Insulating Film", *Journal of Applied Physics*, **34**, 2581 (1963)
[17] K.F. Schuegraf, C.C. King, and C.Hu, "Ultra Thin Silicon Dioxide Leakage Current and Scaling Limit", *VLSI Technical Digest*, 18 (1992)
[18] Z.A. Weinberg, W.C. Johnson, and M.A. Lampert, "Determination of the Sign of Carrier Transported across SiO_2 films on Si", *Applied Physics Letters,* **25**, 42 (1974)
[19] F.J. Grunther and J. Maserjian, *The Physics of SiO_2 and its Interfaces*, edited by S.T. Pantelides, New York: Pergamon, 1978, pp. 389-395
[20] E. Wu, *IRPS Tutorials*, 201 (2005)
[21] M. Rodder, et. al., "A 1.2V, 0.1μm Gate Length CMOS Technology: Design and Process Issues", *IEDM Technical Digest*, 623 (1998)

[22] P.E. Nicollian, et. al., "Extending the Reliability Scaling Limit of SiO_2 through Plasma Nitridation", *IEDM Technical Digest*, 545 (2000)
[23] D.M. Brown, P.V. Gray, F.K. Heumann, H.R. Philipp, and E.A. Taft, "Properties of $Si_xO_yN_z$ Films on Si", *Journal of the Electrochemical Society*, **115**, 311 (1968)
[24] H. Yang and G. Lucovsky, "Integration of Ultra thin (1.6 ~ 2.0 nm) RPECVD Oxynitride Gate Dielectrics into Dual Poly-Si Gate Submicron CMOSFET's", *IEDM Technical Digest*, 245 (1999)
[25] X. Guo and T.P. Ma, "Tunneling Leakage Current in Oxynitride: Dependence on Oxygen/Nitrogen Content", *IEEE Electron Device Letters,* **19**, 207 (1998)
[26] N. Yasuda, K. Muraoka, and H. Satake, "The Relation between Dielectric Constant and Film Composition of Ultra-thin Silicon Oxynitride Films: Experimental Evaluation and Analysis of Nonlinearity", *Solid State Devices and Materials*, 486 (2001)
[27] A.E. Islam, et. al., "Gate Leakage vs. NBTI in Plasma Nitrided Oxides: Characterization, Physical Principles, and Optimization", *IEDM Technical Digest*, 329 (2006)
[28] H. Yang, H. Niimi, J.W. Keister, G. Lucovsky, and J.E. Rowe, "The Effects of Interfacial Sub-Oxide Transition Regions and Monolayer Level Nitridation on Tunneling Currents in Silicon Devices", *IEEE Electron Device Letters,* **21**, 76 (2000)
[29] D.J. DiMaria, *The Physics of SiO_2 and its Interfaces*, edited by S.T. Pantelides, New York: Pergamon, 1978, pp. 160-178, and references contained therein
[30] M. Lax, "Cascade Capture of Electrons in Solids", *Physical Review*, **119**, 1502 (1960)
[31] D.J. DiMaria, F.J. Feigl, and S.R. Butler, "Capture and Emission of Electrons at 2.4eV Deep Trap Level in SiO_2 Films", *Physical Review B*, **11**, 5023 (1975)
[32] T.H. Ning, "High Field Capture of Electrons by Coulomb Attractive Centers in Silicon Dioxide", *Journal of Applied Physics*, **47**, 3203 (1976)
[33] J.J. Tzou, J. Y. Sun, and C. T. Sah, "Field Dependence of Two Large Hole Capture Cross Sections in Thermal Oxide on Silicon", *Applied Physics Letters*, **43**, 861 (1983)
[34] E.H. Nicollian, C.N. Berglund, P.F. Schmidt, and J.M. Andrews, "Electrochemical Charging of Thermal SiO_2 Films by Injected Electron Currents", *Journal of Applied Physics*, **42**, 5654 (1971)
[35] C.H. Henry and D.V. Lang, "Non-Radiative Capture and Recombination by Multi-Phonon Emission in GaAs and GaP", *Physical Review B*, **15**, 989 (1977)
[36] P. E. Blöchl and J.H. Stathis, "Hydrogen Electrochemistry and Stress Induced Leakage Current in Silica", *Physical Review Letters*, **83**, 372 (1999)
[37] J. Maserjian and N. Zamani, "Behavior of the Si/SiO_2 Interface Observed by Fowler-Nordheim Tunneling", *Journal of Applied Physics*, **53**, 559 (1982)
[38] D.J. DiMaria, "Determination of Insulator Bulk Trapped Charge Densities and Centroids from Photocurrent-voltage Characteristics of MOS Structures", *Journal of Applied Physics*, **47**, 4073 (1976)
[39] Y. Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, "Trap Generation and Occupation Dynamics in SiO_2 under Charge Injection Stress", *Journal of Applied Physics*, **60**, 2024 (1986)
[40] D.J. DiMaria, E. Cartier, and D. Arnold, "Impact Ionization, Trap Creation, Degradation, and Breakdown in Silicon Dioxide Films on Silicon", *Journal of Applied Physics*, **73**, 3367 (1993)

[41] J.F. Verwey, "Non-avalanche Injection of Hot Carriers into SiO_2", *Journal of Applied Physics*, **44**, 2681 (1973)
[42] D.J. DiMaria, "Defect Generation under Substrate Hot Electron Injection into Ultra-thin Silicon Dioxide Layers", *Journal of Applied Physics*, **86**, 2100 (1999)
[43] E.H. Nicollian, A. Goetzberger, and C.N. Berglund, "Avalanche Injection Currents and Charging Phenomenon in Thermal SiO_2", *Applied Physics Letters*, **15**, 174 (1969)
[44] P.E. Nicollian, A.T. Krishnan, C.W. Bowen, S. Chakravarthi, C.A. Chancellor, and R.B Khamankar, "The Roles of Hydrogen and Holes in Trap Generation and Breakdown in Ultra-thin SiON Dielectrics", *IEDM Technical Digest*, 403 (2005)
[45] V. Reddy, et. al., "Impact of Negative Bias Temperature Instability on Digital Circuit Reliability", *Proceedings of the IRPS*, 248 (2002)
[46] A.T. Krishnan, et. al., "Material Dependence of Hydrogen Diffusion: Implications for NBTI Degradation", *IEDM Technical Digest,* 705 (2005)
[47] S.M. Sze, Physics of Semiconductor Devices, 2ND edition, New York: Wiley, 1981, pp. 446-448
[48] J.S. Brugler and P.G.A. Jespers, "Charge Pumping in MOS Devices", *IEEE Transactions on Electron Devices*, **ED-16**, 297 (1969)
[49] G. Groeseneken, H.E. Maes, N. Beltran, and R.F. DeKeersmaecker, "A Reliable Approach to Charge Pumping Measurements in MOS Transistors", *IEEE Transactions on Electron Devices*, **ED-31**, 42 (1984)
[50] D.J. DiMaria, D.A. Buchanan, J.H. Stathis, and R.E. Stahlbush, "Interface States Induced by the Presence of Trapped Holes near the Silicon-Silicon Dioxide Interface", *Journal of Applied Physics*, **77**, 2032 (1995)
[51] J.M. Aitken and D.R. Young, "Electron Trapping by Radiation-Induced Charge in MOS Devices", *Journal of Applied Physics*, **47**, 1196 (1976)
[52] Y. Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, "Dynamic Model of Trapping-Detrapping in SiO_2", *Journal of Applied Physics*, **58**, 2252 (1985)
[53] R. Moazzami and C. Hu, "Stress Induced Current in Thin Silicon Dioxide Films", *IEDM Technical Digest*, 139 (1992)
[54] D.J. Dumin, S. Mopuri, S. Vanchinathan, R.S. Scott, and R. Subramoniam, "High Field Emission Related Thin Oxide Wearout and Breakdown", *Proceedings of the IRPS*, 143 (1994)
[55] R.S. Scott and D.J. Dumin, "The Transient Nature of Excess Low-Level Leakage Currents in Thin Oxides", *Journal of the Electrochemical Society*, **142**, 586 (1995)
[56] W.D. Zhang, J.F. Zhang, M. Lalor, D. Burton, G.V. Groeseneken, and R. Degraeve, "Two Types of Neutral Electron Traps Generated in the Gate Silicon Oxide", *IEEE Transactions on Electron Devices*, **49**, 1868 (2002)
[57] D.J. Breed, "Non-ionic Room Temperature Instabilities in MOS Devices", *Solid State Electronics*, **17**, 1229 (1974)
[58] D.R. Young, E.A. Irene, D.J. DiMaria, R.F. De Keersmaecker and H.Z. Massoud, "Electron Trapping in SiO_2 at 295 and 77°K", *Journal of Applied Physics*, **50**, 6366 (1979)
[59] S.K. Lai, "Effects of Gate Metals on Interface Effects in Metal Oxide Semiconductor Systems after Electron Trapping", *Journal of Applied Physics*, **53**, 7104 (1982)
[60] R.A. Gdula, The Effects of Processing on Hot Electron Trapping in SiO_2", *Journal of the Electrochemical Society*, **123**, 42 (1976)

[61] J.F. Zhang, S. Taylor, and W. Eccleston, “A Quantitative Investigation of Electron De-trapping in SiO_2 under Fowler-Nordheim Stress”, *Journal of Applied Physics*, **71**, 5989 (1992)
[62] W.D. Zhang, J.F. Zhang, C.Z. Zhao, M.H. Chang, G. Groeseneken, and R. Degraeve, “Electrical Signature of the Defect Associated with Gate Oxide Breakdown”, *IEEE Electron Device Letters*, **27**, 393 (2006)
[63] D.J. DiMaria and J.W. Stasiak, “Trap Creation in Silicon Dioxide Produced by Hot Electrons”, *Journal of Applied Physics*, **65**, 2342 (1989)
[64] D.J. DiMaria, “Explanation for the Polarity Dependence of Breakdown in Ultra thin Silicon Dioxide Films”, *Applied Physics Letters*, **68**, 3004 (1996)
[65] C.S. Jeng, T.R. Ranganath, C.H. Huang, H.S. Jones, and T.T.L. Chang, “High-Field Generation of Electron Traps and Charge Trapping in Ultra-thin SiO_2”, *IEDM Technical Digest*, 388 (1981)
[66] S. Ogawa, N. Shiono, and M. Shimaya, “Neutral Electron Trap Generation in SiO_2 by Hot Holes”, *Applied Physics Letters*, **56**, 1329 (1990)
[67] L.P. Trombetta, F.J. Feigl, and R.J. Zeto, “Positive Charge Generation in Metal-Oxide-Semiconductor Capacitors”, *Journal of Applied Physics*, **69**, 2512 (1991)
[68] J.F. Zhang, C.Z. Zhao, A.H. Chen, G. Groeseneken, and R. Degraeve, “Hole Traps in Silicon Dioxides – Part I: Properties”, *IEEE Transactions on Electron Devices*, **51**, 1267 (2004)
[69] M.V. Fischetti, “Generation of Positive Charge in Silicon Dioxide during Avalanche and Tunnel Electron Injection”, *Journal of Applied Physics*, **57**, 2860 (1985)
[70] J.F. Zhang, H.K. Sii, G. Groeseneken, and R. Degraeve, “Hole Trapping and Trap Generation in the Gate Silicon Dioxide”, *IEEE Transactions on Electron Devices*, **48**, 1127 (2001)
[71] G. Van den Bosch, G. Groeseneken, and H.E. Maes, ”Oxide and Interface Degradation resulting from Substrate Hot Hole Injection in Metal-Oxide-Semiconductor Field Effect Transistors at 295 and 77K”, *Journal of Applied Physics*, **75**, 2073 (1994)
[72] D.J. DiMaria, Z.A. Weinberg, and J.M. Aitken, “Location of Positive Charges in SiO_2 Films on Si Generated by VUV Photons, X-rays, and High Field Stressing”, *Journal of Applied Physics*, **48**, 898 (1977)
[73] D.J. DiMaria and E. Cartier, “Mechanism for Stress Induced Leakage Currents in Thin Silicon Dioxide Films”, *Journal of Applied Physics*, **78**, 3883 (1995)
[74] P.E. Nicollian, M. Rodder, D.T. Grider, P. Chen, R.M. Wallace, and S.V. Hattangady, “Low Voltage Stress Induced Leakage Current in Ultra thin Gate Oxides”, *Proceedings of the IRPS*, 400 (1999)
[75] S. Takagi, N. Yasuda, and A. Toriumi, “Experimental Evidence of Inelastic Tunneling and New I-V Model for Stress Induced Leakage Current”, *IEDM Technical Digest*, 323 (1996)
[76] C. Chang, C. Hu, and R. Brodersen, “Quantum Yield of Electron Impact Ionization in Silicon”, *Journal of Applied Physics*, **57**, 302 (1985)
[77] R.C. Alig, S. Bloom, and C.W. Struck, “Scattering by Ionization and Phonon Emission in Semiconductors”, *Physical Review B*, **22**, 5565 (1980)
[78] A. Ghetti, M.A. Alam, J. Bude, and F. Ventura, “Assessment of Quantum Yield Experiments via Full Band Monte Carlo Simulations”, *IEDM Technical Digest*, 873 (1997)

[79] A. Ghetti, M.A. Alam, J. Bude, D. Monroe, E. Sangiorgi, and H. Vaidya, "Analysis of Trap Assisted Conduction Mechanisms through Silicon Dioxide Films using Quantum Yield", *IEDM Technical Digest*, 723 (1999)
[80] M.A. Alam, "SILC as a Measure of Trap Generation and Predictor of t_{BD} in Ultra thin Oxides", *IEEE Transactions on Electron Devices*, **49**, 226 (2002). M.A. Alam, Notes from Purdue University course EE650R, Lecture 23.3, *Stress Induced Leakage Current* http://cobweb.ecn.purdue.edu/~ee650/downloads/Lecture23.pdf
[81] T. Nigam, R. Degraeve, G. Groeseneken, M.M. Heyns, and H.E. Maes, "A Fast and Simple Methodology for Lifetime Prediction of Ultra thin Oxides", *Proceedings of the IRPS*, 381 (1999)
[82] J.H. Stathis, "Physical and Predictive Models of Ultra Thin Oxide Reliability in CMOS Devices and Circuits", *Proceedings of the IRPS*, 132 (2001)
[83] W.E. Dahlke and S.M. Sze, "Tunneling in Metal Oxide Silicon Structures", *Solid State Electronics*, **10**, 865 (1967)
[84] A. Ghetti, E. Sangiori, J. Bude, T.W. Sorsch, and G. Weber, "Tunneling into Interface States as a Reliability Monitor for Ultrathin Oxides", *IEEE Transactions on Electron Devices*, **47**, 2358 (2000)
[85] F. Crupi, C. Ciofi, A. Germano, G. Iannaccone, J.H. Stathis, and S. Lombardo, "On the Role of Interface States in Low Voltage Leakage Currents of Metal-Oxide-Semiconductor Structures", *Applied Physics Letters*, **80**, 4597 (2002)
[86] J.H Stathis, "Broad Energy Distribution of NBTI-Induced Interface States in P-MOSFETs with Ultra-Thin Nitrided Oxide", *Proceedings of the IRPS*, 1 (2004)
[87] P.E. Nicollian and A.T. Krishnan, "Two Trap Model for Low Voltage Stress Induced Leakage Current", submitted to the *Journal of Applied Physics* (2007)
[88] A.S. Grove, *Physics and Technology of Semiconductor Devices*, New York: Wiley, 1967, pp. 289-316

CHAPTER 3

Models for dielectric breakdown

3.1 Introduction

In this chapter, we will present our contributions to the understanding of dielectric breakdown models. There has been a decades long controversy in the industry regarding breakdown models. This lack of consensus has been partly due to the failure to develop a comprehensive picture that incorporates the effect of carrier energies in gate oxide failure. Because of the central importance of energy in determining which breakdown mechanisms are operative at a given stress condition, we will begin this chapter with an expansion of the discussion on the time-0 tunneling processes that were presented in Section 2.2 to include the role of transport on carrier energies. We will then review the properties of field-based models for dielectric breakdown. After discussing the limitations of field-based models, we will show that breakdown is gate voltage driven at low stress voltages.

3.2 Tunneling and energy

In this section, we will discuss the three electrode limited thermal tunneling mechanisms that are encountered in virgin, thermally grown device grade SiO_2 films. The operative tunneling mechanism is determined by the oxide thickness, voltage, and field.

3.2.1 Definition of tunneling processes

We define the following terminology regarding tunneling processes:

In an elastic process, there is no change in kinetic or total energy until after the tunneling electron or hole enters the anode.

In a ballistic process, there are no collisions (no change in total energy) until after the tunneling electron or hole enters the anode. The kinetic energy in the oxide may increase during ballistic tunneling

In an inelastic process, both kinetic and total energy of the tunneling electron or hole change in the oxide.

3.2.2 Direct tunneling (DT)

As discussed in Section 2.2.2, direct tunneling occurs through a trapezoidal barrier when the oxide voltage is less than the Si-SiO_2 barrier height and the oxide thickness is less than one tunneling length (about 40Å). As shown in Figure 3.1, direct tunneling is

an elastic process. The electrons that direct tunnel from the cathode are thermal so that the energy distribution is nearly mono-energetic when they arrive at the anode. After entering the anode, the electrons may be accelerated in the anode space charge region. For direct tunneling, the average and maximum kinetic energies of electrons delivered to the anode, prior to any scattering in the anode are respectively [1,2]

$$<KE_{Si}> = qV_{OX} \tag{3.1}$$

$$KE_{MAX,Si} = q|V_G| = |E_{FN}| - |E_{FP}| \tag{3.2}$$

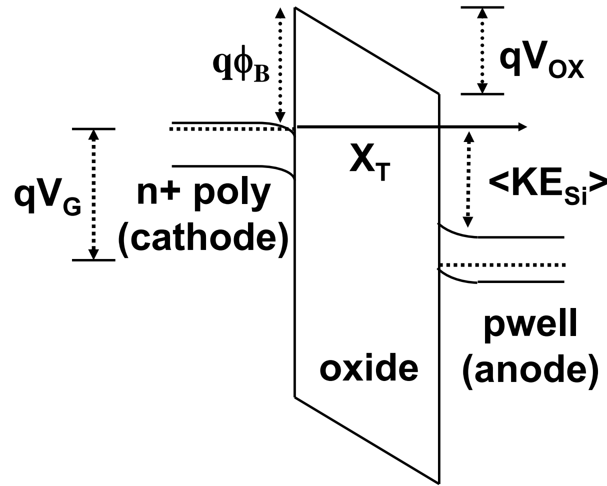

Figure 3.1. Energy band diagram for direct tunneling in an NMOS device. Injection is from n+ poly. The average energy of electrons entering the pwell anode is qV_{OX} and the maximum energy in the pwell anode is qV_G.

3.2.3 Ballistic Fowler-Nordheim tunneling (BFN)

For oxide voltages greater than the barrier height, electrons tunnel from the cathode conduction band into the oxide conduction band through a triangular barrier. Unlike direct tunneling, BFN is not an elastic process because tunneling electrons gain kinetic energy in the oxide conduction band as they traverse the oxide. However, Fowler-Nordheim tunneling is a ballistic process if there are no scattering events prior to the electron entering the anode as illustrated in Figure 3.2, where $<KE_{OX}>$ is the average kinetic energy of electrons with respect to the bottom of the conduction band.

For BFN, the average and maximum kinetic energies of electrons delivered to the anode, prior to any scattering in the anode are respectively [1,2]

$$<KE_{Si}> = qV_{OX} = \Phi_B + <KE_{OX,ANODE}> \tag{3.3}$$

$$KE_{MAX,Si} = q|V_G| = |E_{FN}| - |E_{FP}| \tag{3.4}$$

$<KE_{OX,ANODE}>$ is the average kinetic energy of electrons in the oxide conduction band arriving at the anode interface.

One of the characteristics of BFN is that oscillations may be observed in the I-V characteristics due to constructive interference between incident and reflected electron waves propagating in the oxide conduction band [3]. As the oxide thickness increases, the distance that the electron travels in the oxide conduction band also increases, leading to a higher probability of scattering. Accordingly, electron wave coherence effects are only observable when the oxide thickness is less than about 70Å [4].

Monte Carlo simulations of the average anode kinetic energy in the oxide as a function of (V_{OX} - Φ_B) are shown in Figure 3.3 for oxide thickness ranging from 50Å to 245Å. The straight line is the condition for ballistic Fowler-Nordheim transport. For oxide thickness less than 50Å with V_{OX} < 9V (corresponding to 6V in Figure 3.3), electron transport is ballistic. As V_{OX} and t_{OX} become larger, the number of scattering events rapidly increases, resulting in a transition from ballistic to quasi-ballistic transport. As can be seen in Figure 3.3, when the thickness has been increased to 245Å, the slope flattens out above 6V. In this regime, the transport mechanism is steady-state Fowler-Nordheim tunneling. The transition from ballistic to quasi-ballistic to steady state tunneling occurs at lower voltages as the oxide thickness increases due higher scattering probabilities when the electrons traverse a longer distance in the oxide conduction band. We will discuss the mechanisms leading to SSFN in the next sub-section.

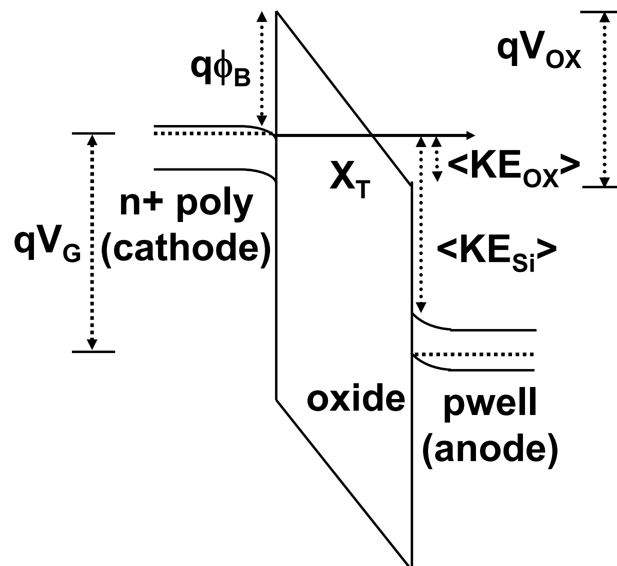

Figure 3.2. Energy band diagram for ballistic Fowler-Nordheim tunneling in an NMOS device. Injection is from n+ poly. The average energy of electrons entering the pwell anode is qV_{OX} and the maximum energy of electrons in the pwell anode is qV_G.

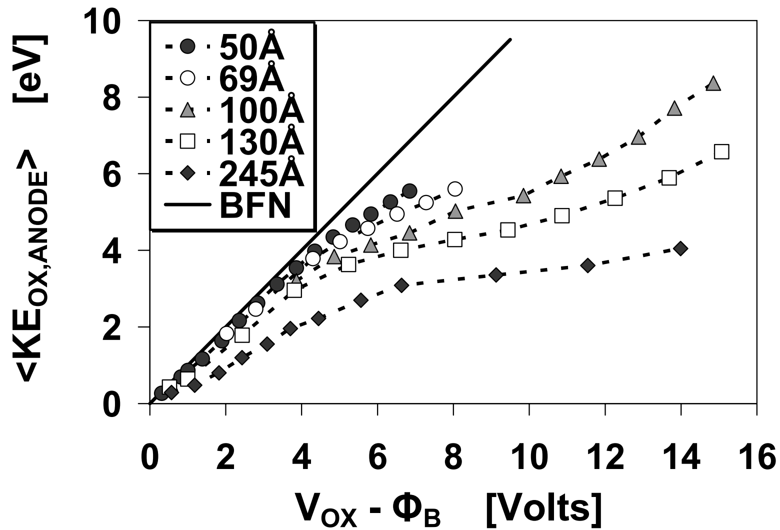

Figure 3.3. Average anode kinetic energy of electrons in the oxide conduction band vs. (V_{OX} - Φ_B) for oxide thickness ranging from 50Å to 245Å. The straight line represents the ballistic tunneling condition. After DiMaria, Ref. [2]. © 1996 American Institute of Physics. Reprinted with permission.

3.2.4 Steady-State Fowler Nordheim tunneling (SSFN)

A band diagram for steady-state Fowler-Nordheim tunneling is shown in Figure 3.4. After entering the oxide, electrons gain kinetic energy until they are scattered by oxide phonons. Accordingly, SSFN is an inelastic process. Figure 3.5 shows the average kinetic energy in the oxide vs. anode field [5]. The average energy is determined by measuring the quantum yield from carrier separation techniques and using the quantum yield universal curves as previously discussed in Section 2.4.2. Below about 5MV/cm, the primary scattering mechanism is longitudinal optical (LO) phonon emission with an energy loss of 0.15eV [6]. In Figure 3.5, it can be seen that electrons have nearly thermal energies below about 1.5MV/cm as they are stabilized by LO phonons. Beyond this field, electron heating begins as the electrons gain more energy from the field than they lose to LO phonons, and the entire electron energy population goes into thermal runaway from the LO phonon modes. However, between about 4MV/cm to 16MV/cm, the rate of increase in average kinetic energy with field reduces and the energy increases very slowly from about 2-3eV to 6eV in this field range [5]. In this regime, electrons are stabilized by acoustic phonon scattering [7] and the average energy does not exceed 6eV all the way to breakdown [5,8]. This is the steady-state transport condition, where the energy gained by electrons from the field is approximately balanced by the energy lost to acoustic phonon emission.

The slope of Figure 3.5 yields the energy relaxation length, which is the average distance that an electron travels in the oxide conduction band before it is stabilized by acoustic phonons and reaches the steady-state condition. For E_{OX} > 5MV/cm, the

relaxation length obtained from Figure 3.5 is ~23Å. A similar analysis of vacuum emission data results in a relaxation length of ~40Å [9].

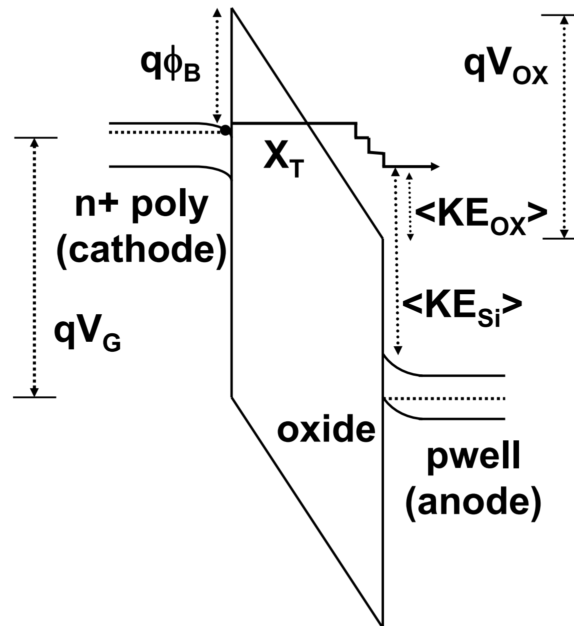

Figure 3.4. Energy band diagram for steady-state Fowler-Nordheim tunneling in an NMOS device. Injection is from n+ poly. The average energy of electrons entering the pwell anode is less than or equal to qV_{OX}.

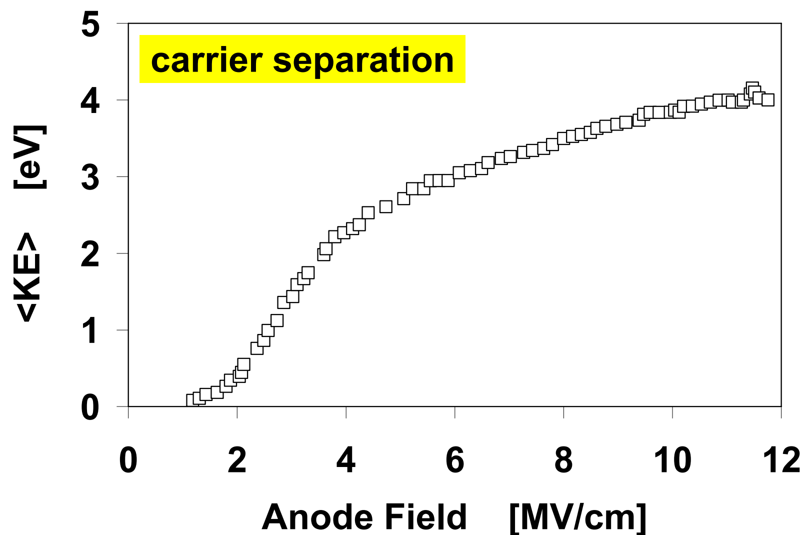

Figure 3.5. Average kinetic energy in the oxide conduction band as a function of anode field determined from carrier separation measurements. After DiMaria, Ref. [5]. © 1993 American Institute of Physics. Reprinted with permission.

The average electron energy for steady state F-N tunneling can be calculated by solving a simple phenomenological equation for the energy lost per unit length [10]

$$d\langle KE(x)\rangle/dx = qE_{OX} - \langle KE(x)\rangle/\lambda \quad (3.5)$$

Where λ is the energy relaxation length and $\langle KE(x)\rangle$ is the average kinetic energy in the oxide at a distance x from the classical turning point (tunneling distance) X_T. The 1ST term on the right hand side is the energy gained per unit distance from the field, and the 2ND term on the right hand side represents the energy loss due to scattering. If λ is constant, which, from the slope in Figure 3.5, seems to be a reasonable assumption for E_{OX} > 5MV/cm, the solution to equation (3.5) for the average kinetic energy acquired in the oxide conduction band at the anode is

$$\langle KE_{OX,ANODE}\rangle = qE_{OX}\lambda(1-\exp(X_T - t_{OX})/\lambda)) \quad (3.6)$$

From (3.6), for $\lambda \approx 30$Å and E_{OX} = 6MV/cm, the average kinetic energy of electrons in the oxide conduction band at the anode is ~2eV for oxide thickness greater than 150Å. For $(t_{OX} - X_T) \gg \lambda$, (3.6) reduces to

$$\langle KE_{OX,ANODE}\rangle = qE_{OX}\lambda \quad (3.7)$$

For steady state transport to occur, the oxide thickness must be greater than the tunneling length at the oxide field of interest, plus the heat up distance (i.e. $t_{OX} > X_T + \lambda$) for the electron kinetic energy to be stabilized by acoustic phonon scattering. Accordingly, we would expect that the minimum oxide thickness for SSFN would be about 60Å. This has been confirmed to be the case from carrier separation experiments, as the quantum yield becomes nearly independent of thickness when t_{OX} exceeds 66Å, up to the maximum thickness of 101Å used in this study [9].

Thus far, we have not considered the occurrence of impact ionization in the oxide in the steady-state transport regime. We will expand our discussion to include this topic in Chapter 4.

3.3 Field based breakdown models

Historically, a controversy regarding the use of the 1/E Model [11-17] and E-Model [18-24] for the analysis of TDDB data raged on in the literature. Both models have a range of thickness, energy, and field where they provide good fits to data. The 1/E Model results in the most optimistic lifetime projections and is applicable at high stress voltages. The E-Model is widely applied at low stress voltages and yields more conservative lifetimes. We will collectively refer to the 1/E and E-Models as field-based models. A property of any field based model is that for a fixed gate area and stress temperature, the time to breakdown at a given oxide electric field is independent of oxide thickness. In this section, we will briefly review field-based models and discuss their limitations.

