Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

NJIT
New Jersey's Science & Technology University

The Van Houten library has removed some of the personal information and all signatures from the approval page and biographical sketches of theses and dissertations in order to protect the identity of NJIT graduates and faculty.

ABSTRACT

RELAXATION AND THE NATURE OF ELECTRICAL STRESS RELATED DEFECTS IN ULTRA-THIN DIOXIDE ON SILICON

by
Lihui Xie

The instability of defects created in ultra-thin insulator, metal-oxide-silicon devices biased in the direct tunnel regime is investigated. For the case of electron injection from the silicon substrate, nearly complete defect relaxation is observed after the bias is removed, allowing the possibility of re-generating the defects. Modeling the defect generation process and examining differences between initial and subsequent degradation periods lead to an improved picture of both the relaxation process and the nature of the involved defects.

RELAXATION AND THE NATURE OF ELECTRICAL STRESS RELATED DEFECTS IN ULTRA-THIN DIOXIDE ON SILICON

by
Lihui Xie

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Engineering Science

Interdisciplinary program in Material Science and Engineering

January 1997

APPROVAL PAGE

RELAXATION AND THE NATURE OF ELECTRICAL STRESS RELATED DEFECTS IN ULTRA-THIN DIOXIDE ON SILICON

Lihui Xie

Dr. Kenneth R. Farmer, Thesis Advisor — Date
Assistant Professor of Physics
Physics Department, NJIT

Dr. Roland A. Levy, Committee Member — Date
Professor of Physics
Materials Science and Engineering Program, NJIT

Dr. Ken Chin, Committee Member — Date
Professor of Physics
Director of Physics and Materials Science and Engineering Program, NJIT

BIOGRAPHICAL SKETCH

Author: Lihui Xie

Degree: Master of Science

Date: January 1997

Undergraduate and Graduate Education:

- Master of Science in Engineering Science
  New Jersey Institute of Technology
  Newark, New Jersey. January 1997

- Bachelor of Science in Analytical Chemistry
  China University of Geosciences
  Wuhan, P.R.China. July 1983

Major: Materials Science and Engineering

Professional Background:

- Research Engineer
  Changsha Research Institute of Mining and Metallurgy
  Changsha, Hunan, P.R.China
  August 1983-January 1996

To my parents

ACKNOWLEDGMENT

It is a great pleasure to have this opportunity to express my deepest appreciation to Dr. Kenneth R. Farmer, who not only served as my research supervisor, providing valuable and countless resources, insight, and intuition, but also constantly gave me support, encouragement and reassurance. Special thanks are given to Dr. Roland A. Levy and Dr. Ken Chin for their friendly concern and support during the period of graduate study.

I want to appreciate my fellow graduate students in our group, such as Chao Sun, Steven Fu, Michael Beggans for their support. I also wish to thank my friend Wei Chen for his invaluable help on formatting this thesis.

Finally, I am grateful to acknowledge that part of this work was made possible through National Science Foundation awards No. ECS-9530984 and ECS-9624798.

TABLE OF CONTENTS

Chapter Page

1 INTRODUCTION 1

1.1 Motivation 1

1.2 Background Information 1

1.3 Introduction to What We Accomplished with This Work..2

2 REVIEW OF LITERATURE 4

2.1 Fundamental Knowledge..4

2.2 What is Already Known 12

3 PRELIMINARY EXPERIMENTAL DETAILS 16

3.1 Device Description 16

3.2 Experimental Setup and Apparatus 20

3.3 Directly Tunneling Regime 20

4 DEVICE CHARACTERISTICS 23

4.1 Current, Capacitance and Conductance vs. Voltage Characteristics..23

4.2 Degradation Characteristics..25

5. RELAXATION EXPERIMENT 29

5.1 Nearly Completely Relaxation 29

5.2 Time Dependent Relaxation..33

5.3 Different Bias Condition 41

TABLE OF CONTENTS
(Continued)

Chapter **Page**

6. MODELS 46

6.1 First Order Kinetics 47

6.2 Linear Time Constant Model 50

6.3 Exponential Time Dependence of Time Constant 52

6.4 Discussion of Fit Parameters from Exponential Model 56

7. SUMMARY AND CONCLUSIONS 61

APPENDIX 62

REFERENCES 69

LIST OF FIGURES

Figure Page

2.1.1 MOS capacitor 4

2.1.2 Band diagram at V=0 of ideal MOS device 4

2.1.3 Band diagram at V≠0 for a MOS structure 7

2.1.4 Silicon-silicon dioxide interface 8

2.1.5 Defects model of (111) Si/SiO_2 interface 9

2.1.6 Defects model of (100) Si/SiO_2 interface 9

3.1.1 Typical structure of MOS device in the experiment 16

3.1.2 Schematic diagram of a chip layout 18

3.1.3 Nomenclature of the device in a chip 19

3.2.1 Experiment circuit of the measurement 20

3.3.1 Energy band diagram under direct tunneling and Fowler-Nordheim tunneling 22

4.1.1 I-V characteristic 24

4.2.1 Degradation measurement(I-t) 27

4.2.2 Fractional increase of current vs. time 28

5.1.1 Nearly completely relaxation 31

5.1.2 Repeated I-t measurement 32

5.2.1 Charging after relaxing of 30 minutes 34

5.2.2 Charging behavior after relaxing of 10 minutes 35

5.2.3 Charging behavior after relaxing of 2 seconds 36

LIST OF FIGURES
(Continued)

Figure **Page**

5.2.4 Comparison of different relaxation time 39

5.2.5 Fit parameter of the Fig. 5.2.5 40

5.3.1 Whole picture of charging behavior after bias sit at different volts 43

5.3.2 Comparison of charging behavior of different Vsit 44

5.3.3 Fit parameters and different sit bias 45

6.1.1 Experiment data and fit by first order kinetics 49

6.2.1 Experiment data and fit by linear time dependent model 51

6.3.1 Experiment data and fit by exponential time dependent model 54

6.3.2 Comparison of three models 55

6.4.1 The effect of parameter N_{max} in the model 58

6.4.2 The effect of parameter K in the model 59

6.4.3 The effect of parameter τ in the model 60

CHAPTER 1

INTRODUCTION

1.1 Motivation

Control of the electrical properties of the metal-oxide-silicon(MOS) system has been one of the major factors that has led to stable and high performance silicon integrated circuits. In recent years, the very large scale integrated circuits have become more and more compact, containing increasingly small individual transistor devices. The scaling of device size requires continued assessment of device properties in each new size regime.

Traditionally, the MOS capacitor is used in both monitoring integrated circuit fabrication and studying the electrical properties of the MOS devices. The MOS capacitor has the advantages of simplicity of fabrication and of analysis. Fabrication of the MOS capacitor uses the same processing used in fabricating the integrated circuit. Therefore, the MOS capacitor provides direct measurement and monitoring of the MOS system as it is actually fabricated and used in the integrated circuit.

1.2 Background Information

By SIA Roadmap predictions, to remain competitive in the global market, the US semiconductor manufacturing industry must reach or exceed the capability to economically produce 16GB DRAM technology by the year 2007. A simple extrapolation of current scaling trends indicates that minimum feature sizes below 0.1μm and gate oxide thickness of approximately 4nm will be required. Thin silicon oxides and

oxynitrides will most likely remain the principle dielectrics in logic devices until the oxide thickness is less than 3-4nm. Recently, circuits with 3.5nm gate oxide high-performance 0.1μm p-channel MOSFET devices, which are important in high-speed CMOS circuits, have been reported[1]. In this thickness regime, the oxide layer is thin enough to permit electrons to tunnel through the oxide directly, providing a possible path for injection-related damage which is not present in thicker oxides. Detailed understanding of the microscopic mechanisms of the device electrical properties becomes very important.

Silicon dioxide (SiO_2) is the first principle gate insulator in the electronic industry because of its near-ideal properties. The ultimate device degradation and failure is still limited by charge buildup on defect sites in the oxide layer[2]. Although many papers have reported on the microscopic phenomena of the degradation and breakdown of thin oxide layers, a universal model capable of explaining all the observations is still being sought. In particular, the study of the degradation in MOS devices at high fields is directly relevant to understanding the mechanisms. In our work, we measure the electrical properties of the ultra-thin (<3.5nm) silicon dioxide on silicon in MOS devices, which pertain to the microscopic dynamic properties of the thin silicon dioxide film.

1.3 Introduction to What We Accomplished in this Work

In the present study we extend the investigation of the defect instability to the case of injection from n-type silicon substrates into degenerate n-type polycrystalline silicon gates. Unlike the situation involving p-type silicon and gate injection, we find that in the

n-type system with substrate injection, nearly complete relaxation occurs, allowing the possibility of re-generating the defects. By modeling the defect generation process and examining differences between original and subsequent defect creation periods, we are able to obtain an improved picture of both the relaxation process and the nature of the involved defects. Besides this increased understanding of the defects, an additional important contribution of this work is the demonstration that a straightforward first-order kinetics process may underlie the degradation. This is surprising since it has been shown previously that the dominant electron tunnel current does not control the rate of degradation[10,13] . The success of our model suggests that a conduction process other than this dominant direct tunnel current component may drive the degradation.

CHAPTER 2

REVIEW OF LITERATURE

2.1 Fundamental Knowledge

2.1.1 MOS Structure and Energy Band Diagram

In this thesis, we have studied the MOS capacitor because it is directly relates to the transistor devices used in integrated circuits. This device is illustrated in Fig. 2.1.1.

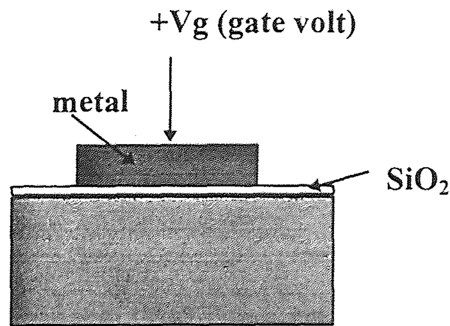


Fig.2.1.1 Cross-section schematic diagram of a MOS capacitor

The MOS capacitor consists of a parallel plate capacitor with one electrode, a metallic plate called the gate, and the other electrode the silicon. The two electrodes are separated by a thin insulating layer of SiO_2.

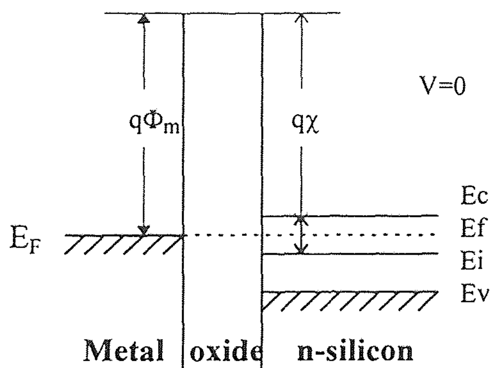


Fig.2.1.2 Energy-band diagram of an ideal MOS diode at V=0.

The energy-band diagram of an ideal MOS structure for V=0 is shown in Fig.2.1.2. An ideal MOS diode is defined as follows. (1) At zero applied bias, energy difference between the metal work function Φ_m and the semiconductor work function is zero, or the work-function difference Φ_{ms} is zero:

$$\phi_{ms} \equiv \phi_m - \left(\chi + \frac{E_g}{2q} - \psi_B \right) = 0 \qquad \text{for n-type}$$

$$\phi_{ms} \equiv \phi_m - \left(\chi + \frac{E_g}{2q} + \psi_B \right) = 0 \qquad \text{for p-type}$$

where Φ_m is the metal work function, χ the semiconductor electron affinity, χ_i the insulator electron affinity, E_g the bandgap, ϕ_B the potential barrier between the metal and the insulator, and ψ_B the potential difference between the Fermi level E_F and the intrinsic Fermi level E_i. (2) The only charges that can exist in the structure under any biasing conditions are those in the semiconductor and those with the equal but opposite sign on the metal surface adjacent to the insulator. (3) There is no carrier transport through the insulator under dc biasing conditions, or the resistivity of the insulator is infinity.

When an ideal MOS diode is biased with positive or negative voltages, basically three cases may exist at the semiconductor surface (Fig.2.1.3). When a negative voltage ($V<0$) is applied to the metal plate, the top of the valence band bends upward and is closer to the Fermi level (Fig.2.1.3a). For an ideal MOS diode, no current flows in the structure, so the Fermi level remains constant in the semiconductor. Since the carrier density depends exponentially on the energy difference (E_F-E_v), this band bending causes an accumulation of majority (in p-type Si) carriers (holes) near the semiconductor surface. This is the "accumulation" case. When a small positive voltage ($V>0$) is applied, the bands bend downward, and the majority carriers are depleted (fig.2.1.3b). This is the "depletion" case. When a larger positive voltage is applied, the bands bend even more downward so that the intrinsic level E_i at the surface crosses over the Fermi level E_F (Fig.2.1.3c). At this point, the number of electrons (minority carriers) at the surface is

larger than that of the holes, the surface is thus inverted, and this is the “inversion” case. Similar results can be obtained for the n-type semiconductor.

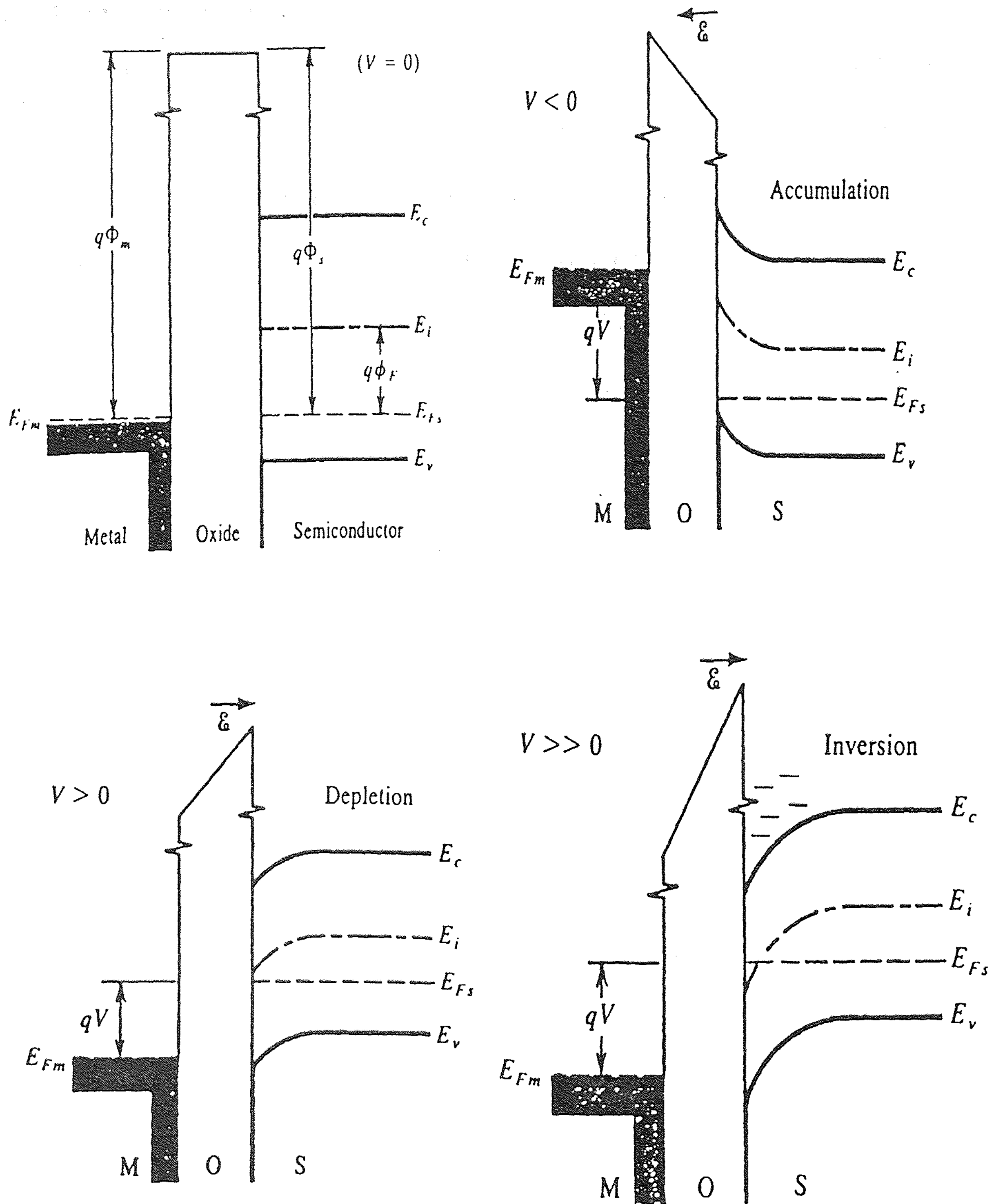


Fig.2.1.3 Energy-band diagrams for ideal MOS diodes for V=0, and, for the following cases when V≠0: accumulation; depletion; inversion.

2.1.2 The Si / SiO_2 Interface

As the film becomes very thin, the Si/SiO_2 interface region becomes very important. Extensive research has been carried out in order to obtain a better understanding of the properties of the Si/SiO_2 interface. Even if there still are a lot of questions to be answered, an increasingly detailed picture has developed in the last years. It is now accepted that the Si/SiO_2 interface has two distinct regions as illustrated in Fig. 2.1.4.

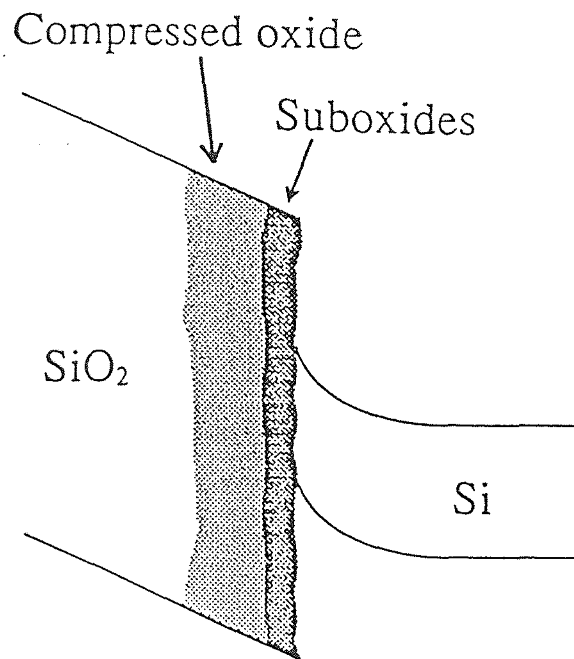


Fig. 2.1.4 The Si/SiO_2 interface region

Ohdomari el al.[4] demonstrated two models of the Si/SiO_2 interface which are shown in Figure 2.1.5. Model a) based on the fact that {111}Si/SiO_2 interface has a low distortion energy shows an Si-O region that is formed between the {111} facets and the amorphous SiO_2 layer. This model gives simulated high-resolution transmission electron microscopy (HRTEM) images which are close to the observed images. However, the problem is that this model does not contain the intermediate chemical bonding states

which have been observed with x-ray photoelectron spectroscopy. The second model shows the intermediate binding states of Si atoms at the {100}Si /SiO_2 interface

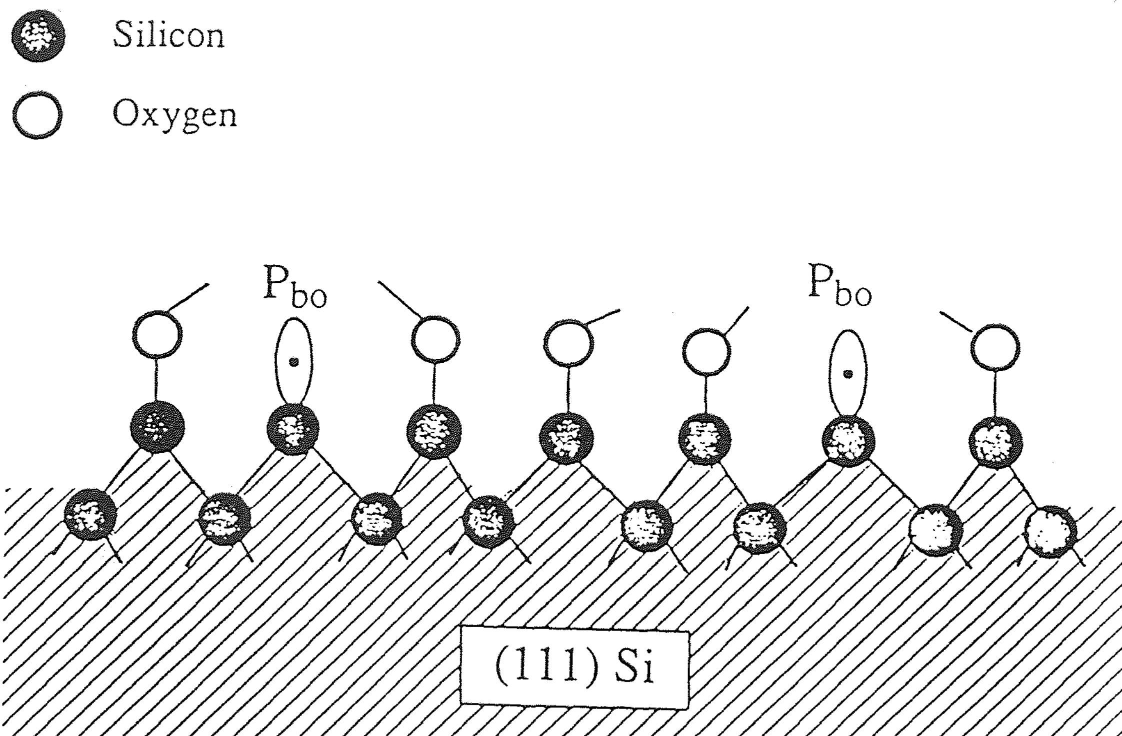


Fig. 2.1.5 Model of the P_b center at the (111) Si/SiO_2 interface.

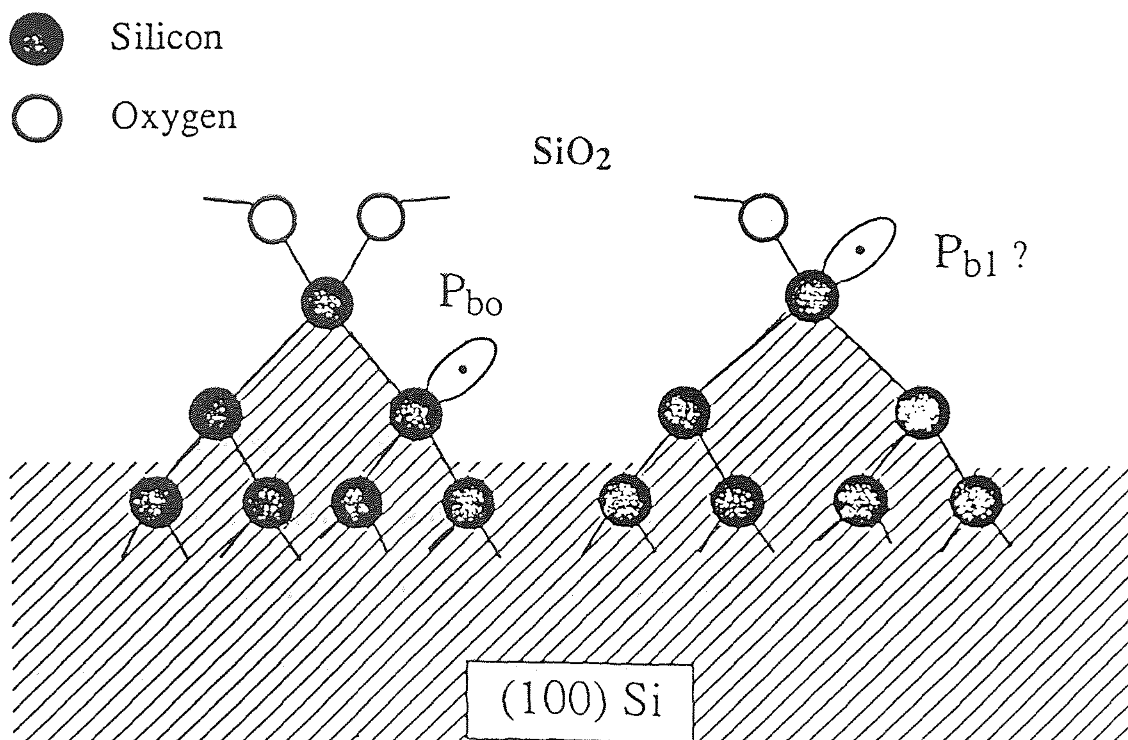


Fig. 2.1.6 Models for P_b centers at the (100) Si/SiO_2 interface.

2.1.3 Defects

Defects within or near the interfacial region cause interface trap levels. Such defects range from stacking faults and micro-pores to various atomic or molecular fragments left as residue of imperfect oxidation. A problem with modeling the interface layer is that the details of these defects are not known. However the results of several workers indicate the presence of nonstoichometry. No relation has been experimentally established between nonstoichiometry at the interface and the electrically active centers.

Several specific defects may be responsible for interface traps produced by thermal oxidation. Four different types of defect that could exist at or near the Si-SiO_2 interface traps are (1) excess silicon (trivalent silicon), (2) excess oxygen (Nonbridging oxygen) (3) impurities and (4) states in oxide charge induced potential wells.

As a consequence of oxidation, there may be *trivalent silicon atoms* or silicon in excess of the stoichiometric amount in the SiO_2 at the interface, probably related to incomplete oxidation of the silicon or to generation of vacancies in the silicon during oxidation.[18] In this type of defect the silicon atom shares three of its four valence electrons with neighboring silicon atoms. Its fourth unsatisfied valence bond acts as a hole trap becoming positively charged after capturing a hole and remaining neutral when empty.

Nonbridging oxygen, oxygen in excess of the stoichiometric amount in the oxide may be present near the interface for three reasons: (1) excess oxygen associated with the oxidation reaction must be present and may form nonbridging oxygen defects; (2) the strain in the region near the interface might be relieved by formation of this defect; and

(3) there are water-related electron traps near the Si-SiO_2 interface that might be related to nonbridging oxygen defects.

There is some tentative experimental evidence that trivalent silicon exists in SiO_2 and may act as interface traps. Nishi and Poindexter et al.[18] interpret their electron spin resonance signal as due to trivalent silicon. In their measurements the observed spin resonance signal is found to increase or decrease in the same manner as midgap interface trap level density under oxidation or annealing variations. Thus trivalent silicon appears to be a likely source of the spin resonance signal observed in their work. However, conclusive evidence for the species responsible for the observed spin resonance signal awaits the detection of hyperfine structure which is a difficult measurement. The trivalent silicon defect is consistent with a bond disorder model.

The third type of defect is an impurity atom at the interface. Because strain at the interface creates a potential minimum for impurities, the interface can accommodate many of the impurities incorporated in the relatively open SiO_2 lattice. There is also a strained region at the metal-SiO_2 interface. Strained regions at either interface are likely to act as sinks for impurities incorporated in the oxide.

The fourth type of defect is an oxide charge near the interface that induces an attractive coulombic potential well in the silicon.[18] The interface trap model based on this defect has been discussed earlier in this section. There is evidence from low temperature measurements that sodium ions which have drifted to the Si-SiO_2 interface induce an interface trap band near the conduction band edge.

2.2 What is Already Known

Silicon dioxide (SiO_2) is the principle gate insulator in the electronics industry. Ultimately, device degradation and failure is related to the build up of charge on defect sites in this oxide. Most of these sites appear when charge carriers are injected from the contacts into the oxide layer. In order to understand the microscopic mechanisms of this effect, significant work has been done in recent years.

Charge and trap generation in Metal-Oxide-Silicon(MOS) structures has been heavily studied in conventional thicker oxide devices. These studies have revealed a complex dynamic process occurring in the transition region between the bulk SiO_2 and the silicon. There exist energy thresholds for the generation of both positive charge[5] and electron traps,[6,7] the electron trap creation is suppressed below 150K,[7] and a turn-around effect has suggested that positive charge created near the oxide-silicon interface is neutralized by a process which does not lead back to the original state of the oxide, but instead eventually can lead to a net negative oxide charge.[8] The commonly accepted picture is that electrons enter the oxide conduction band and are accelerated by an applied electric field.

By our previous work, we know that very thin (<3.5nm) oxide MOS devices show similar charging effects to the thicker oxide devices, but possibly generation occurs at a much lower electron energy level. In the thin oxide system, the injected electrons can tunnel directly between the metal and the silicon, never entering the oxide conduction band. The charging effects are mainly the result of the presence of defects sites, and these defects have similar spatial distributions within the oxide layer since the charging curves

are mainly the same for a given area device, biased at a given voltage, independent of thickness in the sub-3.5 nm regime.

In the paper of K. R. Farmer et al., in 1991[10], the authors described the positive and negative charging effects for small area (0.008-20μm^2), very thin dielectric (~2.5nm) MOS devices. According to that work, we know that the negative charge is strongly temperature dependent. In contrast, the positive charging arises only above a threshold bias level, is strongly voltage dependent, but weakly sensitive to temperature particularly below 150K. Subsequent additional work [11] showed that positive charge generation could also be found in very thin silicon-oxides even after the device received a post-metallization anneal (PMA) treatment at low temperature.

By the study of time-dependent positive charge generation, the fractional increase in the current density is independent of the oxide thickness and the rate of charge generation is not controlled by the flux of tunneling electrons over a five order of magnitude range in current density, and the charging rate increases strongly with bias voltage.

Since the positive charging is universal, we paid more attention to its investigation. It has been demonstrated that this degradation mode exists for devices fabricated at different facilities, under different process conditions, with both aluminum and polycrystalline silicon gates, and undergoing either substrate or gate electron injection.[9-13] The positive charging has the following universal characteristics: 1) threshold voltage for onset of degradation, 2) degradation rate that is relatively independent of oxide thickness below ~3.5nm, 3) strong voltage dependence above the

threshold, 4) weak temperature dependence below ~150K and thermal activation above ~ 150K.[14]

Knowledge and the understanding of damage in Metal-Oxide-Silicon tunnel diodes is still very limited. It has been shown that direct tunnel oxides can endure very high electron fluence without experiencing dielectric breakdown, for example, greater than 3×10^5 C/cm^2 through a ~2.5 nm silicon oxide at approximately 9MV/cm [9] compared with the typical charge to breakdown of ~10 C/cm^2 due to Fowler-Nordheim (FN) injection through thicker oxides at comparable fields. Thus in contrast to FN tunneling in thicker oxides the direct tunnel current in very thin oxides only weakly influences the oxide quality. However, even in tunnel diodes, the charging effects introduced above, "negative charging" and positive charging, have been identified. The negative charging has been associated with water-related traps in devices which could been eliminated by PMA (post- metallization anneal). [15] The origin of the net positive charge and the generation mechanism have not been identified.

On the other hand, positive charging is an important degradation effect related to breakdown [16] in direct tunneling devices. The recent study of hundreds of breakdowns shows that there are two types of breakdowns even in these ultra-thin oxides. One is "defect-related breakdown", the other is so called "intrinsic breakdown" which is related to the build-up of the positive charge. This relationship is demonstrated by comparing the thickness dependence of the two breakdown modes with the thickness dependence of the positive charge generation. Experiment shows that there is a significant difference in the thickness dependence of defect-related breakdown compared to that of positive charging,

whereas there is little when comparing intrinsic breakdown to positive charging. In particular, the positive charge generation rate and the time to breakdown are both relatively independent of oxide thickness below ~2.9nm, while for the 3.4nm oxide, the positive charge generation rate is decreased while the time to breakdown is increased. This correlation suggests that the positive charge may be at the root of intrinsic oxide failure in ultra-thin oxides.

CHAPTER 3

PRELIMINARY EXPERIMENTAL DETAILS

3.1 Device Description

An MOS capacitor is the typical structure used in most investigations of electrical properties of oxide films on silicon. This structure is shown in Figure 3.1 for our devices which have an ultra-thin dielectric layer.

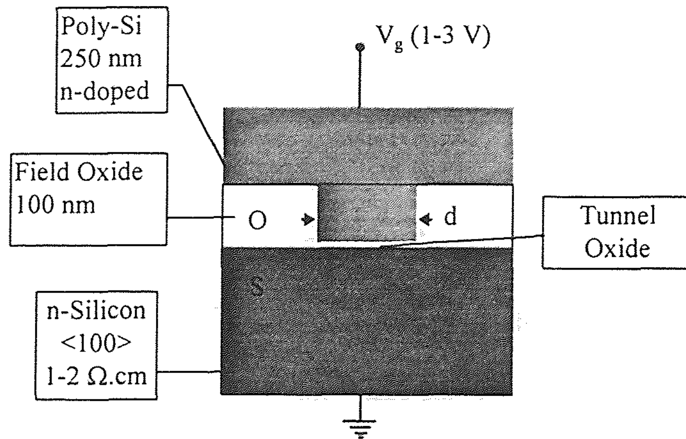


Fig.3.1.1: Typical MOS capacitor structure

It consists of a parallel plate capacitor with one electrode a metallic plate, called the gate (here polycrystalline silicon) and the other ground electrode, the silicon. The two electrodes are separated by a thin insulating layer of silicon dioxide with small diameter window areas. The substrates were (100) n-type silicon with a resistivity of ~0.1 to 10 ohm-cm. After a standard cleaning process, a recessed oxide layer was grown to a thickness of ~450nm at 950°C to form a field oxide after lithographic definition of a LPCVD Si_3N_4. After removing the LPCVD Si_3N_4, the gate oxide was then grown to a thickness of ~3.5nm at 700-750°C. This was quickly followed by the deposition of a 100 nm layer of LPCVD polycrystalline silicon. Phosphorus was then

implanted at 10 kev to a dose of 10^{15} ions/cm^2 to form an *n*-type gate. This was followed by the sputtering of a diffusion barrier layer of 30nm TiN and 50nm of TiSi for a gate contact.[2] The samples were annealed at 1000°C for ~5 sec. in a rapid thermal processing tool to activate the dopant in the polycrystalline silicon. After the patterning and etching of the gate stack, aluminum was thermally evaporated onto the backside to produce an electrical contact.

The devices used in this study have been fabricated on the silicon line at IBM, Yorktown Heights. Using the fabrication process described above, active tunnel oxides of four different nominal thickness, 2.0, 2.4, 2.8 and 3.4 nm are formed in dry O_2. Although windows of various sizes ranging from 1 μm^2 to 4 mm^2 are available on each chip, measurements for this work have been carried out using mainly 250 μm × 250 μm devices.

The chip layout schematic diagram is shown in the Fig.3.1.2. Various devices with different dimensions were fabricated on each ~1×1cm^2 chip. The region of the chip that contains the 250μm×250μm devices used in this work is shown in Fig.3.1.3. Each device on each chip was given a name. For example, in Fig.3.1.3, from right side to left, different devices with a certain areas from 2000× 2000 μm^2 to 250 × 250 μm^2 were named as A, B, C, D separately. There are three lines in D group, separately called line D1, D2, and D3. In the vertical direction, the devices were labeled using letters in alphabetical order, thus for example,the third device in the line D1 was called device D1C.