3.3.1 The 1/E Model

The formulation for the 1/E Model given in equation (1.18) will be re-stated:

$$t_{BD} = t_{0C}\exp(B_C/E_{OX}) \qquad (3.8)$$

This relationship arises from the assumption that breakdown occurs after a critical hole fluence (Q_P) has passed through the oxide [12]. For a fixed stress temperature and oxide thickness, Q_P is independent of E_{OX} [12]. The holes are generated through either (a) impact ionization in the oxide, or (b) impact ionization in the anode, resulting in the injection of a hole from the anode into the oxide. In both cases, the holes are generated by tunneling electrons. Of all of the breakdown mechanisms, impact ionization processes require the highest energies. The threshold energy is approximately $<KE_{OX}> > 9eV$ ($V_{OX} > 12V$) for impact ionization in the oxide [5,25] and $V_{OX} > 6V$ for impact ionization in the anode [2,15]. For the case of impact ionization in the oxide, the time to breakdown will have the following dependence [12]:

$$t_{BD} = Q_P/J_H \qquad (3.9)$$

where J_H is the hole current, and is the product of the F-N tunneling current (J_{FN}) and the impact ionization rate (γ):

$$J_H = \gamma J_{FN} \qquad (3.10)$$

Combining (3.9) and (3.10),

$$t_{BD} = Q_P/\gamma J_{FN} \qquad (3.11)$$

The impact ionization rate [26] and F-N tunneling current are, respectively:

$$\gamma = \gamma_0\exp(-H/E_{OX}) \qquad (3.12)$$

$$J_{FN} = AE_{OX}^2\exp(-B/E_{OX}) \qquad (3.13)$$

Inserting (3.12) and (3.13) into (3.11) and dropping the E_{OX}^2 pre-factor in (3.13),

$$t_{BD} = Q_P C\exp((B+H)/E_{OX}) \qquad (3.14)$$

Where C is a constant. The temperature dependence is primarily carried in the Q_P term. Inspection of (3.8) and (3.14) shows that

$$B_C = B+H \qquad (3.15)$$

When ln t_{BD} is plotted vs.$1/E_{OX}$, the slope is B_C and is approximately independent of field at high fields in thick oxides. In this regime, the Fowler-Nordheim B, which is approximately 240MV/cm, is the major contribution to B_C, which is about 320MV/cm [13]. The value of H is approximately 80 MV/cm [13,26] between 7 MV/cm to 14 MV/cm

[14], which nearly covers the field range from the onset of F-N tunneling to destructive breakdown. The acceleration factor for the 1/E Model is equal to B_C/E_{OX}^2 so it increases rapidly with decreasing electric field.

Since $t_{BD} = Q_{BD}/J_{FN}$, the charge to breakdown for the 1/E Model is

$$Q_{BD} = Q_P C' \exp(H/E_{OX}) \quad (3.16)$$

Where C' is a constant. In experiments, Q_{BD} has been observed to have the same field and thickness dependence of γ [14], in agreement with (3.16). Accordingly, the 1/E Model formulation resulting from impact ionization in the oxide correctly captures the energy and fluence dependence of breakdown.

One of the puzzling aspects of this model is that breakdown due to holes is still observed at voltages below where impact ionization in the oxide would be expected to occur. One explanation for this observation is that tunneling electrons impact ionize in the anode (rather than in the oxide) and a hole is injected into the oxide. This is known as the anode hole injection model (AHI) [15-17]. To comprehend AHI, another term is needed in (3.11) to account for the transmission probability (P_H) for a hole generated in the anode to tunnel into the oxide:

$$t_{BD} = Q_P/\gamma' P_H J_{FN} \quad (3.17)$$

Where γ' is the impact ionization rate in the anode material rather than in the oxide. Equation (3.17) is often expressed as

$$t_{BD} = Q_P/G_H J_{FN} \quad (3.18)$$

Where G_H is the product of the impact ionization rate and the hole transmission probability. Q_{BD} is then

$$Q_{BD} = Q_P/G_H \quad (3.19)$$

Accordingly, the 1/E Model formulation resulting from impact ionization in the anode also correctly captures the energy and fluence dependence of breakdown.

Unlike the oxide impact ionization regime, B_C is not a constant 320MV/cm for anode hole injection, but rather increases with decreasing field [16,17]. This is due to a sharp decrease of G_H with decreasing field, and ultimately leads to an E-Model, then to a voltage Model rather than a 1/E Model dependence at low voltages [17]. The two impact ionization mechanisms that give rise to the 1/E Model behavior will be discussed in detail in Chapter 4.

3.3.2 The E-Model

The formulation for the 1/E Model given in equation (1.18) will be re-stated:

$$t_{BD} = t_{0E}\exp(-B_E{*}E_{OX}) \tag{3.20}$$

The mechanisms that lead to an E-Model dependence on the time to breakdown are anode hydrogen release (AHR), anode hole injection (AHI), and the thermo chemical model (TC). The AHR model can result in an E-Model dependence when the stress voltage is above the V_{OX} > 5-6V threshold energy for hydrogen release [27]. The AHI model can result in an E-Model dependence at voltages below the threshold energy for anode hole injection when G_H decays exponentially with voltage [17]. The TC Model treats breakdown as a thermodynamic process in which the Gibbs free energy (ΔG) between time-0 and breakdown states drives the time to breakdown [22,23]. In this framework, the physical mechanism for breakdown is an alignment of dipoles in the oxide resulting from the reaction [23,24].

To derive (3.20), we will use an Eyring Model, comprehending polarization for a system where the pressure is constant [23].

$$\Delta G = \Delta H_0 - E_{OX}\Delta P - T\Delta S \tag{3.21}$$

Where ΔH_0 is the enthalpy and P is the dielectric polarization. Assuming that ΔP and B_E are temperature dependent,

$$\Delta P = k_B T B_E(T) \tag{3.22}$$

Using the following equations (3.23) and (3.24), which are respectively

$$t_{BD} = a\exp(\Delta G/k_B T) \tag{3.23}$$

$$B_E(T) = b + c/T \tag{3.24}$$

and combining (3.21) through (3.23) yields the complete Eyring formulation for the E-Model:

$$t_{BD} = a\exp(\Delta H_0/k_B T)\exp[-\Delta S/k_B]\exp(-B_E(T){*}E_{OX}) \tag{3.25}$$

Subsuming the entropy term into the pre-factor, the temperature and field dependence of the E-Model reduces to the familiar form

$$t_{BD} = t_0\exp[\Delta H_0/k_B T]\exp(-B_E(T){*}E_{OX}) \tag{3.26}$$

Note that with (3.24) inserted into (3.26), the activation energy is dependent on E_{OX}. The Eyring formulation (3.26) has been applied to thick oxides to model temperature dependent field acceleration factors and field dependent activation energies [22,23]. However, as mentioned in Chapter 1, voltage acceleration is temperature-independent in ultra-thin dielectrics [28]. If B_E is not a function of temperature, and ΔH_0 is not a function of field, then the field dependent part of (3.26) reduces to (3.20).

In contrast to the 1/E Model, the field acceleration factor for the E-Model is constant with respect to field. Accordingly, the E-Model yields lifetime projections that are more conservative than the 1/E Model.

While the TC model gives rise to a physics based derivation of the E-Model, its fundamental flaw is that it does not comprehend the experimentally observed dependence of breakdown on fluence and energy. Therefore, we will not further consider the thermo chemical model in this dissertation.

3.3.3 Applications and limitations of field-driven breakdown models

While the applicability of the 1/E and E-Models can be mechanism driven to some degree, in practice, the choice of model is often empirically driven by the data at hand [29,30]. This could be due to a number of factors, e.g. incorrect calculation of the electric field, or effects due to processing. TDDB data from several sources [11,23,29,31] are plotted in Figure 3.6. It can be seen that the E-Model best fits the data at low fields whereas the 1/E Model provides a better fit at high fields.

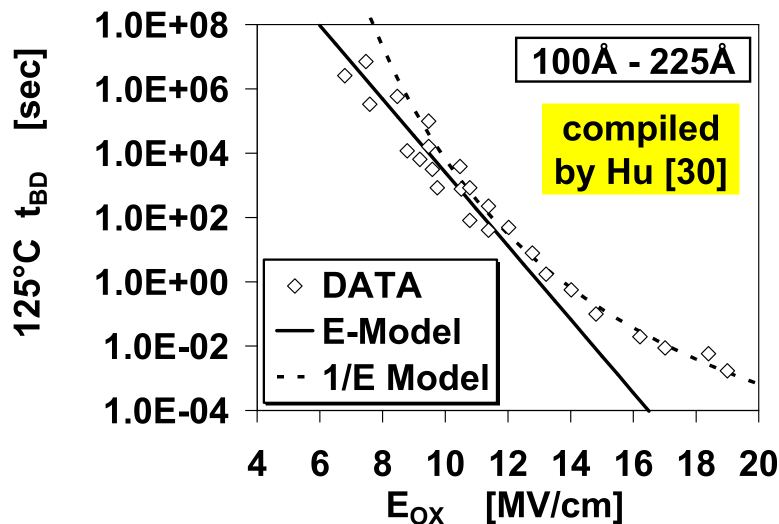

Figure 3.6. TDDB data from Chen, Ref. [11], Kimura, Ref. [23], Suehle, Ref. [29], and Shiono, Ref. [31]. Compiled by Hu, Ref. [30]. The E-Model is a better fit to the data at low fields whereas the 1/E Model is a better fit at high fields.

As previously stated, a property of any field based model is that for a fixed gate area and stress temperature, the time to breakdown at a given oxide electric field is independent of oxide thickness. In Figure 3.7, t_{BD} is plotted vs. oxide thickness for E_{OX} = 8MV/cm at 125°C from several sources in the literature [23,31-36]. Under these conditions, the time to breakdown is independent of oxide field when the oxide is thicker than 50Å, consistent with the behavior expected for either the 1/E or E-Model. However, because t_{BD} becomes a function of thickness below 50Å, field-based models no longer apply in this regime.

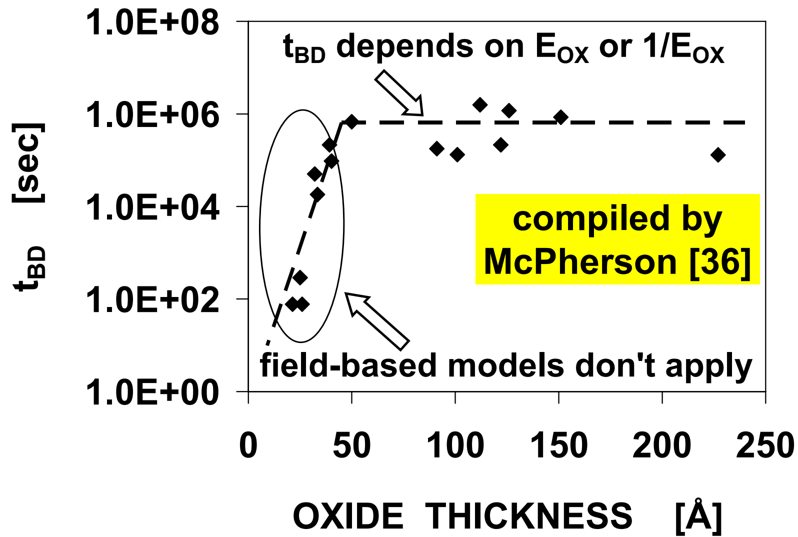

Figure 3.7. TDDB data from Kimura, Ref. [23], Shiono, Ref. [31], Stathis, Ref. [32], McPherson, Ref. [33], Suehle, Ref. [34], Prendergast, Ref. [35], and McPherson, Ref. [36]. Compiled by McPherson, Ref. [36].

In Figure 3.8, the mean charge to breakdown is plotted as a function of oxide thickness under constant current stress conditions [37]. Because the F-N tunneling gate current is independent of thickness at a given electric field, constant current stress is equivalent to constant field stress. From Figure 3.8, it can be seen that Q_{BD} becomes increasingly polarity dependent as the oxide thickness is reduced, which further indicates that field based models do not apply under these stress conditions, i.e. there is a "polarity gap" for Q_{BD}. In the next section, we will explain the meaning of this effect.

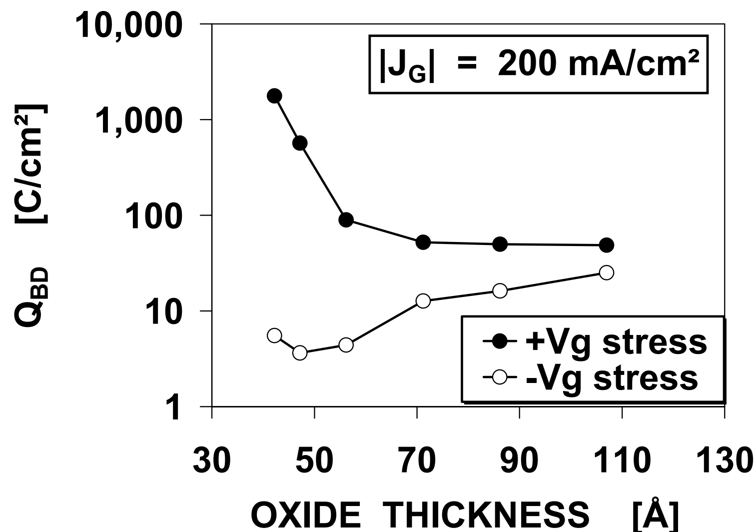

Figure 3.8. Q_{BD} vs. t_{OX}, showing polarity gap effect. After Han, Ref. [37]. © 1994 IEEE. Reprinted with permission.

3.4 Gate voltage driven breakdown models

In this section, we will present our original contributions that have been instrumental in ending the long running controversy in the industry regarding breakdown models at low voltage stress conditions. We will begin with an expansion of the discussion on the polarity gap that was introduced in the last section. We will then present our results that definitively prove that dielectric breakdown is gate voltage rather than field driven at low voltages. We will then examine experiments from other labs that further verify that breakdown is voltage (or energy) driven. We conclude this section with a discussion of the two mathematical formulations for a Gate Voltage-Model for TDDB.

3.5.1 The polarity gap

Figures 3.7 and 3.8 showed that below 50Å, the data do not behave as expected for field driven breakdown. Figure 3.7 showed that the time to breakdown at a given field is dependent on thickness and Figure 3.8 showed the charge to breakdown under constant current stress is polarity asymmetric. Recalling from equation (3.3) that $\langle KE_{OX,ANODE} \rangle$ is a field-driven quantity, Q_{BD} is plotted vs. $\langle KE_{OX,ANODE} \rangle$ for both $+/-V_G$ stress in Figure 3.9 [1]. While the data in Figures 3.8 and 3.9 are from two different laboratories, the conclusion is the same: There is a polarity gap in the charge to breakdown data. The same data in Figure 3.9 are re-plotted in Figure 3.10 with gate voltage as the x-axis. It can be seen that the polarity gap disappears, indicating that breakdown is controlled by the maximum energy (gate voltage driven) rather than by the average energy (field driven). This is the origin of the concept of gate voltage driven breakdown [1]. While it has been pointed out that a polarity gap may still exist due to differences between poly-SiO_2 and Si-SiO_2 interfaces [38], the polarity gap is always minimized when the x-axis is gate voltage or energy.

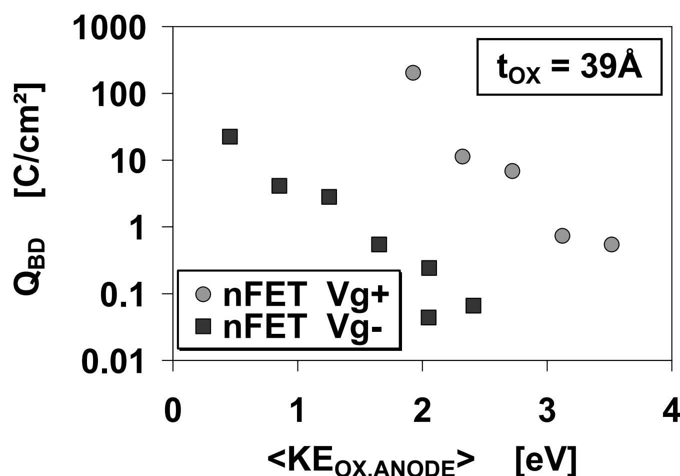

Figure 3.9. Q_{BD} vs. $\langle K_{EOX,ANODE} \rangle$, showing a polarity gap. After DiMaria, Ref. [1]. *© 1996 American Institute of Physics. Reprinted with permission.*

A “polarity gap” has also been reported for NMOS t_{BD}, which is also reduced when the time to breakdown is plotted as a function of gate voltage rather than field [39]. TDDB data for 27Å NMOS and PMOS devices stressed with $+V_G$ applied are plotted vs. field in Figure 3.11 [40]. At a given field, the time to breakdown is much longer for NMOS compared to PMOS. These data are plotted vs. gate voltage in Figure 3.12 [40]. The NMOS/PMOS “polarity gap” has been eliminated, as the time to failure is the same for NMOS and PMOS at the same gate voltage. Band diagrams for NMOS and PMOS at the same electric field are shown in Figure 3.13 [40]. While the average energy is the same, the maximum energy delivered to the anode is higher for PMOS.

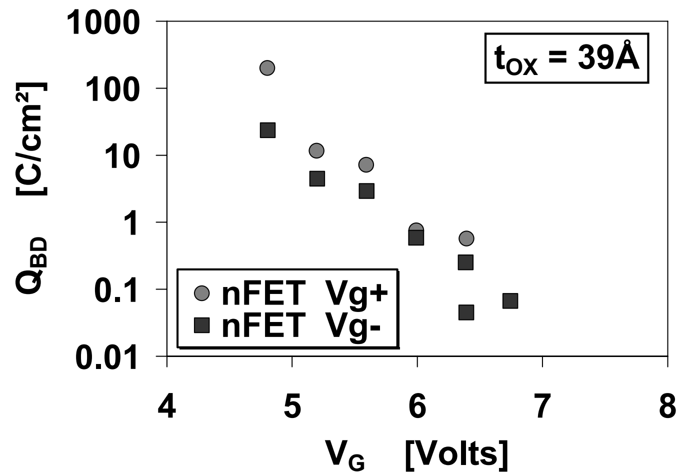

Figure 3.10. Q_{BD} vs. V_G using the same data as in Figure 3.8. Plotted against V_G, the polarity gap disappears. After DiMaria, Ref. [1]. © 1996 American Institute of Physics. Reprinted with permission.

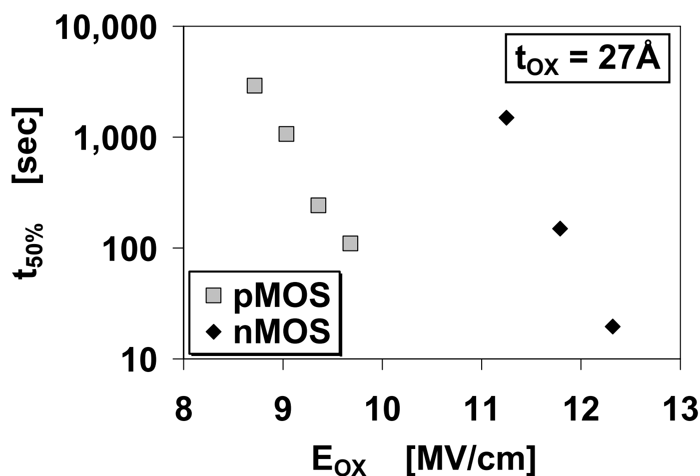

Figure 3.11. $t_{50\%}$ vs. E_{OX} for 27Å NMOS and PMOS oxides for $+V_G$ stress. A “polarity gap” for NMOS and PMOS TDDB is seen. After Nicollian, Ref. [40]. © 2000 IEEE.

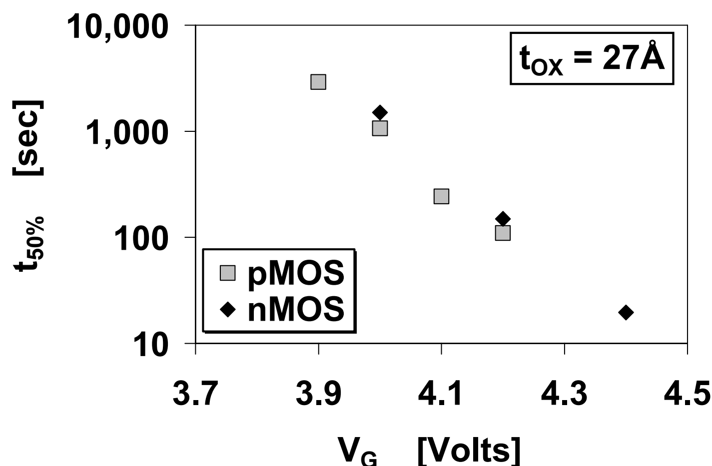

Figure 3.12. $t_{50\%}$ vs. V_G using the same data as in Figure 3.11. Plotted against V_G, the NMOS/PMOS "polarity gap" disappears. After Nicollian, Ref. [40]. © 2000 IEEE.

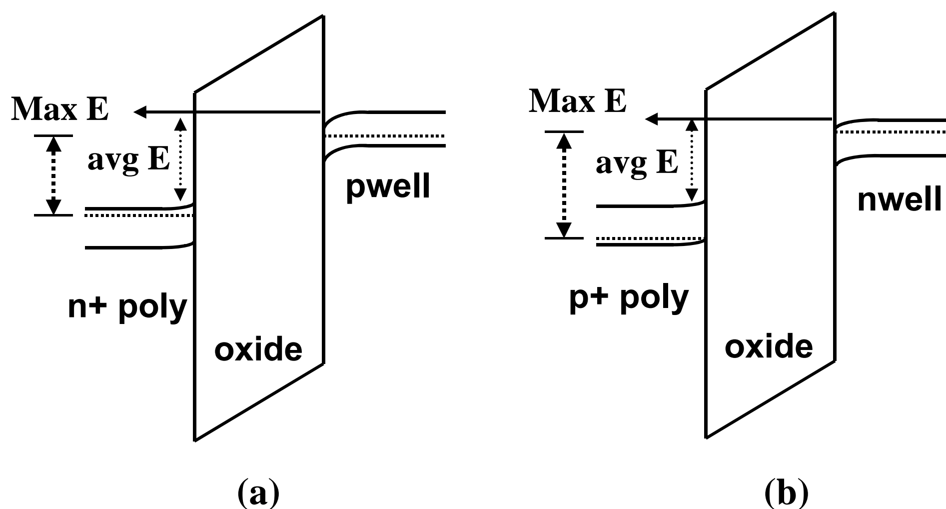

Figure 3.13. Band diagrams for (a) NMOS and (b) PMOS at the same E_{OX}. The average energy delivered to the anode (qV_{OX}) is the same, but the maximum energy dissipated at the anode (qV_G) is higher for PMOS. After Nicollian, Ref. [40]. © 2000 IEEE.

It can be seen that the data fit a straight line in both Figures 3.11 and 3.12. Linearity is not a clear indicator of which lifetime model is physically correct. Straight lines result in Figures 3.11 and 3.12 because of the linear relationship between E_{OX} and V_G in the region of interest [41]. However, the E-Model and V_G-Models cannot both be physically correct. In the next sub-section, we will prove that the V_G-Model is the correct description of breakdown at low voltages in ultra thin gate oxides.

3.5.2 Poly doping experiments

To resolve the roles of E_{OX} and V_G in the breakdown process, a series of NMOS devices with 26Å oxide thickness and different poly doping were fabricated. These variations in poly doping allow the oxide field to be varied under stress at a fixed $+V_G$ and fixed t_{OX}. Due to poly depletion, E_{OX} can be modulated by 2.5MV/cm in inversion with fixed $+V_G$ and t_{OX} over the space of this experiment as shown in Figure 3.14. The oxide field was determined through C-V matched quantum device simulations using the method in [42].

The E-Model predicts that the time to fail will increase as the poly doping is reduced, while the voltage model predicts that the lifetime will be independent of poly doping. The results of constant voltage direct tunneling stress at V_G = +3.6V are shown in Figure 3.15. Devices with the lowest poly doping that can be inverted during stress are not included in this plot, and will be analyzed in Chapter 4. All data in Figure 3.15 are normalized to the highest poly doping split. It can be seen that the lifetime is nearly independent of poly doping, whereas the E-Model predicts a 20,000X change. This proves that breakdown is not field-driven. Breakdown is driven by the maximum energy rather than average energy when the voltage is lower than the threshold energy for trap generation below about 5V to 6 V, corresponding to the steep roll-off in the generation rate previously shown in Figure 2.26.

As was discussed in Chapter 2, gate current varies more slowly with field for direct tunneling compared to F-N tunneling. In our experiments, both t_{BD} and Q_{BD} are voltage driven since the current during stress only varies by a factor of 4X between the devices with the lowest and highest poly doping.

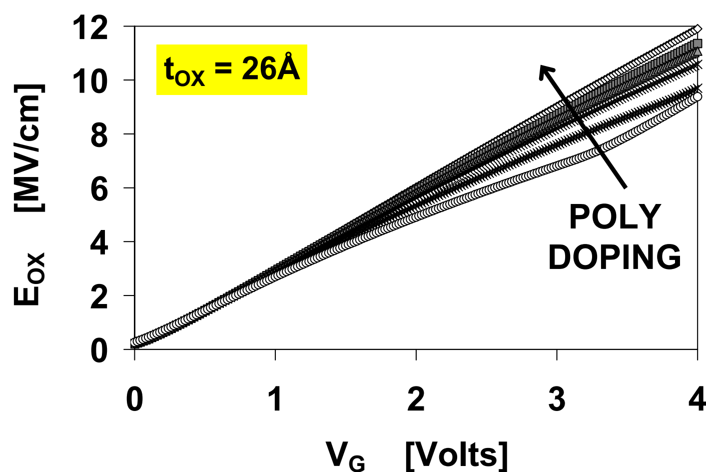

Figure 3.14. Oxide electric field vs. gate voltage with poly doping as a parameter. Due to poly depletion, E_{OX} can be modulated by 2.5MV/cm at a fixed $+V_G$. After Nicollian, Ref. [40]. © 2000 IEEE.

The ramped voltage breakdown distributions for these devices are shown in Figure 3.16. The breakdown voltages are also nearly independent of poly doping. This

is also in contradiction with the E-Model, which predicts a 1 volt excursion in breakdown voltage over the space of this experiment.

We have shown that the E-Model is an incorrect description of breakdown at low voltages in ultra thin oxides. We will review an experiment from another lab in the next sub-section that further confirms the invalidity of the E-Model for TDDB.

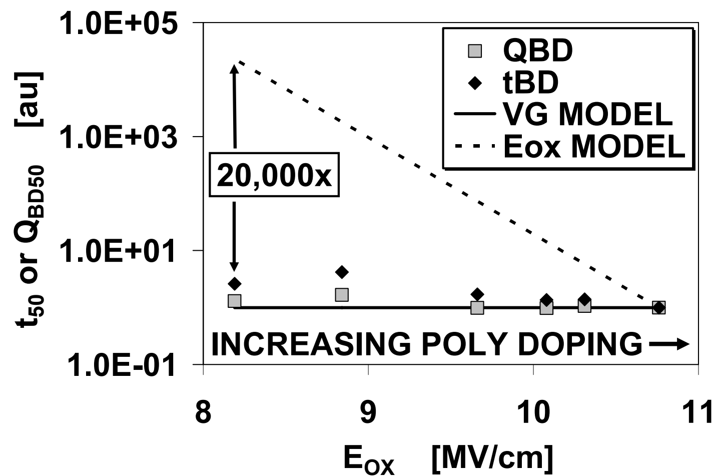

Figure 3.15. Q_{BD} and t_{BD}, normalized to the highest poly doping, vs. oxide field. E_{OX} was varied at a fixed $+V_G$. The oxide thickness is 26Å. Breakdown is not field driven in this regime. After Nicollian, Ref. [40]. © 2000 IEEE.

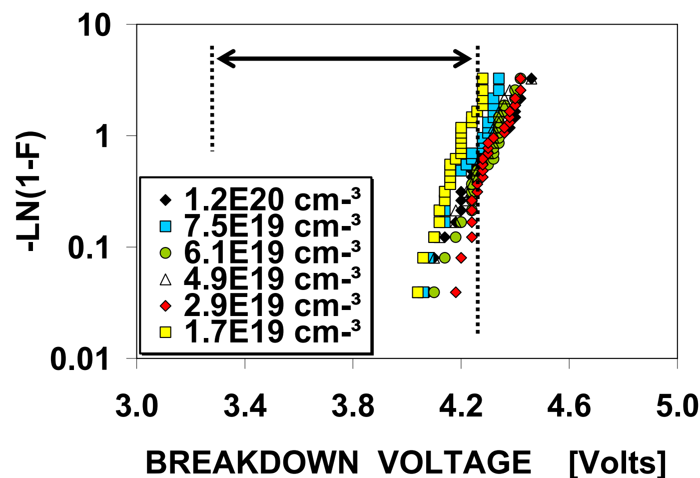

Figure 3.16. Weibull distributions of breakdown voltage as a function of poly doping. The arrow demarcates the variation in breakdown voltage predicted by the E-Model. These data further confirm that breakdown is V_G-driven. After Nicollian, Ref. [40]. © 2000 IEEE.

3.5.3 Substrate hot electron (SHE) experiments

As previously discussed in Chapter 2, one of the properties of SHE is that the electron energy, electron fluence, and oxide field can be separately controlled. This makes this technique a useful tool to further investigate the validity of breakdown models. The bias configuration and band diagram for SHE were shown earlier in Figure 2.13 but will be reproduced here in Figure 3.17 to facilitate the reading of this sub-section.

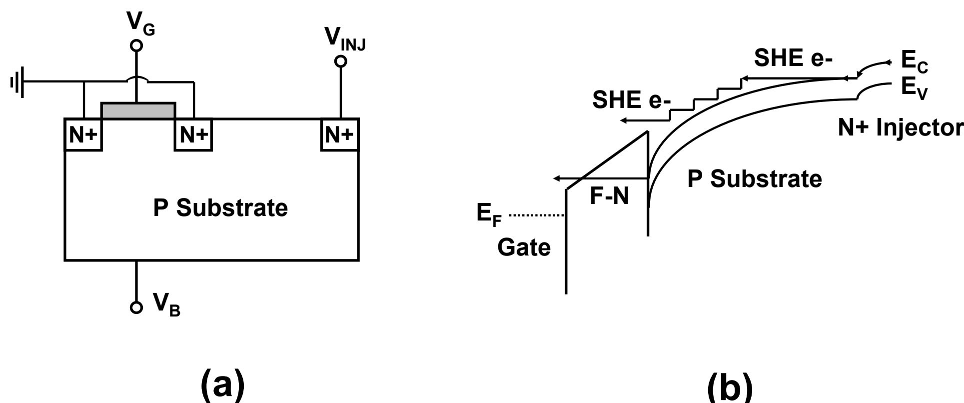

Figure 3.17. (a) Bias configuration for SHE. (b) Band diagram for SHE. Electrons are injected from the forward biased n+p junction into the reverse biased pwell. The electrons are heated in the pwell space charge region prior to being injected into the oxide after they arrive at the Si-SiO$_2$ interface. F-N tunneling is shown for comparison.

For constant voltage stress (CVS), the maximum energy delivered to the anode is qV_G (as was discussed in section 3.2.2), while the electron energy at the cathode is on the order of k_BT for thermal injection. However, this is not the case for SHE. Noting in Figure 3.17-b that there is a contact potential ($\varphi_{N/P}$) of 1.1V between the p-type substrate and the electron inversion layer, the maximum energy of SHE electrons arriving at the cathode and interfaces are respectively [43]

$$E_{MAX}(\text{cathode}) = q(V_B + \varphi_{N/P}) \quad (3.27)$$

$$E_{MAX}(\text{anode}) = q(V_G + V_B + \varphi_{N/P}) \quad (3.28)$$

From (3.27) and (3.28), it can be seen that the effect of SHE is to increase the maximum electron energy by $q(V_B + \varphi_{N/P})$ relative to CVS.