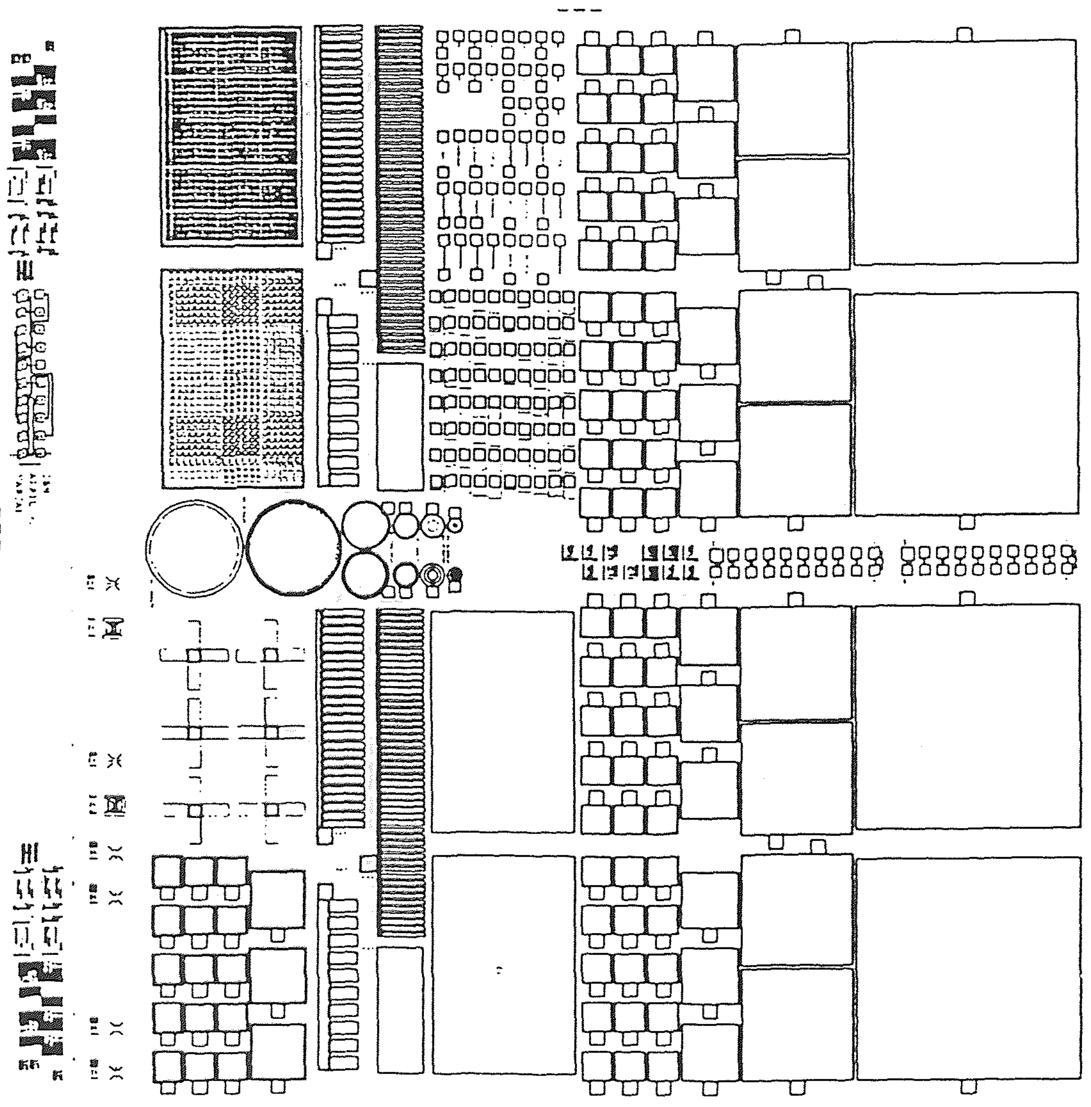


Fig. 3.1.2 Schematic diagram of the IBM chip

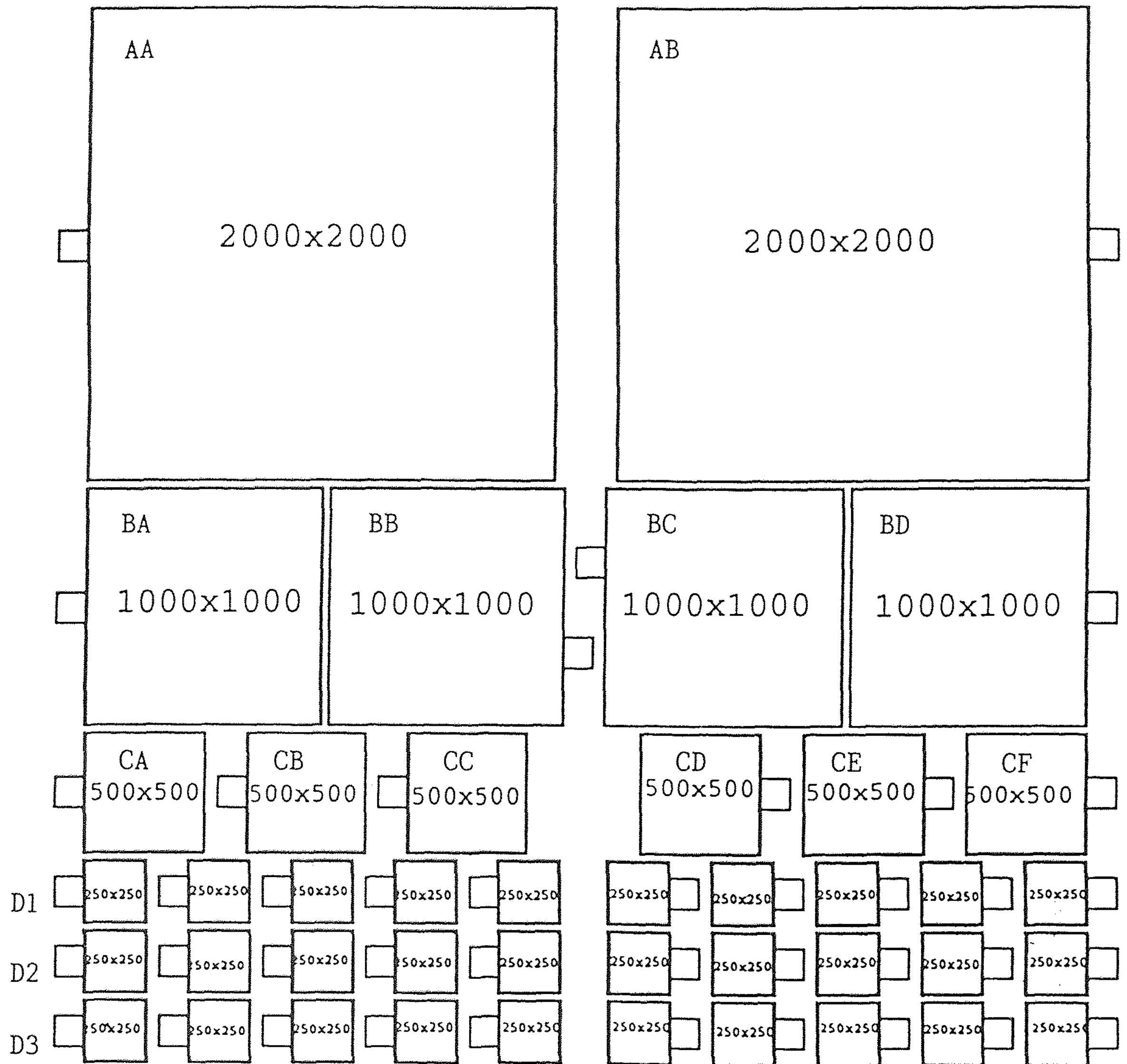


Fig. 3.1.3 Nomenclature for devices studied. (the areas are shown in μm^2)

3.2 Experimental Setup and Apparatus

Usually, to measure I-V and C-V characteristics, an *hp 4140B* PA meter/DC voltage source and an *hp 4284A* 20Hz-1MHz Precision LCR meter were used. In this study, the measurement system was set up as shown in Fig.3.2.1.

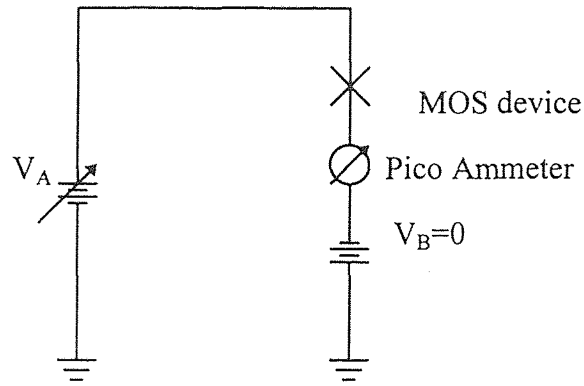


Fig.3.2.1 Schematic circuit of the measurement system

The measurement procedure was controlled by the computer system. In order to investigate the degradation and relaxation process, we designed several versions of programming to take the data. The procedure will be illustrated in the following chapter.

3.3 Direct Tunneling Regime

In solid state devices, the quantum mechanical tunneling of an electron injected through a dielectric barrier of a small thickness is very important to the performance of small dimensional devices. In our study, an ultra-thin (2.8nm) silicon dioxide MOS device was investigated.

Since the oxide in the devices is so thin (<3.5nm) direct tunneling is the dominant conduction mechanism. The difference between a direct tunneling (DT) and a Fowler-Nordheim tunneling (FNT) electron injection mechanism is explained in Fig.3.3.1.

For a positive applied voltage Vg, the n-Si substrate will be in accumulation. If the voltage is sufficiently high, the SiO_2 tunnel barrier changes from a trapezoidal to a triangular shape and causes electron injection by so called "Fowler-Nordheim" tunneling in which the electrons are injected into the oxide conduction band. In contrast, if the oxide barrier is sufficiently thin, an electron can tunnel from the Si accumulation layer directly to the poly-Si condition band, without entering the SiO_2 conduction band.

Tunnel oxide thicknesses have been estimated electrically from capacitance-voltage characteristics. For all of the different oxides, devices exhibit current-voltage (I-V) characteristics which scale well both with device area, and with oxide thickness in good agreement with first principles tunneling theory.[17]

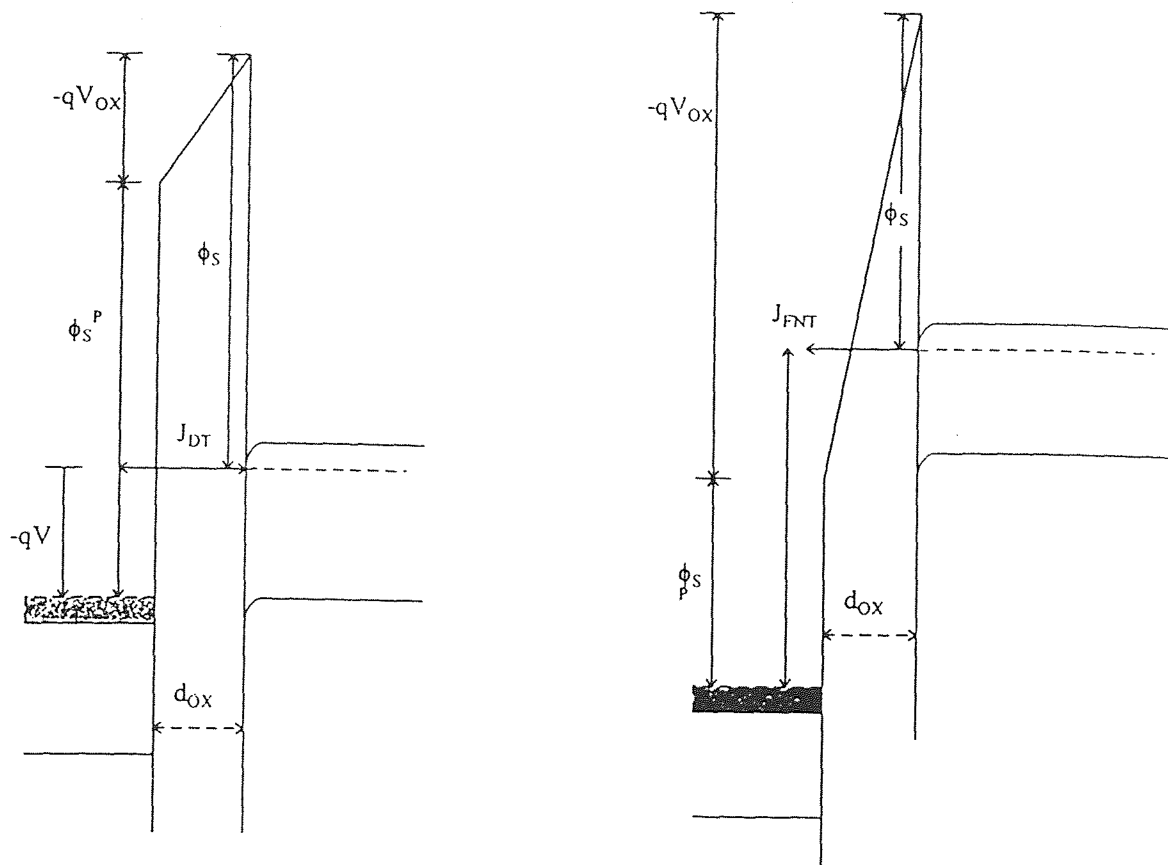


Fig. 3.3.1 Energy band diagram of direct tunneling and Fowler-Nordheim tunneling

CHAPTER 4

DEVICE CHARACTERISTICS

4.1 Current, Capacitance and Conductance vs. Voltage Characteristics

In order to understand the effectiveness of the oxide reliability on the performance of the MOS diodes, researchers have developed a number of electrical measurement techniques. Among the most useful techniques are the current versus voltage measurement (I-V) and the frequency dependent capacitance and conductance versus voltage (C-V and G-V) techniques. Monitoring the current is particularly useful in ultra-thin oxide devices, where capacitance and conductance measurements become increasingly difficult due to the presence of the direct tunneling current. In capacitance measurements, the direct tunnel current greatly exceeds the displacement current required for quasistatic C-V's, while in the conductance measurements, the tunnel current can add a large frequency independent offset to G-V's. In this work, we report exclusively the measurement of current passing through the devices, both current versus voltage to compare different devices, and current versus time to monitor device dynamics. For the measurements reported in this work, care has been taken to select devices for each thickness which display almost identical initial I-V characteristics.

Fig 4.1.1 shows a typical I-V for our 2.8 nm oxide. The current passing through the oxide layer increases with increasing positive gate voltage. The device has area of 250μm×250μm. The current is plotted on both a linear scale (circular dots) and a common logarithm scale (square dots) to accentuate the bias dependence. The increment between voltage steps is 0.05 V.

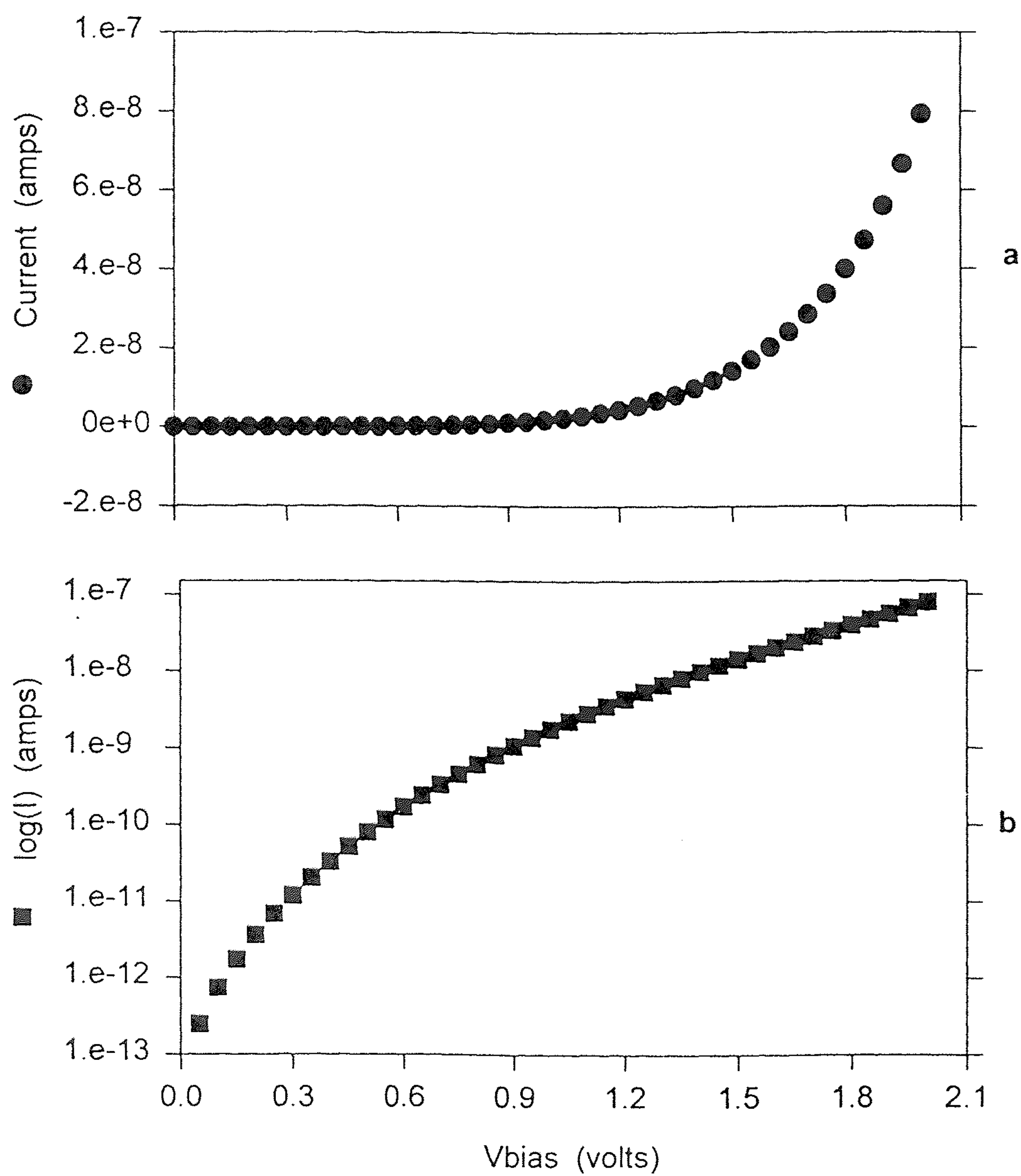


Fig. 4.1.1 Measured I-V using different current scales.

4.2 Degradation Characteristics

The ultimate device degradation and failure is limited by the charging behavior in the insulator layer. For silicon-based devices such as dynamic-random-access memories (DRAM), both electrons and holes can be inadvertently injected locally into the insulator. In nonvolatile memories such as electrically programmable read-only memories (EPPROM), the insulator is actually used to transport carriers to buried electrodes or charge trapping layers. Because of their ease of fabrication MOS diodes can be used to study the oxide degradation effects related to charge injection in the more complex systems.

As discussed previously, degradation occurs in ultra-thin silicon dioxide in MOS devices biased in the direct tunnel regime, and this degradation is different in character as compared with what is observed in thick oxides. To characterize the oxide degradation in our ultra-thin silicon dioxide device, we measured the increase in the tunnel current with time which represented the build up charges under the fixed DC bias. When negative bias was applied to the gate, electrons were injected from the gate to substrate. Under positive gate bias, electrons were injected from the substrate to the gate. P-type substrates have been used for gate injection studies. In this work we use n-type substrates for substrate injection investigations. Under both conditions, the devices show degradation effects.[4] For electron injection from the gate, depending on the strength of the bias, both net negative charging (a decreasing tunnel current over time) and net positive charging (an increasing current) can be observed. The negative charging can be seen at low bias levels, the positive charging arises only above an apparent threshold bias level.[9]

It is believed that the effective positive charge that builds up after a DT stress is the result of physical damage that is done to the oxide[9]. Since our purpose is to shed light on the mechanism of the oxide failure, the positive charge will be the focus of our discussion. For this work, we have carefully selected the gate bias level to be 2.1 V, high enough to be above the threshold for the onset of degradation, but not so high that the devices broke down catastrophically even after 10^4 seconds of stress. We measured the change in current versus time at room temperature. Fig.4.2.1 shows a typical degradation curve obtained in this manner using a D-group device. The positive charging (an increasing current) can be observed, consistent with the results of previous work.[9]

Another way to indicate the positive charging in silicon dioxide is the fractional change in current (or current density) as a function of time. This is given by the equation:

$$\frac{\Delta I}{I_0} = \frac{I - I_0}{I_0} \quad \text{or} \quad \frac{\Delta J}{J_0} = \frac{J - J_0}{J_0} \qquad (4\text{-}1)$$

Where I is the current, I_0 is the initial current at the fixed bias. J is the current density. A typical plot of the fractional change in current density is given in Fig.4.2.2. The nature of this positive charging effect will be the subject of the rest of this investigation.