Charge to breakdown is plotted as a function of gate voltage for both CVS and SHE in Figure 3.18 for oxide thickness ranging from 20Å to 34Å [44]. In this regime, Q_{BD} is a weak function of thickness [32] so that $\ln(Q_{BD})$ vs. V_G should be approximately linear. In Figure 3.18, for a given t_{OX} and V_G, E_{OX} will be the same for both CVS and SHE, (i.e.) a vertical line represents constant field. It can be seen that Q_{BD} under SHE stress is significantly lower compared to CVS. While the electric fields are the same, the electron

energies are much higher for SHE as shown in equations (3.27) and (3.28). Therefore, for V_G < 6V, this experiment verifies that breakdown is energy rather than field driven.

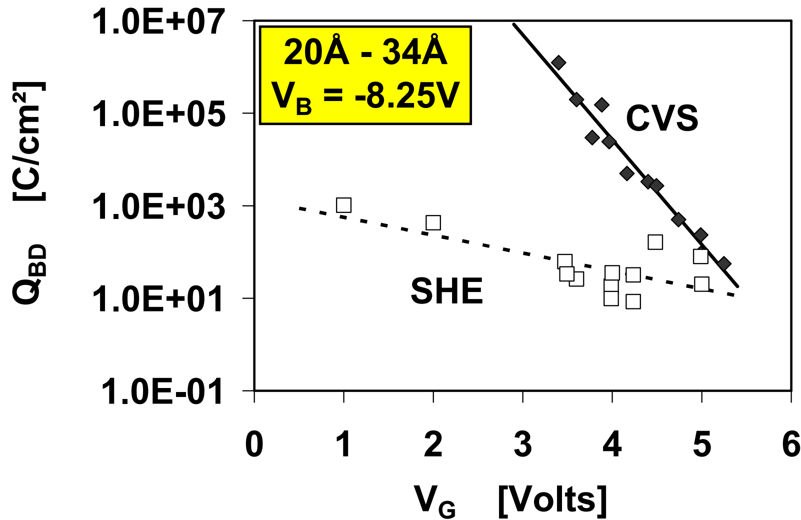

Figure 3.18. Q_{BD} vs. V_G for SHE and CVS for oxide thickness ranging from 20Å to 34Å. For V_B = -8.25V and V_G < 6V, breakdown is energy rather than field driven. After Vogel, Ref. [44]. © 2000 IEEE. Reprinted with permission.

3.5.4 Formulations of the V_G Model

Now that we have shown that ultra-thin dielectric breakdown is driven by the maximum energy dissipated at the anode (qV_G for thermal injection), we will review the functional forms for voltage model reliability projections. As introduced in Chapter 1, there are two formulations for gate voltage driven breakdown. They are known as the exponential law and power law respectively [45]:

$$t_{BD} = t_{0V}\exp(-B_V{*}V_G) \qquad (3.29)$$

$$t_{BD} = aV_G^{-N} \qquad (3.30)$$

Similar relationships can also be written for Q_{BD}. From (3.29) and (3.30), data following the exponential model will result in a straight line on a semi-log scale whereas the power law model yields a straight line on a log-log scale. Data are plotted using exponential and power law model scaling in Figures 3.19 and 3.20 respectively. In Figure 3.19, it can be seen that the lines corresponding to the least squares fits of the data are not parallel for different areas. Extrapolation to low voltages yield the unrealistic result that devices with large areas have longer lifetime than smaller areas. This violates Weibull scaling of the time to breakdown with gate area. In contrast, using the same data as in Figure 3.19, devices with different gate areas form parallel lines in the log-log scale in Figure 3.20. Accordingly, the power law model is consistent with Weibull scaling.

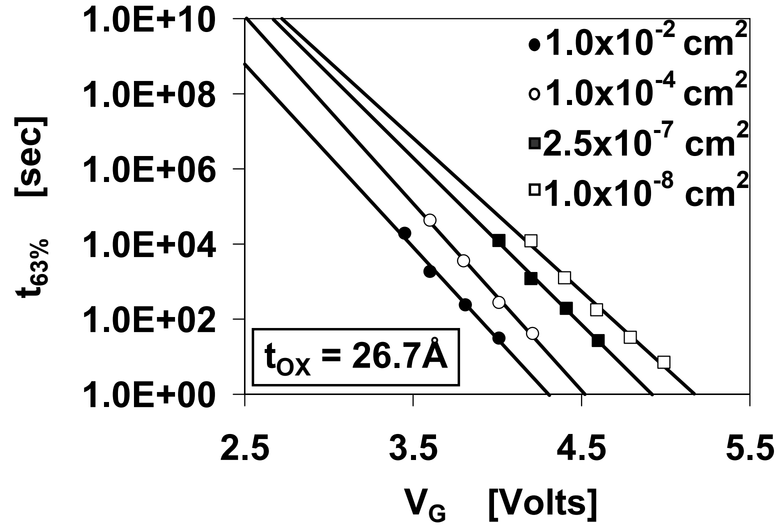

Figure 3.19. Semi-log plot of time to breakdown vs. gate voltage with area as a parameter, showing that Weibull scaling is violated. After Wu, Ref. [45]. © 2000 IEEE. Reprinted with permission.

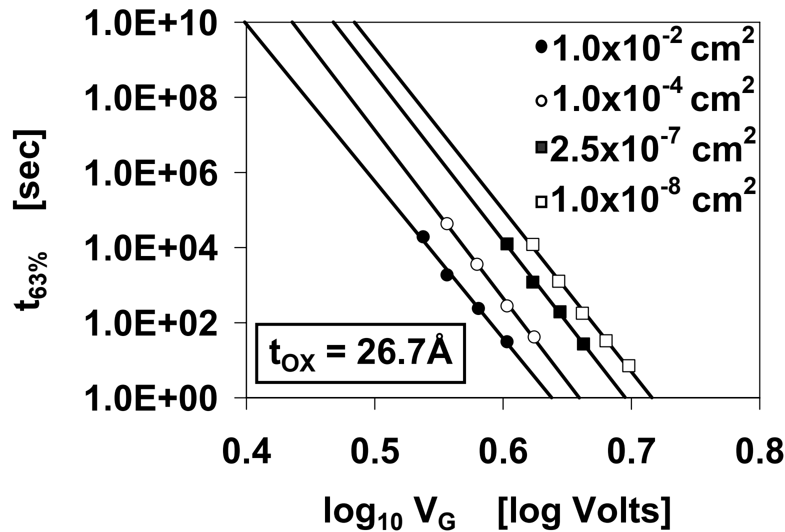

Figure 3.20. Log-log plot of time to breakdown vs. gate voltage with area as a parameter, showing that Weibull scaling is obeyed. After Wu, Ref. [45]. © 2000 IEEE. Reprinted with permission.

As discussed in Chapter 1, the voltage acceleration factor is defined as

$$AF = -\partial \ln(t_{BD})/\partial V_G \qquad (3.31)$$

Inspection of (3.29) – (3.31) shows that the acceleration factor for the exponential law is constant, while the power law voltage acceleration factor is N/V_G, which increases with decreasing voltage. In Figure 3.21 [46], the voltage acceleration factors for SiON films

ranging from 10Å to 32Å show the V_G^{-1} dependence expected for a power law model. Accordingly, from Figures 3.19-3.21, the power law appears to be the more plausible model. However, the power law exponent "N" is inexplicably high; approximately 48 as shown in Figure 3.21. We will analyze this puzzling behavior in Chapter 4.

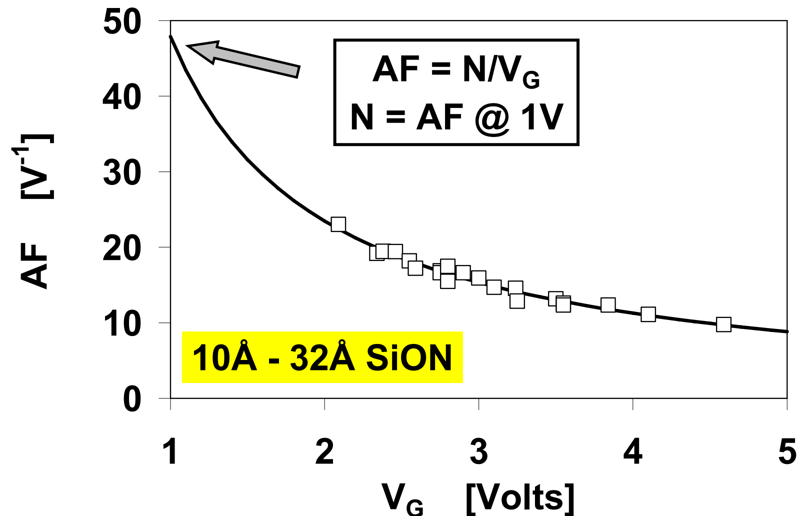

Figure 3.21. Voltage acceleration factors for SiON films ranging from 10Å to 32Å EOT. AF follows the V_G^{-1} dependence expected for a power law model. The power law exponent "N" is 48 for this dataset. After Nicollian, Ref. [46]. © 2007 IEEE.

3.5 Chapter summary

Due to the central importance of carrier energy on the operative breakdown mechanisms, we began this chapter with a discussion on the role of transport on electron energies. For direct and ballistic F-N tunneling of thermal electrons injected from the cathode, the carrier energies are easily determined. The average energy of electrons entering the anode is qV_{OX} and the maximum energy of electrons in the anode is qV_G. At higher voltages in thicker oxides, steady-state Fowler-Nordheim tunneling occurs once the electron energy distribution has been stabilized by phonons. The average electron energy can be approximated once steady state transport has ensued.

The 1/E Model can be readily derived from an analysis of the two impact ionization mechanisms that lead to this dependence. The threshold energies for impact ionization are the largest of all of the defect generation processes in the MOS system. At lower fields, while anode hydrogen release, anode hole injection, and thermo chemical mechanisms can all lead to an E-Model behavior, only the thermo chemical model can be used to readily derive the E-Model. However, the thermo chemical model does not correctly incorporate the effects of carrier fluence on breakdown.

Below about 5V to 6V, the behavior of TDDB data is no longer consistent with either 1/E or E-Models. Field-based models have the property that the time to breakdown at a

given oxide field is independent of oxide thickness. However, this is not observed at low stress voltages in oxides with thickness less than about 50Å. Moreover, a polarity gap arises, where the charge to breakdown becomes polarity dependent, also in contradiction with field-based models. The polarity gap can be explained if breakdown is voltage rather than field driven. This theory has been confirmed in two separate experiments. First, by processing devices with varying poly doping, it is possible to modulate the oxide field at a fixed oxide thickness and gate voltage. In this experiment, both time to breakdown and charge to breakdown were shown to be dependent on gate voltage rather than field, proving that the E-Model is an incorrect description of breakdown at low voltages. A second experiment compared the charge to breakdown resulting from constant voltage stress and substrate hot electron injection. The results showed that at a given oxide field, the devices subjected to substrate hot electron stress had much lower charge to breakdown compared to constant voltage stress, also showing that the E-model is not an adequate description of breakdown.

There are two formulations for the V_G-Model; an exponential law and a power law. Only the power law obeys Weibull statistics. Moreover, the voltage scaling of the acceleration factor follows the behavior expected for a power law. Accordingly, the power law model appears to be the more plausible representation for the V_G-model. One puzzling aspect of this model is that it has a power law exponent that is greater than 40, which requires additional explanation. We will address this issue in Chapter 4.

3.6 References

[1] D.J. DiMaria, “Explanation for the Polarity Dependence of Breakdown in Ultra thin Silicon Dioxide Films”, *Applied Physics Letters*, **68**, 3004 (1996)
[2] D.J. DiMaria, E. Cartier, and D.A. Buchanan, “Anode Hole Injection and Trapping in Silicon Dioxide”, *Journal of Applied Physics*, **80**, 304 (1996)
[3] J. Maserjian, “Tunneling in Thin MOS Structures”, *Journal of Vacuum Science and Technology*, **11**, 996 (1974)
[4] S. Zafar, Q. Liu, and E.A. Irene, “Study of Tunneling Current Oscillation Dependence on SiO_2 Thickness and Si Roughness at the Si/SiO_2 Interface”, *Journal of Vacuum Science and Technology*, **13**, 47 (1995)
[5] D.J. DiMaria, E. Cartier, and D. Arnold, “Impact Ionization, Trap Creation, Degradation, and Breakdown in Silicon Dioxide Films on Silicon”, *Journal of Applied Physics*, **73**, 3367 (1993)
[6] W.T. Lynch, “Calculation of Electric Field Breakdown in Quartz as Determined by Dielectric Dispersion Analysis”, *Journal of Applied Physics*, **46**, 998 (1975)
[7] B.K. Ridley, “Mechanism of Electrical Breakdown in SiO_2 Films”, *Journal of Applied Physics*, **43**, 3274 (1972)
[8] M.V. Fischetti, D.J. DiMaria, S.D. Brorson, T.N. Theis, and J.R. Kirtley, “Theory of High-Field Electron Transport in Silicon Dioxide”, *Physical Review B*, **31**, 8124 (1985)
[9] D.J. DiMaria, T.N. Theis, J.R. Kirtley, F.L. Pesavento, D.W. Duong, and S.D. Brorson, “Electron Heating in Silicon Dioxide and Off-Stoichiometric Silicon Dioxide Films”, *Journal of Applied Physics*, **57**, 1214 (1985)

[10] C. Chang, C. Hu, and R. Brodersen, “Quantum Yield of Electron Impact Ionization in Silicon”, *Journal of Applied Physics*, **57**, 302 (1985)
[11] I.C. Chen, S.E. Holland, and C. Hu, “Electrical Breakdown in Thin Gate and Tunneling Oxides”, *IEEE Transactions on Electron Devices*, **ED-32**, 413 (1985)
[12] I.C. Chen, S. Holland, K.K. Young, C. Chang, and C. Hu, “Substrate Hole Current and Oxide Breakdown”, *Applied Physics Letters*, **49**, 669 (1986)
[13] I.C. Chen and C. Hu, “Accelerated Testing of Time-Dependent Breakdown of SiO_2”, *IEEE Electron Device Letters*, **EDL-8**, 140 (1987)
[14] I.C. Chen, S. Holland, and C. Hu, “Hole Trapping and Breakdown in Thin SiO_2”, *IEEE Electron Device Letters*, **EDL-7**, 164 (1986)
[15] K.F. Schuegraf and C. Hu, “Hole Injection Oxide Breakdown Model for Very Low Voltage Lifetime Extrapolation”, *Proceedings of the IRPS*, 7 (1993)
[16] K.F. Schuegraf and C. Hu, “Hole Injection SiO_2 Breakdown Model for Very Low Voltage Lifetime Extrapolation”, *IEEE Transactions on Electron Devices*, **41**, 761 (1994)
[17] M.A. Alam, J. Bude, and A. Ghetti, “Field Acceleration for Oxide Breakdown – Can an Accurate Anode Hole Injection Model Resolve the E vs. 1/E Controversy?”, *Proceedings of the IRPS*, 21 (2000)
[18] P. Solomon, N. Klein, and M. Albert, “A Statistical Model for Step and Ramp Voltage Breakdown Tests in Thin Insulators”, *Thin Solid Films*, **35**, 321 (1976)
[19] D.L. Crook, “Method of Determining Reliability Screens for Time Dependent Dielectric Breakdown”, *Proceedings of the IRPS*, 1 (1979)
[20] E.S. Anolick and G.R. Nelson, “Low Field Time Dependent Dielectric Integrity”, *Proceedings of the IRPS*, 8 (1979)
[21] A. Berman, “Time-0 Dielectric Reliability Test by a Ramp Method”, *Proceedings of the IRPS*, 204 (1981)
[22] J.W. McPherson and D.A. Baglee, “Acceleration Factors for Thin Gate Oxide Stressing”, *Proceedings of the IRPS*, 1 (1985)
[23] M. Kimura, “Oxide Breakdown Mechanism and Quantum Physical Chemistry for Time-Dependent Dielectric Breakdown”, *Proceedings of the IRPS*, 190 (1997)
[24] B. Schlund, J.S. Suehle, C. Messick, and P. Chaparala, “A New Physics-Based Model for Time-Dependent Dielectric Breakdown”, *Proceedings of the IRPS*, 84 (1996)
[25] T.H. DiStefano and M. Shatzkes, “Impact Ionization Model for Dielectric Instability and Breakdown”, *Applied Physics Letters*, **25**, 685 (1974)
[26] Y. Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, “High-Field and Current-Induced Positive Charge in Thermal SiO_2 Layers”, *Journal of Applied Physics*, **57**, 2830 (1985)
[27] D.J. DiMaria and E. Cartier, “Mechanism for Stress Induced Leakage Currents in Thin Silicon Dioxide Films”, *Journal of Applied Physics*, **78**, 3883 (1995)
[28] E. Wu, J.M. McKenna, W. Lai, E. Nowak, and A. Vayshenker, “The Effect of Change of Voltage Acceleration on Temperature Activation of Oxide Breakdown for Ultra Thin Oxides”, *IEEE Electron Device Letters,* **22**, 603 (2001)
[29] J.S. Suehle, P. Chaparala, C. Messick, W.M. Miller, and K. Boyko, “Field and Temperature Acceleration of Time-Dependent Dielectric Breakdown in Intrinsic Thin SiO_2”, *Proceedings of the IRPS*, 120 (1994)
[30] C. Hu and Q. Lu, “A Unified Gate Oxide Reliability Model”, *Proceedings of the IRPS*, 47 (1999)

[31] N. Shiono and M. Itsumi, "A Lifetime Projection Method Using Series Model and Acceleration Factors for TDDB Failures of Thin Gate Oxide", *Proceedings of the IRPS*, 1 (1993)
[32] J.H. Stathis and D.J. DiMaria, "Reliability Projection for Ultra thin Oxides at Low Voltage", *IEDM Technical Digest*, 167 (1998)
[33] J.W. McPherson, V. Reddy, K. Banerjee, and H. Le, "Comparison of E and 1/E TDDB Models for SiO_2 Under Long-Term/Low-Field Test Condition", *IEDM Technical Digest*, 171 (1998)
[34] J.S. Suehle and P. Chaparala, "Low Electric Field Breakdown of Thin SiO_2 Films Under Static and Dynamic Stress", *IEEE Transactions on Electron Devices*, **44**, 801 (1997)
[35] J. Prendergast, J. Suehle, P. Chaparala, E. Murphy, and M. Stephenson, "TDDB Characterization of Thin SiO_2 Films with Bimodal Failure Populations", *Proceedings of the IRPS*, 124 (1995)
[36] J.W. McPherson and R.B. Khamankar, "Molecular Model for Intrinsic Time Dependent Dielectric Breakdown in SiO_2 Dielectrics and the Reliability Implications for Hyper Thin Gate Oxide", *Semiconductor Science and Technology*, **15**, 462 (2000)
[37] L.K. Han, M. Bhat, D. Wristers, J. Fulford, and D.L. Kwong, "Polarity Dependence of Dielectric Breakdown in Scaled SiO_2", *IEDM Technical Digest*, 617 (1994)
[38] E. Wu, et. al., "Polarity-Dependent Oxide Breakdown of NFET Devices for Ultra-Thin Gate Oxide", *Proceedings of the IRPS*, 60 (2002)
[39] R. Degraeve, N. Pangon. B. Kaczer, T. Nigam, G. Groeseneken, and A. Naem, "Temperature Acceleration of Oxide Breakdown and its Impact on Ultra thin Gate Oxide Reliability", *VLSI Technical Digest*, 59 (1999)
[40] P.E. Nicollian, W.R. Hunter, and J.C. Hu, "Experimental Evidence of Voltage Driven Breakdown Models in Ultra thin Gate Oxides", *Proceedings of the IRPS*, 7 (2000)
[41] E.Y. Wu, W.W. Abadeer, L.K. Han, S.H. Lo, and G.R. Hueckel, "Challenges for Accurate Reliability Projections in the Ultra Thin Oxide Regime", *Proceedings of the IRPS*, 57 (1999)
[42] C. Bowen, et. al., 'Physical Oxide Thickness Extraction and Verification Using Quantum Mechanical Simulation", *IEDM Technical Digest*, 869 (1997)
[43] D.J. DiMaria, "Defect Generation under Substrate-Hot-Electron Injection into Ultra thin Silicon Dioxide Layers", *Journal of Applied Physics*, **86**, 2100 (1999)
[44] E.M. Vogel, J.S. Suehle, M.D. Edelstein, B. Wang, Y. Chen, and J.B. Bernstein, "Reliability of Ultra thin Silicon Dioxide Under Combined Substrate Hot-Electron and Constant Voltage Tunneling Stress", *IEEE Transactions on Electron Devices*, **47**, 1183 (2000)
[45] E.Y. Wu, et. al., "Voltage Dependent Voltage Acceleration of Oxide Breakdown for Ultra-Thin Oxides", *IEDM Technical Digest*, 541 (2000)
[46] P.E. Nicollian, A.T. Krishnan, C.A. Chancellor, R.B. Khamankar, S. Chakravarthi, C. Bowen, and V.K. Reddy, "The Current Understanding of the Trap Generation Mechanisms that Lead to the Power Law Model for Gate Dielectric Breakdown", *Proceedings of the IRPS*, 197 (2007)

CHAPTER 4

Mechanisms for trap generation and breakdown

4.1 Introduction

In this chapter, we will discuss the physical mechanisms that lead to trap generation and breakdown in SiO_2 and SiON dielectrics. Since the traps that are formed are ultimately the result of an electrochemical reaction, we will begin this chapter with a discussion of Reaction-Diffusion (R-D) theory. We will then review the physical mechanisms in (roughly) descending order of the energy range in which they are operative. Threshold energies for impact ionization, anode hole injection, and anode hydrogen release are tabulated in Figure 4.1. We will present our novel contributions to the understanding of the roles of anode hydrogen release and anode hole injection in trap generation and breakdown. We will also present our original findings regarding the stress induced defects that result in breakdown.

MECHANISM	E_{TH}
impact ionization	**9 eV**
anode hole injection (majority ionization)	**6 eV**
anode hole injection (minority ionization)	**3 - 3.5 eV**
anode hydrogen release (electronic excitation)	**5 - 7 eV**
anode hydrogen release (coherent vibrational excitation)	**2.5 - 3 eV**
anode hydrogen release (multi-electron vibrational excitation)	**0.25 eV**

Figure 4.1. Threshold energies for trap generation in devices with SiO_2 gate dielectrics with silicon substrates and poly silicon gate electrodes, ignoring band gap narrowing effects in highly doped silicon and poly silicon.

4.2 Reaction-diffusion theory

In this section, we will use R-D theory to derive the time (fluence) dependence of the trap density on the hydrogen species released in a reaction. Much of the understanding of R-D theory has been developed by groups researching negative-bias-temperature-instability (NBTI) mechanisms [1-8]. We will analyze the kinetics for reactions where H^+, H^-, H_2^+, H_2^-, H^0, or H_2 is one of the products and the reaction is in quasi-equilibrium.

4.2.1 Quasi-equilibrium

Let us consider a chemical reaction with reactants A and B, with products C and D:

$$[A] + [B] \leftrightarrow [C] + [D] \quad (4.1)$$

The time dependence of the creation of the product species C is:

$$d[C]/dt = k_F[A][B] - k_R[C][D] \quad (4.2)$$

where k_F and k_R are the reverse and forward reaction constants respectively. If the reaction is in equilibrium, then the left hand side of (4.2) is zero. However, it the left hand side of (4.2) is small, we can assume that the reaction is in quasi-equilibrium. The assumption of quasi-equilibrium simplifies the analysis, and is valid when both forward and reverse reactions are operative and are similar in magnitude so that

$$k_F[A][B] \approx k_R[C][D] \quad (4.3)$$

For our purposes, the forward reaction corresponds to trap generation and the reverse reaction corresponds to the recovery process. Equations (1.2) and (1.3) will be re-introduced as (4.4) and (4.5) respectively. The trap generation power law exponent is defined by (4.6) in the time domain.

$$N(t) = b_t t^m \quad (4.4)$$

$$N(Q) = b_Q Q^m \quad (4.5)$$

$$m = \partial \log N(t) / \partial \log t \quad (4.6)$$

In practice, the test for quasi-equilibrium is to compare the trap generation exponent "m" in equation (4.4), where the trap density is a power law in time, or in (4.5), where the trap density is a power law in fluence, for interrupted vs. uninterrupted stress [7]. The same value of "m" results whether (4.4) or (4.5) is used, as will be shown below. If the reaction is in quasi-equilibrium, then a delay time between sense and stress will result in a higher power law exponent due to the recovery that occurs after the stress is removed. The trap generation power law exponents for the reaction products that we will consider are tabulated in Figure 4.2. Note that charged species result in $m > ¼$.

Recalling that $Q = Jt$,

$$\partial \log Q / \partial \log t = 1 \quad (4.7)$$

Taking the derivative of log N(Q) with respect to log Q in (4.5) and inserting (4.6) and (4.7),

$$\partial \log N(Q) / \partial \log Q = [\partial \log N(t) / \partial \log t][\partial \log t / \partial \log Q] = m \quad (4.8)$$

Therefore, the trap generation exponent “m” is the same whether working in the time domain (4.4) or fluence domain (4.5). However, the pre-factors b_t and b_Q in (4.4) and (4.5) respectively are not equal.

reaction product	m	reference
H_2	1/6	[5]
H^0	1/4	[1]
H_2^+	1/3	[6]
H_2^-	1/3	[6]
H^+	1/2	[3]
H^-	1/2	[3]

Figure 4.2. Trap generation power law “m” for various hydrogen reaction products. Compiled by Alam, Ref. [6].

4.2.2 Trap generation kinetics when the reaction product is a charged species

We will first consider a reaction where silicon-hydrogen bonds are broken after interacting with holes, resulting in interface traps Si^0 and H^+ as reaction products:

$$[SiH] + [h^+] \leftrightarrow [Si^0] + [H^+] \quad (4.9)$$

In quasi-equilibrium,

$$k_F[SiH][h^+] \approx k_R[Si^0][H^+] \quad (4.10)$$

To solve (4.10) for the build-up of the interface trap density $[Si^0]$ as a function of time (or fluence), we need to determine the hydrogen profile resulting from the reaction. Since the reaction that releases the charged species into the oxide occurs with a bias applied, we will assume that H^+ moves through drift and that diffusion is negligible. The H^+ profile can then be approximated as rectangular [6] as illustrated in Figure 4.3. The height of the rectangle is the hydrogen concentration $[H^+]$ at the interface where the reaction occurs. The length of the rectangle is the distance that the hydrogen has drifted into the dielectric and is given by:

$$x(t) = \mu_H E_{OX} t \tag{4.11}$$

where u_H is the drift mobility of hydrogen in the oxide. With the assumption that one interface trap is generated for each hydrogen atom released into the oxide, i.e. $[Si^0] = [H^+]$,

$$[Si^0] = \int_0^{\mu_H E_{OX} t} H(x,t)dx \tag{4.12}$$

The solution to (4.12) is:

$$[Si^0] = [H^+]\mu_H E_{OX} t \tag{4.13}$$

Placing (4.13) into (4.10) yields

$$[Si^0] = (k_F[SiH][h^+]/k_R)(\mu_H E_{OX} t)^{1/2} \tag{4.14}$$

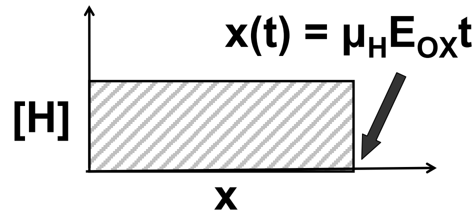

Figure 4.3. Hydrogen profile of a charged species released from the interface into the dielectric. After Alam, Ref. [6].

From (4.14), the interface trap density increases as $t^{1/2}$ for a reaction where H^+ is liberated [3]. The same result is obtained for H^-. To determine the temperature dependence of (4.14), we assume that the forward and reverse reaction rates follow Arrhenius relationships as per equations (4.15) and (4.16). We also assume that the temperature dependence of diffusivity follows Arrhenius behavior as defined in equation (4.17). The net activation energy of the reaction is the sum of the activation energies for k_F, k_R, and μ_H in equation (4.18). We use the Einstein relationship in (4.19) for the mobility.

$$k_F = k_{F0}\exp(-E_F/k_B T) \tag{4.15}$$

$$k_R = k_{R0}\exp(-E_R/k_B T) \tag{4.16}$$

$$D = D_0\exp(-E_D/k_B T) \tag{4.17}$$

$$\Delta H = (E_F + E_R + E_D) \tag{4.18}$$

$$\mu_H = (k_B T/q)D \tag{4.19}$$

Inserting (4.15) through (4.19) into (4.14) yields the temperature dependence for a reaction where interface traps are created from the release of H^+.

$$\Delta H = 1/2(E_F - E_R) + 1/2E_D + (k_B T/q) \tag{4.20}$$

We will next consider a reaction where interface traps and H_2^+ are the products of a reaction between silicon-hydrogen bonds and holes:

$$[SiH] + [h^+] \leftrightarrow [Si^0] + ½[H_2^+] \tag{4.21}$$

$$k_F[SiH][h^+] \approx k_R[Si^0][H_2^+]^{1/2} \tag{4.22}$$

Inserting the solution to (4.12), and (4.15) – (4.19) into (4.22), noting that the number of interface traps created is twice the number of H^+ atoms, the time and temperature dependence of trap creation from the release of H_2^+ are respectively:

$$[Si^0] = (k_F[SiH][h^+]/k_R)^{2/3}(\mu_H E_{OX} t)^{1/3} \tag{4.23}$$

$$\Delta H = 2/3(E_F - E_R) + 1/3E_D + 1/3(k_B T/q) \tag{4.24}$$

From (4.23), the interface trap density increases as $t^{1/3}$ for a reaction where H_2^+ is liberated [6]. The same result is obtained for H_2^-.

4.2.3 Trap generation kinetics when the reaction product is a neutral species

We will now consider a reaction where silicon-hydrogen bonds are broken after interacting with holes, resulting in interface traps Si^+ and H^0 as reaction products:

$$[SiH] + [h^+] \leftrightarrow [Si^+] + [H^0] \tag{4.25}$$

In quasi-equilibrium,

$$k_F[SiH][h^+] \approx k_R[Si^+][H^0] \tag{4.26}$$

To solve (4.26) for the build-up interface trap density $[Si^+]$ as a function of time (or fluence), the hydrogen profile resulting from the reaction must be determined. For a neutral diffusing species, the profile in the oxide will be given by the solution to Fick's differential equation. While the result has an erfc(x,t) dependence, it can be approximated as a triangular profile as shown in Figure 4.4 [4]. This approach was successfully used to model impurity profiles resulting from solid-state diffusion in the 1960's [9]. The height of the triangle is $[H^0]$ and the length of the triangle is $(Dt)^{1/2}$, where D is the diffusivity. For NBTI, the triangular approximation underestimates the true profile by about 6% to 10% [8]. Accordingly, the diffusion distance (the x-axis intercept in Figure 4.4) is $p(Dt)^{1/2}$, where p is an empirical constant:

$$x(t) = p(D_H t)^{1/2} \quad (4.27)$$

With the assumption that one interface trap is generated for each hydrogen atom released in to the oxide, i.e. $[Si^+] = [H^0]$,

$$[Si^+] = \int_0^{p(D_H t)^{1/2}} H(x,t)dx \quad (4.28)$$

The solution to (4.28) yields:

$$[Si^+] = ½[H^0]p(D_H t)^{1/2} \quad (4.29)$$

Inserting (4.29) into (4.26),

$$[Si^+] = (pk_F[SiH][h^+]/2k_R)^{1/2}(Dt)^{1/4} \quad (4.30)$$

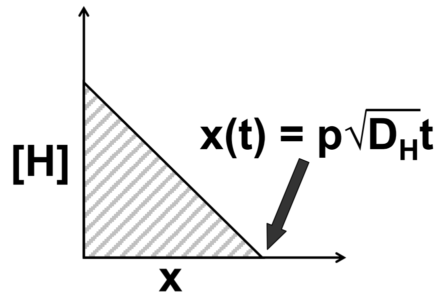

Figure 4.4. Hydrogen profile of a neutral species released from the interface into the dielectric. After Alam, Ref. [6], Krishnan, Ref. [8], and Grove, Ref. [9].