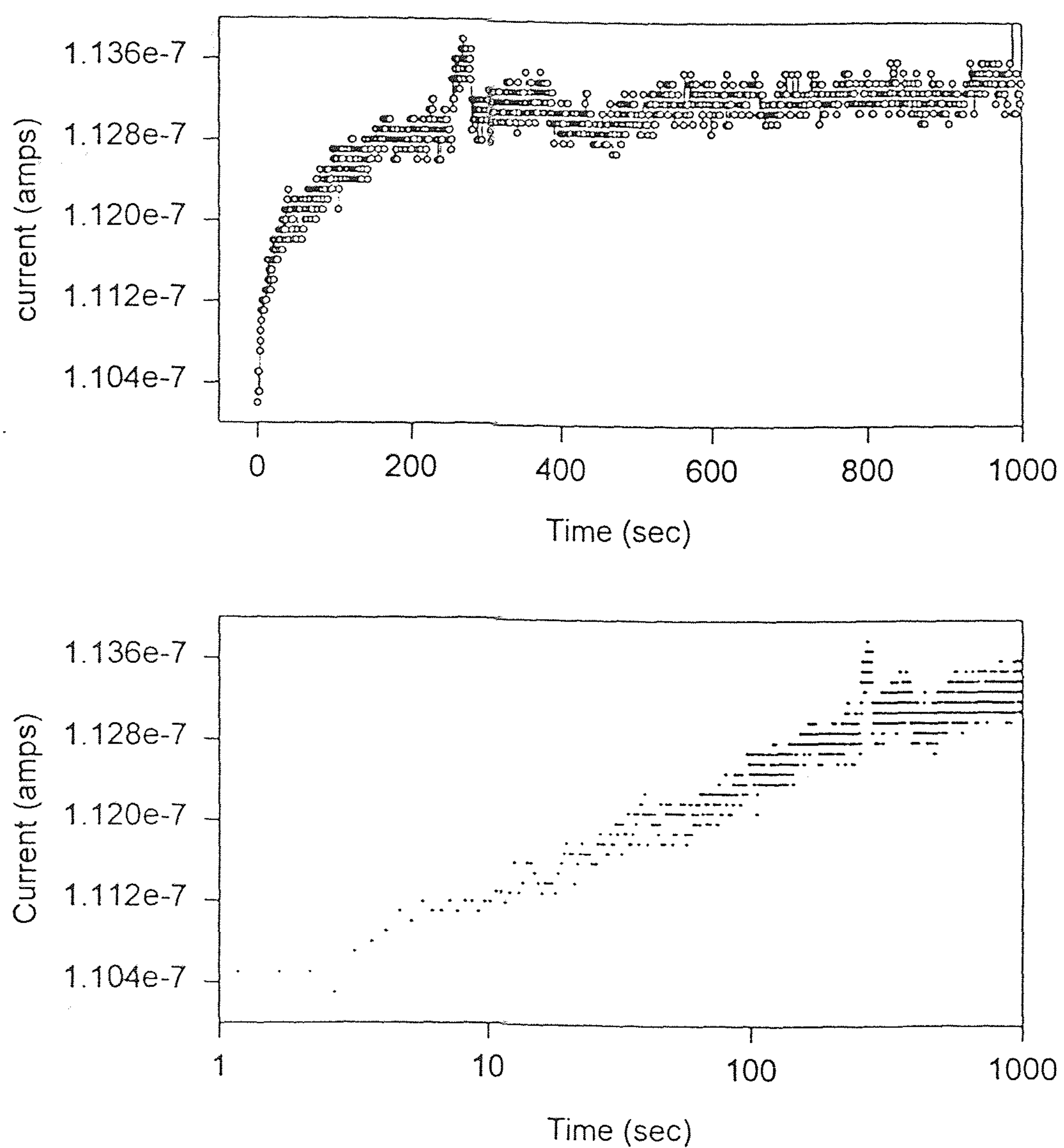
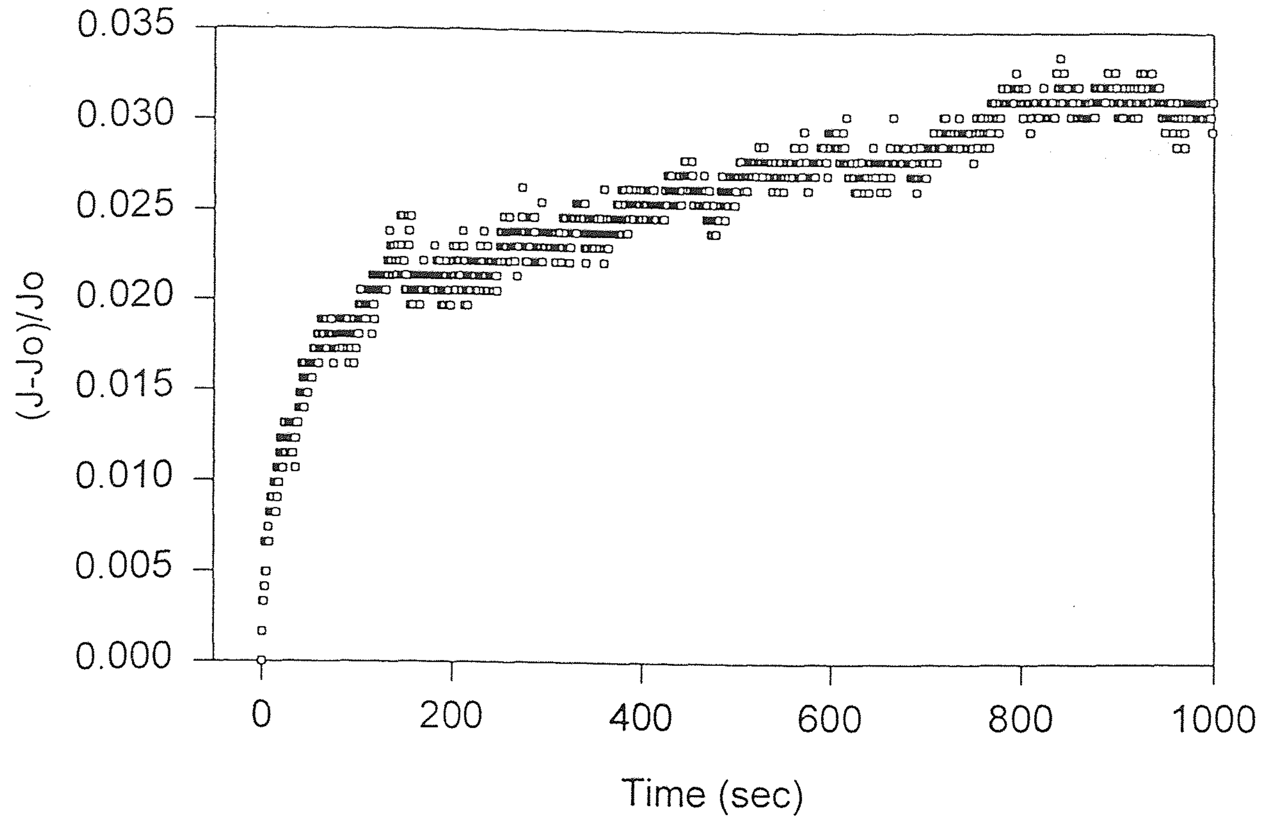


Fig. 4.2.1 Positive charging at Vg=2.1 V for a 2.8 nm oxide device



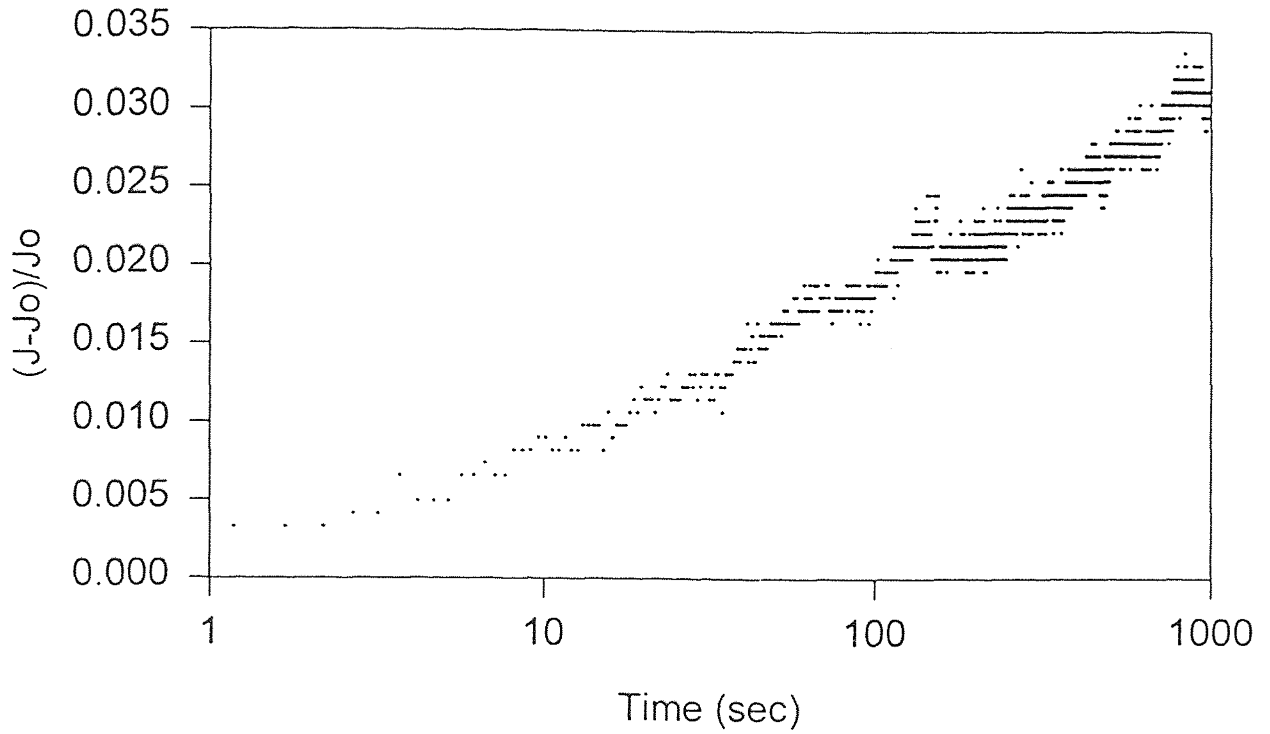


Fig. 4.2.2 The fractional change in current through 2.8 nm oxide device

CHAPTER 5

RELAXATION EXPERIMENT

In this chapter we present detailed observations of the relaxation of stress-induced positive charge in thin oxides on n-type silicon substrates. We find that in this system, the relaxation can be nearly complete, compared to the partial relaxation seen in p-type devices.[12] This provides an important possibility to observe the re-generation of the positive charge. Our observations lead to an improved understanding of the nature and origin of the charge.

5.1 Nearly Complete Relaxation

The procedure for measuring the positive charging effect is to bias a device at some constant voltage, and then to monitor the current passing through the device, as described in Chapter 4. The increase in positive charge is evident as an increase in the current passing through the device.

In becoming familiar with the IBM set of devices, and with the technique of measuring the increase in positive charge, we noticed that a device which had been previously stressed at 2.1 V at room temperature, the next day showed no signs of having been stressed. During the first stress, the current rose from an initial value of 1.124×10^{-7} amps to a final value of 1.152×10^{-7} amps after 1000 s of stress. During the second stress, after the device had sat unbiased overnight, the initial and final current values were roughly the same as those for the first stress. This is illustrated in Figure 5.1.1 which shows the fresh I-t curve (Fig.5.1.1a), the second repeat I-t curve (Fig.5.1.1b) and a comparison of the fit curves (Fig.5.1.1c) with respect to time. The solid dots are the

fitting curve, and the open circles and squares are the experimental data of the fresh and repeated I-t. The fitting procedure will be outlined in Chapter 6. The important observation at this point is that the appearance of the initial I-t curve and the subsequent repeated I-t curve is nearly identical, as are the starting and ending current values in Figure 5.1.1(a) and (b). The similarity between the initial and second I-t curves is unusual and unexpected, given the fact that such nearly total "relaxation" of the positive charge has not been observed previously in devices fabricated on p-type substrates, measured at room temperature. Confirmation of this effect is provided in Figure 5.1.2 which shows sequential I-t measurements on a device, repeated not after an overnight sit unbiased as in Figure 5.1.1, but after less than one hour unbiased. Note the similarity in the curves, even for this shorter period unbiased. Small differences between the curves in Figures 5.1.1 and 5.5.2 will be the subject of the rest of this study.

The present work is the first detailed work using n-type substrates and poly-crystalline silicon gates. This observation of strong relaxation is the starting point of our investigation using the n-type devices. Because the relaxation can be nearly complete, and more importantly because the re-generation of the charge appears to be so similar to the initial generation, we can investigate the nature and origin of the charge by studying the time, bias and temperature dependence of both the relaxation and the re-generation.

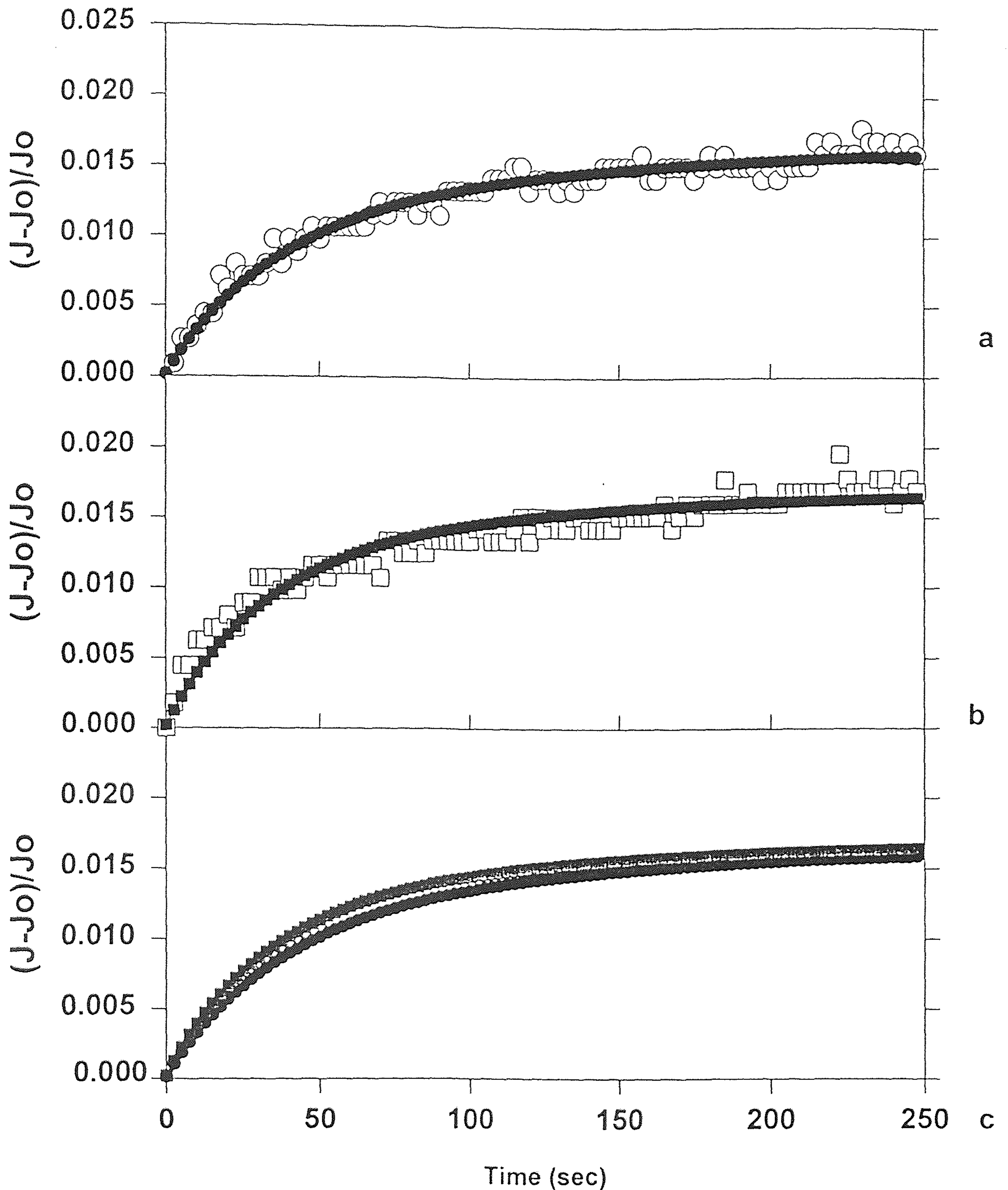


Fig. 5.1.1 The I-t measurement results of nearly completely relaxation in a thin oxide layer. (a) Initial charge build up in a fresh, previously unstressed device (circles) and fitted curve (dots). (b) Second stress (open square) and fitted curve (solid square). (c) Comparison of two fitting curve to underscore similarity.

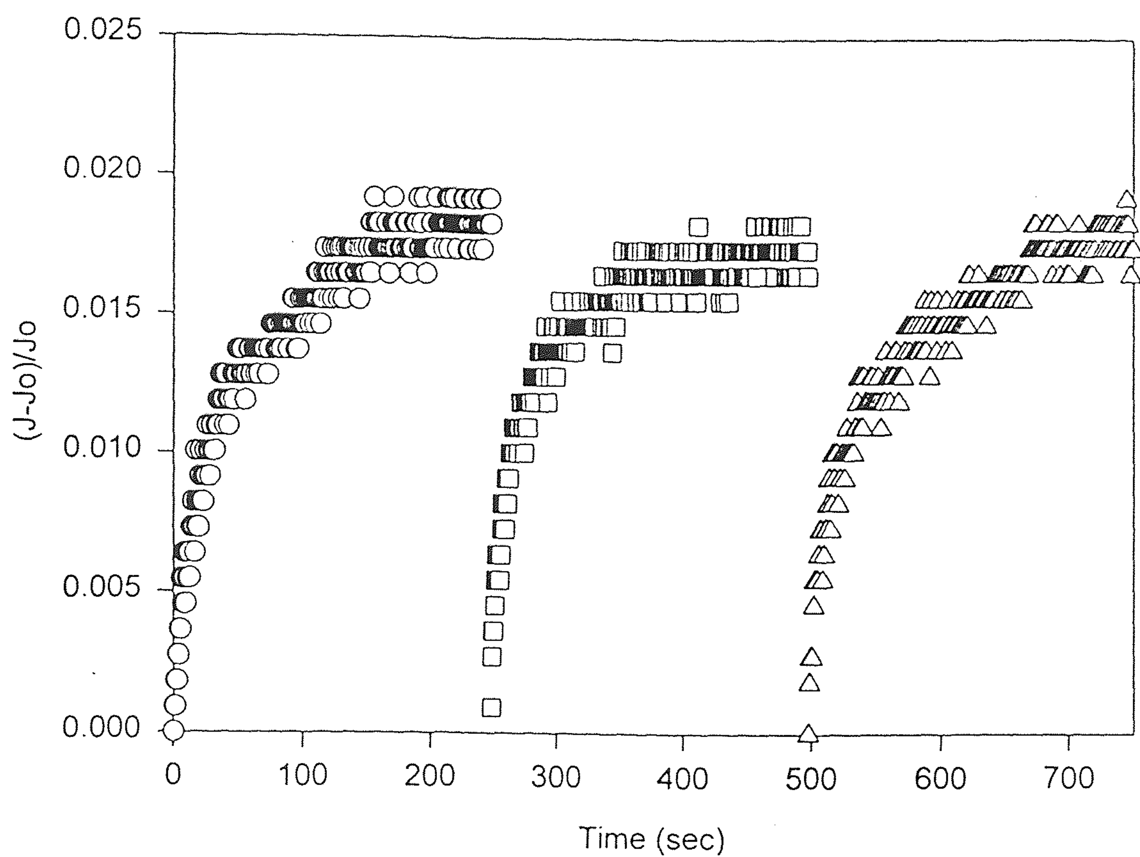


Fig. 5.1.2 The results of repeated I-t measurement under the bias of V_g = 2.1 V. The thickness (t_{ox}) is 2.8 nm, the line of circles denotes the initial charge build-up, the lines of squares and triangles denote the second and third passes, respectively.