From (4.30), the interface trap density increases as $t^{1/4}$ for a reaction where H^0 is liberated [1]. Inserting (4.15) – (4.18) into (4.30), the temperature dependence for a reaction where interface traps are released following the liberation of H^0 is:

$$\Delta H = 1/2(E_F - E_R) + 1/4E_D \quad (4.31)$$

For a reaction where interface traps and H_2 are the products of a reaction between silicon-hydrogen bonds and holes:

$$[SiH] + [h^+] \leftrightarrow [Si^+] + ½[H_2] \quad (4.32)$$

$$k_F[SiH][h^+] \approx k_R[Si^+][H_2]^{1/2} \quad (4.33)$$

Inserting the solution to (4.28), and (4.15) – (4.18) in (4.33), noting that the number of interface traps created is twice the number of H^0 atoms, the time and temperature dependence of interface trap creation resulting from the release of H_2 are respectively:

$$[Si^+] = [½p(k_F[SiH][h+]/k_R)^2]1/3(Dt)^{1/6} \quad (4.34)$$

$$\Delta H = 2/3(E_F - E_R) + 1/6E_D \quad (4.35)$$

From (4.34), the interface trap density increases as $t^{1/6}$ for a reaction where H_2 is liberated [5].

In section 4.6, we will use R-D theory to show that two hydrogen species are involved in trap generation and breakdown in PBTI stress of NMOS devices with SiON dielectrics.

4.3 Impact ionization in the gate dielectric

In this section, we will describe the impact ionization process in oxides along with its role stabilizing electron energies from thermal runaway at high fields. We will then review data that provide evidence that impact ionization is a mechanism for trap generation and breakdown in thick oxides.

4.3.1 Stabilization of the electron energy distribution at high fields

In Section 3.2.4, we introduced steady-state F-N tunneling a transport mechanism at high fields in thick oxides. In this process, electron energies remain thermal below 1.5MV/cm, as they are stabilized by LO phonons emitted at 0.15eV. Above 1.5MV/cm, electron heating begins as the electrons gain more energy from the field than they lose to LO phonons, and the entire electron population goes into thermal runaway from the LO phonon modes. Above 4MV/cm - 5MV/cm, electron energies become stabilized by acoustic phonons. The average energy does not exceed 6eV all the way to breakdown. This is the steady-state transport condition, where the energy gained by electrons from the field is approximately balanced by the energy lost to acoustic phonon emission. However, Monte-Carlo simulations show that the acoustic phonon scattering rate peaks at 6eV and decreases at higher energies [10]. Accordingly, it is puzzling that average electron energies above 6eV (due to acoustic phonon runaway) are not observed.

While the carrier separation measurements presented in Chapter 3 sense the average electron energy, the vacuum emission technique detects the entire electron energy distribution [11]. The energy distribution from vacuum emission experiments is shown in Figure 4.5 [12]. It can be seen that high energy tails with kinetic energies up to 15eV are observed. Kinetic energies as high as 20eV have been reported [11]. From Figure 4.5, since only a small fraction of electrons become heated beyond 6eV, there must be another scattering process that stabilizes the electron energy distribution from acoustic phonon runaway. This mechanism is impact ionization, where hot electrons in the oxide conduction band lose their energy in the creation of electron hole pairs. Electrons must gain a kinetic energy of 9eV (with respect to the bottom of the oxide conduction band) to initiate this process and only occurs for fields greater than 7MV/cm in oxides thicker than about 200Å - 300Å [12]. Monte-Carlo simulations illustrating electron scattering via LO phonons, acoustic phonons, and impact ionization are shown in Figure 4.6 [13]. With all three of these stabilizing mechanisms present, the model predicts that the electron

energy distribution does not go into thermal runaway, in agreement with carrier separation data [14].

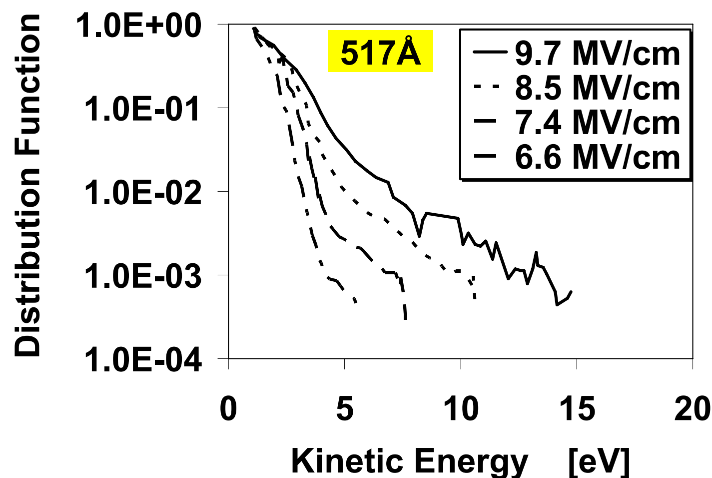

Figure 4.5. Electron energy distribution from vacuum emission experiments. After DiMaria, Ref. [12]. © 1992 American Institute of Physics. Reprinted with permission.

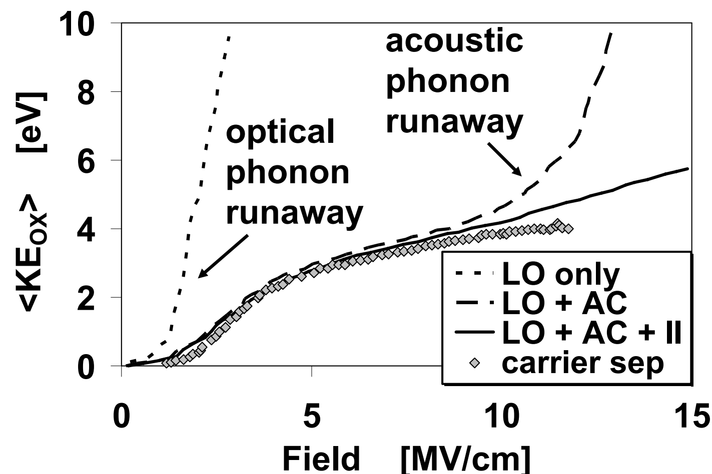

Figure 4.6. Monte Carlo simulations of the average electron energy in the oxide compared to carrier separation data. Inclusion of scattering via LO phonons, acoustic (AC) phonons, and impact ionization (II) into Monte-Carlo simulations results in agreement between theory and experiment. After Arnold, Ref. [13] and DiMaria, Ref. [14]. © 1992, 1994 American Institute of Physics. Reprinted with permission.

4.3.2 Experimental evidence for impact ionization

It is clear from the discussion in sub-section 4.3.1 that impact ionization in the oxide is not an important degradation process in ultra-thin oxides stressed at low voltages. However, in addition to providing the reader with a broader perspective of trap generation mechanisms, the inclusion of a discussion on impact ionization is relevant

because there are some experimental similarities with anode hole injection, which is operative at lower voltages. Both mechanisms introduce holes into the oxide through an impact ionization process, and lead to the 1/E Model for TDDB at high oxide voltages. For impact ionization in the oxide, the ionization rate and field dependence are significantly higher than for anode hole injection [14]. It has been postulated that trapped electrons rather than holes lead to breakdown [15]. This theory was based on the observation that the larger the change in holding voltage resulting from electron trapping during stress, the larger the time to breakdown [15]. However, in the regime where the 1/E Model applies, breakdown occurs when a critical hole fluence has passed through the oxide [16]. Moreover, the average oxide field decreases during stress as a result of electron trapping, while the anode field increases during stress due to hole trapping. Breakdown is observed to occur at the same anode field for oxides with different thickness [17], supporting hole trapping as the mechanism for breakdown. These holes trapped at the anode are introduced into the oxide either from impact ionization near the anode (where the probability is the highest for this to occur) or from injection of holes from the anode electrode into the oxide valence band.

When impact ionization in the oxide is an operative mechanism, some of the holes created from the band gap ionization process are trapped near the anode. Others migrate towards the cathode, where they are trapped in as-fabricated sites or recombine with electrons to form slow states and interface traps. The build-up trapped charge as a function of electron fluence is shown for oxide thickness ranging from 93Å to 957Å in Figure 4.7 [14]. The key aspects of this plot are (1) The positive trapped charge density increases until a steady state level is reached due recombination between free electrons and trapped holes. (2) The steady state positive charge density can be accurately predicted from values of the ionization rate that are generated from Monte-Carlo simulations of impact ionization in the oxide and equation (2.41) [10,14]. (3) The steady state hole density is approximately independent of oxide thickness for the three thickest films. This is expected if holes are generated from impact ionization in the oxide since the high energy electron tails that create band gap ionization events are stabilized above 300Å - 400Å [10].

The charge to breakdown vs. average oxide field for oxide thickness ranging from 55Å to 500Å is shown in Figure 4.8 [18]. The key features of this plot are (1) The field dependence of Q_{BD} for the thickest oxides is the strongest and the magnitude of Q_{BD} is the lowest when impact ionization in the oxide is an operative process. (2) As thickness and field decrease, steady transport is still observed, but the oxide is no longer thick enough for impact ionization to develop. In this regime, Q_{BD} is a slowly decreasing function of electric field and thickness.

Another experiment evaluated impact ionization and anode hole injection as mechanisms for positive charge formation by evaluating the polarity dependence of the post-stress charge centroid in 700Å oxides [19]. The positive charge centroid was found to depend on the distance from the injecting electrode and not on the distance from the positive electrode. The centroid occurred at a fixed distance of 250Å - 300Å from the

electrode that was the cathode during stress. The authors argued that this result supports impact ionization as the mechanism for bulk charge generation.

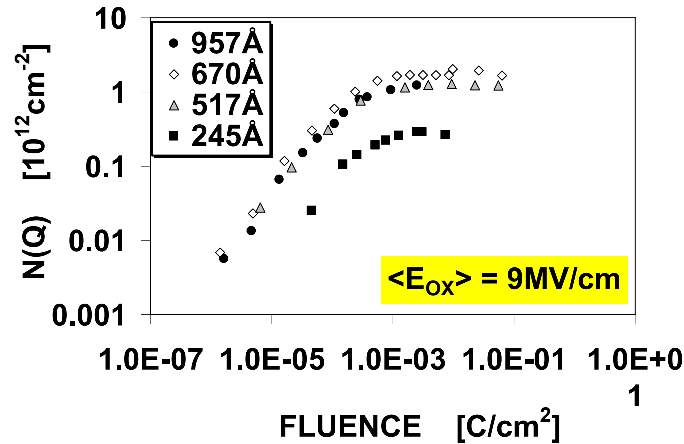

Figure 4.7. Trapped hole density as a function of electron fluence at 9MV/cm. After DiMaria, Ref. [14]. © 1994 American- Institute of Physics. Reprinted with permission.

In a study of the oxide field dependence of post-stress detrapping rates of trapped electrons in 245Å films, a sharp drop in the electron trap occupancy was observed at fields above 8MV/cm [14]. This corresponds with the onset of impact ionization in the oxide, which results in electron detrapping via recombination with the generated holes as described by equation (2.40). Accordingly, this experiment also points to impact ionization in the oxide as a mechanism for positive charge formation in the oxide.

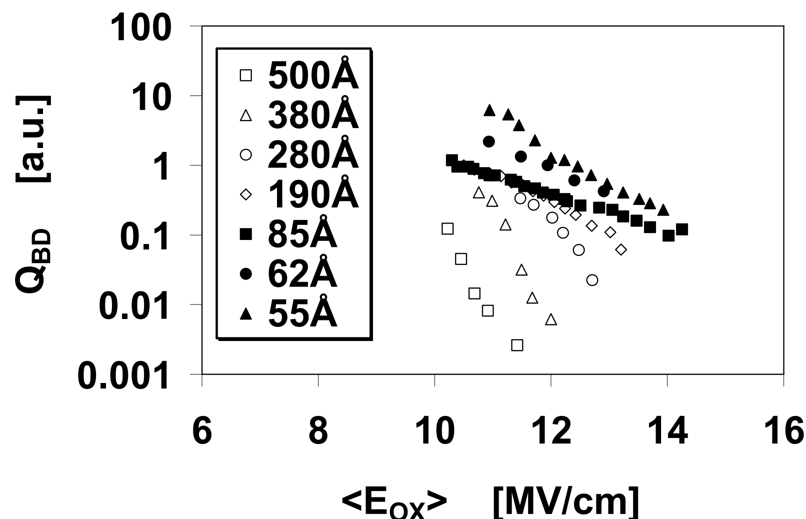

Figure 4.8. Charge to breakdown vs. average field in the oxide during stress. Open fill denotes impaction ionization at high fields and dark fill denotes steady-state transport at high fields. After DiMaria, Ref. [18]. © 1992 American Institute of Physics. Reprinted with permission.

4.4 Features of anode hole injection

In this section, we will review the anode hole injection model at high and present our findings that show that this mechanism remains a plausible explanation for breakdown at least down to 3.6V.

4.4.1 Majority and minority ionization

Impact ionization is observed in oxides thicker than about 200Å - 300Å when the electric field is greater than 7MV/cm and electron kinetic energies in the oxide are greater than 9eV. However, positive charging is still observed in thinner oxides at lower fields and energies than required for impact ionization in the oxide [12,14,19-22]. The mechanism leading to this behavior is anode hole injection (AHI), where electrons tunnel from the cathode to the anode and impact ionize in the anode rather than in the oxide to produce electron hole pairs. These energetic holes are subsequently injected into the oxide where they can generate electron traps, hole traps, interface traps, or be captured by as-grown centers. This process has a threshold voltage of V_{OX} = 6V and is illustrated in Figure 4.9, where E_{IN} is the kinetic energy of tunneling electrons dissipated in the anode. The mechanism shown in Figure 4.9 is known as majority ionization [21,23,24], where the final states of the tunneling and generated electron lie in the anode conduction band. This process occurs only when the kinetic energy of electrons entering the anode is greater than the 6eV threshold for anode hole injection. The 6eV threshold energy for AHI is the sum of the hole barrier height φ_H and the energy required to create an electron-hole pair (which is at least as large as the band gap of the anode material). For a silicon or poly silicon anode,

$$E_{TH} \approx \varphi_H + E_G(Si) \approx 4.8eV + 1.1eV \approx 6eV \tag{4.36}$$

The average and maximum kinetic energy of holes in the anode created through majority ionization are respectively

$$KE(h)_{AVG} \approx qV_{OX} - E_G(Si) \approx qV_{OX} - 1.1eV \tag{4.37}$$

$$KE(h)_{MAX} \approx qV_G - E_G(Si) \approx qV_G - 1.1eV \tag{4.38}$$

Majority ionization will be most efficient when V_{OX} is greater than 6V but can still occur, albeit at a reduced rate, as long as V_G is greater than 6V.

If holes are already present in the anode electrode, then a process known as minority ionization can result in anode hole injection, where electrons tunneling into the anode transfer their kinetic energy to free holes [23,24]. While this effect occurs at a lower rate than majority ionization [23], it enables anode hole injection at much lower voltages because tunneling electrons do not have to lose 1.1eV to generate electron-hole pairs when free holes are already present. There are two types of minority ionization: (i) One of the electrons has its final state in the anode valence band, which we shall call minority-1. For this mechanism, an electron with a distribution f_1 collides with and transfers its kinetic energy to a thermal hole with distribution $(1-f_4)$. The hole scatters

into an occupied state f_2 in the anode valence band, becoming hot in the process. The incident electron ends up in the anode conduction band in a state $(1-f_3)$.The net scattering rate for minority-1 is:

$$R_{MIN1} \sim [f_1(1-f_4)]*[f_2(1-f_3)] \quad (4.39)$$

The minority-1 process is illustrated in Figure 4.10 [23,24]. In contrast to majority ionization, the possibility of anode hole injection through the barrier rather than over the barrier cannot be ruled out. The hole currents resulting from minority ionization are too small to provide adequate data for a model. The average and maximum kinetic energy of holes in the anode created through minority-1 ionization are respectively [23]

$$KE(h)_{AVG} \approx qV_{OX} + |E_F - E_V| \quad (4.40)$$

$$KE(h)_{MAX} \approx qV_G + |E_F - E_V| \quad (4.41)$$

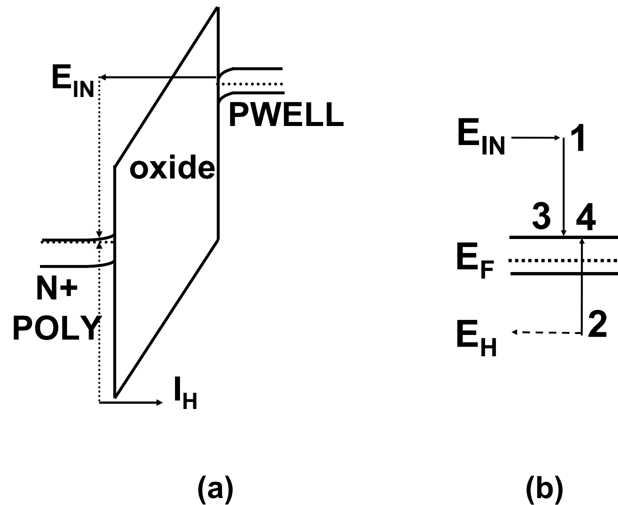

Figure 4.9. (a) Band diagram for majority ionization in an inverted NMOS device. An electron tunnels into the anode and impact ionizes to produce an electron-hole pair. If the hole acquires > 5eV kinetic energy, it can be injected into the oxide valence band where it can generate trap states. (b) Schematic for the initial and final electron states. For majority ionization, both electrons have their final states in the anode CB.

(ii) In the minority-2 process, two electrons have their final states in the anode valence band. For this mechanism, an electron with a distribution f_1 collides with and transfers its kinetic energy to a hot hole with distribution $(1-f_4)$. The hole scatters into an occupied state f_2 deep in the anode valence band, becoming hotter in the process. The incident electron ends up in the anode valence band in a state $(1-f_3)$. The net scattering rate for minority-2 is:

$$R_{MIN2} \sim [f_1(1-f_4)]*[f_2(1-f_3)] \quad (4.42)$$

The minority-2 process is illustrated in Figure 4.11 [23-25]. The average and maximum kinetic energy of holes in the anode created through minority-2 ionization are respectively [23]

$$KE(h)_{AVG} \approx qV_{OX} + 2|E_F - E_V| + E_G(Si) \quad (4.43)$$

$$KE(h)_{MAX} \approx qV_G + 2|E_F - E_V| + E_G(Si) \quad (4.44)$$

An alternate explanation for the minority-2 process is that the incident electron at f_1 scatters off electron f_2 deep within the valence band. Both electrons end up in empty valence band states $(1-f_3)$ and $(1-f_4)$. The hole left behind by the valence band electron is hot. For this path, the net scattering rate is [26]

$$R_{MIN2} \sim [f_1 f_2]*[(1-f_3)(1-f_4)] \quad (4.45)$$

Inspection of (4.42) and (4.45) shows that both explanations result in the same net scattering rate and the same hole kinetic energy. Accordingly, the two different interpretations of minority-2 given in (4.42) and (4.45) are indistinguishable.

Equation (4.44) predicts that AHI is still operative down to V_G ~ 3.0V - 3.5V. Later in this section, we will present experimental evidence that confirms that AHI through minority ionization is a plausible theory.

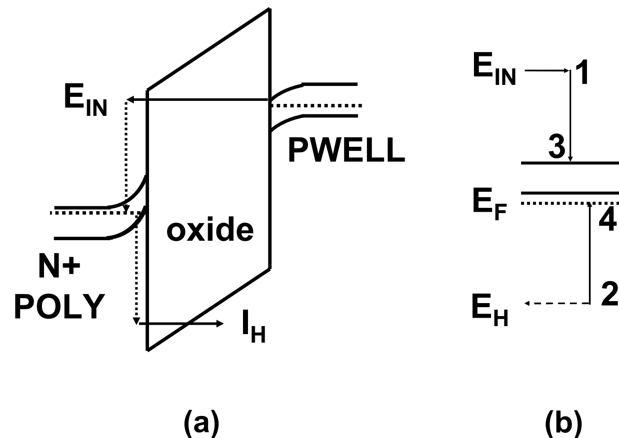

Figure 4.10. (a) Band diagram for minority-1 ionization in an NMOS device where both pwell and poly are inverted. An electron tunnels into the anode and collides with a thermal hole. If the hole acquires > 5eV kinetic energy, it will be injected into the oxide VB where it can generate trap states. (b) Schematic for the initial and final electron states. For minority-1 ionization, one electron has a final state in the anode VB. After Bude, Ref. [23] and Alam, Ref. [24]. © 1998, 2000 IEEE. Reprinted with permission.

4.4.2 Experimental evidence for anode hole injection

Historically, there was some controversy on the origins of positive charging in the realm where impact ionization is not an operative mechanism. A dependence of the ionization

rate on the work function of metal gate anodes was observed at low fields, although the differences diminished at higher fields [20]. This indicated that holes are injected from the anode in into the oxide. However, carrier separation measurements showed that the substrate hole current at a given field remains dependent on oxide thickness down to 55Å [16]. This observation is not consistent with the behavior expected for a tunneling process, which anode hole injection was presumed to be. Subsequently, substrate hole currents were successfully modeled as an anode hole injection process where the holes were injected into the oxide via thermionic emission rather than tunneling [27]. Another group found that the threshold energy for anode hole injection did not have a strong dependence on thickness or field (for the same anode material), which also supports thermionic emission as the transport mechanism for anode hole injection [28].

Another experiment showed that the gate voltage threshold for both the hole generation efficiency and the steady-state trapped hole density are the same and can be modulated by the work function of the anode material as shown in Figure 4.12 [29]. The threshold gate voltages observed in Figure 4.12 are consistent with the majority ionization threshold energies expected from equation (4.36).

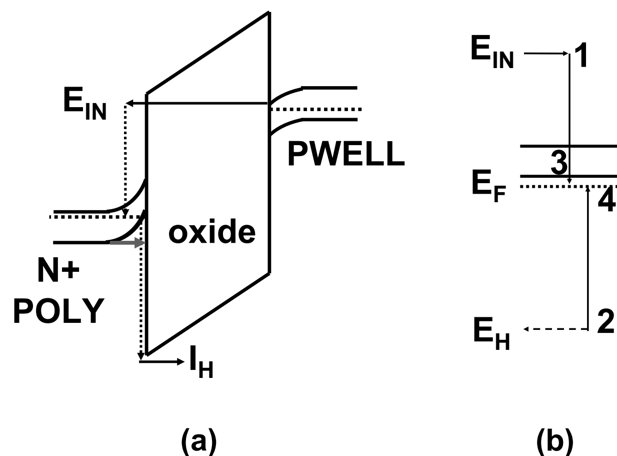

Figure 4.11. (a) Band diagram for minority-2 ionization in an NMOS device where both pwell and poly are inverted. An electron tunnels into the anode and collides with a hot hole. If the hole acquires > 5eV kinetic energy, it will be injected into the oxide valence band where it can generate trap states. (b) Schematic for the initial and final electron states. For minority-2 ionization, both electrons have their final states in the anode VB. After Bude, Ref. [23], Alam, Ref. [24] and Nicollian, Ref. [25]. © 1998, 2000 IEEE.

Inverted NMOS devices were used for Figure 4.12, where electrons are injected from the pwell inversion layer into the oxide. Accordingly, the hole generation efficiency I_B/I_D is not the same as the metric I_D/I_G that characterizes the quantum yield of electron hole pair creation in the substrate when electrons are injected from the gate. In inverted NMOS, electrons tunneling from the pwell enter the anode and impact ionize, resulting in anode hole injection. The hole generation efficiency I_B/I_D then senses the collection of these anode hot holes back in the pwell. This complicates the interpretation of the hole

generation efficiency since any source of substrate current can contribute to the measurement. Some of these mechanisms (in the absence of impact ionization in the oxide) include VB electron tunneling, generation-recombination, trap assisted tunneling, and photo-generation of electron hole pairs in the substrate by a photon originating in the anode after absorbing the energy of tunneling electrons [28,30].

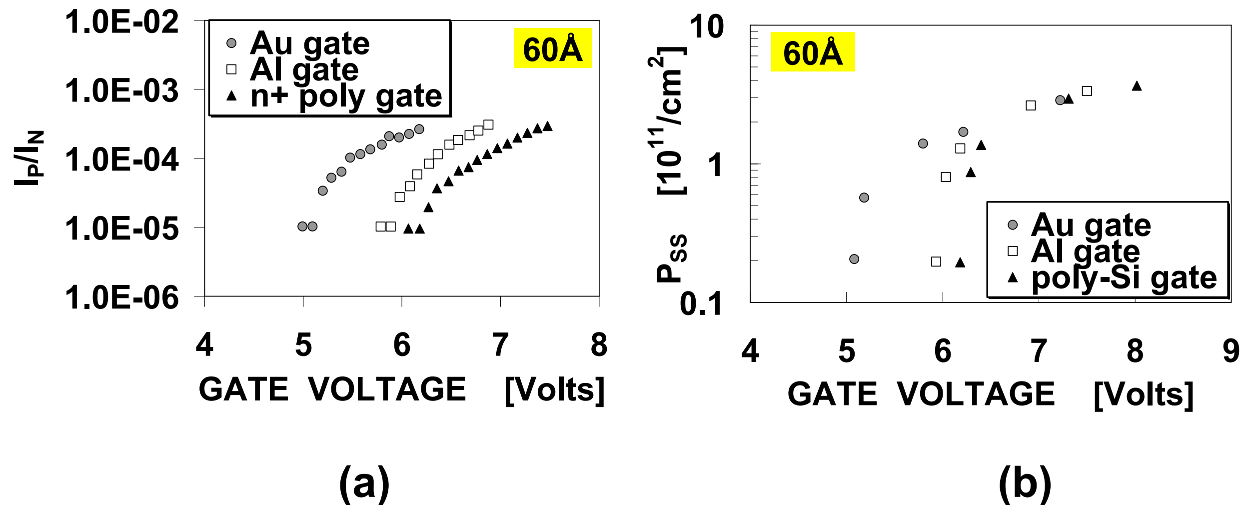

Figure 4.12. (a) Hole generation efficiency vs. gate voltage. (b) Steady-state trapped hole density vs. gate voltage. After Gao, Ref. [29]. © 1994 American Institute of Physics. Reprinted with permission.

Another investigation showed an increase in trap generation rate and a slight decrease in time to breakdown when a back bias was applied to PMOS devices stressed in inversion without an injector [23]. Since a back bias does not affect the gate current, but increases the energy of holes created in the anode, the authors claimed that this experiment supports anode hole injection. The post stress quantum yield was found to be similar with and without back bias, indicating that the same type of trap was created.

An issue with majority ionization is that it does not explain the polarity gap. Also, it does not account for higher hole current when holes are present. Accordingly, any viable low voltage breakdown mechanism must be dependent on the position of the anode Fermi level. Simulations incorporating minority ionization comprehend the dependence of the ionization rate on the position of the Fermi level [23] and correctly reproduce the polarity gap for the time to breakdown [24] as shown in Figures 4.13 and 4.14 respectively.

A shortcoming of minority ionization studies is the inability to measure the hole current in the substrate. To address this, the NMOS devices with different poly doping densities used in Chapter 3 were utilized with the process space expanded to include poly doping low enough to invert the poly. Since the poly is degenerately doped, the surface potential required to invert the poly is equal to the band gap. We will use these devices to obtain a readily measurable anode hole injection current by increasing the supply of both hot and thermal holes in the poly anode. Substrate current vs. poly band bending

for these devices are shown in Figure 4.15 [25]. For the lighter doped poly devices, the current in the substrate sharply increases when the poly inverts. This indicates that the origin of the substrate current is hole injection from the poly anode. The oxide voltage does not exceed 2.8V for any of the devices in Figure 4.15. Accordingly, the substrate current is not due to majority ionization. In Section 4.6, we will discuss whether this substrate current is "polluted" by mechanisms other than anode hole injection.

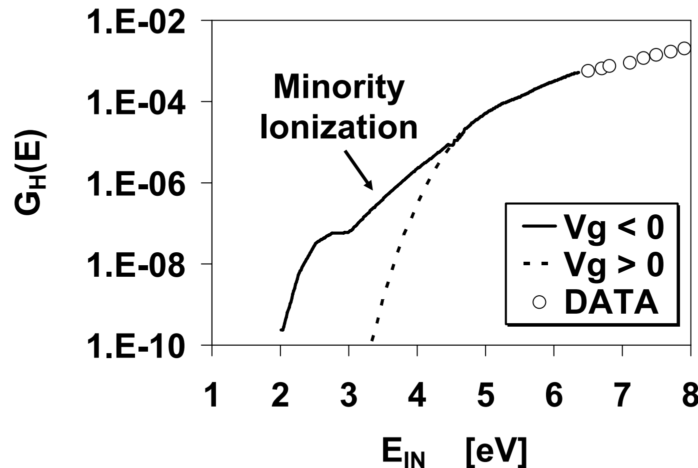

Figure 4.13. $G_H(E)$ vs. E_{IN}, showing polarity asymmetry when minority ionization is invoked. After Bude, Ref. [23]. © 1998 IEEE. Reprinted with permission.

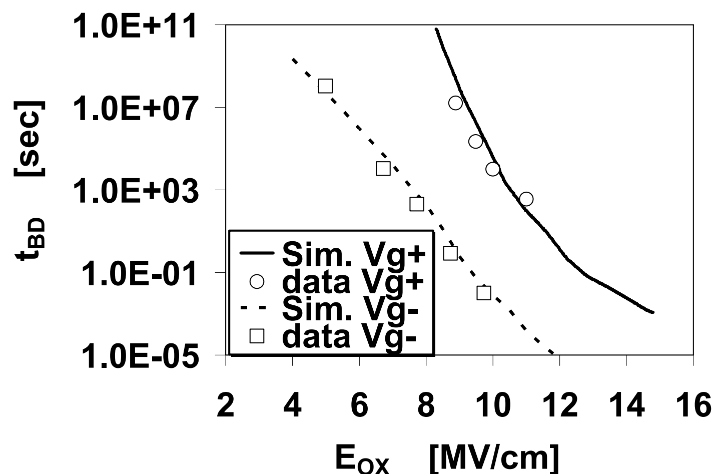

Figure 4.14. AHI simulations incorporating majority and minority ionization reproduce the polarity gap. After Alam, Ref. [24]. © 2000 IEEE. Reprinted with permission.

Band diagrams for minority ionization in NMOS devices with inverted poly were previously shown in Figures 4.10 and 4.11. From equation (4.44), the maximum hole kinetic energy in the anode is sufficiently high for the minority-2 process to occur when the poly inverts. Charge to breakdown as a function of poly band bending at a stress

voltage V_G = +3.6V is shown in Figure 4.16. It can be seen that Q_{BD} drops sharply after the poly inverts. Therefore, we have shown that anode hole injection through minority ionization is a plausible model for oxide breakdown at least down to 3.6V. However, explaining breakdown as due to anode hole injection below 3.0V – 3.5V is problematic. We will revisit anode hole injection at lower voltages in Section 4.9.

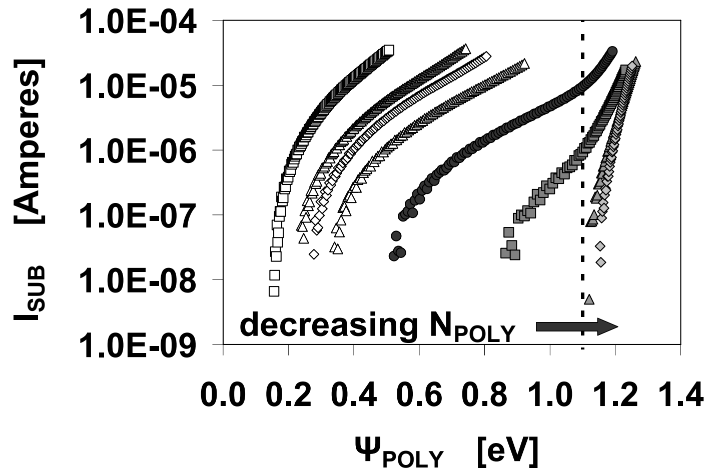

Figure 4.15. I_{SUB} vs. Ψ_{POLY}. I_{SUB} sharply increases when the poly inverts. After Nicollian, Ref. [25]. © 2000 IEEE.