5.2 Time Dependent Relaxation

The relaxation of the positive charge is more difficult to study than the re-generation of the charge, because the relaxation is a very quick process compared to the re-generation. In this section we describe the salient observations concerning the relaxation.

In order to determine the time scale on which the relaxation is occurring, we have measured charge generation and re-generation curves with various intervals of time between the initial and subsequent I-t's. These measurements are shown in Figures 5.2.1, 5.2.2 and 5.2.3. Each figure plots initial and subsequent charge generation curves for a single device biased at 2.1 V, with wait times between each curve of 30 minutes, 10 minutes and 2 seconds for each figure, respectively. The open circles and squares are the experimental data for the initial (plot a) and 2nd (plot b) stress measurement, the solid lines are the fit curves. It is obvious, that the fit curves follow the data quite well. The difference between the initial and 2nd stress period is illustrated in each figure (plot c). Consider Fig. 5.2.2 as an example. Note that the first stress curve (a) seems to increase relatively slow compared to the second stress measurement (b) in the initial part of each curve (first ~50 seconds). In the latter part of each curve (last ~100 seconds) the second curve (b) is flatter than the first (a). However, finally, the fractional increase of current density will reach approximately the same maximum value as shown clearly in each plot (c). The bending of the two curves is different in the 0 to 150 second time period which indicates that the charge build up is faster in previously stressed devices as compared to fresh ones. The degree of the curve bending is related to the time constant in our model which will be discussed in Chapter 6.

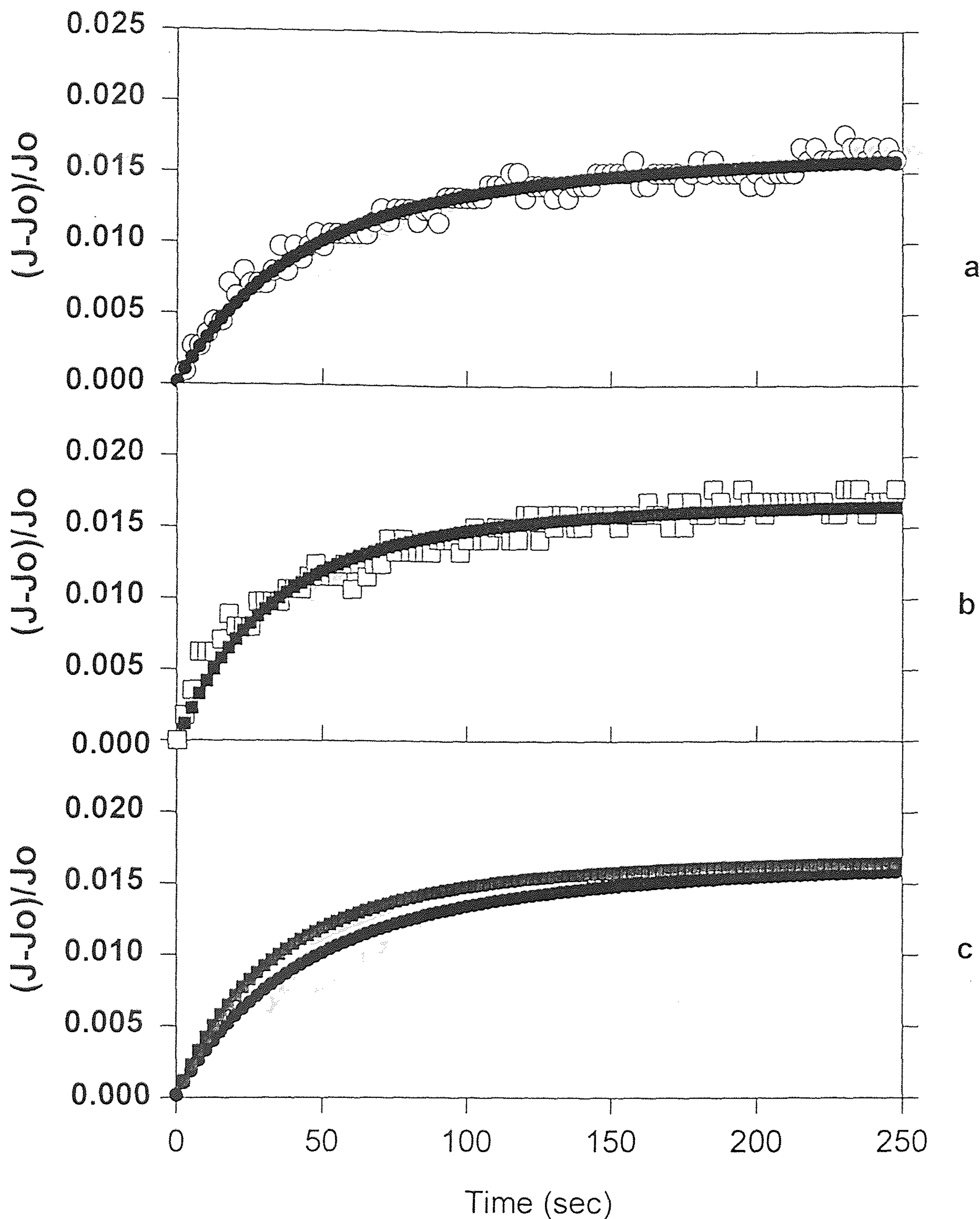


Fig. 5.2.1: Incomplete relaxation during 30 minutes unbiased. During the stress period, V_{bias}=2.1 Volts. (Wafer W8, Chip J2, Device D1E, t_{ox} = 2.8 nm, T = 300 K) (a): initial stress. (b): 2nd stress after 30 min. sit unbiased. (c): fit comparison.

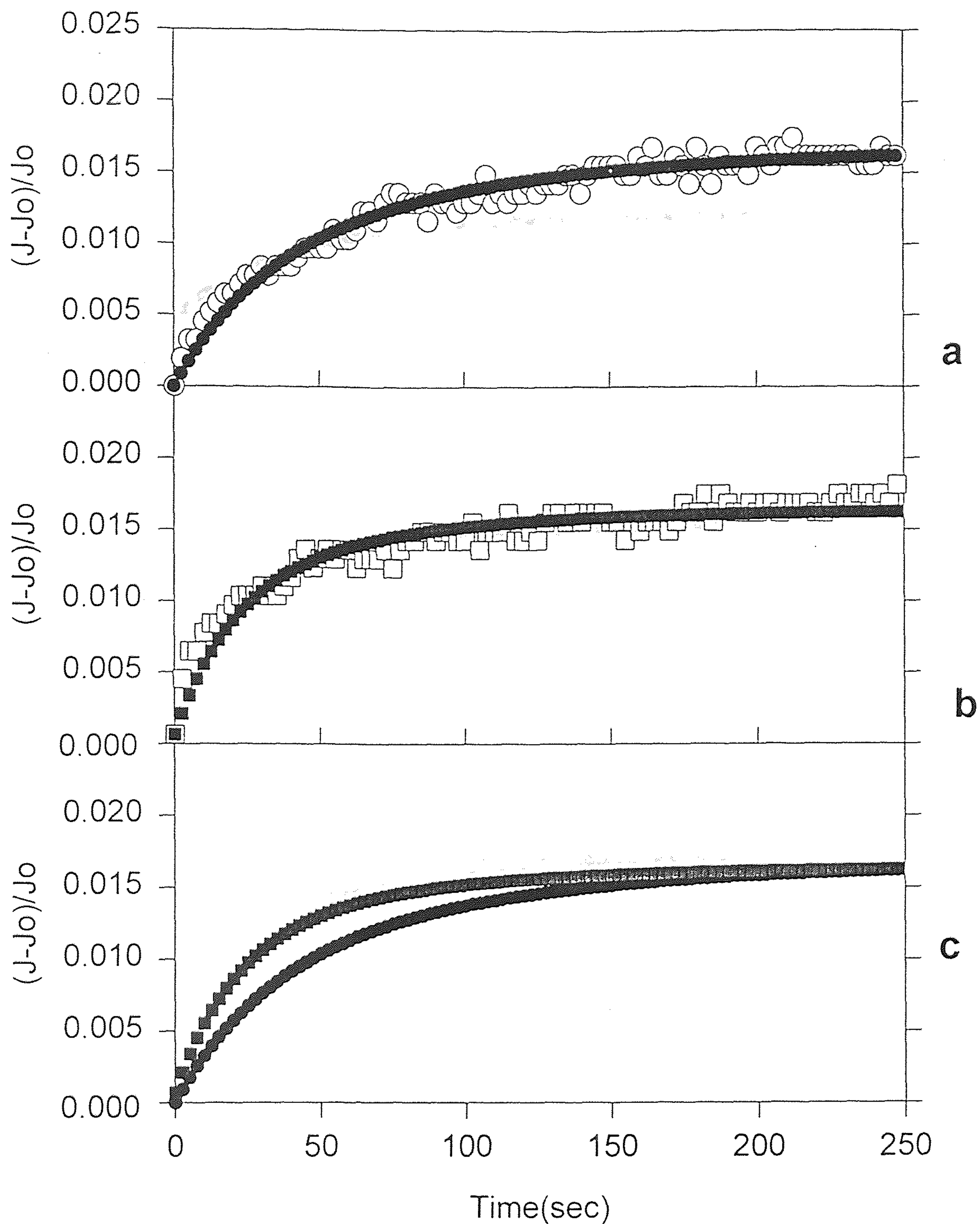


Fig. 5.2.2: Incomplete relaxation during 10 minutes unbiased. During the stress period, V_{bias}=2.1 Volts. (Wafer W8, Chip I3, Device D2C, t_{ox} = 2.8 nm, T = 300 K) (a): initial stress. (b): 2nd stress after 30 min. sit unbiased. (c): fit comparison.

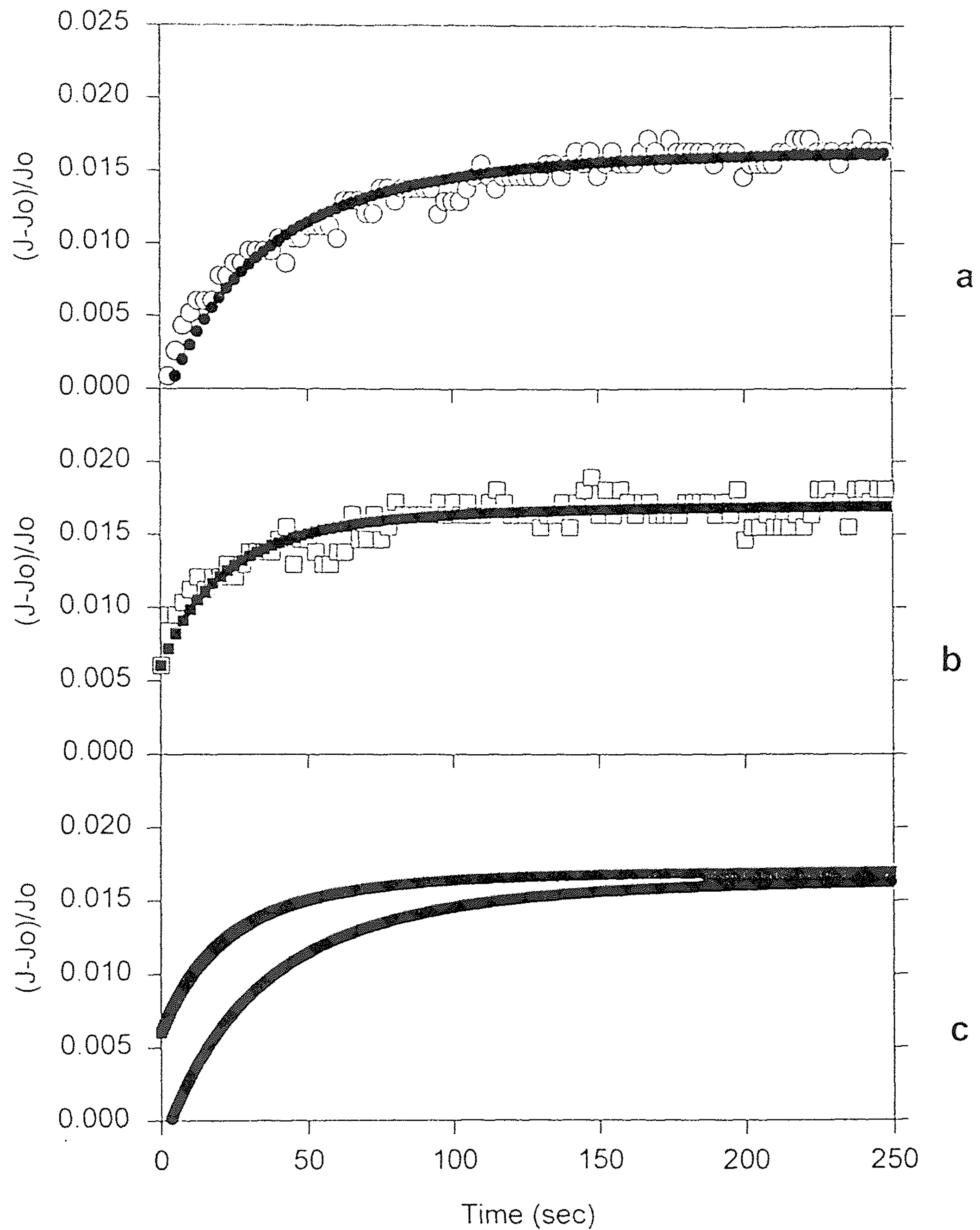


Fig. 5.2.3: Incomplete relaxation during 2 seconds unbiased. During the stress period, V_{bias}=2.1 Volts. (Wafer W8, Chip I3, Device D2A, t_{ox} = 2.8 nm, T = 300 K) (a): initial stress. (b): 2nd stress after 30 min. sit unbiased. (c): fit comparison.

The shorter the device rest at zero bias, the less relaxation occurred. This result is demonstrated again in Fig. 5.2.4 which summarizes the data in the previous three figures, and includes data for an additional 12 hour sit unbiased. For clarity, only the fits to the degradation data are shown. The time at 0 V, t_{sit}, ranges from 2 sec to 12 hr for four different devices. Notice that after 12 hours, nearly complete relaxation occurs. In other words, the degradation curve measured on a fresh device is nearly the same as the curve measured on the same device after the degradation followed by a 12 hour sit at $V_g = 0$ V.

The dynamics of the relaxation can be studied by analyzing the small differences between the curves in each section of Figure 5.2.4. This comparison is made in Figure 5.2.5 which plots each of the four fit parameters versus t_{sit}. These fit parameters are obtained from the model developed in Chapter 6. Notice that for the fresh devices, the parameters do not vary with t_{sit}, showing that the initial degradation of each device is similar, as expected. The following observations can be made concerning the previously stressed devices: 1) The parameter N_{max} is essentially unchanged from that for the initial degradation period, implying that the second degradation period is mainly a recovery of the original number of defects which relaxed during the interval t_{sit}. (A slight amount of additional new defect creation does occur during the recovery period, as will be shown below in the discussion of Fig. 5.3.4) 2) The parameter N_{min} is already less than 35% of N_{max} after only 2 seconds of relaxation, and decreases to zero after several hours. 3) The time constant τ increases with increasing t_{sit}, while the prefactor K decreases with increasing t_{sit}. Both parameters approach their original value for fresh devices, indicating that during relaxation the shallowest defects relax first. It is important

to note that neither τ nor K reach their original values even after a 12 hour sit at $V_g = 0$ V, but N_{min} reaches its original value sooner than this, showing that relaxation involves more than just elimination of defects as sources of increased tunnel current. Evidently additional structural relaxation occurs long after the defect is eliminated as a current source, placing the relaxed defect precursors physically deeper into the oxide.

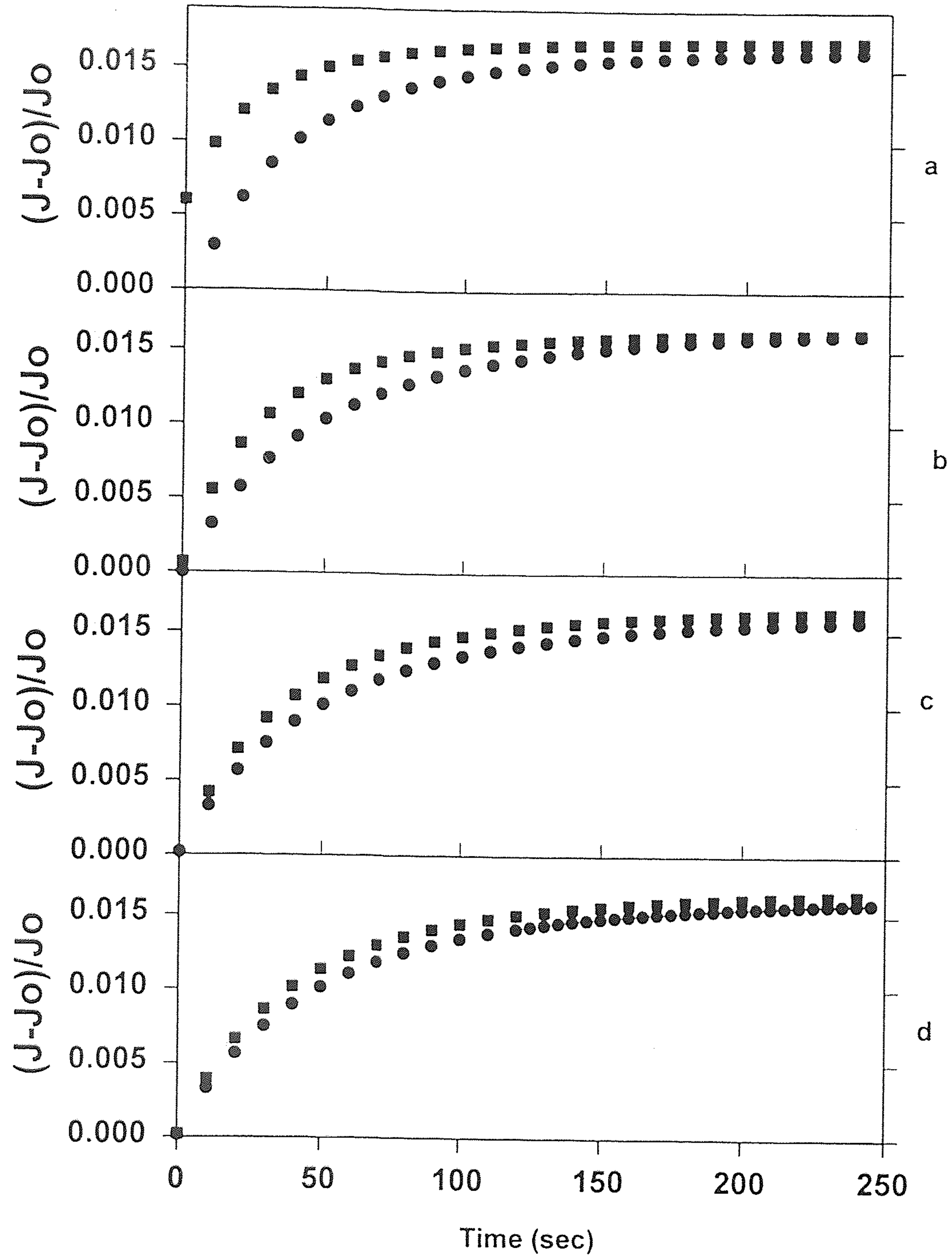


Fig. 5.2.4: Initial and second degradation curves measured on four different devices. The relaxation period between the two curves during which the device is biased at Vg = 0 Volts is 2 sec., 30 min., 2 hr., and 12 hr. for figures a, b, c, d, respectively.