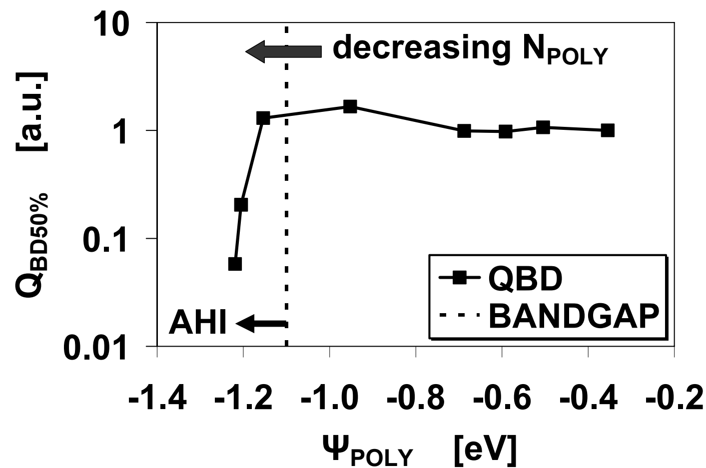

Figure 4.16. Q_{BD} vs. poly band bending for 26Å oxides stressed at V_G = +3.6V. Q_{BD} decreases sharply after the poly inverts. After Nicollian, Ref. [25]. © 2000 IEEE.

4.5 Features of anode hydrogen release

In this section, we will briefly describe anode hydrogen release as a mechanism for trap generation. We will then review experimental findings in the literature that support this model.

4.5.1 Description of the mechanism

Anode hydrogen release (AHR) occurs when the dissipation of the energy of electrons of electrons entering the anode results in the desorption of hydrogen [31]. This process results in the creation of bulk electron traps and interface states at both interfaces as shown in Figure 4.17. AHR is strongly dependent on temperature and weakly dependent on thickness [31] (which are the opposite dependencies of impact ionization in the oxide). Similar to anode hole injection, the threshold energy for AHR is thought to be about 6eV. However, scanning tunneling microscopy (STM) experiments have shown that the threshold energy ranges from about 0.25eV to 7eV; depending on the details of the desorption mechanism [32]. Another gap in the understanding of anode hydrogen release is that until recently, neither the reaction leading to AHR nor the hydrogen species desorbed from the anode were elucidated. In Section 4.6, we will present our novel contributions to the resolution of these issues.

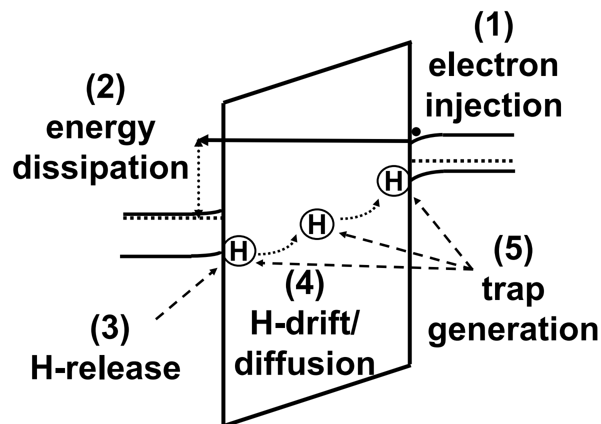

Figure 4.17. Band diagram for AHR. A hydrogen species is released in the processes that dissipate the energy of tunneling electrons that have entered the anode. The desorption of hydrogen results in bulk and interface trap generation.

4.5.2 Experimental evidence for anode hydrogen release

Much of the evidence that hydrogen causes trap generation and breakdown come from experiments that deliberately introduce hydrogen into the oxide. Figure 4.18 compares the impact of high temperature (900°C) post oxidation anneals in nitrogen vs. hydrogen on the charge to breakdown [33]. The devices went through gate oxidation together so that the oxide thicknesses are the same (250Å). It can be seen that Q_{BD} is about a factor of 10 lower when the annealing ambient is hydrogen.

Figure 4.19 shows the normalized increase in gate current in the direct tunneling regime (SILC) following exposure of gate oxides to atomic hydrogen in a remote plasma [34]. The gate leakage increases as a power law in hydrogen dose. The degradation is severe after long exposure times.

Another experiment utilized internal photoemission to analyze the redistribution of hydrogen in Al gate capacitors after exposure to UV radiation [35]. The Al gate deposition process provided a source of hydrogen at the Al-SiO_2 interface (prior to stress). The increase in interface trap density at the Si-SiO_2 interface and decrease in doping density of the p-type substrate are shown in Figure 4.20 [35]. The change in substrate doping is due to boron deactivation by hydrogen [36], so that it is a measure of the amount of hydrogen migrating from the Al gate into the substrate. It can be seen that hydrogen build-up at the Si-SiO_2 interface correlates with trap generation after UV exposure. Both trap generation and hydrogenation of the substrate show the same 5eV – 6eV threshold energy.

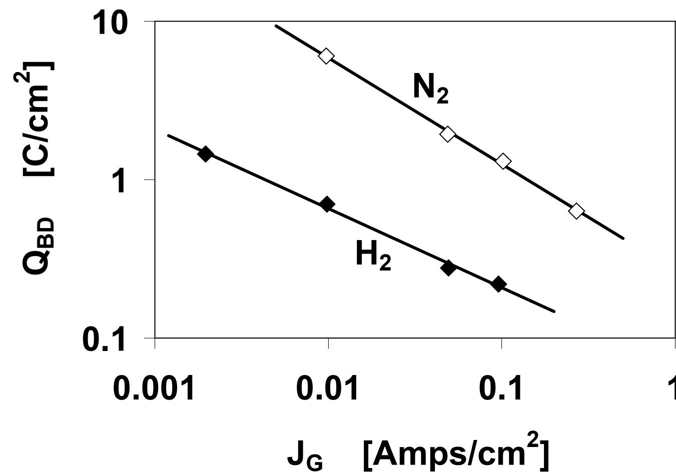

Figure 4.18. Charge to breakdown under constant current stress comparing post oxidation anneals in H_2 and N_2. After Nissan-Cohen, Ref. [33]. © 1988 IEEE. Reprinted with permission.

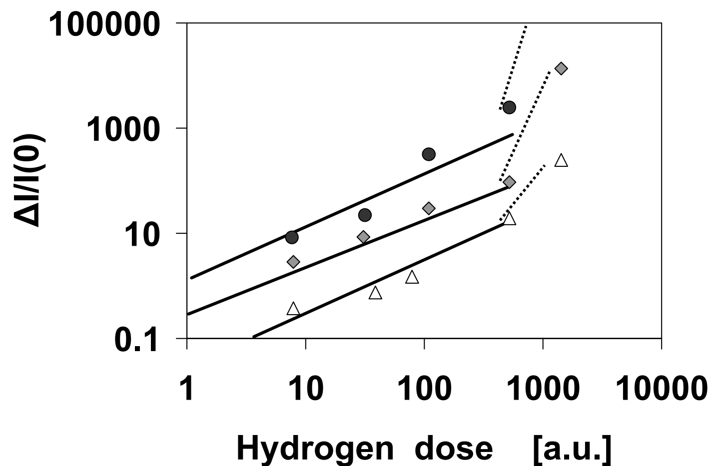

Figure 4.19. Increase in leakage current as a function of hydrogen dose in a remote plasma. The oxide thickness is 34Å (circles), 42Å (diamonds), and 52Å (triangles). After DiMaria, Ref. [34]. © 1995 American Institute of Physics. Reprinted with permission.

In reference [37], passivated silicon surfaces probed by scanning tunneling microscopy showed significantly lower desorption yields for deuterium passivated interfaces compared to hydrogen passivated interfaces. This finding stimulated studies of the effect of isotopes on channel hot carrier (CHC) and TDDB reliability. CHC experiments showed that the incorporation of deuterium significantly improved hot carrier lifetimes [38,39]. However, no concurrent improvement in TDDB was observed [39]. It was subsequently shown that if even the desorption yield is as much as 100X higher for hydrogen compared to deuterium, the exchange ratio between hydrogen and deuterium must be greater than 0.8 (i.e. > 80% of the hydrogen must be replaced by deuterium) to observe a statistically significant increase in Q_{BD} [40]. In the experiments in references [38,39], deuterium and hydrogen were introduced in low temperature (<450°C) anneals. A subsequent investigation of the SILC increase during TDDB stress showed that trap generation rates were reduced only when deuterium was incorporated during a high temperature pyrogenic gate oxidation process [41]. This work further supports a correlation between hydrogen release and trap generation.

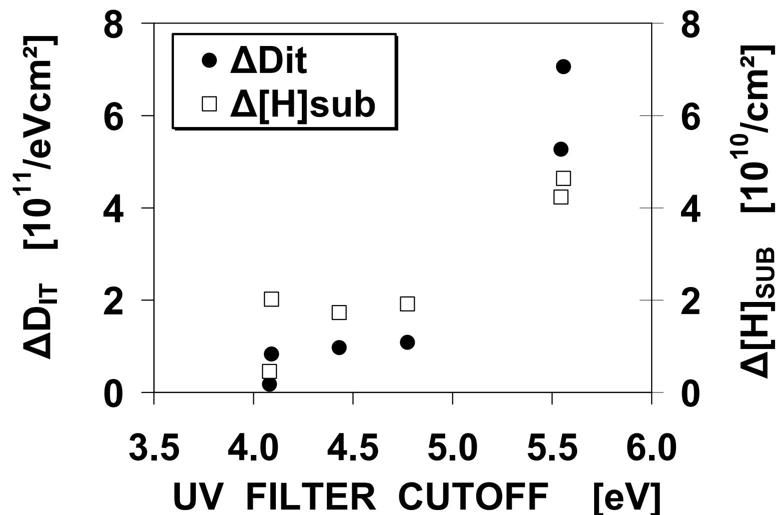

Figure 4.20. Change in Si-SiO_2 interface trap density and p-substrate doping following exposure to UV radiation. After Buchanan, Ref. [35]. © 1994 American Institute of Physics. Reprinted with permission.

4.6 Identification of the hydrogen species that cause trap generation

In the previous section, we reviewed data in the literature supporting hydrogen release as a mechanism for trap generation and breakdown. However, neither the reaction leading to AHR nor the hydrogen species involved has been elucidated. In this section, will present our original findings that trap generation and breakdown at low stress voltages in ultra thin SiON gate dielectrics are triggered by the release of two hydrogen species (H^+ and H^0) [42].

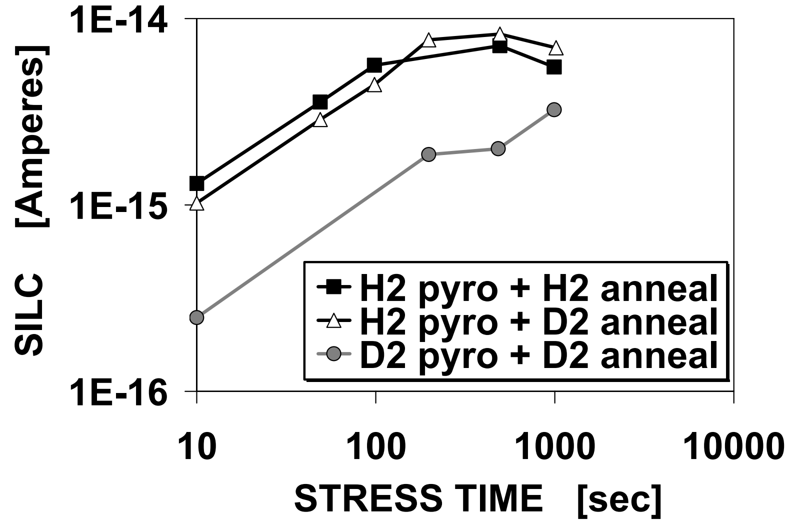

Figure 4.21. Effect of low temperature post-metallization annealing and high temperature pyrogenic oxidation ambients on the SILC increase. High temperature incorporation of deuterium reduces the trap generation rate. The oxide thickness was 62Å. After Mitani, Ref. [41]. © 2001 IEEE. Reprinted with permission.

4.6.1 The stress method

Reliability models for the NMOS on-state are of particular interest because in an inverter, the entire gate oxide area is stressed. A problem with the analysis of degradation mechanisms for this mode is that the anode, where the reactions initiate, is the poly-SiON interface, which is difficult to characterize. To circumvent this problem, a known reaction is initiated at the readily characterized Si-SiON interface and used as a probe of the anode reactions [42]. This is accomplished by applying a back-bias to the pwell to induce to induce the desorption of cathode hydrogen by substrate hot electrons as shown in Figure 4.22 [42]. No n+ injector is used to avoid both uncontrolled flooding of the dielectric with the cathode hydrogen species and high level injection of hot electrons into the anode.

The experiments presented on this section are performed on 12Å SiON (PNO) dielectrics. All devices are NMOS. C-V sweeps are at 25°C on 10^{-6} cm^2 gate areas; all other measurements are performed at 105°C on 10^{-7} cm^2 gate areas. Time-0 I-V characteristics are shown in Figure 4.23 [42]. Since the gate current is independent of V_{BB}, the devices remain in strong inversion during stress. The maximum $|V_{BB}|$ during stress is 6V. The increase in I_B from band to band tunneling is < 2μA (20 Amps/cm^2) below 6V $|V_{BB}|$. Accordingly, the hot electron flux at the cathode interface resulting from back bias is less than 2% of the gate current for all stress conditions in this experiment, verifying that the back-bias stress is a low level injection condition. The maximum kinetic energy of substrate hot electrons dissipated at the cathode is [42]

$$E_{CATH}(max) \approx qV_{BB} + 2\Psi_S(inv)|_{Vbb=0V} - \Psi_S(inv)|_{Vbb} \qquad (4.46)$$

From (4.46), the electron kinetic energy at the cathode is sufficiently high for all of the hydrogen desorption mechanisms listed in Figure 4.1 to be operative in this experiment.

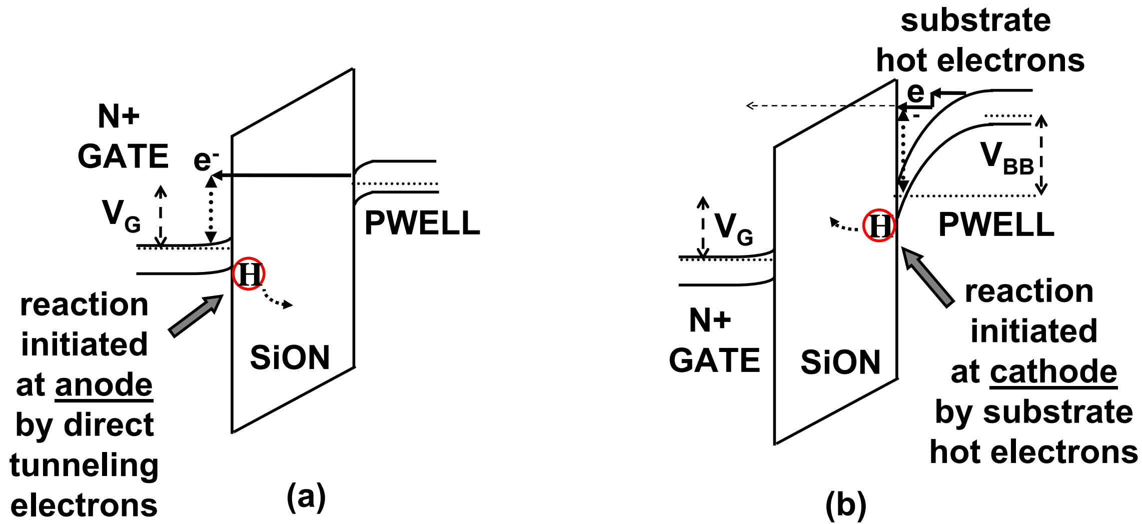

Figure 4.22. Band diagram for stress. Reaction initiated: (a) At anode, drawn at V_{BB} = 0V; (b) at cathode, $|V_{BB}|$ > 0V. Dashed horizontal line in (b) denotes transmission through the barrier. After Nicollian, Ref. [42]. © 2005 IEEE.

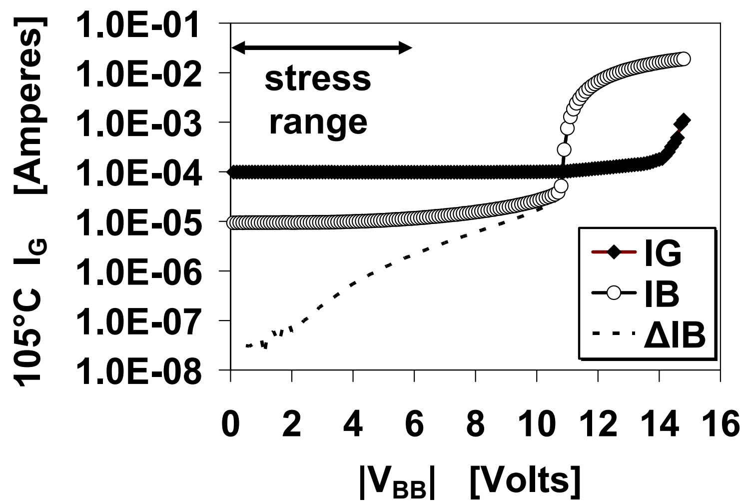

Figure 4.23. Time-0 gate and substrate currents vs. $|V_{BB}|$ with V_G = +2.3V. The increase in I_B, where $\Delta I_B = I_B(V_{BB}) - I_B(0)$ is < 2μA below 6V $|V_{BB}|$. After Nicollian, Ref. [42]. © 2005 IEEE.

4.6.2 Cathode interface reactions induced with a back bias

The LV-SILC characteristics for V_{BB} = 0V and V_{BB} = -6V are shown in Figure 4.24 [42]. A back-bias increases the density of both trap states, but the effect is more pronounced for the peak at -1V sense. Forward and reverse C-V characteristics before and after

stress with V_{BB} = -6V are shown in Figure 4.25 [42]. There is little hysteresis or V_{FB} shift, indicating that charge trapping is insignificant so that the C-V shifts are primarily due to the generation of interface traps. The change in capacitance ΔC resulting from stress with V_{BB} = -6V, V_G = +2.32V is also shown in Figure 4.25 [42]. Two trap peaks are also seen in the C-V characteristics. The area under the ΔC vs. V_G curve is charge. As the sum of the charge $Q_1 + Q_2 \sim Q_{\Delta Vt}$, both trap states are acceptor like (negatively charged when occupied, neutral when empty). Because holes create only donor states in SiON at low stress voltages [7,43], holes are not involved in the reactions that create interface traps in inverted NMOS when a back bias is applied during stress. Therefore, the reactants are electrons and Si-H bonds. The dissociation of N-H bonds are not considered because they have not been observed in PNO films by ESR [44].

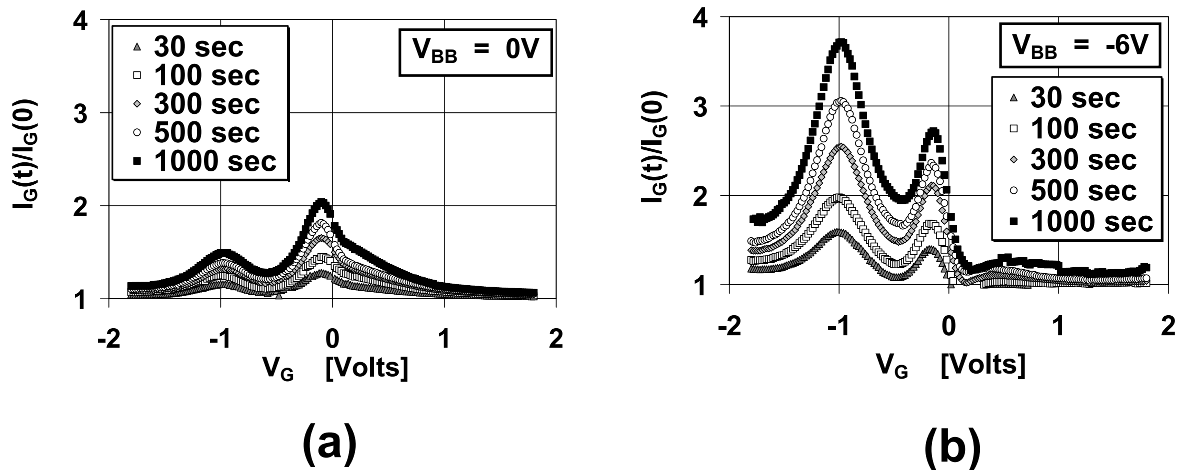

Figure 4.24. LV-SILC after stress at (a) V_G = +2.32V, V_{BB} = 0V. (b) V_G = +2.32V, V_{BB} = -6V. The two peaks are due to interface traps. After Nicollian, Ref. [42]. © 2005 IEEE.

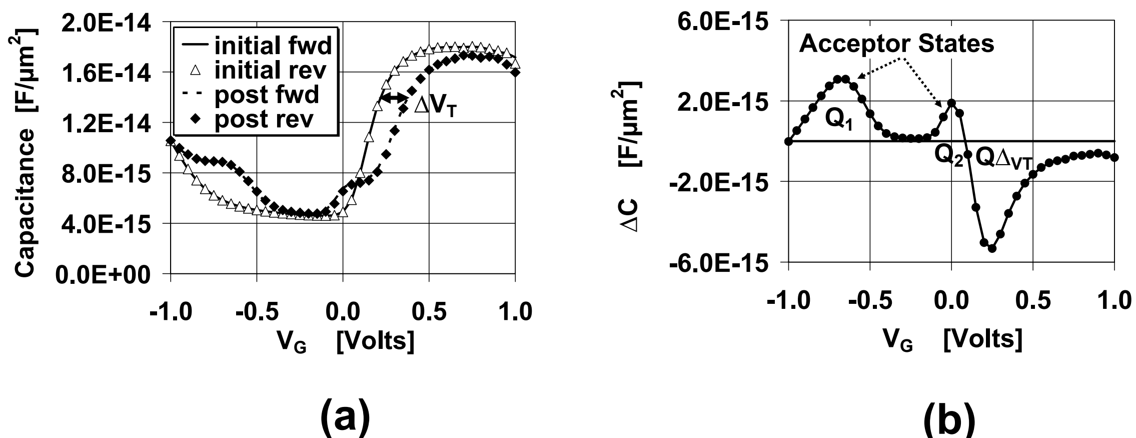

Figure 4.25. (a) Pre and post stress forward and reverse C-V characteristics. The devices were stressed for 1,000 seconds at V_G = +2.32V, V_{BB} = -6V. (b) Change in capacitance ΔC = C(t) – C(0) resulting from stress using the data in (a). The two trap peaks are acceptor states. After Nicollian, Ref. [42]. © 2005 IEEE.

Using the change in ΔI_{DLIN} as a measure for interface state generation, interface trap generation power laws for our 12Å SiON films are plotted with V_{BB} as a parameter in Figure 4.26 [42]. The interface trap generation power law "m" increases to 0.38 at V_{BB} = -6V, indicating that a charged species is involved (see Figure 4.2). Figure 4.27 shows the effect of delay time on interface trap generation when a back bias is applied [42]. The interrupted stress and uninterrupted stress curves are determined from LV-SILC and ΔI_{DLIN} respectively. Since the power law exponent "m" is significantly larger for interrupted stress, recovery effects are present and the system is in quasi-equilibrium. Therefore, reaction-diffusion theory may be applied to the interface trap generation reactions resulting from the application of a back-bias during stress. It should be noted that the recovery observed when a back bias is applied implies that LV-SILC is not a steady-state current. However, when no back bias is applied, no interface trap recovery effects are observed [42] so that LV-SILC is a steady state effect for NMOS SiON under PBTI stress conditions for V_{BB} = 0V.

From Figures 4.25 and 4.26: (i) two acceptor traps are generated, (ii) the reactants are electrons and Si-H bonds, (iii) a charged species is released at the cathode. A model for the reaction steps that are consistent with these observations is shown in Figure 4.28 [42]. The 1/3 power law is driven by the desorption of $H^-$and H^0 by cathode hot electrons. Note that the steps leading to the final product H_2^- in Figure 4.28 are an "effective" reaction, as our TCAD simulations show that a 1/3 power law can also arise from equal amounts of H^- and H^0. "m" slightly higher than 1/3 may be due to a higher desorption rate for H^- over H^0 when hot electrons impinge on the cathode interface. The existence of one form of charged H_2 in SiO_2 (H_2^+) has been proposed [45].

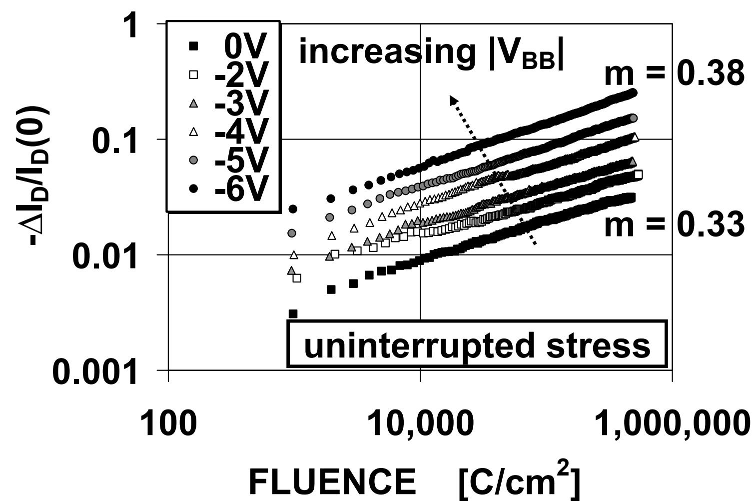

Figure 4.26. Change in I_{DLIN} after stress, with V_{BB} as a parameter. After Nicollian, Ref. [42]. © 2005 IEEE.

It is also possible to directly observe the reactions that initiate trap generation and breakdown at the Si-SiON interface by stressing in accumulation. However, this mode is

more complicated to analyze because large numbers of holes are present, which introduce additional reaction possibilities.

4.6.3 Bulk trap generation

Having established that application of a back bias releases H^0 and H^- from the cathode into the dielectric, the effects of these species on SiON bulk trap creation can be evaluated. Figure 4.29 compares bulk trap generation with and without back bias applied during stress [42]. The bulk trap generation "m" increases from 0.26 (V_{BB} = 0V) to 0.51 (V_{BB} = -6V). The bulk trap generation pre-factor "b" is lower at -6V V_{BB}, indicating that the -6V V_{BB} cathodic species released suppresses nominal bulk trap generation through a reaction that depletes the anodic species. Otherwise, the trap density would be the same or higher at all fluence. Accordingly, nominal bulk trap generation is due to the release of a positively charged anode species [42].

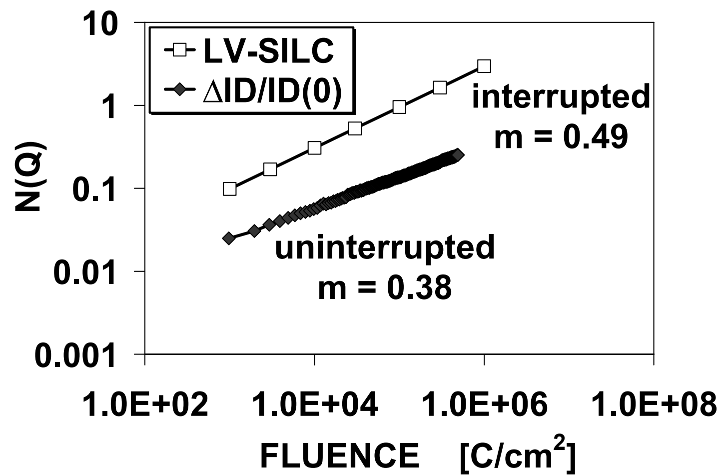

Figure 4.27. Interface trap generation power laws from LV-SILC (-1V sense) and I_{DLIN} for interrupted vs. uninterrupted stress respectively. The devices were stressed at V_G = +2.32V and V_{BB} = -6V. Larger "m" for interrupted stress is due to recovery effects. After Nicollian, Ref. [42]. © 2005 IEEE.

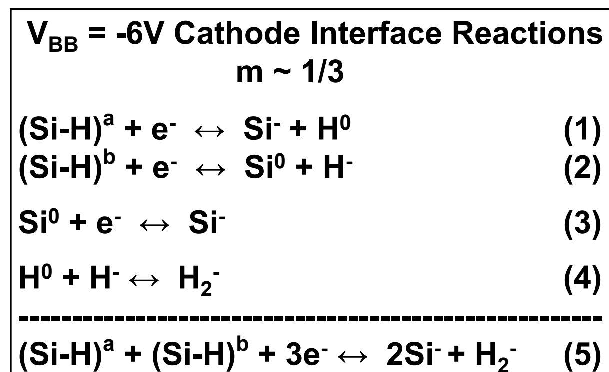

Figure 4.28. Model for cathode reactions for acceptor interface trap creation at -6V V_{BB} stress. (5) is the effective net reaction. After Nicollian, Ref. [42]. © 2005 IEEE.

In Figure 4.30, the V_{BB} dependence of the Weibull slope, which is a geometric quantity, rules out the transmission of substrate hot electrons into the anode (dashed horizontal line in Figure 4.22-b) as the primary cause for the modifications in trap generation seen in Figure 4.29 [42]. The changes of Weibull slopes in Figure 4.30 are proportional to the changes in "m", consistent with predictions from the cell-based percolation model [46], which we will discuss in Section 4.8. While breakdown becomes "softer" with back bias due to lower inversion charge [47], we are still able to accurately detect the 1ST breakdown event at V_{BB} = -6V using noise variance as a trigger as shown in Figure 4.31 [42]. This finding is confirmed in Figure 4.32 [42], which shows that the trends in trap generation power law parameters, which are not extracted from the measurement of breakdown, follow similar trends with back-bias as the Weibull slope in Figure 4.30. In addition, the similarity of the trends with back bias in Figures 4.30 and 4.32 show that our trap generation power laws are tracking the traps that cause SBD [42].

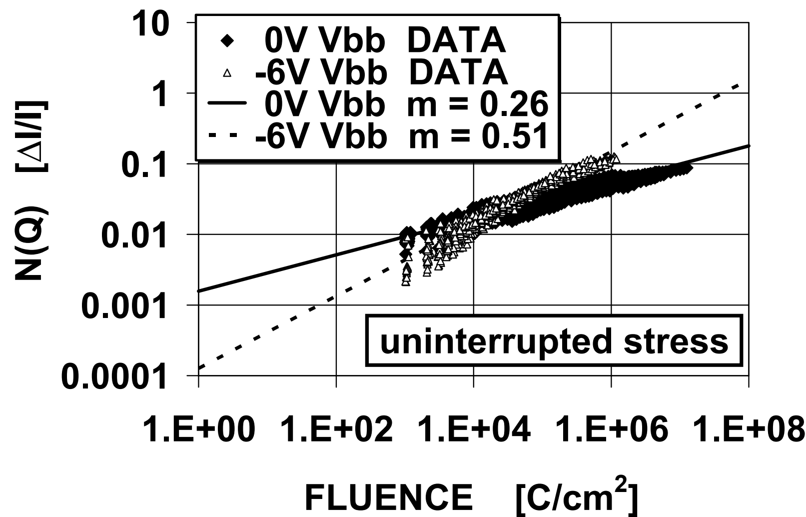

Figure 4.29. Bulk trap generation power laws from SILC, showing the effect of V_{BB}. V_G = +2.32V during stress and sense. The lines are statistical fits to the data. After Nicollian, Ref. [42]. © 2005 IEEE.

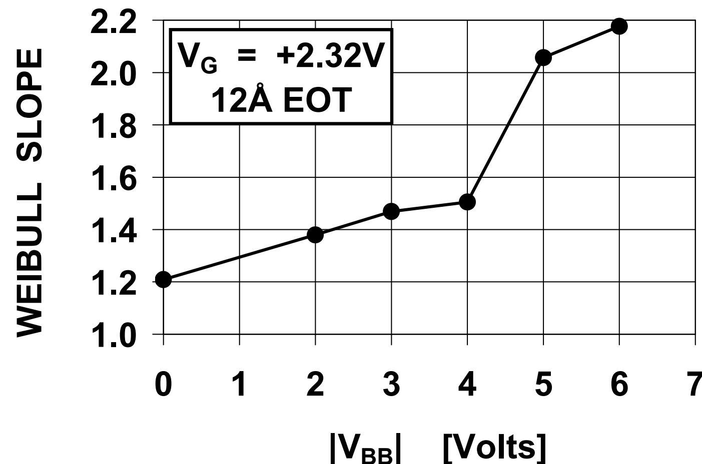

Figure 4.30. Weibull slope vs. $|V_{BB}|$ for stress at V_G = +2.32V. After Nicollian, Ref. [42]. © 2005 IEEE.

4.6.4 Two reaction model

Figure 4.33 shows the effect of interrupted stress on bulk trap generation measured by SILC sensed at V_G = +2.32V [42]. Regardless of whether a back bias is applied, interrupted stress results in a larger power law "m". Therefore, the reaction that gives rise to bulk trap generation is in quasi-equilibrium [42]. Accordingly, the detrapping of holes, which increases with increasing field [48], cannot be the cause for the larger "m" for interrupted stress since the field is much smaller when the stress is turned off.

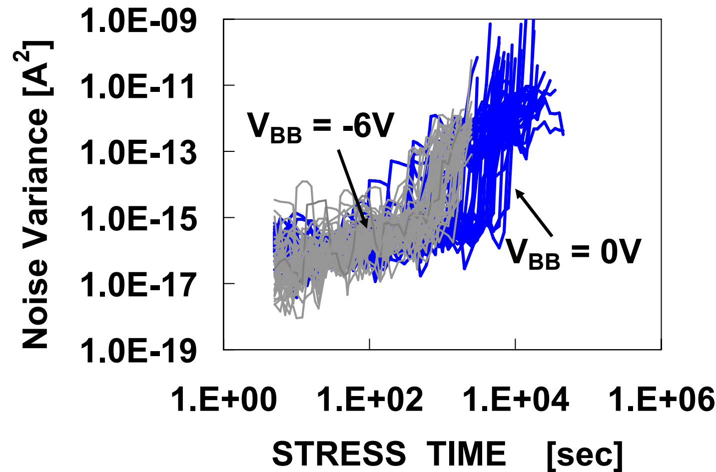

Figure 4.31. Noise variance sensed at V_G = +1.0V during TDDB stress. The stress conditions were V_G = +2.32V at 0V V_{BB} (dark lines) and -6V V_{BB} (grey lines). After Nicollian, Ref. [42]. © 2005 IEEE.

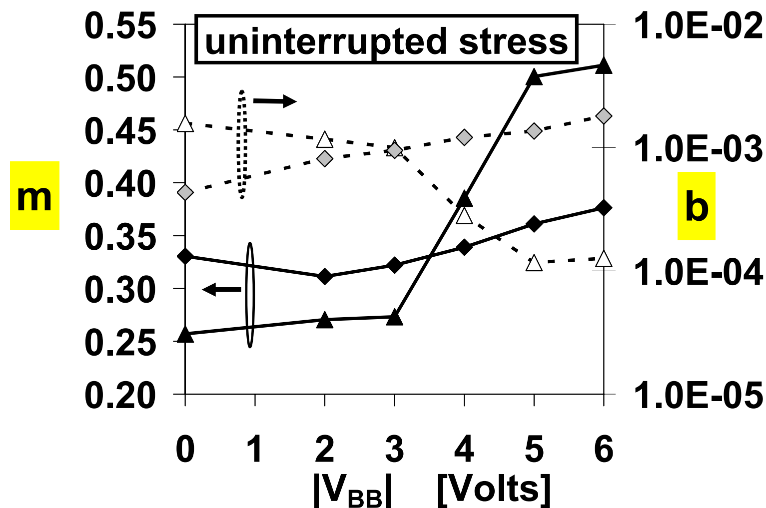

Figure 4.32. Trap generation power law parameters vs. $|V_{BB}|$ for uninterrupted stress at V_G = +2.32V. Bulk traps are sensed by SILC and interface traps are sensed by ΔI_{DLIN}. Dark fill symbols are "m", light fill symbols are "b", triangles are bulk trap parameters, and diamonds are interface trap parameters. After Nicollian, Ref. [42]. © 2005 IEEE.

Models for the nominal (V_{BB} = 0V) anode reactions are shown in Figure 4.34 [42]. Bulk traps are generated in SiON by H^+ or H_2^+ following the desorption of Si-H bonds by anode hot holes. The hot holes originate through impact ionization of electrons tunneling into the anode. A 2ND anode reaction that releases H^0 is also needed to account for nominal cathode interface trap creation because only neutral or positive species can be released from the anode into the dielectric and H^+ is not highly reactive at Si-SiO_2 interfaces [45]. Moreover, positive species create donor traps at the Si-SiON interface, but acceptor states are still the dominant cathode interface traps when V_{BB} = 0V as shown in Figure 4.35 [42]. As we will show in Section 4.8, the differences in V_G dependence between SILC and LV-SILC indicate that bulk and interface trap generation are not triggered by the same reaction [42]. While these experiments cannot differentiate whether anodic H^0 is desorbed by holes (Figure 4.34-a) or electrons (Figure 4.34-b), the results that we will present in Section 4.8 will show that the correct reaction sequence is that in Figure 4.34-b (H^0 is desorbed by electrons).

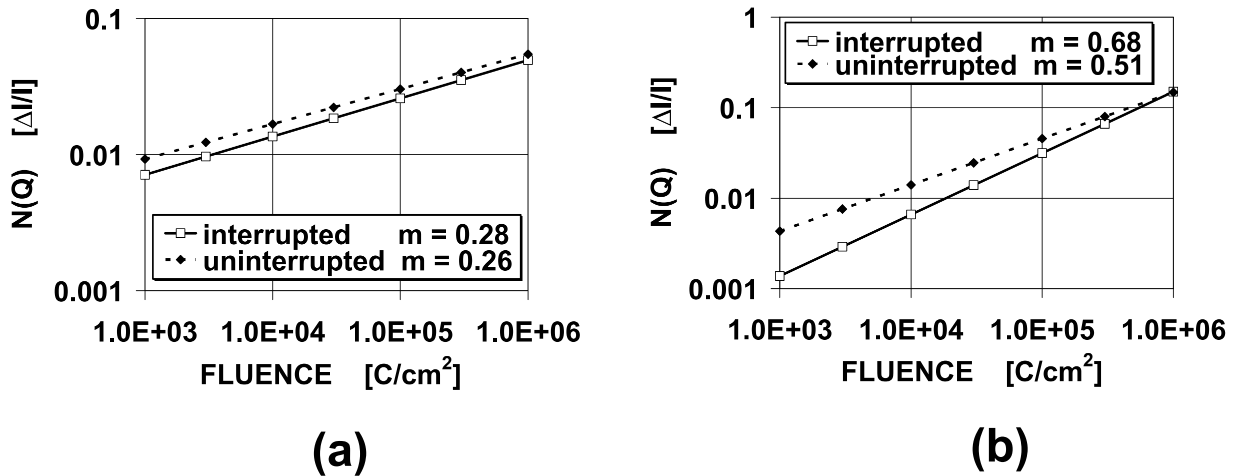

Figure 4.33. Bulk trap generation power laws from SILC for interrupted vs. interrupted stress respectively. The lines are statistical fits to the data. The devices were stressed and sensed at V_G = +2.32V. Larger "m" for interrupted stress is due to recovery effects. (a) V_{BB} = 0V, (b) V_{BB} = -6V. After Nicollian, Ref. [42]. © 2005 IEEE.

Nominal Anode Interface Reactions – (a)		**Nominal Anode Interface Reactions – (b)**	
$(Si\text{-}H)^c + h^+ \leftrightarrow Si^0 + H^+$	(1)	$(Si\text{-}H)^c + h^+ \leftrightarrow Si^0 + H^+$	(1)
$(Si\text{-}H)^d + h^+ \leftrightarrow Si^+ + H^0$	(2a)	$(Si\text{-}H)^f + e^- \leftrightarrow Si^- + H^0$	(2b)
$H^0 + H^+ \leftrightarrow H_2^+$	(3)	$H^0 + H^+ \leftrightarrow H_2^+$	(3)
$(Si\text{-}H)^c + (Si\text{-}H)^d + 2h^+ \leftrightarrow Si^+ + Si^0 + H_2^+$	(4a)	$(Si\text{-}H)^c + (Si\text{-}H)^f + h^+ + e^- \leftrightarrow Si^- + Si^0 + H_2^+$	(4b)

(a) **(b)**

Figure 4.34. Models for nominal (V_{BB} = 0V) anode reactions. (a) Desorption of both H^+ and H^0 by holes. (b) Desorption of H^+ by holes and H^0 by electrons. (4a) or (4b) is the net effective reaction. After Nicollian, Ref. [42]. © 2005 IEEE.

Although H^0 may be liberated from nominal bulk trap creation (m ~ 0.25 for V_{BB} = 0V; see Figures 4.2 and 4.29), this cannot be the sole source of H^0 since the generation rate is higher for interface traps as shown in Figure 4.36 [42]. A band diagram of our model for trap creation is shown in Figure 4.37. Both H^+ and H^0 create interface traps at the anode when they are released. After entering the dielectric, H^+ subsequently creates SiON bulk traps and H^0 creates cathode interface traps [42].

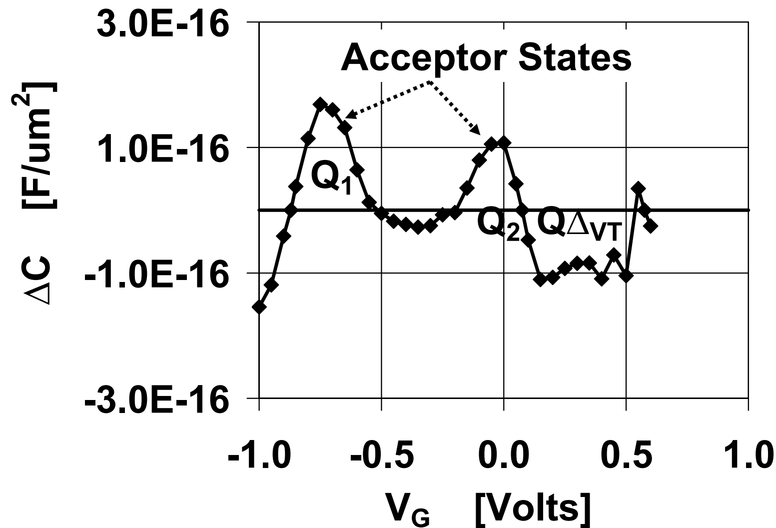

Figure 4.35. ΔC after 1,000 second stress at V_G = +2.32V, V_{BB} = 0V. Two acceptor trap peaks are present. After Nicollian, Ref. [42]. © 2005 IEEE.

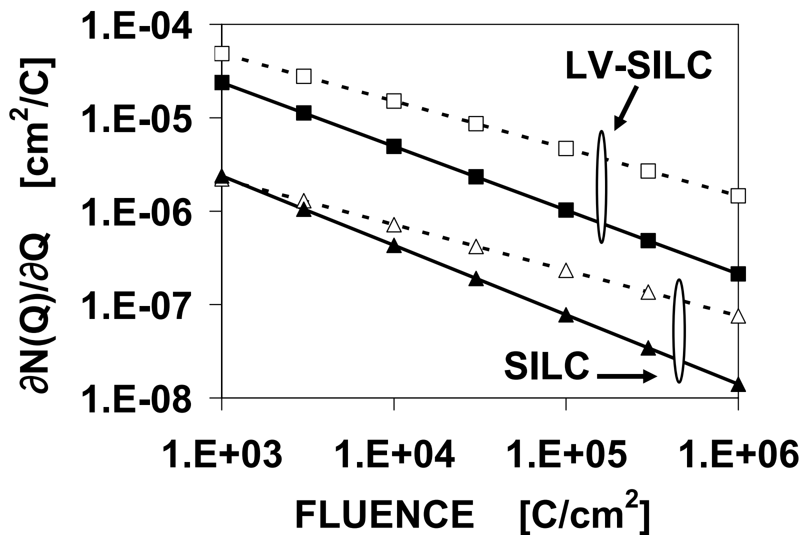

Figure 4.36. Bulk trap (SILC) and interface trap (LV-SILC) generation rates calculated using equation (2.66). The devices were stressed at V_G = +2.32V. V_{BB} = 0V (dark fill), V_{BB} = -6V (open fill). After Nicollian, Ref. [42]. © 2005 IEEE.

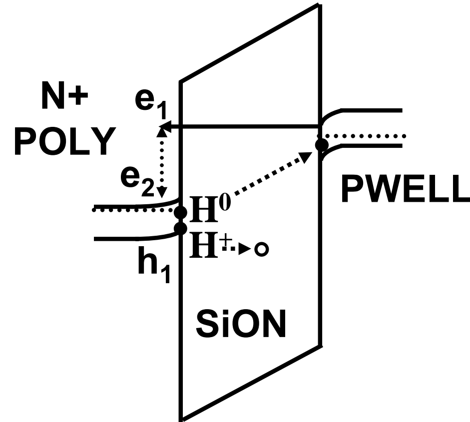

Figure 4.37. Model for SiON trap generation. e_1 tunnels into anode, and impact ionizes. H^+ and H^0 are subsequently desorbed. H^+ creates bulk and poly interface traps. H^0 creates traps at both interfaces. After Nicollian, Ref. [42]. © 2005 IEEE.

4.7 Comparison of breakdown data with scanning tunneling microscopy results

Many of the recent breakthroughs regarding the role of hydrogen in trap generation and breakdown in MOS devices have been leveraged from scanning tunneling microscopy (STM) experiments. In the STM technique, hydrogen is desorbed by placing a biased probe in close proximity to a silicon surface [32]. In this section, we review the features of STM that are central to the current understanding of oxide breakdown and show that the results of these studies leads to an explanation to the TDDB Power Law Model.

4.7.1 STM induced hydrogen desorption from silicon surfaces

There are two silicon-hydrogen bond desorption mechanisms that we will consider: Electrical excitation (EE) and vibrational excitation (VE). Electrical excitation is a process where hydrogen is desorbed via field emission [32]. EE has a threshold energy of 6eV to 7eV and is weakly dependent on energy and current. [32]. Vibrational excitation is a process where hydrogen is desorbed via excitation of phonon modes [32,49-52]. VE is an important mechanism in the Si-SiO_2 system because the Si-H bond has a long vibrational lifetime; on the order of 10^{-8} seconds [32]. VE can occur at energies below the threshold energy for EE since E_{TH} for VE is about 2.5eV to 3.0eV. Since the phonon modes are about $\hbar\omega$ = 0.25eV apart, there E_{TH}/0.25 ~ 10 to 12 levels between the ground state and the top of the potential well [32]. The mechanisms for vibrational excitation are shown in Figure 4.38 as follows: (a) one electron coherent excitation. This occurs when a single electron desorbs the Si-H bond from the ground state [32]. This process is strongly dependent on voltage and weakly dependent on current. (b) Incoherent vibrational excitation [50,51]. In this process, the Si-H bond is desorbed via 10 -12 electrons (1 per energy level) and occurs when incoming electrons have energies < $2\hbar\omega$. The rate of this process is extremely small. Incoherent vibrational excitation is strongly dependent on voltage and current. The strong current dependence

for multi-vibrational hydrogen release (MVHR) arises from the higher probability of desorption when large numbers of electrons are present. (c) Multi-electron excitation, where only a few electrons are needed for hydrogen release [52]. For M electrons, the energy per electron for desorption to occur is ≥ E_{TH}/M. Accordingly, MVHR also occurs at energies below the threshold energy for VE, albeit at a lower probability than single electron desorption. However, MVHR involving only 2 or 3 electrons is significantly more probable than incoherent excitation and can play an important role below 2.5eV. The voltage and current dependences of EE and VE processes are tabulated in Figure 4.39.

Yield vs. voltage for hydrogen desorption measured in STM experiments is shown in Figure 4.40. The plateau of the curve is the EE regime. In the region where the yield is rapidly decreasing with voltage, VE becomes an important mechanism. Since the TDDB Power Law Model has a strong voltage dependence, electrical excitation is ruled out as the mechanism that results in the TDDB Power Law Model.

The voltage dependence of the fraction of tunneling electrons that excite a silicon hydrogen phonon mode is shown in Figure 4.41. It can be seen that this inelastic tunneling fraction has a power law dependence of V^4 (per level) [32]. Accordingly, the inelastic tunneling fraction has a net voltage dependence of V^{4n}, which is V^{40-48} for n = 10 to 12 phonon levels. Therefore, vibrational excitation of hydrogen provides a plausible mechanism for the TDDB Power Law Model [51]. The general form of the voltage and current dependence for the generation efficiency of vibrational excitation processes involving M electrons is [52]

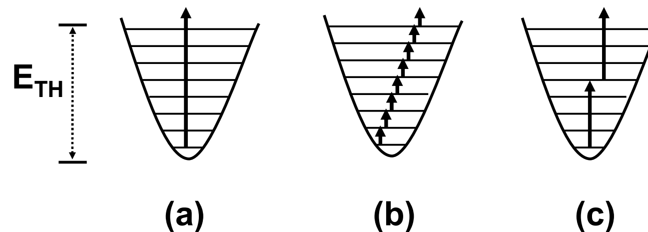

Figure 4.38. Vibrational excitation processes. (a) Single electron coherent excitation. (b) Incoherent excitation. (c) Two electron excitation.

HR Mechanism	Current Dependence	Voltage Dependence
EE	WEAK	WEAK
COHERENT VE	WEAK	STRONG
MULTI-ELECTRON VE	STRONG	STRONG
INCOHERENT VE	STRONG	STRONG

Figure 4.39. Summary of voltage and current dependence of electrical and vibrational excitation processes.

$$\xi(V,I) = I_0^{M-1}V^{4n} \quad (4.47)$$

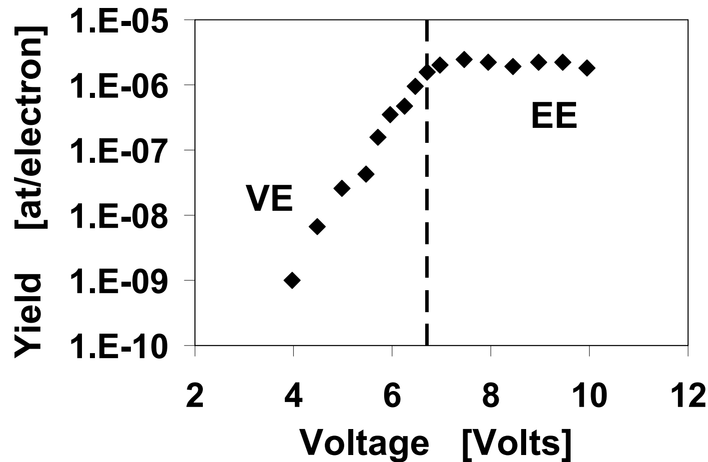

Figure 4.40. STM Yield vs. voltage for hydrogen desorption. After Shen, Ref. [32]. © 1995 AAAS. Reprinted with permission.

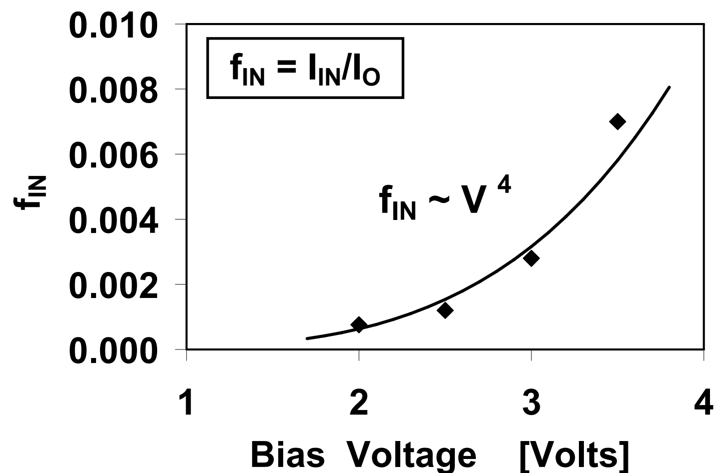

Figure 4.41. Inelastic tunneling fraction vs. bias voltage from STM experiments. After Shen, Ref. [32]. © 1995 AAAS. Reprinted with permission.

4.7.2 Evidence of vibrational excitation of hydrogen from TDDB studies

We will now compare trap generation and breakdown data to STM results. One of the difficulties in resolving single vs. multi-electron desorption is that the effect of current, which is a function of voltage, must be decoupled from the voltage. Because substrate hot electron injection (SHE) can be used to independently control oxide electric field, electron energy, and electron fluence [53], this technique can be applied to the studies vibrational excitation mechanisms. Using SHE, interface trap generation has been reported to be dependent on gate current at a fixed energy at low voltages [54],

indicating the presence of MVHR as a mechanism for interface trap generation at the cathode interface. These findings indicate that behavior reported in STM experiments also occurs under electrical stress in MOS devices.

The current and voltage dependence of trap generation and breakdown can also be separated by stressing oxides of different thickness at the same voltage [55]. From equation (4.47), the trap generation efficiency ξ should be independent of current for single electron VE so that ξ vs. E_{MAX} should be independent of thickness. For two electron VE, ξ/J vs. E_{MAX} should be independent of thickness [55]. For E_{MAX} > 2.5eV, the data corresponding to ξ(E_{MAX}) for different thickness fall on the same curve in Figure 4.42-a. Accordingly, these data are consistent with single electron vibrational excitation. For E_{MAX} < 3.0eV, the data corresponding to ξ/J(E_{MAX}) fall on the same curve in Figure 4.42-b. Therefore, below the 2.5eV to 3.0eV threshold energy for vibrational excitation, trap generation and breakdown are consistent with 2 electron VE.

Figure 4.42-b yields a power law $\xi/J \sim E_{MAX}^{41}$ for E_{MAX} < 2.5eV and Figure 4.42-a yields a power law $\xi \sim E_{MAX}^{35}$ for E_{MAX} > 2.5eV. The power law exponent decreases as E_{MAX} increases and goes to zero as HR transitions towards the EE regime. This suggests that EE and VE may cooperate in the intermediary region between 2.5eV and 7.0eV [55]. Since a vibrational excitation increases the S-H bond distance, the threshold energy for desorption should decrease proportionally [37]. In this scenario, one electron excites the bond to a vibrational state, followed by an electrical excitation event that desorbs the hydrogen. While the symmetric harmonic oscillator potentials in Figure 4.38 provide a useful visualization of vibrational excitation processes, they do not capture cooperative processes between EE and VE via an increase in bond length. A Morse potential, which is asymmetric, has been used to model this behavior [55].

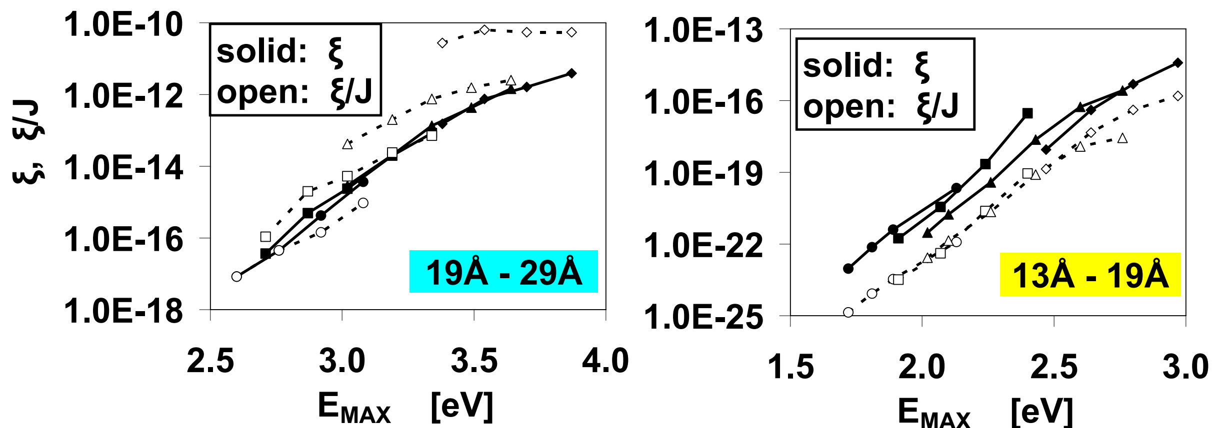

Figure 4.42. Trap generation efficiency ξ and ξ/J vs. E_{MAX} for (a) E_{MAX} > 2.5eV. The oxide thicknesses are 19Å (circles), 23Å (squares), 25Å (triangles), 29Å (diamonds). The data are consistent with single electron VE. (b) E_{MAX} < 3.0eV. The oxide thicknesses are 13.8Å (circles), 15Å (squares), 16.7Å (triangles), 18.7Å (diamonds). The data are consistent with two electron VE. After Suñé, Ref. [55]. © 2005 IEEE. Reprinted with permission.

4.8 Identification of the stress-induced defects that cause breakdown

In this section, we review geometric models for the formation of percolation (SBD) paths in thin dielectrics. We will then present our original work on the defects that cause breakdown in SiO_2 and SiON films, along with the implications for SiON thickness scaling. We have included this discussion in this chapter because the defects that are created during stress are consequent of the operative trap generation mechanisms.

4.8.1 Percolation models

It is desirable to construct a quantitative modeling capability that is predictive of thickness scaling trends. This was first demonstrated with Monte-Carlo simulations [56], where bulk traps of a fixed radii are generated at random positions in the dielectric until a sufficient number of traps have been created to form path of overlapping states that connect the gate and substrate electrodes as shown in Figure 4.43-a. This method correctly reproduces the experimental trend of decreasing Weibull slope with decreasing oxide thickness. This approach was later extended to include the possibility of spatially non-uniform trap generation in the oxide [57] as illustrated in Figure 4.43-b for gradient percolation, where the generation probability decreases exponentially with the distance from one of the interfaces. For oxides thicker than about 20Å - 30Å, the maximum Weibull slope occurs when the trap generation probability is spatially uniform [57]. This prediction is consistent with experiments where PNO films with non-uniform nitrogen profiles have polarity dependent Weibull slopes and higher trap generation rates compared to films with uniform nitrogen incorporation [58].

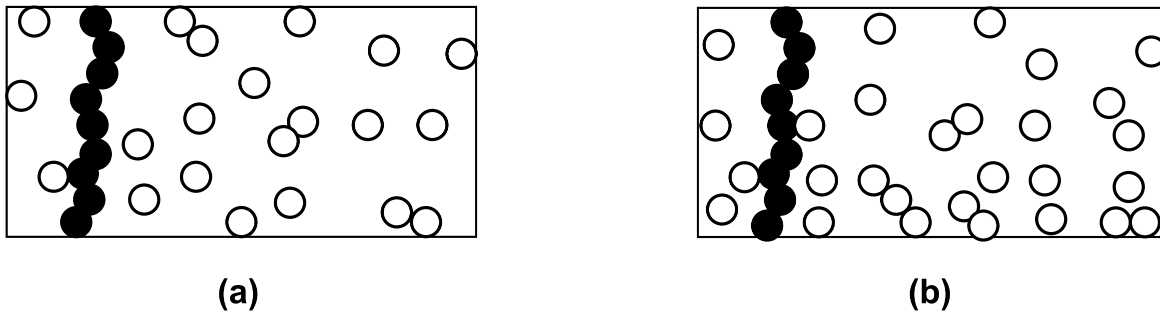

Figure 4.43. Spatial distribution of bulk traps for: (a) Uniform defect generation. (b) Exponential distribution where the trap generation probability decreases with the distance from the bottom interface. Dark fill traps are part of the percolation path.

An analytical cell-based model has been introduced that captures the observed dependence of Weibull slope on thickness in a simple equation [46]. In this cell based approach, the dependence of the Weibull slope on t_{OX} for trap size a_0 is [46]

$$\beta = m t_{OX}/a_0 \quad (4.48)$$

In this framework, the cell matrix has n = t_{OX}/a_0 rows, with SBD occurring when a straight column of bulk traps is formed between the top and bottom electrodes, as shown in Figure 4.44-a. The cell based approach has recently been extended to include non-columnar conduction, which allows the formation of percolation paths that include any nearest neighbor cell (as shown in Figure 4.44-b) and comprehends pre-existing defects, correlated defect generation, consequences of quasi-equilibrium in the reactions that generate traps, and misaligned columns (e.g. two different types of traps in the dielectric) [59]. We will apply the cell based approach to study the thickness scalability of SiON films in sub-section 4.8.4.

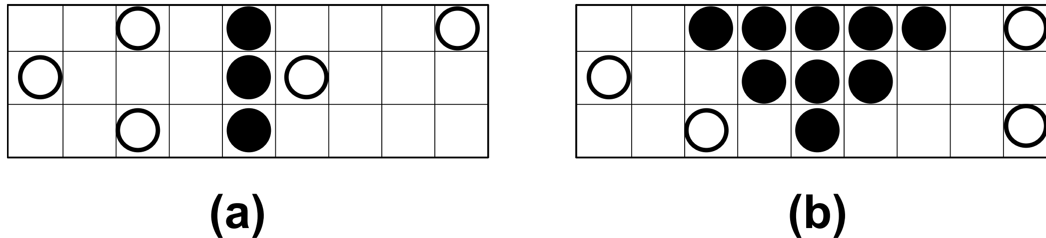

Figure 4.44. Cell-based approach for the formation (in 2 dimensions) of (a) columnar percolation path. (b) Non-columnar percolation paths. In (a), the trap in the bottom row can participate in only 1 SBD path while in (b) it can give rise to 9 different SBD paths.

4.8.2 SiO_2 gate dielectrics below 30Å

The percolation models discussed in the previous section successfully model the thickness scaling of the Weibull slope based on the build-up of bulk traps in the oxide. However, it has not been unambiguously proven in experiments since the critical bulk trap density at breakdown is tuned to C-V extracted interface trap densities [60]. Figure 4.45 compares the critical trap density at breakdown of interface states and bulk traps extracted from LV-SILC and SILC respectively for 27Å NMOS and PMOS oxides [25]. Although interface trap generation is higher [25,61], it is not clear whether breakdown is controlled by bulk or interface traps.

To solve this problem, we once again return to our experiment containing NMOS poly doping variations. Since the poly is degenerately doped, the surface potential required to invert the poly is equal to the band gap. The build-up of bulk and interface traps in 26Å NMOS oxides is shown as a function of poly band bending in Figure 4.46 [25]. The bulk trap generation rate (SILC) increases when the poly inverts and anode hole injection becomes an operative process. In contrast, the effect of AHI on interface trap generation (LV-SILC) is weak, with the possible exception being the devices with the lowest poly doping. Since the devices are stressed in the direct tunneling regime, interface trap creation through electron-hole recombination [62] would not be expected to be significant because there are no electrons in the oxide conduction band to recombine with holes. Since the presence of trapped holes by themselves can generate slow states [63], it is possible that LV-SILC is detecting these states in the lowest doped poly devices. Another possibility is that the apparent increase in interface trap density in the low doped poly devices is a result of holes trapped near the poly interface when AHI

becomes operative. While the poly interface is the anode during stress, it is the cathode during the LV-SILC sense operation, where the field may have increased slightly from the trapped holes.