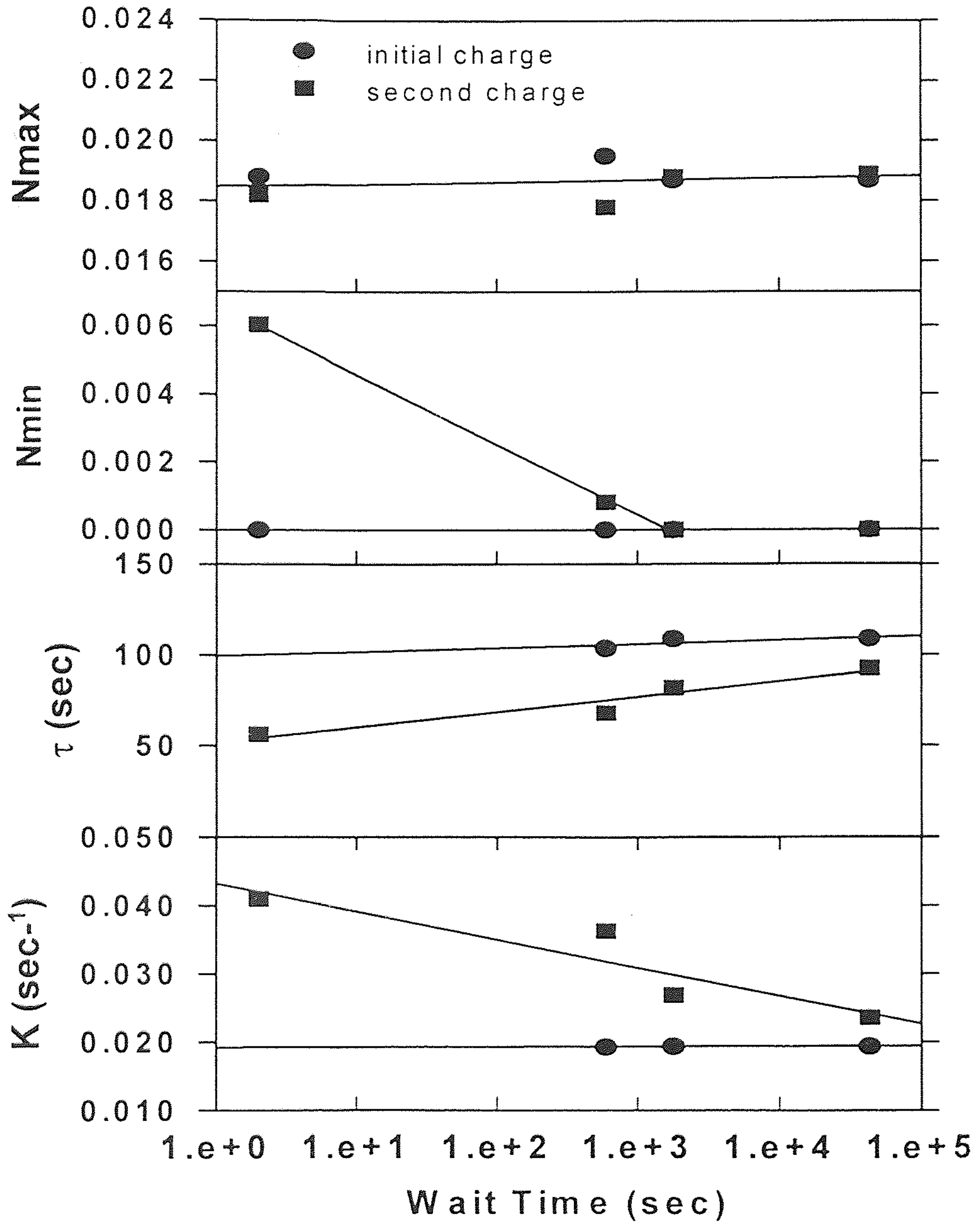


Fig. 5.2.5: Relaxation time dependence of the fit parameters resulting from the data in Fig. 5.2.4.

5.3 Different Bias Conditions

The relaxation effect in these devices is also investigated under different bias conditions. In order to get more information about the effectiveness of the electric field on the relaxation, we also designed a series of experiments to measure the charge build-up process with the relaxation sit period occurring not at zero bias as in the previous section, but at different bias levels. Figure 5.3.1 shows the fractional increase in current at Vg=2.1 V measured after 2 hour relaxation periods at various bias levels ranging from 0.0 to 1.8 V. The open squares are the experimental data and the solid squares are fit curves. To make the graph more compact the Figure does not show the time periods of relaxation. In order to make the comparison more clearly, we replot every (J-Jo)/Jo vs. time on the same scale and fit every curve with our model ignoring the points which exceed the normal range. Figure 5.3.2 shows these curves in the order of V_{sit} as labeled. Again, the open symbols represent the experimental data, while the solid symbols represent the fit. Despite some fluctuation, it is clear that the final values reached during the degradation periods are always about the same, indicating that N_{max} remains relatively constant after the initial stress. However, the values of N_{min} vary significantly.

Since the experimental data shows fluctuations, we used the fit curves to analyze the voltage dependent behavior. The parameters extracted from these fit curves are plotted with respect to the sit voltages in Fig. 5.3.3. All these parameters come from the model described in Chapter 6.

From Figure 5.3.3, the following observations can be made: 1) There is a slight increase in N_{max} with increasing V_{sit} indicating that a small amount of defect creation

occurs during each of the recovery periods. 2) The parameters τ and K do not change significantly with V_{sit} indicating that after the initial degradation period, the subsequent regeneration periods sample roughly the same distribution of defect precursors. 3) The parameter N_{min} clearly increases with increasing V_{sit}, especially above $V_g \sim 1.0$ V. If the degradation occurs mainly above a threshold voltage of $V_g \sim 1.0$ V, as has been reported for substrate injection[12,13], then one would expect to observe nearly complete relaxation below the threshold (i.e., $N_{min} = 0$), and relaxation of a decreasing number of defects at increasing voltages above 1.0 V (i.e., N_{min} increases with increasing V_g). The N_{min} data in Figure 5.3.2 are consistent with this interpretation.

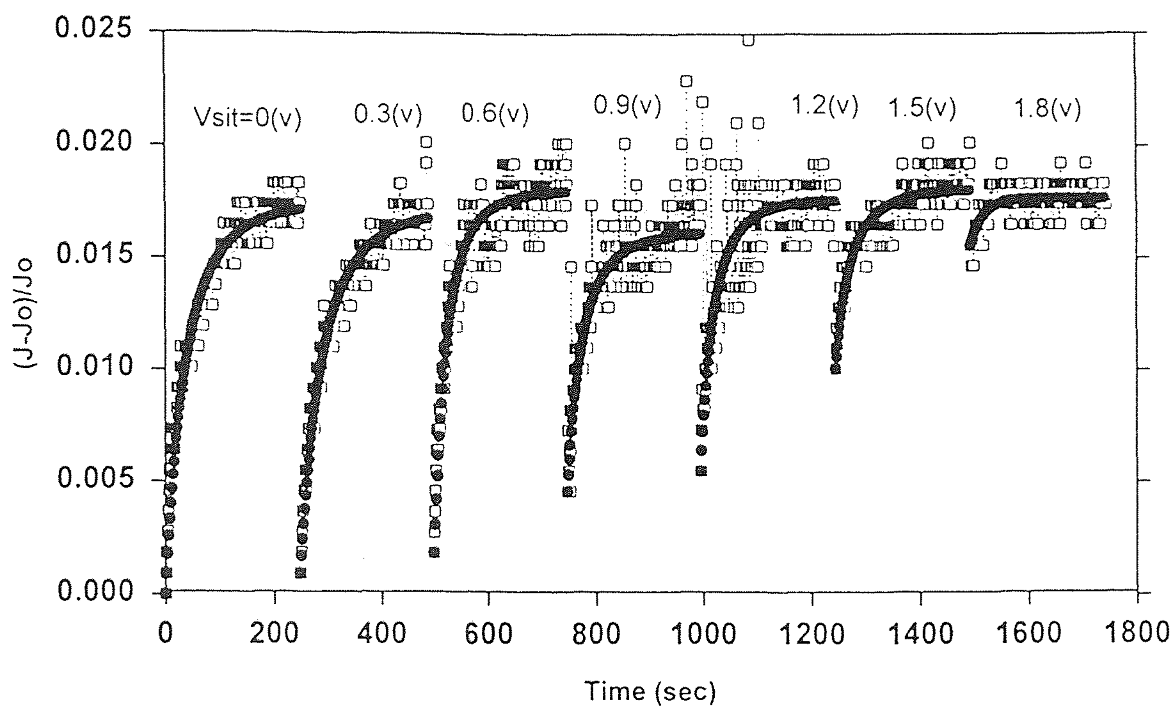


Fig. 5.3.1: Fractional increase in current at Vg=2.1 V versus time, measured after 2 hours biased at different levels ranging from 0.0 to 1.8 V.

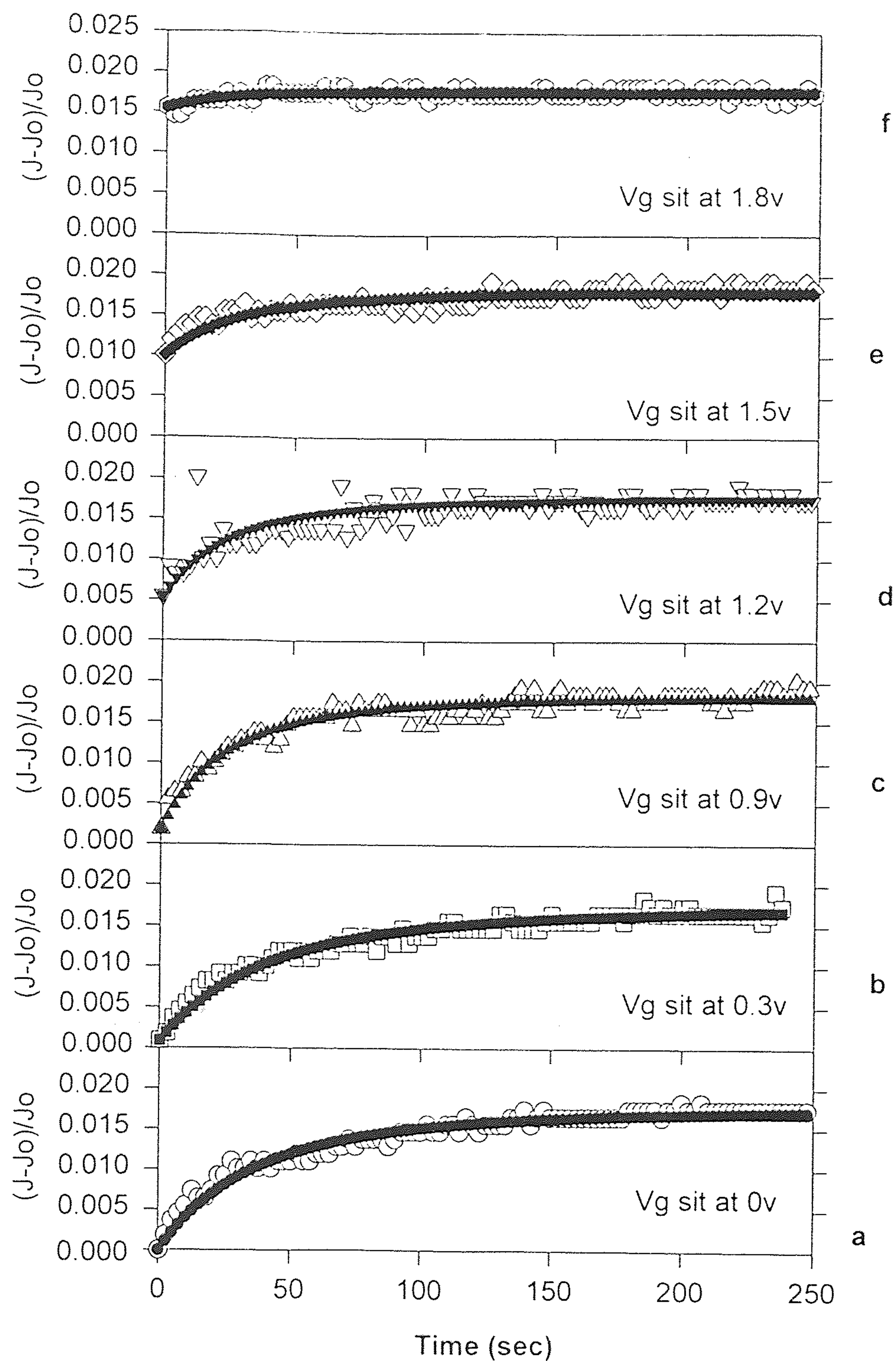


Fig. 5.3.2: Replot of the data of Figure 5.3.1.

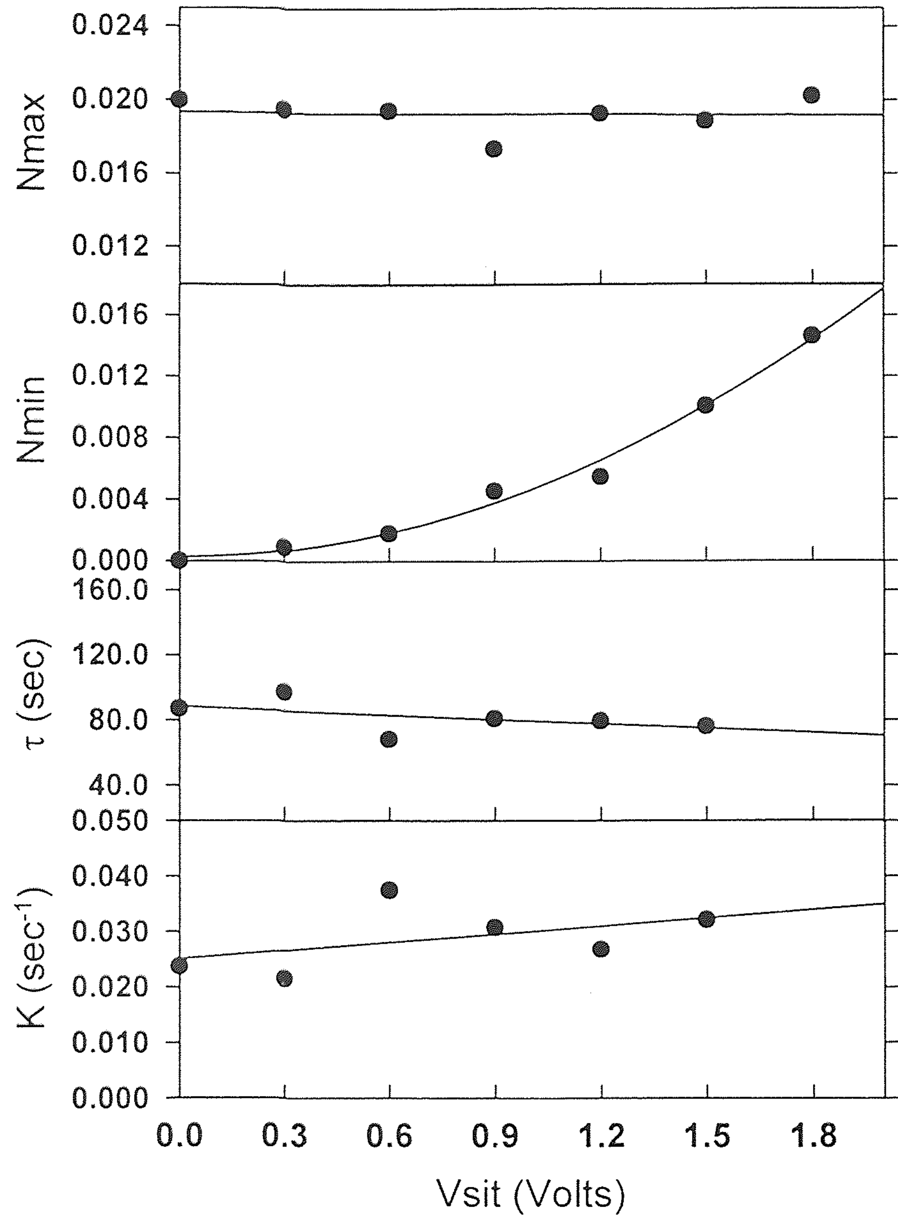


Fig. 5.3.3: Relaxation voltage dependence of the fit parameters resulting from the data in Fig.5.3.2. The solid dots are the parameters of every fit curve. The lines show the regression fits to these dots.

CHAPTER 6

MODELS

The degradation effect in MOS devices is manifested by the fractional increase in current density versus time. The gradually increasing current over time has previously been shown to correspond both to an increase in positive charge trapped in the ultra-thin oxide which increases the tunnel probability for DT (direct tunneling) electrons[12] and to the growth of a defect-assisted conduction path through the oxide.[14] If the magnitude of the current enhancement arising from newly created individual defects is randomly distributed in time about some mean value, then because a large number of defects is created in our large area device, the fractional increase in the current can be used as a measure of the number of these defects, with each defect assumed to contribute the average current enhancement. Thus we take the fractional increase in current density $(J-J_O)/J_O$, to be a measure of the number of defects created during the degradation period. In this chapter we attempt to model this degradation effect described in the two previous chapters. First, because of the shape of the data, we consider a first order kinetics model for the fractional increase in current, i.e., number of defects. Because this model does not fit our measured data very well, we go on to consider two modified first order kinetics models. Both of these models fit our data very well, but only one of them seems to have a good physical basis. At the end of the chapter we interpret the parameters that result from this physical model, and show that they have reasonable values.

6.1 First Order Kinetics

At first, based on the curvature of our experimental data, we tried fitting our data by the model of first order kinetics:

$$\frac{dn}{dt} = \frac{1}{\tau}(N_{max} - n) \qquad (6.1)$$

where n is the density of conductive units, i.e., precursors turned into active defects. N_{max} is the maximum density of defect precursors, and t is the time constant for defect creation. Integration of equation 6.1 leads to the function used for fitting our data.

$$\int_{N_{min}}^{N} \frac{dn}{N_{max} - n} = \int_{t_o}^{t} \frac{dt}{\tau} \qquad (6.2)$$

In equation 6.2, N(t) is the number of precursors converted to active defects after a time t. In this integration we usually assume that zero converted defects are present initially, though we could just as well take this lower limit to be some non-zero value N_{min}. The result of this integration is:

$$-\ln \frac{N\max - N}{N\max - N\min} = \frac{1}{\tau}(t - t_0) \qquad (6.3)$$

Rewriting this equation leads to eq 6.4 and 6.5:

$$1 - \frac{N}{N_{max}} = ke^{-\frac{t}{\tau}} \qquad (6.4)$$

$$N = N_{max}(1 - ke^{-\frac{t}{\tau}}) \qquad (6.5)$$

The equation 6.5 is the final form of our first kinetics model, where the parameter N in the model is taken to be equivalent to our measured values of $(J\text{-}J_o)/J_o$. The

parameter k is approximately 1 for small values of N_{min} and $t_o << \tau$. When we used this equation to fit our experimental data, it failed to fit the initial part of the curve as shown in Fig. 6.1.1. If we constrain the parameters to make the function fit the initial part of the data, the final part of fit curve will flatten out at a value that is significantly below the experimental data. These fitting results imply that the first order kinetics process is more rapid during the initial part of the charging process, and slows down as the degradation proceeds. In other words the time constant for the first order process in not "constant" but decreases with increasing time. Such a finding is not unreasonable. For example, it is possible that shallower defect precursors are converted to defects more efficiently than deeper ones. This reasonable possibility could easily give rise to a time dependent time constant. Based on this possibility, we attempted a modified first order kinetics fit as described in the next section.