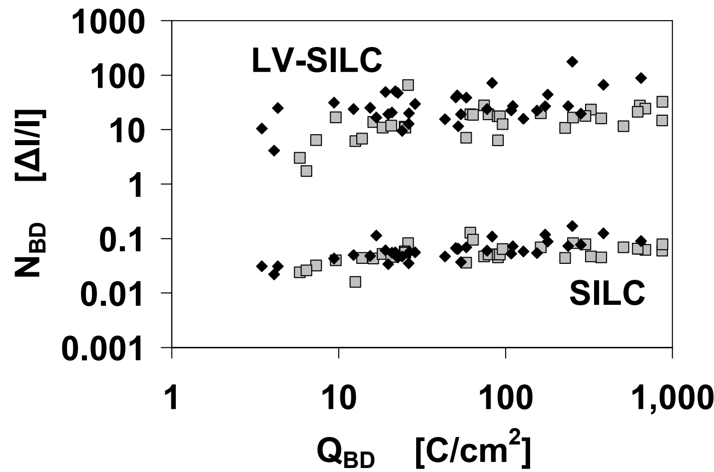

Figure 4.45. Critical density of interface states and bulk traps at breakdown in 27Å oxide films. Gray fill is NMOS, dark fill is PMOS. After Nicollian, Ref. [25]. © 2000 IEEE.

The critical trap densities at breakdown are plotted in Figure 4.47 [25]. While the interface trap generation rates are not significantly affected by AHI, the critical interface trap density at breakdown drops when AHI becomes active. In contrast, while the bulk trap generation rates are higher in the presence of anode hole injection, the critical bulk trap density at breakdown is unchanged. Therefore, Figures 4.46 and 4.47 unequivocally show that oxide breakdown is controlled by bulk traps.

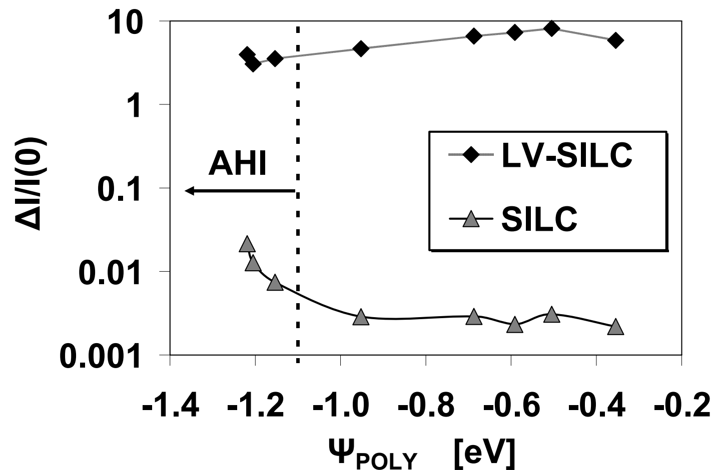

Figure 4.46. Normalized SILC increase in 26Å NMOS oxides after 100 C/cm^2 fluence at +3.6V stress. AHI increases the bulk trap generation rate while having a minimal effect on interface state creation. After Nicollian, Ref. [25]. © 2000 IEEE.

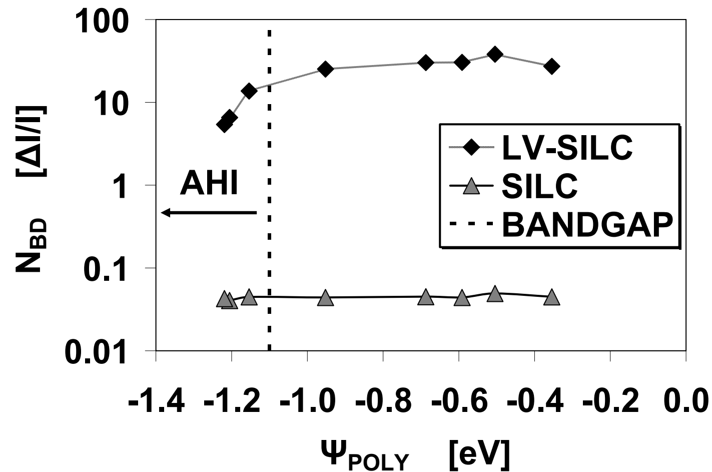

Figure 4.47. Critical density of bulk and interface traps at breakdown for the devices used in Figure 4.46. Breakdown occurs when a critical density of bulk traps is reached. After Nicollian, Ref. [25]. © 2000 IEEE.

We will now resume the discussion in sub-section 4.4.2 regarding the "pollution" of the substrate current with sources other than anode hole injection in our experiment. Photo-generation of an electron-hole pair in the substrate by photons originating in the poly anode may be a significant source of substrate current [30]. However, if photo-generation were the primary source of the substrate current, then it would not have been possible to separately generate bulk and interface traps. The increased bulk trap generation rate would have been due to the injection of the photogenerated electron back to the anode, which would have also resulted in an increased interface trap generation rate, but was not observed in our experiment.

Valence band tunneling occurs when an electron in the pwell valence band tunnels into the poly conduction band, leaving behind a free hole in the pwell that results in a substrate current. This process may also be trap assisted, as shown in Figure 4.48 for inelastic tunneling [25]. When the poly inverts and the pwell valence band edge moves above the poly Fermi Level, and additional transport path is created. In this event, tunneling from the pwell valence band to poly valence band via trap assisted tunneling through oxide bulk traps can result in an additional substrate current component. However, this process only occurs if the electron emitted from the bulk trap is at an electrostatic potential within the poly band gap when it arrives at the anode. Since this process can only deliver a maximum energy of 1.1eV to the anode, this mechanism does not explain the increase in trap generation rates that are observed when the poly inverts. Therefore, having ruled out other mechanisms, the hole currents measured in our experiments when the poly is inverted appear to be due to anode hole injection.

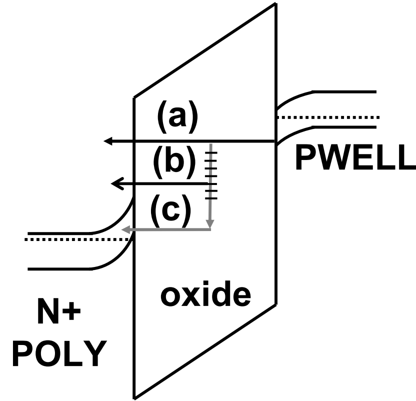

Figure 4.48. Band diagram for (a) tunneling from pwell valence band to poly conduction band. (b) Tunneling from the pwell valence band to poly conduction band via inelastic trap assisted tunneling through bulk traps. (c) Tunneling from pwell valence band to poly valence band via inelastic trap assisted tunneling through bulk traps after the poly inverts. After Nicollian, Ref. [25]. © 2000 IEEE.

4.8.3 SiON films below 20Å

In this sub-section, we will explore the roles of bulk and interface traps in the breakdown of ultra-thin SiON (PNO) dielectrics. Our approach is to link the voltage and temperature dependence of trap generation to breakdown. We will also provide additional evidence that two different hydrogen species are released during TDDB stress as discussed in Section 4.6. We will also show additional results that support a transition from single to multiple electron vibrational excitation as the mechanism that causes breakdown below the 2.5eV – 3.0eV threshold energy for vibrational excitation.

We will begin this section with a discussion on the voltage dependence of trap generation and breakdown. The voltage acceleration factor for Q_{BD} is written as:

$$AF(Q_{BD}) = -\partial \ln(Q_{BD})/\partial V_G \quad (4.49)$$

Solving equation (4.5) for $Q = Q_{BD}$, where N_{BD} is the trap density at breakdown:

$$Q_{BD} = (N_{BD}/b)^{1/m} \quad (4.50)$$

Inserting (4.50) into (4.49) and assuming that N_{BD} is independent of V_G, we get:

$$AF(Q_{BD}) = (1/m)\partial \ln b/\partial V_G \quad (4.51)$$

The trap density N(Q) in (4.51) is obtained from SILC and LV-SILC measurements. Note that it is not necessary to stress to breakdown to extract the voltage dependence of trap generation because the V_G dependence is carried solely in the trap generation pre-factor “b”. The voltage acceleration factors for Q_{BD}, SILC, and LV-SILC are plotted in Figure 4.49 [64]. For all stress conditions, the acceleration factors for SILC and

LV-SILC are sharply different. This confirms the two reaction model proposed in Figures 4.34 and 4.37 A transition from bulk trap limited to interface state limited breakdown occurs at about 2.7V, as AF(SILC) tracks AF(Q_{BD}) above 2.7V and AF(LV-SILC) tracks AF(Q_{BD}) below 2.7V [64]. This transition coincides with E_{TH} for multi-vibrational hydrogen release. As bulk traps are generated by H^+, the reduced role of H^+ relative to H^0 may be due to the higher bias and current required for hole induced multi-carrier vibrational excitation [65]. Accordingly, below 2.7V, breakdown is controlled by electron induced desorption of H^0 [64].

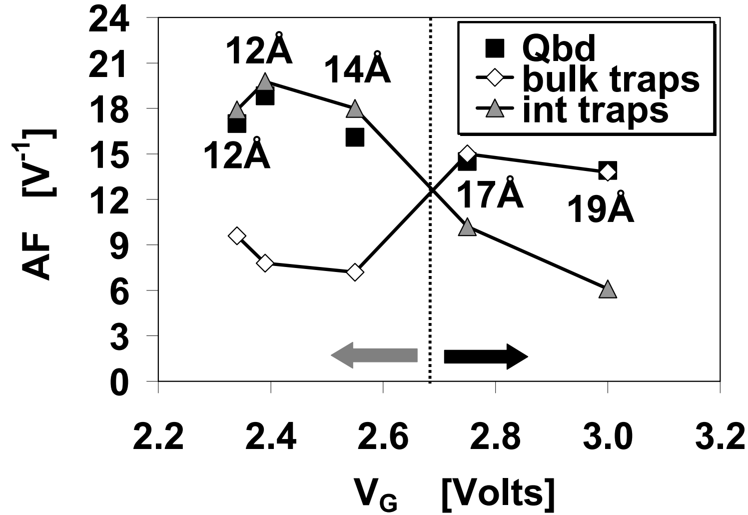

Figure 4.49. Voltage acceleration factors for Q_{BD}, SILC, and LV-SILC. A transition from bulk to interface trap controlled breakdown occurs below about 2.7V. After Nicollian, Ref. [64]. © 2006 IEEE.

We now link the temperature dependence of breakdown to trap generation. We will assume that Q_{BD}, “m”, and “b” follow Arrhenius relationships with activation energy ΔH:

$$\Delta H(N_{BD}) = -k_B*\partial \ln N_{BD}/\partial(1/T) \quad (4.52)$$

$$\Delta H(b) = -k_B*\partial \ln b/\partial(1/T) \quad (4.53)$$

$$\Delta H(Q_{BD}) = -k_B*\partial \ln Q_{BD}/\partial(1/T) \quad (4.54)$$

Inserting (4.50), (4.52), and (4.53) into (4.54), the activation energy for breakdown becomes [64]

$$\Delta H(Q_{BD}) = (1/m)*[\ \Delta H(N_{BD}) - \Delta H(b)] \quad (4.55)$$

Unlike the V_G dependence, oxides must be stressed to breakdown to tie together the temperature dependence of trap generation and breakdown. The temperature

dependence of Q_{BD} for 13Å SiON films with 10^{-7} cm^2 gate areas stressed at +2.2V is shown in Figure 4.50. The thickness, area, and stress voltage were kept the same so that only extrapolations in temperature are performed. The data fit an Arrhenius relationship between 75°C to 150°C. The temperature dependence of N_{BD} and "b" also follow an Arrhenius relationship in this temperature range as shown in Figure 4.51 [64]. For LV-SILC, the trap generation pre-factor "b" has a larger contribution to the temperature dependence than N_{BD}. The results are tabulated in Figure 4.52 [64] and confirm that interface traps are the defects that control breakdown at low V_G.

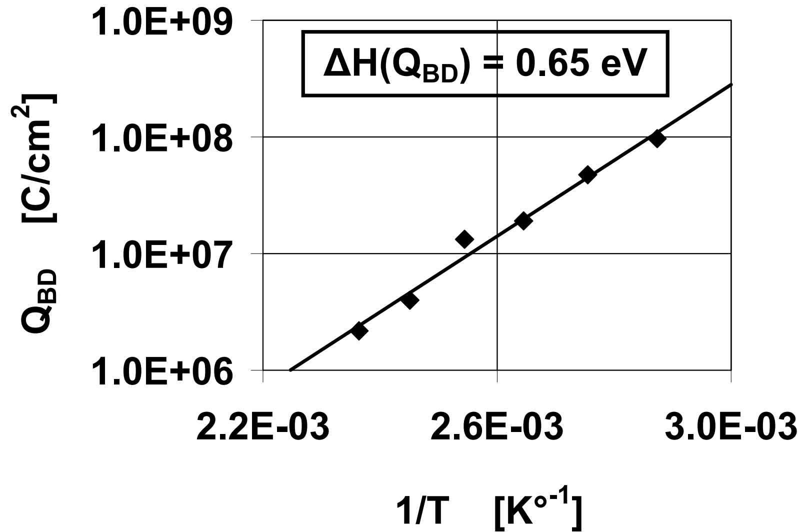

Figure 4.50. Arrhenius plot for Q_{BD}. 13Å SiON films with 10^{-7} cm^2 gate areas are stressed at +2.2V from 75°C to 150°C. After Nicollian, Ref. [64]. © 2006 IEEE.

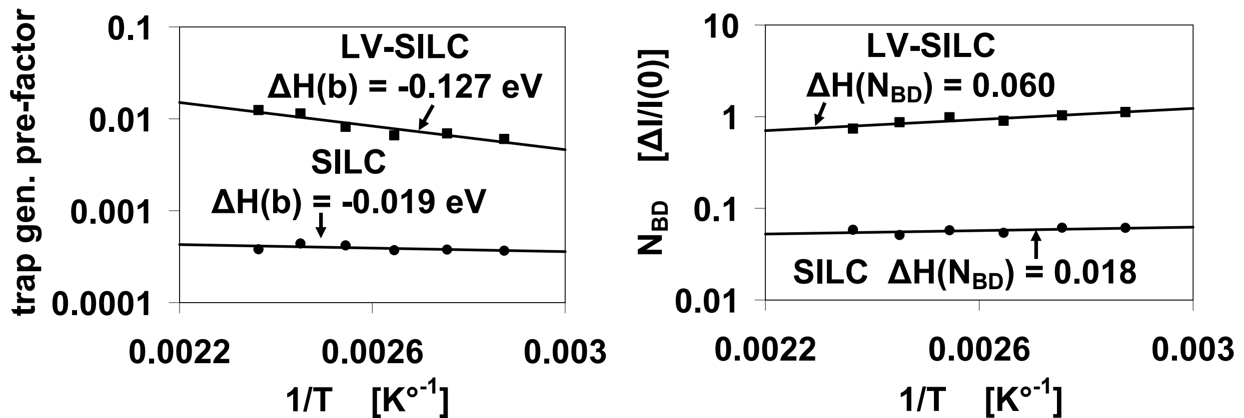

Figure 4.51. Arrhenius plot for Q_{BD}. 13Å SiON films with 10^{-7} cm^2 gate areas are stressed at +2.2V from 75°C to 150°C. (a) ΔH for "b". (b). ΔH for N_{BD}. After Nicollian, Ref. [64]. © 2006 IEEE.

ΔH directly from Q_{BD}	ΔH from SILC (bulk traps)	ΔH from LV-SILC (interface traps)
0.65 eV	0.12 eV	0.66 eV

Figure 4.52. ΔH from breakdown and trap generation measurements using equations (4.52) – (4.55). The temperature data confirm that interface traps control breakdown at low voltages. After Nicollian, Ref. [64]. © 2006 IEEE.

Our experimental values of the trap generation power law exponent 'm" are about 0.3 [42,64]. This is smaller than the 0.5 to 1.0 values that are often reported in the literature [46,56]. However, our experimentally obtained m ~ 0.3 provides a match between the voltage and temperature dependence of trap generation with breakdown [64]. It has been reported that bulk trap N_{BD} is time dependent when $t_{63\%} > 10^6$ seconds [66]. However, our mean stress times are significantly less than this. Moreover, the maximum stress times shown in [66] are on the order of 10^9 seconds, so the longest term data are extrapolations rather than actual measurements.

4.8.4 Implications for SiON thickness scaling

We will now explore the roles of bulk and interface traps in the formation of SBD percolation paths in SiON. Weibull slopes for EOT < 32Å and EOT < 22Å are shown in Figure 4.53-a and 4.53-b respectively [64]. While the t_{OX} dependence weakens below 22Å, it is still apparent down to 10Å. The continuing scalability of β with thickness implies that at least two traps are still needed to form a SBD path down to 10Å. The t_{OX} dependence of β is not unique because both the physical thickness and nitrogen profiles are modified to scale EOT. We will apply the cell-based approach [46] to analyze this problem using the simple relation between β and t_{OX} in (4.48).

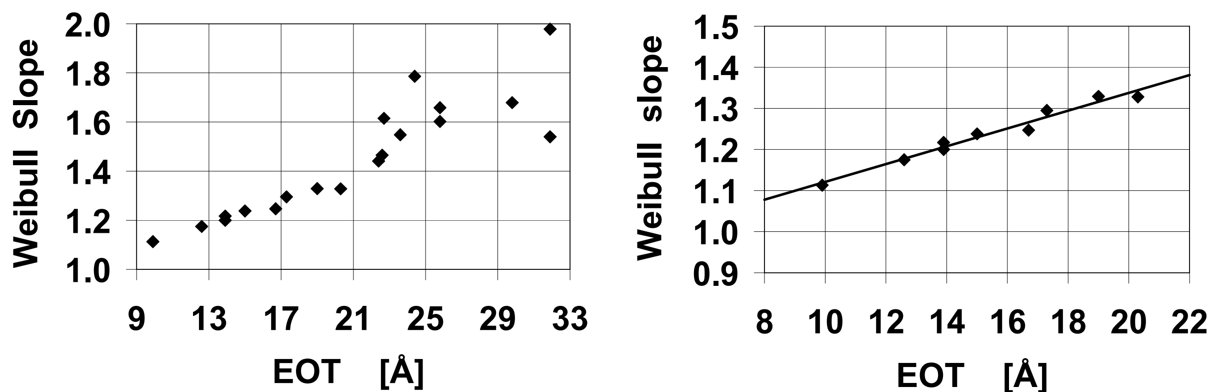

Figure 4.53. TDDB Weibull slope vs. SiON thickness. (a) EOT < 32Å. (b) Enlarged view of (a) for EOT < 22Å. After Nicollian, Ref. [64]. © 2006 IEEE.

Recalling that a_0 is the defect size in equation (4.48), a percolation path formed by 2 interface traps at the maximum oxide thickness (arbitrarily, $5a_0$ that this can occur) is shown in Figure 4.54-a. Figure 4.54-b shows the possible percolation paths involving only interface traps with $t_{OX} = 4a_0$ [64]. From (4.48) and Figure 4.54, a model involving only interface traps does not capture the observed thickness dependence of β shown in Figure 4.53 since the number of traps in the percolation path is always two. Therefore, while the generation of interface traps is the rate limiting step, at least one bulk trap must be involved to capture the correct thickness scaling trend [64]. A possible scenario involving both bulk and interface traps in the percolation path and exhibiting the correct t_{OX} scalability of β is shown in Figure 4.55 [64].

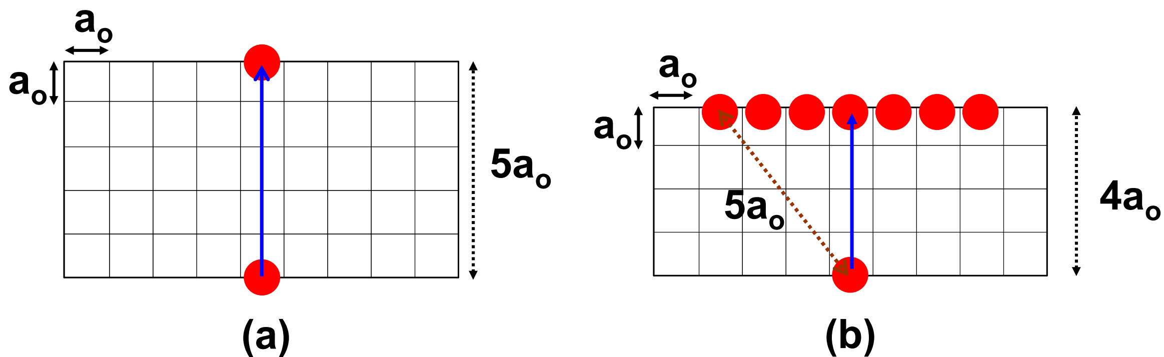

Figure 4.54. Schematic for 2-trap percolation involving only 2 interface traps for (a) $t_{OX} = 5a_0$. (b) $t_{OX} = 4a_0$. This model does NOT explain the thickness scaling of the Weibull slope shown in Figure 4.53-b. After Nicollian, Ref. [64]. © 2006 IEEE.

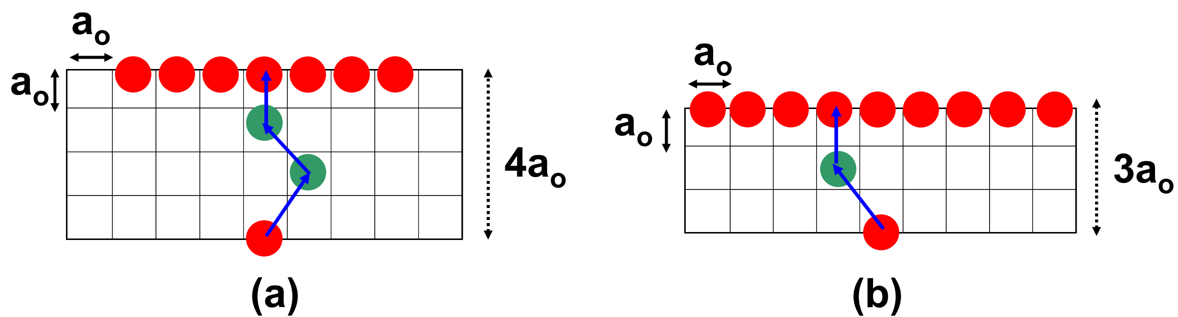

Figure 4.55. Schematic for percolation involving both interface and bulk traps for (a) $t_{OX} = 4a_0$. (b) $t_{OX} = 3a_0$. This model is consistent with t_{OX} dependence (Fig. 4.53), temperature dependence (Fig.4.52), and V_G dependence (Fig. 4.49). After Nicollian, Ref. [64]. © 2006 IEEE.

In the cell based approach, the trap size a_0 is determined by the slope of β vs. t_{OX}. Since $\beta \neq 0$ at $t_{OX} = 0$, the existence of an interfacial layer that offsets the oxide thickness has been proposed [46]. However, since a mixture of bulk and interface traps is required for breakdown, a single defect size will not be extracted from the slope. Instead, we find the instantaneous value of a_0 for each β. When plotted against thickness, a_0 asymptotically approaches its bulk value as t_{OX} increases, as shown in Figure 4.56 [64]. Using

m = 0.26 (see Figure 4.29) [42], the bulk trap diameter is about 4Å, resulting in the Weibull slope remaining thickness dependent and > 1 down to 10Å EOT.

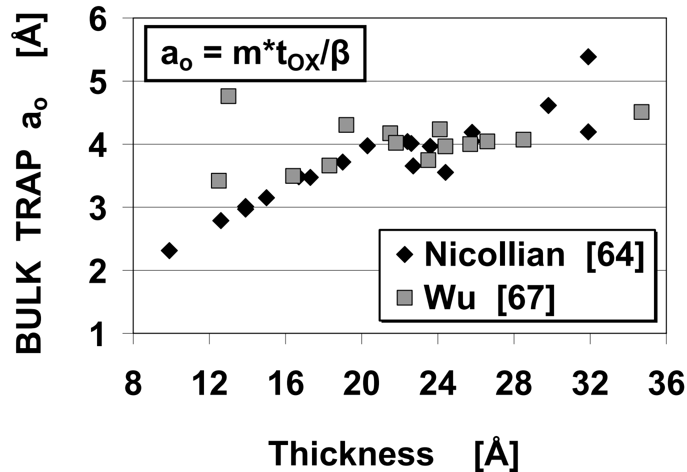

Figure 4.56. Bulk trap diameter vs. thickness using the cell based approach. For m = 0.26 [42,64], the asymptotic value of a_0 at large thickness is about 4Å. After Nicollian, Ref. [64]. *© 2006 IEEE.*

4.9 Anode hole injection below 3V

We now evaluate whether AHI is a viable mechanism for breakdown in inverted NMOS devices below the 3V threshold for this mechanism.

4.9.1 Minority ionization at low voltages

Since the n+ doped poly silicon anode is depleted under low voltage stress conditions, we examine whether there are sufficient numbers of free holes available for minority ionization. The number of holes is plotted vs. temperature in Figure 4.57. The effects of band gap narrowing due to high doping [68] are included and it is assumed that the bands are bent to the intrinsic condition ($\Psi = \Psi_B$) and the poly depletion region width is X_{DMAX}. Since the poly band bending at stress conditions is less than $E_G/2$ even for degenerate doping and since X_D is actually less than X_{DMAX}, the intrinsic concentration of holes shown in Figure 4.57 is an overestimate of the actual number. It can be seen that even for the highest doping concentration, the temperature must exceed 80°C for there to be 1 free hole in a 10μm^2 gate area that is typical of the devices that were stressed. Therefore, anode hole injection through minority ionization in inverted NMOS is improbable at low V_G.

4.9.2 Majority ionization at low voltages

We will now investigate whether AHI through majority ionization is a viable mechanism for breakdown at low gate voltage. A band diagram for AHI for NMOS stressed in inversion is shown in Figure 4.58. It can be seen that there is insufficient energy to inject

holes over the barrier. However, this by itself does not rule out AHI because inelastic scattering could result in a high energy tail. As shown in Figure 4.35, the interface traps created are primarily acceptor states. Because holes create donor states, AHI is ruled out as the mechanism for interface trap generation. Since we have shown that interface traps control breakdown in Figures 4.49 and 4.52, AHI is ruled out as the mechanism for low voltage breakdown.

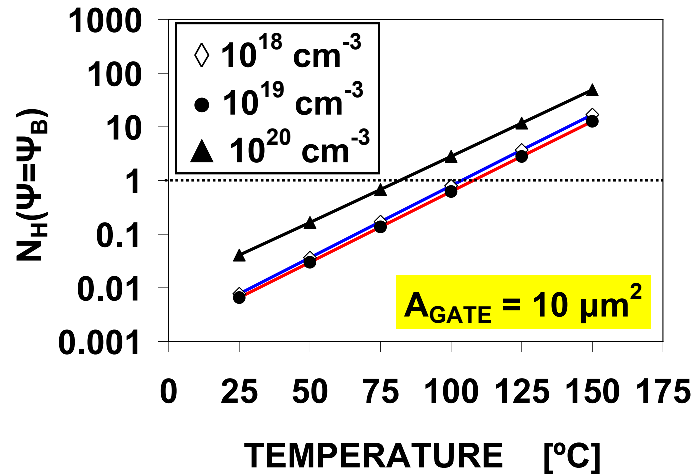

Figure 4.57. Calculated number of holes vs. temperature for inverted NMOS devices with 10µm^2 gate area and the poly doping density as a parameter. After Nicollian, Ref. [64]. © 2006 IEEE.

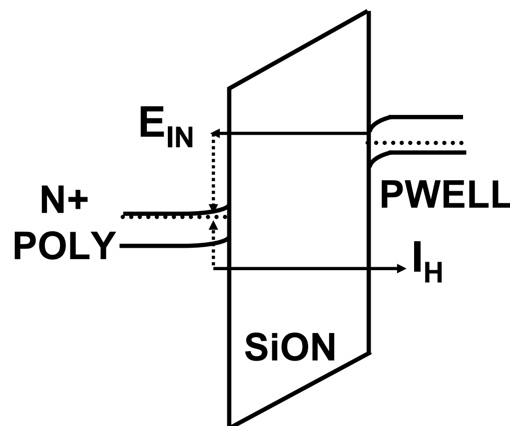

Figure 4.58. Band diagram for majority ionization in depleted NMOS poly with the device stressed in inversion.

4.10 Chapter summary

In this chapter, we discussed the three primary mechanisms that lead to trap generation and breakdown in gate oxide materials: Impact ionization in the oxide, anode hole injection, and anode hydrogen release. While there have been many controversies over

the past 30 to 40 years regarding the applicability of these processes, there is significant evidence that all three mechanisms can result in dielectric breakdown in the regime that they are operative. The dominant breakdown mechanism depends on oxide thickness, electric field, and electron energy. At least one trap generation process is always active down to 0.25eV.

Since all breakdown mechanisms are ultimately the consequence of a chemical reaction that leads to defect states in the dielectric system, we began this chapter with a discussion of reaction-diffusion theory. This framework is particularly useful since it relates the trap generation power law to the species that is the product of the reaction that results in trap creation.

In Chapter 3, we introduced the concept of steady state F-N transport at high fields in thick oxides. For this process, electron energies are stabilized by LO and acoustic phonons below 6eV. At higher energies, impact ionization in the oxide limits acoustic phonon runaway as electrons lose 9eV kinetic energy in creating electron hole pairs. The average electron energy does not exceed 6eV all the way to breakdown; only a small fraction of electrons in high energy tails exceed the 9eV required for impact ionization in the oxide. This process only occurs at oxide fields greater than 7MV/cm in oxides thicker than about 200Å to 300Å. Of all the trap generation mechanisms, impact ionization in the oxide has the highest threshold energy.

Impact ionization in the oxide can be difficult to resolve from anode hole injection, because both mechanisms introduce holes into the oxide through an impact ionization process. Accordingly, both mechanisms can lead to a 1/E dependence of the time to breakdown on electric field. Breakdown occurs after a critical fluence of holes has passed through the oxide. Holes in the oxide lead to neutral bulk traps, trapped holes, slow states, and if electrons are present in the oxide conduction band, interface traps are generated through the recombination of holes with electrons. Of these defects, all can contribute to breakdown in the appropriate regime except for slow states, because they can be completely discharged from the oxide.

Anode hole injection results from tunneling electrons impact ionizing in the anode electrode, resulting in the injection of holes into the oxide valence band. The two types of anode hole injection are majority ionization and minority ionization. For majority ionization, electron hole pairs are created through silicon band gap ionization. This process has a 6eV threshold energy. Minority ionization occurs when free holes are present in the anode, so that the silicon band gap energy does not have to be lost by the electrons initiating the impact ionization event. Accordingly, minority ionization has lower threshold energy (3.0eV to 3.5eV) than majority ionization, although minority ionization is a much lower probability event.

By using NMOS devices with different poly doping, we showed that when the doping was sufficiently low so that the poly was inverted during stress, an extra hole current appeared in the silicon cathode, accompanied by a sharp drop in the charge to

breakdown. This experiment provides experimental confirmation that anode hole injection through minority ionization is a plausible mechanism down to 3.6V.

However, in small area devices stressed in the on-state at lower voltages, our calculations show that there are insufficient numbers of holes in depleted NMOS poly for anode hole injection through minority ionization to be a significant degradation mechanism.

The two types of hydrogen desorption mechanisms are electrical excitation and vibrational excitation. Electrical excitation has a threshold energy of 5eV to 7eV and occurs through tunneling emission. Vibrational excitation is a process where silicon-hydrogen bond desorption occurs through phononic excitation. Coherent (1 electron) vibrational excitation has a 2.5eV to 3.0eV threshold energy. A two electron desorption path has a 1.25eV to 3.0eV threshold. In general, the threshold energy for vibrational excitation is E_{TH}/M, where M electrons are involved in the desorption process. Since there are about 10 silicon-hydrogen bond phonon levels, vibrational excitation can occur down to an energy of 0.25eV.

We showed that reaction-diffusion theory applies to $+V_G$ stress of ultra-thin NMOS SiON films. Measurable recovery effects are present, showing that quasi-equilibrium exists for NMOS TDDB. This allows us to identify the anode reactions that result in anode hydrogen release. By using a substrate bias to initiate a known reaction that liberates hydrogen at the cathode interface, the anode reactions leading to trap generation were probed by observing the effects of the cathode hydrogen on bulk trap creation. We showed that trap generation is triggered by the release of two hydrogen species (H^+ and H^0) in two separate anode reactions. Both species create interface traps at the poly interface when they are released. After migrating into the dielectric, H^+ subsequently creates bulk traps and H^0 creates pwell interface traps. Both holes and electrons are involved in the desorption of hydrogen at the anode. We found that the hydrogen species that controls breakdown is voltage dependent.

The interface trap recovery effects observed show that LV-SILC is not a steady state current when a back-bias is applied during stress. However, LV-SILC is a steady-state effect when no back bias is applied during PBTI stress of NMOS SiON films since no interface trap recovery effects are observed under these conditions.

We unequivocally showed that bulk traps are the defects controlling breakdown in SiO_2 films below 30Å thickness. This is accomplished by using our poly doping splits to inject holes from the anode into the oxide to preferentially increase the generation rate of bulk traps, while having a negligible effect on interface trap generation. Breakdown always occurred when the same critical density of bulk traps was attained, while the density of interface traps at breakdown was typically lower when holes were injected.

However, bulk traps are not always the defects that control breakdown in SiON films below 20Å thickness. By linking the voltage and temperature dependence of trap generation to breakdown, we discovered that a transition from bulk to interface trap

limited breakdown occurs at 2.7V in ultra thin SiON films. The interface traps controlling breakdown are acceptor-like. Accordingly, breakdown is controlled by the desorption of H^0 from the anode below 2.7V. This voltage coincides with the threshold voltage for vibrational excitation of silicon-hydrogen bonds. Other workers have shown that the trap generation efficiency during electrical stress tracks the results from STM experiments of H^0 desorption. A transition from a one electron to two electron VE process below about 2.5V has been deduced from Q_{BD} data.

Accordingly, since holes create donor states, anode hole injection through either majority ionization or minority ionization is further eliminated as the mechanism for NMOS SBD in SiON films below 2.7V when the rate limiting step for breakdown is the generation of acceptor-like interface traps. Therefore, hydrogen plays a central role in dielectric breakdown at low stress voltages.

A power law dependence of the time to breakdown on gate voltage only occurs when the hydrogen desorption mechanism is vibrational excitation. The TDDB power law exponent $N > 40$ can arise from the voltage dependence of the fraction of tunneling electrons that excite a silicon-hydrogen phonon mode. The experiments summarized in the previous paragraphs confirm that the origin of the TDDB power law model is indeed vibrational excitation of silicon-hydrogen bonds.

While the generation of interface states becomes the rate limiting step below the threshold energy for vibrational excitation, both bulk and interface traps are still required for breakdown to occur. A minimum of two traps is needed to cause breakdown down to 10Å EOT. At least one trap must be an interface state and at least one trap must be a bulk state.

While there are some discrepancies in the literature regarding bulk trap size, our results lead to a bulk trap diameter of about 4Å. This is the smallest estimate bulk trap size reported to date, and results in the Weibull slope remaining thickness dependent and greater than 1 down to 10Å EOT (for $t_{PHYSICAL} \geq 12$Å). Our extracted bulk trap size of 4Å is in part consequent of our experimentally obtained trap generation power law exponent "m" being about 0.3. This is smaller than the values that are typically quoted in the literature, and provides a consistent explanation for the voltage, temperature, and thickness dependence of breakdown.

4.11 References

[1] K.O. Jeppson and C.M. Svensson, "Negative Bias Stress of MOS Devices at High Electric Fields and Degradation of MNOS Devices", *Journal of Applied Physics*, **48**, 2004 (1977)
[2] C.E. Blat, E.H. Nicollian, and E.H. Poindexter, "Mechanism of Negative-Bias-Temperature Instability", *Journal of Applied Physics*, **69**, 1712 (1991)
[3] S. Ogawa and N. Shiono, "Generalized Diffusion-Reaction Model for the Low-Field Charge-Buildup Instability at the Si-SiO_2 Interface", *Physical Review B*, **51**, 4218 (1995)

[4] M.A. Alam, "A Critical Examination of the Mechanisms of Dynamic NBTI for PMOSFETs," *IEDM Technical Digest*, 345 (2003)
[5] S. Chakravarthi, A.T. Krishnan, V. Reddy, C.F. Machala, and S. Krishnan, "A Comprehensive Framework for Predictive Modeling of Negative Bias Temperature Instability", *Proceedings of the IRPS*, 273 (2004)
[6] M.A. Alam, "NBTI Modeling", *IRPS Tutorials*, 211 (2005)
[7] A.T. Krishnan, et. al., "Material Dependence of Hydrogen Diffusion: Implications for NBTI Degradation", *IEDM Technical Digest*, 705 (2005)
[8] A.T. Krishnan, S. Chakravarthi, P.E. Nicollian, V. Reddy, and S. Krishnan, "Negative Bias Temperature Instability Mechanism: The role of Molecular Hydrogen", *Applied Physics Letters*, **88**, 153518 (2006)
[9] A.S. Grove, *Physics and Technology of Semiconductor Devices*, New York: Wiley, 1967, pp. 35-88
[10] D. Arnold, E. Cartier, and D.J. DiMaria, "Acoustic-Phonon Runaway and Impact Ionization by Hot Electrons in Silicon Dioxide", *Physical Review B*, **45**, 1477 (1992)
[11] S.D. Brorson, D.J. DiMaria, M.V. Fischetti, F.L. Pesavento, P.M. Solomon, and D.W. Dong, "Direct Measurement of the Energy Distribution of Hot Electrons in Silicon Dioxide", *Journal of Applied Physics*, **58**, 1302 (1985)
[12] D.J. DiMaria, D. Arnold, and E. Cartier, "Impact Ionization and Positive Charge Formation in Silicon Dioxide Films on Silicon", *Applied Physics Letters*, **60**, 2118 (1992)
[13] D. Arnold, E. Cartier, and D.J. DiMaria, "Theory of High-Field Electron Transport and Impact Ionization in Silicon Dioxide", *Physical Review B*, **49**, 10278 (1994)
[14] D.J. DiMaria, E. Cartier, and D. Arnold, "Impact Ionization, Trap Creation, Degradation, and Breakdown in Silicon Dioxide Films on Silicon", *Journal of Applied Physics*, **73**, 3367 (1993)
[15] E. Harari, "Dielectric Breakdown in Electrically Stressed Thin Films of Thermal SiO_2," *Journal of Applied Physics*, **49**, 2478 (1978)
[16] I.C. Chen, S. Holland, K.K. Young, C. Chang, and C. Hu, "Substrate Hole Current and Oxide Breakdown", *Applied Physics Letters*, **49**, 669 (1986)
[17] D.J. DiMaria, T.N. Theis, J.R. Kirtley, F.L. Pesavento, and D.W. Dong, "Electron Heating in Silicon Dioxide and Off-Stoichiometric Silicon Dioxide Films", *Journal of Applied Physics*, **57**, 1214 (1985)
[18] D.J. DiMaria, D. Arnold, and E. Cartier, "Degradation and Breakdown of Silicon Dioxide Films on Silicon", *Applied Physics Letters*, **61**, 2329 (1992)
[19] Y. Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, "High-Field and Current-Induced Positive Charge in Thermal SiO_2 Layers", *Journal of Applied Physics*, **57**, 2830 (1985)
[20] Z.A. Weinberg, M.V. Fischetti, and Y. Nissan-Cohen, "SiO_2 Induced Substrate Current and its Relation to Positive Charge in Field Effect Transistors", *Journal of Applied Physics*, **59**, 824 (1986)
[21] K.F. Schuegraf and C. Hu, "Hole Injection SiO_2 Breakdown Model for Very Low Voltage Lifetime Extrapolation", *IEEE Transactions on Electron Devices*, **41**, 761 (1994)
[22] D.J. DiMaria, "Hole Trapping, Substrate Currents, and Breakdown in Thin Silicon Dioxide Films", *IEEE Electron Device Letters*, **16**, 184 (1995)
[23] J.D. Bude, B.E. Weir, and P.J. Silverman, "Explanation of Stress-Induced Damage in Thin Oxides", *IEDM Technical Digest*, 179 (1998)

[24] M.A. Alam, J. Bude, and A. Ghetti, “Field Acceleration for Oxide Breakdown – Can an Accurate Anode Hole Injection Model Resolve the E vs. 1/E Controversy?”, *Proceedings of the IRPS*, 21 (2000)
[25] P.E. Nicollian, W.R. Hunter, and J.C. Hu, “Experimental Evidence for Voltage Driven Breakdown Models in Ultrathin Gate Oxides”, *Proceedings of the IRPS*, 7 (2000)
[26] M.A. Alam, *Private Communication* (2002)
[27] K. Kobayashi, A. Teramoto, and M. Hirayama, “Model for the Substrate Hole Current Based on Thermionic Hole Emission from the Anode During Fowler-Nordheim Electron Tunneling in N-Channel Metal-Oxide-Semiconductor Field-Effect Transistors”, *Journal of Applied Physics*, **77**, 3277 (1995)
[28] D.J. DiMaria, E. Cartier, and D.A. Buchanan, “Anode Hole Injection and Trapping in Silicon Dioxide”, *Journal of Applied Physics*, **80**, 304 (1996)
[29] X. Gao and S. Yee, “Threshold Energies of High-Field-Induced Hole Currents and Positive Charges in SiO_2 Layers of Metal-Oxide-Semiconductor Structures”, *Journal of Applied Physics*, **76**, 5795 (1994)
[30] M. Rasras, I. DeWolf, G. Groeseneken, B. Kaczer, R. Degraeve, and H.E. Maes, “Photo-Carrier Generation as the Origin of Fowler-Nordheim-Induced Substrate Hole Currents in Thin Oxides”, *IEDM Technical Digest*, 465 (1999)
[31] D.J. DiMaria and J.W. Stasiak, “Trap Creation in Silicon Dioxide Produced by Hot Electrons”, *Journal of Applied Physics*, **65**, 2342 (1989)
[32] T.C. Shen, et. al., ”Atomic Scale Desorption through Electronic and Vibrational Excitation Mechanisms”, *Science*, **268**, 1590 (1995)
[33] Y. Nissan-Cohen and T. Gorczyca, “The Effect of Hydrogen on Trap Generation, Positive Charge Trapping, and Time-Dependent Dielectric Breakdown of Gate Oxides”, *IEEE Electron Device Letters*, **9**, 287 (1988).
[34] D.J. DiMaria and E. Cartier, “Mechanism for Stress Induced Leakage Currents in Thin Silicon Dioxide Films”, *Journal of Applied Physics*, **78**, 3883 (1995)
[35] D. A. Buchanan, A.D. Marwick, and D.J. DiMaria, “Hot-Electron-Induced Hydrogen Redistribution and Defect Generation in Metal-Oxide-Semiconductor Capacitors”, *Journal of Applied Physics*, **76**, 3595 (1994)
[36] C.T. Sah, et. al., “Deactivation of the Boron Acceptor in Silicon by Hydrogen”, *Applied Physics Letters*, **43**, 204 (1983)
[37] Ph. Avouris, et. al., “STM-Induced H Atom Desorption from Si(100): Isotope Effects and Site Selectivity”, *Chemical Physics Letters*, **257**, 148 (1996)
[38] J.W. Lyding, K. Hess, and I.C. Kizilyalli, “Reduction of Hot Electron Degradation in Metal Oxide Semiconductor Transistors by Deuterium Processing”, *Applied Physics Letters*, **68**, 2526 (1996)
[39] J. Wu, et. al., “Anode Hole Injection versus Hydrogen Release: The Mechanism for Gate Oxide Breakdown”, *Proceedings of the IRPS*, 27 (2000)
[40] W.F. Clark, E. Cartier, and E.Y. Wu, “Hot Carrier Lifetime and Dielectric Breakdown in MOSFETs Processed with Deuterium”, *Proceedings of the P2ID*, 80, (2001)
[41] Y. Mitani, H. Satake, and A. Toriumi, “Experimental Evidence of Hydrogen-Related SILC Generation in Thin Gate Oxide”, *IEDM Technical Digest*, 129 (2001)
[42] P.E. Nicollian, A.T. Krishnan, C. Bowen, S. Chakravarthi, C.A. Chancellor, and R.B. Khamankar, “The Roles of Hydrogen and Holes in Trap Generation and Breakdown in Ultra-thin SiON Dielectrics”, *IEDM Technical Digest*, 403 (2005)

[43] V.K. Reddy, et. al., “Impact of Negative Bias Temperature Instability on Digital Circuit Reliability”, *Proceedings of the IRPS*, 248 (2002)
[44] S. Fujieda, Y. Miura, M. Saitoh, E. Hasegawa, S. Koyama, and K. Ando, “Interface Defects Responsible for Negative-Bias Temperature Instability in Plasma-Nitrided SiON/Si(100) Systems”, *Applied Physics Letters*, **82**, 3677 (2003)
[45] S. Pantelides, S.N. Rashkeev, R. Buczko, D.M. Fleetwood, and R.D. Schrimpf, ”Reactions of Hydrogen with Si-SiO_2 interfaces”, *IEEE Transactions on Nuclear Science*, **47**, 2262 (2000)
[46] J. Suñé, “New Physics Based Analytic Approach to the Thin Oxide Breakdown Statistics”, *IEEE Electron Device Letters*, **22**, 296 (2001)
[47] S. Lombardo, F. Crupi, and J.H. Stathis, “Softening of Breakdown in Ultra-thin Gate Oxide nMOSFETs at Low Inversion Layer Density”, *Proceedings of the IRPS*, 163 (2001)
[48] Y. Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, “Trap Generation and Occupation Dynamics in SiO_2 under Charge Injection Stress”, *Journal of Applied Physics*, **60**, 2024 (1986)
[49] G.P. Salam, M. Persson, and R.E. Palmer, “Possibility of Coherent Multiple Excitation in Atom Transfer with a Scanning Tunneling Microscope”, *Physical Review B*, **49**, 10655 (1994)
[50] R.E. Walkup, D.M. Newns, and Ph. Avouris, “Role of Multiple Inelastic Transitions in Atom Transfer with the Scanning Tunneling Microscope”, *Physical Review B*, **48**, 1858 (1993)
[51] W. McMahon, A. Haggag, and K. Hess, “Reliability Scaling Issues for Nanoscale Devices”, *IEEE Transactions on Nanotechnology*, **2**, 33 (2003)
[52] B.C. Stipe, M.A. Rezaei, and W. Ho, “Single Molecule Dissociation by Tunneling Electrons”, *Physical Review Letters*, **78**, 4410 (1997)
[53] D.J. DiMaria, “Defect Generation under Substrate Hot Electron Injection into Ultra thin Silicon Dioxide Layers”, *Journal of Applied Physics*, **86**, 2100 (1999)
[54] G. Ribes, S. Bruyere, M. Denais, D. Roy, and G. Ghibaudo, ”MVHR (Multi-Vibrational Hydrogen Release): Consistency with Bias Temperature Instability and Dielectric Breakdown”, *Proceedings of the IRPS*, 377 (2005)
[55] J. Suñé and E.Y. Wu, “Mechanisms of Hydrogen Release in the Breakdown of SiO_2 Based Gate Oxides”, *IEDM Technical Digest*, 399 (2005)
[56] R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, and H.E. Maes, “A Consistent Model for the Thickness Dependence of Intrinsic Breakdown in Ultra-thin Oxides”, *IEDM Technical Digest*, 863 (1995)
[57] J.H. Stathis, “Percolation Models for Gate Oxide Breakdown”, *Journal of Applied Physics*, **86**, 5757 (1999)
[58] P.E. Nicollian, et. al., ”Extending the reliability Scaling Limit of SiO_2 through Plasma Nitridation”, *IEDM Technical Digest*, 545 (2000)
[59] A.T. Krishnan and P.E. Nicollian, “Analytic Extension of the Cell-Based Oxide Breakdown Model to Full Percolation and its Implications” *Proceedings of the IRPS*, 232 (2007)
[60] J.H. Stathis and D.J. DiMaria, “Reliability Projection for Ultra thin Oxides at Low Voltage”, *IEDM Technical Digest*, 167 (1998)

[61] P.E. Nicollian, M. Rodder, D.T. Grider, P. Chen, R.M. Wallace, and S.V. Hattangady, “Low Voltage Stress Induced Leakage Current in Ultra thin Gate Oxides”, *Proceedings of the IRPS*, 400 (1999)
[62] S.K. Lai, “Interface Trap Generation in Silicon Dioxide when Electrons are Captured by Trapped Holes”, *Journal of Applied Physics*, **54**, 2540 (1983)
[63] D.J. DiMaria, D.A. Buchanan, J.H. Stathis, and R.E. Stahlbush, “Interface States Induced by the Presence of Trapped Holes near the Silicon-Silicon Dioxide Interface”, *Journal of Applied Physics*, **77**, 2032 (1995)
[64] P.E. Nicollian, A.T. Krishnan, C.A. Chancellor, and R.B. Khamankar, “The Traps that cause Breakdown in Deeply Scaled SiON Dielectrics”, *IEDM Technical Digest*, 743 (2006)
[65] K. Stokbro, et. al., “STM Induced Hydrogen Desorption via a Hole Resonance”, *Physical Review Letters*, **80**, 2618 (1998)
[66] J.H. Stathis, “Physical and Predictive Models of Ultra Thin Oxide Reliability in CMOS Devices and Circuits”, *Proceedings of the IRPS*, 132 (2001)
[67] E.Y. Wu, E. Nowak, and W. Lai, “Off-State Mode TDDB Reliability for Ultra-thin Gate Oxides: New Methodology and the Impact of Oxide Thickness Scaling”, *Proceedings of the IRPS*, 84 (2004)
[68] S.M. Sze, *Physics of Semiconductor Devices*, 2ND edition, New York: Wiley, 1981, pp. 143 - 144

CHAPTER 5

Closing remarks

The core of dielectric reliability physics is the carrier energy. It is driven from the transport properties and determines the trap generation and degradation mechanisms that will be operative. This in turn drives the stress generated defects that will control breakdown, as well as which TDDB lifetime model the system will follow. To a certain extent, some of the controversies in the dielectric reliability physics literature over the past 50 years may have been due to a lack of appreciation of the symbiotic relationships involving energy, transport, defects, and models. Due to the rapid pace of technology scaling over the past 10 to 20 years, a broader perspective has arisen due to the availability of data over a widened range of processing and stress conditions.

In particular, the role of the operative trap generation mechanism in driving which stress induced defects will control breakdown is a concept that we have brought forth in this work. One of the reasons why this connection was not fully recognized in the past may be the manner in which traps and trapped charge are characterized. In ultra-thin oxides, we have found stress induced leakage current (SILC) measurements to be a power tool in the analysis of device degradation resulting from stress. While SILC techniques are widely applied, to this day, their value is not fully appreciated.

In our research, the lion's share of our breakthroughs have resulted from our extensive use of both SILC, which senses the build-up of bulk traps, along with LV-SILC, which is our discovery on sensing the generation of interface traps. Some of the insights that these techniques have provided us with are the regimes where anode hydrogen release and anode hole injection are operative trap creation mechanisms, the hydrogen species responsible for trap generation, the conditions where bulk vs. interface traps control breakdown, and implications for scaling.

One of the outcomes of our research, in addition to the findings of other groups is that multi-electron hydrogen excitation is a viable trap generation mechanism in SiON gate dielectric materials. Because this process has a threshold energy of only 0.25eV, the physics of trap generation and breakdown will remain an active field for the foreseeable future. This will be true whether the industry standard for failure is the 1ST breakdown event or a post-breakdown specification.

LIST OF PUBLICATIONS

1. P.E. Nicollian, A.T. Krishnan, C.A. Chancellor, R.B. Khamankar, S. Chakravarthi, C. Bowen, and V.K. Reddy, “The Current Understanding of the Trap Generation Mechanisms that lead to the Power Law Model for Gate Dielectric Breakdown”, in *Proceedings of the IRPS*, 2007, pp. 197-208
2. A.T. Krishnan and P.E. Nicollian, “Analytic Extension of The Cell Based Oxide Breakdown Model to Full Percolation and its Implications”, in *Proceedings of the IRPS*, 2007, pp. 232-239
3. P.E. Nicollian, A.T. Krishnan, C.A. Chancellor, and R.B. Khamankar, “The Traps that cause Breakdown in Deeply Scaled SiON Dielectrics”, in *IEDM Technical Digest*, 2006, pp. 743-746
4. P.E. Nicollian, “SiON Gate Dielectric Reliability”, *IRW*, 2006
5. A.T. Krishnan, S. Chakravarthi, V. Reddy, P.E. Nicollian, S. Krishnan, “Negative Bias Temperature Instability Mechanism: The Role of Molecular Hydrogen”, *Applied Physics Letters*, Volume 88, 2006. pp. 1535518 1-3
6. P.E. Nicollian, A.T. Krishnan, C. Bowen, S. Chakravarthi, C. Chancellor, R. Khamankar, “The Roles of Hydrogen and Holes in Trap Generation and Breakdown in Ultra-thin SiON Dielectrics”, in *IEDM Technical Digest*, 2005, pp. 403-406
7. A.T. Krishnan, C. Chancellor, S. Chakravarthi, P.E. Nicollian, V. Reddy, A. Varghese, R.B. Khamankar, S. Krishnan, “Material Dependence of Hydrogen Diffusion and its Implications for NBTI Degradation”, in *IEDM Technical Digest*, 2005, pp. 705-708
8. R. Khamankar, H. Bu, C. Bowen, S. Chakravarthi, P. Chidambaram, M. Bevan, A. Krishnan, B. Smith, J. Blatchford, B. Hornung, J. Lu, P. Nicollian, B. Kirkpatrick, D. Miles, M. Hewson, D. Farber, L. Hall, H. AlShareef, A. Varghese, A. Gurba, V. Ukraintsev, B. Rathsack, J. Deloach, J. Tran, C. Kaneshige, M. Somervell, S. Aur, C. Machala, T. Grider, “An Enhanced 90nm High Performance Technology with Strong Performance Improvements from Stress and Mobility Increase through Simple Process Changes”, in *VLSI Symposium Technical Digest*, 2004, pp. 162-163
9. B. Hornung, R. Khamankar, H. Niimi, M. Goodwin, L. Robertson, D. Miles, B. Kirkpatrick, H. AlShareef, A. Varghese, M. Bevan, P.E. Nicollian, P.R. Chidambaram, S. Chavravarthi, A. Gurba, X. Zhang, J. Blatchford, B. Smith, J.P. Lu, J. Deloach, B. Rathsack, C. Bowen, G. Thakar, C. Machala, T. Grider, “A High Performance 90nm Logic Technology with a 37nm Gate Length, Dual Plasma Nitrided Gate Dielectric and Differential Offset Spacer”, in *VLSI Symposium Technical Digest*, 2003, pp. 85-86
10. P.E. Nicollian, “Dielectric Reliability Models and Mechanisms”, in *IRPS Tutorials*, 2003
11. A. Shanware, M.R. Visokay, J.J. Chambers, A.L.P. Rotondaro, H. Bu, M.J. Bevan, R. Khamankar, S. Aur, P.E. Nicollian, J. McPherson, L. Colombo, “Evaluation of the Positive Biased Temperature Stress Stability in HfSiON Gate Dielectrics”, in *Proceedings of the IRPS*, 2003, pp. 208-213

12. A.T. Krishnan, S. Krishnan, P.E. Nicollian, "Impact of Gate Area on Plasma Charging Damage: The Reverse Antenna Effect", in *IEDM Technical Digest*, 2002, pp. 525-528
13. M. Mehrotra, J. Wu, A. Jain, T. Laaksonen, K. Kim, W. Bather, R. Koshy, J. Chen, J. Jacobs, V. Ukraintsev, L. Olsen, J. Deloach, J. Mehigan, R. Agarwal, S. Walsh, D. Sekel, L. Tsung, M. Vaidyanathan, B. Trentman, K. Liu, S. Aur, R. Khamankar, P. Nicollian, Q. Jiang, Y. Xu, B. Campbell, P. Tiner, R. Wise, D. Scott, M Rodder, "60 nm Gate Length Dual Dual-Vt CMOS for High Performance Applications", in *VLSI Symposium Technical Digest*, 2002, pp. 124-125
14. P.E. Nicollian, "Dielectric Reliability Scaling Trends: Models and Mechanisms", in *IRPS Tutorials*, 2002
15. P.E. Nicollian, G.C. Baldwin, K.N. Eason, D.T. Grider, S.V. Hattangady, J.C. Hu, W.R. Hunter, M. Rodder, A.L.P. Rotondaro, "Extending the Reliability Scaling Limit of SiO_2 through Plasma Nitridation", in *IEDM Technical Digest*, 2000, pp. 545-548
16. P.E. Nicollian, W.R. Hunter, J.C. Hu, "Experimental Evidence for Voltage Driven Breakdown Models in Ultrathin Gate Oxides", *ESREF* 2000
17. P.E. Nicollian, W.R. Hunter, J.C. Hu, "Experimental Evidence for Voltage Driven Breakdown Models in Ultrathin Gate Oxides", in *Proceedings of the IRPS*, 2000, pp. 7-15 **(BEST PAPER AWARD)**
18. P.E. Nicollian, M. Rodder. D.T. Grider, P. Chen, R.W. Wallace, S.V. Hattangady, "Low Voltage Stress-Induced-Leakage-Current in Ultrathin Gate Oxides", in *Proceedings of the IRPS*, 1999, pp. 400-404
19. Y. Hu, D. Anderson, A. Rotondaro, S. Obrien, W. Hsu, R. Kraft, P. Tiner, P.E. Nicollian, S. Aur, "W/TiN/Poly Stack Gate Technology for High Density and Embedded DRAM Technology", in *International Symposium on VLSI Technology, Systems, and Applications*, 1999, pp. 247-250
20. M. Rodder, S. Hattangady, N. Yu, W. Shiau, P.E. Nicollian, T. Laaksonen, C.P. Chao, M. Mehrota, C. Lee, S. Murtaza, S. Aur, "A 1.2V, 0.1μm Gate Length CMOS Technology: Design and Process Issues", in *IEDM Technical Digest*, 1998, pp. 623-626
21. S.V. Hattangady, D.T. Grider, R. Kraft, W-T. Shiau, M. Douglas, P.E. Nicollian, M. Rodder, A. Chatterjee, J. Hu, S. Aur, H-L. Tsai, R.H. Eklund, I-C. Chen, M.F. Pas, "Remote Plasma Nitrided Oxides for Ultrathin Gate Dielectric Applications, in *Proceedings of the SPIE*, 1998, pp. 30-40
22. T.C. Holloway, G.A. Dixit, D.T. Grider, S.P. Ashburn, R. Agarwal, A. Shih, X. Zhang, G. Misium, A.L. Esquivel, M, Jain, S. Madan, T. Breedijk, A. Singh, G. Thakar, G. Shinn, B. Riemenschneider, S. O'Brien, D. Frystak, J. Kittl, A. Amerasekera, S. Aur, P.E. Nicollian, D. Aldrich, B. Eklund, "0.18 μm CMOS Technology for High-Performance, Low-Power, and RF Applications", in *VLSI Symposium Technology Digest*, 1997, pp. 13-14
23. D.T. Grider, S.V. Hattangady, R. Kraft, P.E. Nicollian, J. Kuehne, G. Brown, S. Aur, R.H. Eklund, M.F. Pas, W.R. Hunter, M. Douglas, "A 0.18 μm CMOS Process Using Nitrogen Profile Engineered Gate Dielectrics", in *VLSI Symposium Technology Digest*, 1997, pp. 47-48
24. S. Murtaza, A. Chatterjee, P. Mei, A. Amerasekera, P.E. Nicollian, J. Kittl, T. Breedijk, M. Hanratty, S. Nag, I. Ali, D. Rogers, I.-C. Chen, "A Shallow Trench

Isolation Study for 0.18 μm CMOS technology with Emphasis on the Effects of Well Design, Channel-Stop Implants, Trench Depth, and Salicide Process", in *VLSI Symposium Technical Digest*, 1997, pp. 133-137

25. S.V. Hattangady, R. Kraft, D.T. Grider, G.A. Brown, P.A. Tiner, J.W. Kuehne, P.E. Nicollian, M.F. Pas, "Ultrathin Nitrogen Profile Engineered Gate Dielectric Films", in *IEDM Technical Digest*, 1996, pp. 495-497
26. P.E. Nicollian, W.R. Hunter, "Model for the Leakage Instability in Unprogrammed Amorphous Silicon Antifuse Devices", in *Proceedings of the IRPS*, 1995, pp. 42-47
27. P.K. Chatterjee, W.R. Hunter, A. Amerasekera, S. Aur, C. Duvvury, P.E. Nicollian, L.M. Ting, P. Yang, "Trends for Deep Submicron VLSI and their Implications for Reliability", in *Proceedings of the IRPS*, 1995, pp. 1-11
28. G.V. Thakar, S.K. Madan, C.M. Garza, W.L. Krisa, P.E. Nicollian, J.L. Wise, C.K. Lee, J.A. Mckee, A. T. Appel, A.L. Esquivel, V.M. McNeil, D.A. Prinslow, B.R. Riemenschneider, T. Utsumi, R.H. Eklund, R.A. Chapman, "High Performance 0.3 μm CMOS using I-Line Lithography and Barc", in *VLSI Symposium Technology Digest*, 1995, pp. 75-76

Paul E. Nicollian

Paul E. Nicollian was born on September 1ST, 1961 in Summit, New Jersey (USA). He received the B.S. degree in Physics from The Pennsylvania State University in 1983 and the M.S. degree in Physics from The University of Texas at Dallas in 1990. He was employed by United Technologies Mostek Corporation in 1984. He joined Texas Instruments in 1985, and was elected Senior Member of the Technical Staff in 2001. He currently works in the Silicon Technology Development center, where his research is focused on the reliability physics of ultra thin gate dielectrics. He has authored or co-authored 28 publications, holds 2 US patents, and is a recipient of the 2000 IRPS Best Paper Award. He served on the IRPS Technical Program Committee from 2000 - 2005, co-chairing Device Dielectrics in 2001 and Process Induced Damage in 2000. He also served on the 2004 – 2005 IEDM CMOS and Interconnect Reliability Committee. Mr. Nicollian is a Senior Member of the IEEE.

Propositions

Paul E. Nicollian

Placing lawyers in charge of US health care reform is akin to asking a pyromaniac to put out the fire that she started.

Someday, China will outsource low paying technology jobs to the United States.

At the current rate of progress, high-k gate dielectrics will go into high volume production on the day before Christ reappears.

The world needs another Star Trek TV series, because Hollywood producers seem to be more effective than semiconductor industry marketing executives in determining the next hot new electronic gadget.

The pace of progress quickens in a complex scientific field once the most distinguished researchers assume unyielding partisan positions on the subject.

A mentor is measured by the number of his protégés that eclipse his achievements.

The answer to the question on whether breakdown is due to holes or hydrogen is that both participate.

Two hydrogen species are involved in the processes resulting in trap generation.

The dominant off-state mechanism for stress induced leakage current in ultra-thin dielectrics is tunneling via interface traps.

Both bulk and interface traps are required for breakdown to occur at low voltages, but the generation of interface traps is the rate limiting step.