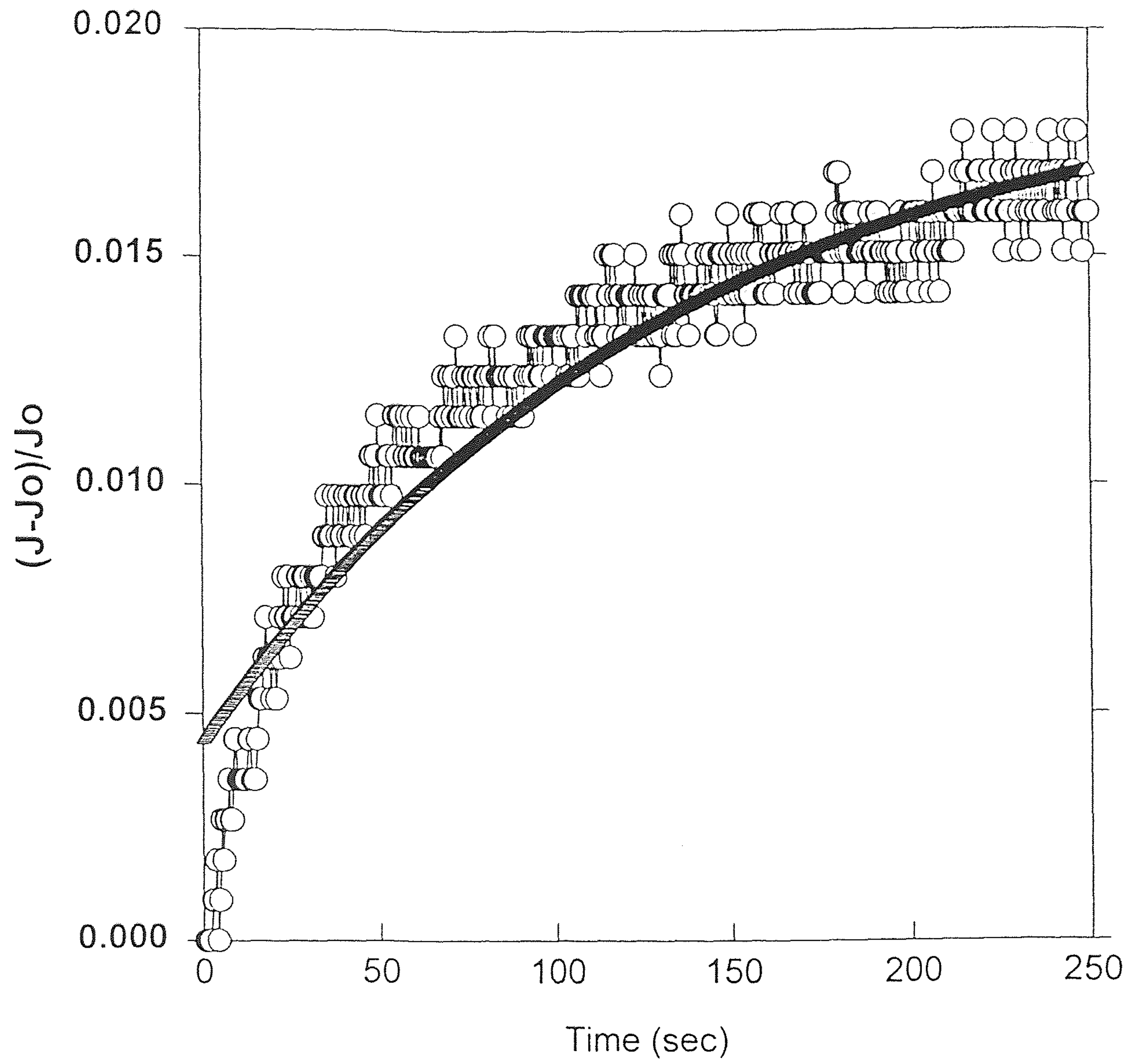


Fig. 6.1.1: First order kinetics model fit to measured data.

6.2 Linear Time Constant Model

To test our hypothesis that a time dependent time constant might better fit our data, we first assumed a linear time dependence. In this interpretation:

$$\frac{dn}{dt} = \frac{k}{t}\left(N_{max} - n\right) \tag{6.6}$$

where $\tau = t/k$ is the explicit time dependence of the time "constant." The parameter k is unitless. It is important to note that no justification is given to support the choice of a linear time constant other than the fact that it provides the simplest increase in τ with time. This choice is merely one of convenience, used to test our hypothesis. This differential equation is solved by the integral:

$$\int_{N_{min}}^{N} \frac{dn}{N_{max} - n} = k \int_{t_0}^{t} \frac{dt}{t} \tag{6.7}$$

Here, the lower limit N_{min} is included explicitly. The solution is:

$$\ln \frac{N_{max} - N_{min}}{N_{max} - N} = k \cdot \ln \frac{t}{t_0} \tag{6.8}$$

$$\frac{N_{max} - N_{min}}{N_{max} - N} = \left(\frac{t}{t_0}\right)^k \tag{6.9}$$

Expressing this result in its final form:

$$N = N_{max} - (N_{max} - N_{min})\left(\frac{t}{t_0}\right)^{-k} \tag{6.10}$$

As shown in Figure 6.2.1, this model provides a good fit to our measured data. However, at this point there is no physical justification for the choice of the assumed linear time dependence. In the next section we develop a model which justifies the choice of an exponential time dependence, which to first order is approximately linear in time.

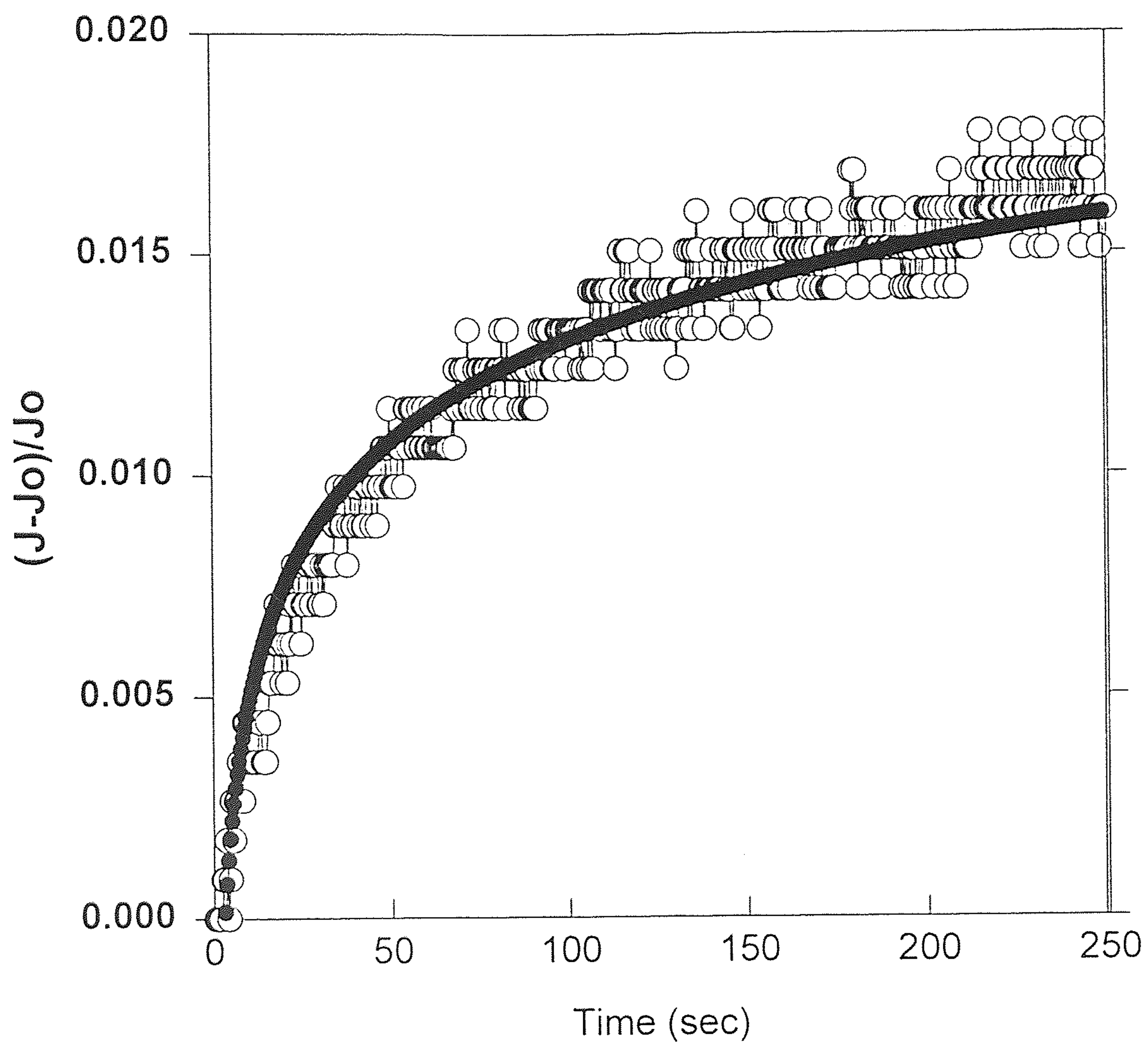


Fig. 6.2.1: Modified first order kinetics model with explicit linear time dependence of reaction time "constant", fit to measured data.

6.3 Exponential Time Dependence of Time Constant

In this section we make the reasonable assumption that the defect creation probability is exponentially dependent on the depth of the defect precursor in the oxide, implying that tunneling is an important aspect of the defect creation mechanism. Such an argument has been made previously in connection with oxide defects, for example for interpreting conductance measurements.[19] The exponential dependence on depth means that the defect creation rate will have an exponential time dependence, i.e., the precursors closer to the oxide surface will be converted to defects exponentially faster than those located farther from the oxide surface. Thus we account for the depth dependence explicitly by adding an exponential time dependence to the first order kinetics relationship. Hence the defect creation rate is given by:

$$\frac{dn}{dt} = n_h v \sigma_o e^{-t/\tau}(N\max - n) = Ke^{-t/\tau}(N\max - n) \qquad (6.11)$$

In this equation, n_h is the density of charge carriers which control the degradation, written suggestively with the subscript "h" to imply that these carriers may be holes, ν is the charge carrier thermal velocity, and $\sigma_0 e^{-t/\tau}$ is the capture cross section for these charges, including the explicit dependence on the oxide depth by way of the exponential time dependence with time constant τ. The parameter Nmax is again the total number of defect precursors and n is the number of precursors transformed to active defects. For simplicity, in equation 6.11 we combine the constant prefactors into one parameter K. This equation is solved by integration:

$$\int_{N\min}^{N} \frac{dn}{N\max - n} = K\int_{t_0}^{t} \frac{dt}{e^{t/\tau}} \qquad (6.12)$$

The solution to this equation can be written in its final form as:

$$N = N\max - (N\max - N\min)\exp[K\tau(e^{-t/\tau} - e^{-t_0/\tau})] \qquad (6.13)$$

Here N_{min} is again the number of defects which are already present at the start of the degradation period, assumed to be zero for the initial stressing of the device, but depending on the extent of relaxation, not necessarily zero for subsequent stressing periods. The last exponential in equation 6.13 is approximately 1 for values of t_o which are small compared to τ. As shown in Figure 6.3.1, the agreement between the fit and the data is quite good. In Figure 6.3.2 we compare the fit results of all three models. While both the linear and exponential models give good fit results, the exponential model gives the best fit as measured by correlation coefficients. Also the exponential model is developed with a sound physical basis. A comparison of the fit parameters and correlation coefficients for the three models is presented in Table 6.1. In the next section we will discuss the physical interpretation of the parameters determined using the exponential model.

Table 1: Fit parameters and correlation coefficients for the three models. Note that the parameters τ and k have different meaning in the different models.

	first order model		linear model		exponential model	
	parameter value	correlation coefficient	parameter value	correlation coefficient	parameter value	correlation coefficient
Nmin	0	1	0	1	0	1
Nmax	1.910×10^{-2}	0.7685	5.000×10^{-2}	0.99903	1.867×10^{-2}	0.99998
τ (sec)	132.696	0.78522	--	--	108.0	0.99999
k(unitless) or K(sec^{-1})	0.7685	0.5353817	0.0868	0.99946	4.208×10^{-2}	0.99996

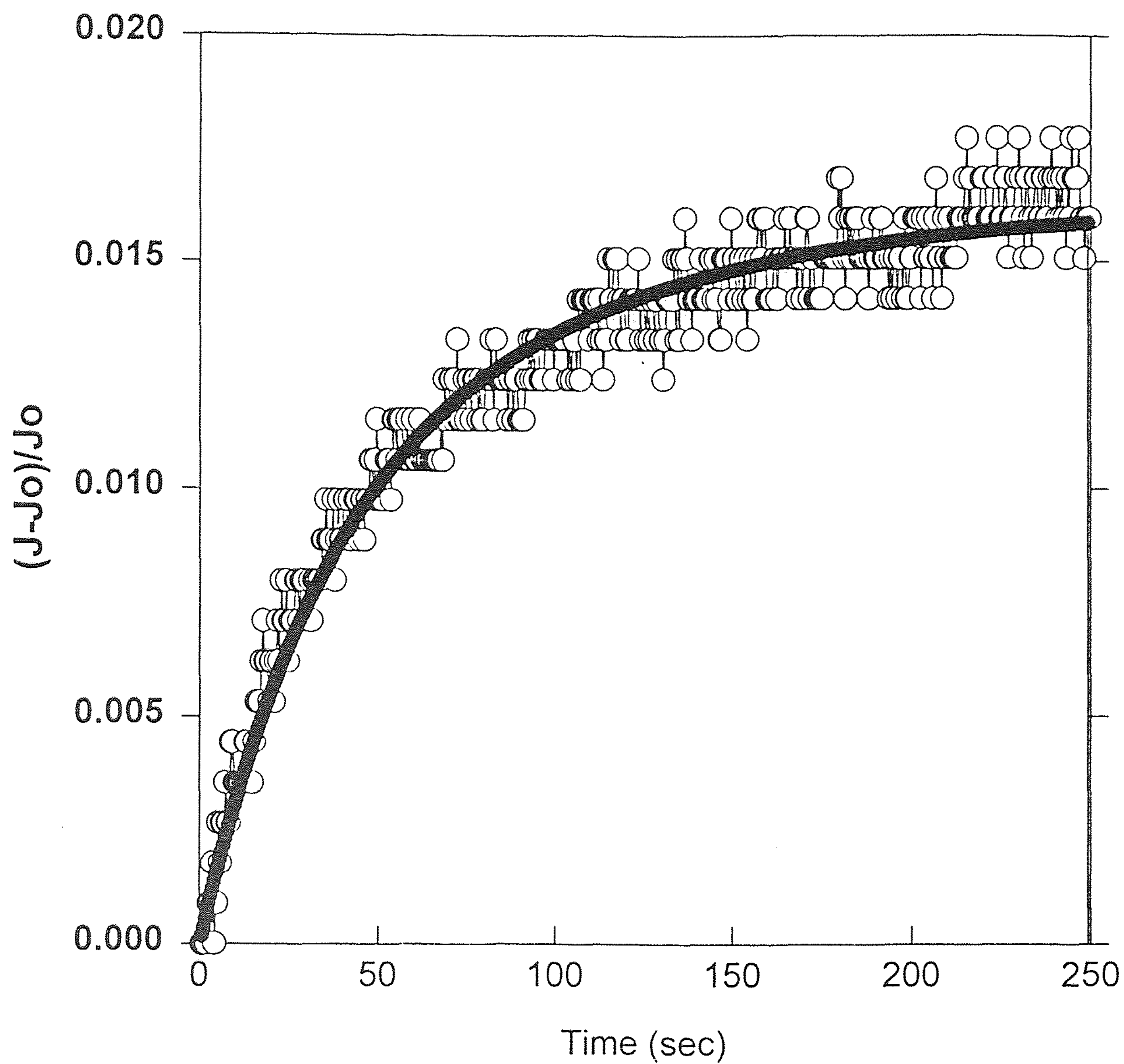


Fig. 6.3.1: Modified first order kinetics model with explicit exponential time dependence of reaction time "constant", fit to measured data.

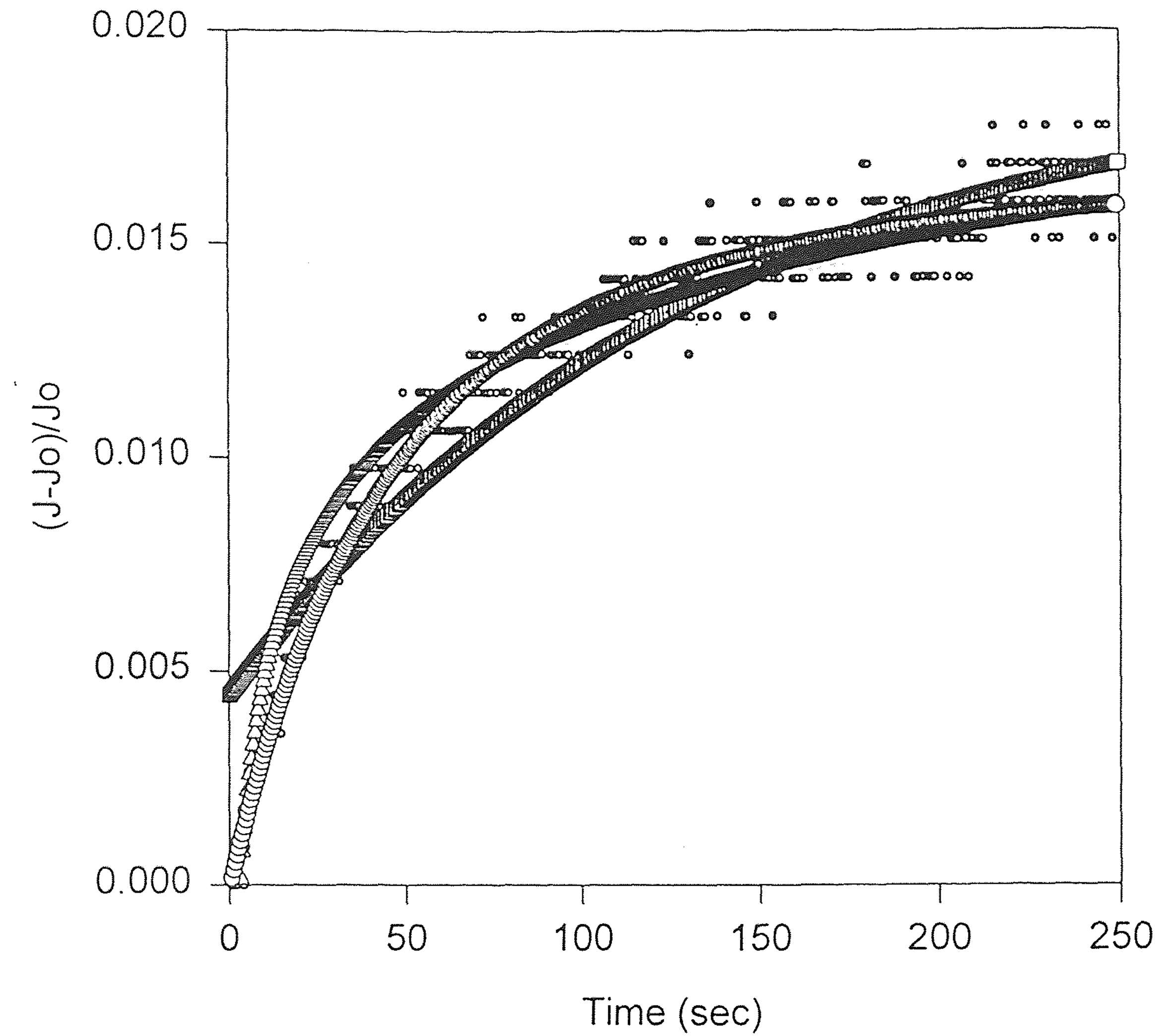


Fig. 6.3.2: Comparison of the three models. The small open circle is the experimental data. The open square line is the fit to the first order kinetic model, the open triangular line is the fit to the linear time constant model, and the large open circle line is the fit to the exponential time constant model.

6.4 Discussion of Fit Parameters from Exponential Model

The effect of the fit parameter N_{max} on the behavior of the predicted degradation curves is straight forward to understand, but the effects the other parameters, τ and K in particular are not as easy to understand without an illustration. Figures 6.4.1, 6.4.2 and 6.4.3 show the effects of these three parameters, N_{max}, K and τ, respectively on the degradation predictions. In Figure 6.4.1, an increase in N_{max} merely increases the long time saturation level of the curves. The knee in the curve shifts to slightly higher times as N_{max} is increased. In Figure 6.4.2, an increase in $K\tau$ sharpens initial rise in the curve, and flattens the long time saturation level. The knee in these curves shifts to lower times for higher values of $K\tau$. Finally, in Figure 6.4.3, The an increase in the parameter τ alone lowers the slope of the initial degradation increase, as might be expected. The overall behavior of the predicted curves is a convolution of the effects of these parameters.

The physical interpretation of the parameters N_{min} and N_{max} has already been discussed in previous chapters. Essentially these are the minimum and maximum numbers of converted precursors in the device. We interpret the values of ν and n_h by assuming that ν is the drift velocity of the charge carriers responsible for the degradation. In this interpretation, the current component of these charge carriers is given by

$$J_h = n_h q \nu \qquad (6.13)$$

where q is the absolute charge of an electron. Thus a reasonable value of the cross-section, for example $\sigma_o \sim 10^{-16}$ cm^2, gives a charge carrier current density of $J_h \sim 10^{-12}$ A/cm^2. This current density is very low suggesting that roughly one defect is created per

injected charge carrier. If a first order kinetics process is involved, it is highly efficient. Perhaps the entire charge creation current is due to the charge flowing into the individual defect precursors.

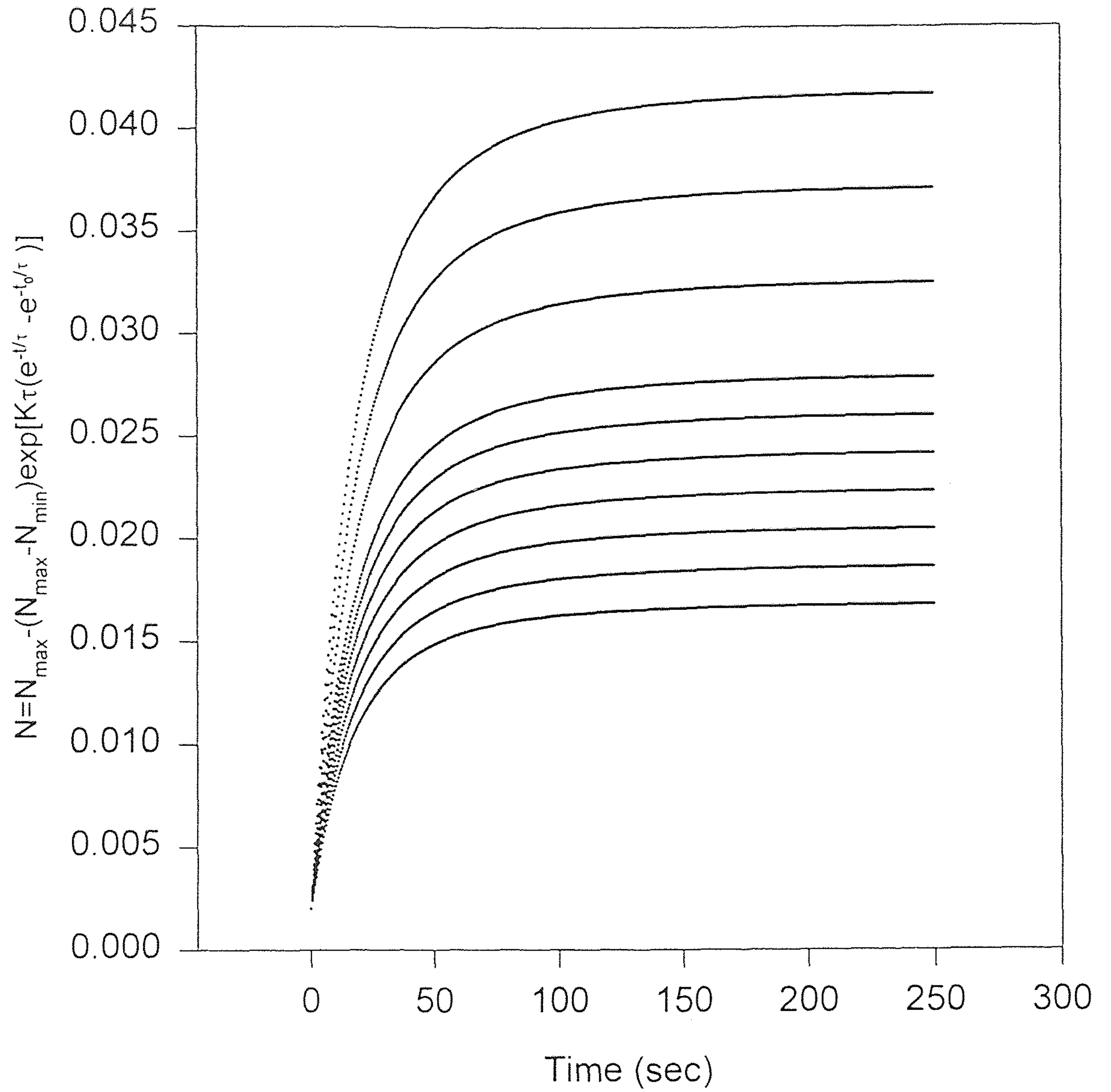


Fig. 6.4.1: Behavior of the exponential model function in response to variations in N_{max} ranging from 0.018 to 0.045. The parameters τ, N_{min} and K are held fixed with values 50 sec, 0.0020, 2.6, respectively.

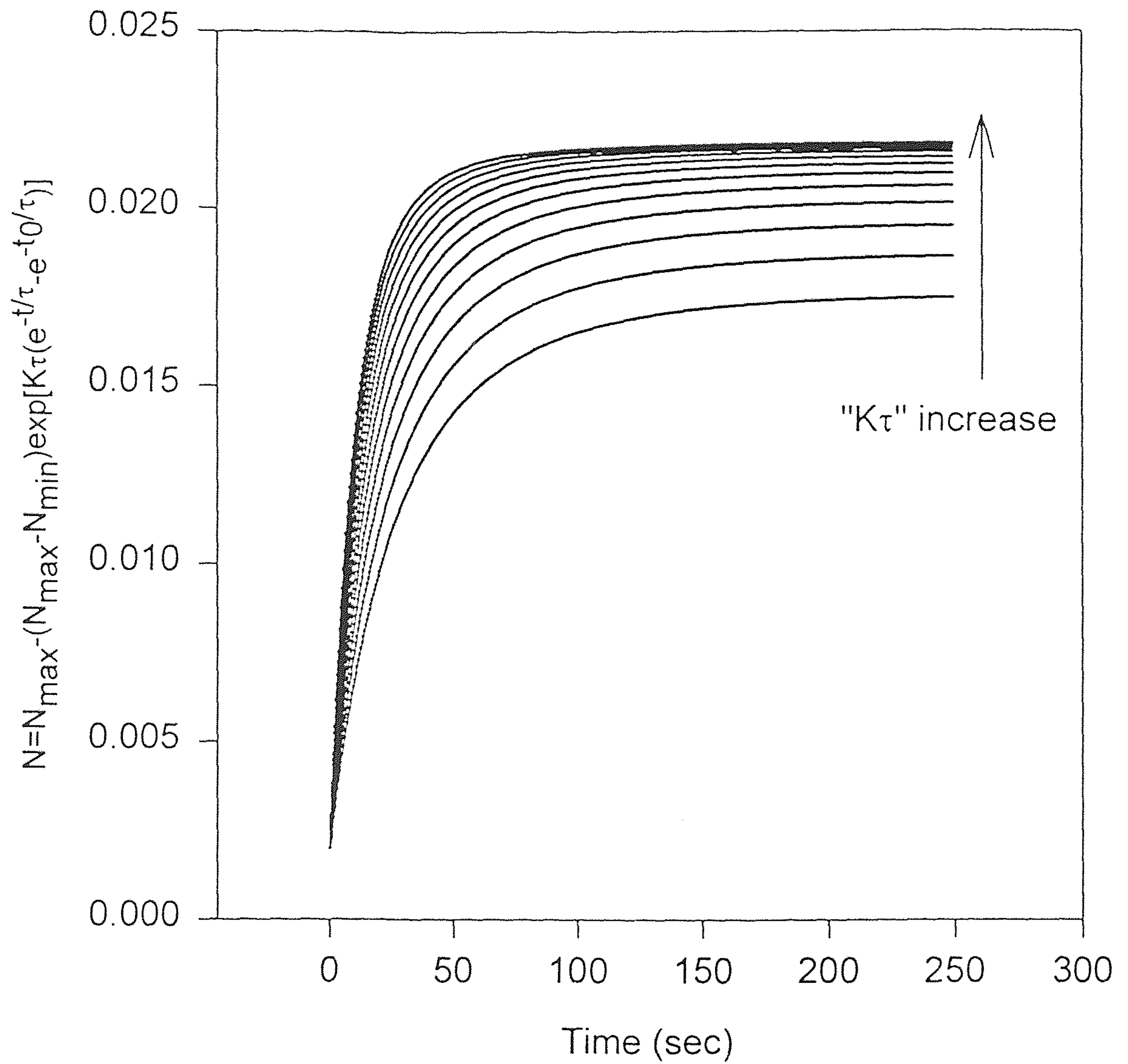


Fig. 6.4.2: Behavior of the exponential model function in response to variations in $K\tau$ ranging from 1.5 to 4.8 in increments of 0.3. The parameters τ, N_{min} and N_{max} are held fixed with values 50 sec, 0.0020, 0.022, respectively.

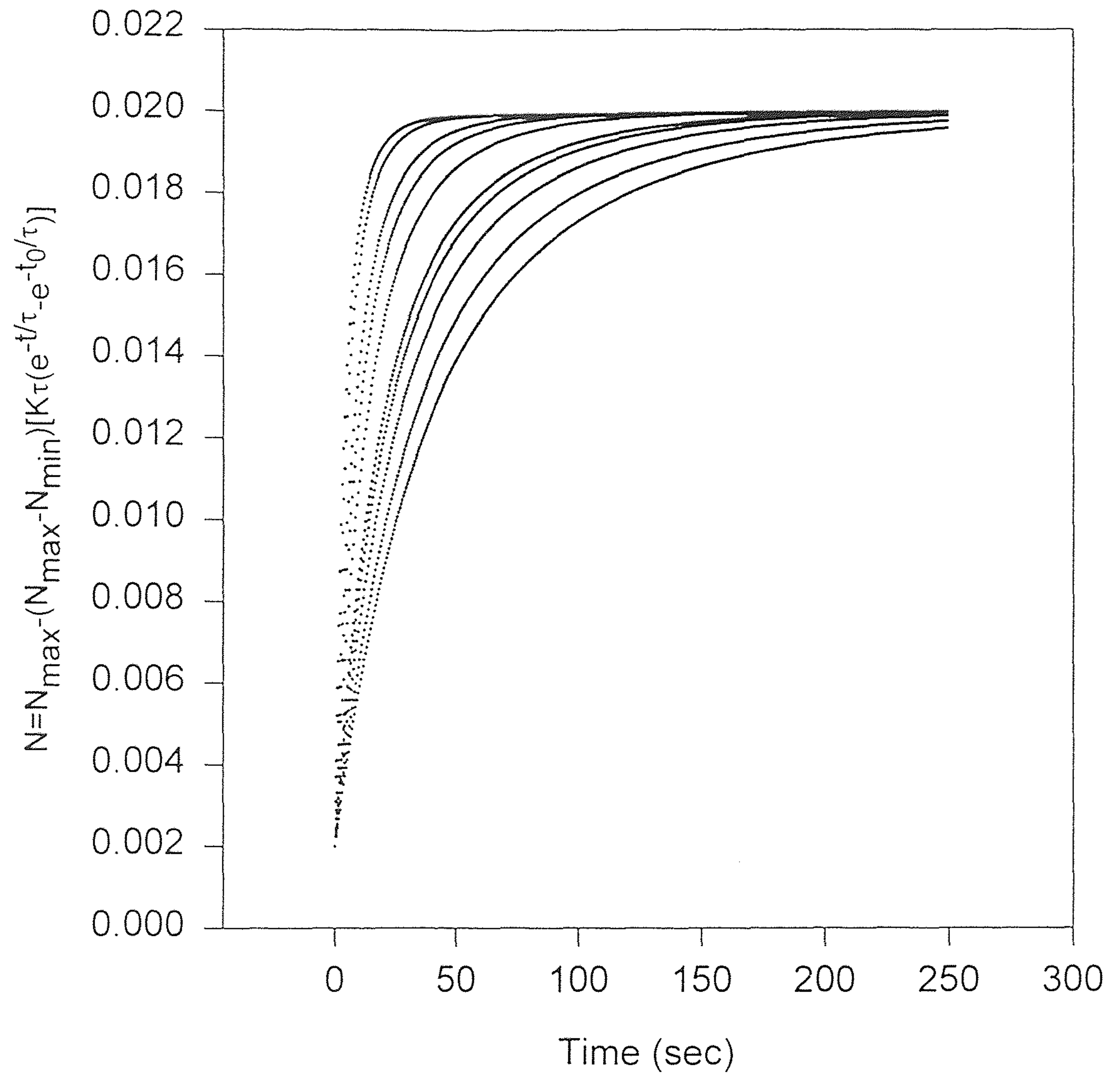


Fig. 6.4.3: Behavior of the exponential model function in response to variations in τ ranging from 10 to 100 sec. The parameters $K\tau$, N_{min} and N_{max} are held fixed with values 2.3, 0.0020, 0.022, respectively.

CHAPTER 7

CONCLUSION

In conclusion, we have investigated defect instability in ultra-thin oxides stressed by injection from n-type silicon substrates into degenerate n-type polycrystalline silicon gates. In this system nearly complete relaxation occurs, allowing the possibility of regenerating the defects. By modeling the defect generation process and examining differences between original and subsequent defect creation periods, we have obtained an improved picture of the relaxation process and the nature of the involved defects. Also, we have demonstrated that a straightforward first-order kinetics process may underlie the degradation. The success of our model suggests that a conduction process other than the dominant direct tunnel current component may drive the degradation with high efficiency.

APPENDIX

1) Two-page Abstract of a presentation at INFOS 97, June 1997, Gothenburg, Sweden.

2) Four-page Paper to be published in the Journal of Non-Crystalline Solids, 1997.

Defect Instability in Ultra-Thin Oxides on Silicon

L. Xie[a], K. R. Farmer[a] and D. A. Buchanan[b]

[a]Dept. of Physics, New Jersey Institute of Technology, University Heights, Newark, NJ 07102
[b]IBM Research Dividsion, T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598

The importance of studying defect dynamics in ultra-thin oxides on silicon has increased recently due to reports of logic elements being fabricated using sub-3.5 nm gate oxides.[1,2] In such thin dielectric layers, significant direct tunneling (DT) through the trapezoidal oxide barrier occurs, and with it, a new defect creation mode, characterized by an increasing current over time at constant voltage.[3] It has been demonstrated that this DT degradation is a universal defect creation mode, found in devices fabricated at different facilities and undergoing either substrate or gate electron injection.[4,5] One recent study has focussed on the instability of these defects, reporting a partial relaxation of the degradation for the case of injection from the gate into p-type silicon substrates.[6] In the present study we extend the investigation of the defect instability to the case of injection from n-type silicon substrates into degenerate n-type polycrystalline silicon gates. Unlike the situation involving p-type silicon and gate injection, we find that in the n-type system with substrate injection, nearly complete relaxation occurs, allowing the possibility of regenerating the defects. By modelling the defect generation process and examining differences between original and subsequent defect creation periods, we are able to obtain an improved picture of both the relaxation process and the nature of the involved defects. An additional contribution of this work is the demonstration that a straightforward first-order kinetics process may underlie the degradation. This is surprising since it has been shown previously that the dominant electron tunnel current does not control the rate of degradation.[3,5] The success of our model suggests that a conduction process other than this dominant direct tunnel current component may drive the degradation.

The DT degradation effect is illustrated in Figure 1 for a 2.8 nm device biased at constant gate voltage, V_g = 2.1 V. The gradually increasing current over time has previously been shown to correspond both to an increase in positive charge trapped in the ultra-thin oxide which increases the tunnel probability for DT electrons[6], and to the growth of a defect-assisted conduction path through the oxide[7], with the vertical axis in Figure 1, $(J-J_0)/J_0$, being a measure of the number of defects created during the degradation period.

The thick line passing through the data in Figure 1 is the result of a fit to a modified first-order kinetics model. In this model we make the reasonable assumption that the defect creation probability is exponentially dependent on the depth of the defect precursor in the oxide, implying that tunneling is an important aspect of the defect creation mechanism. Such an argument has been made previously in connection with oxide defects, for example for interpreting conductance measurements.[8] The exponential dependence on depth means that the defect creation rate will have an exponential time dependence, *i.e.*, the precursors closer to the oxide surface will be converted to defects exponentially faster than those located farther from the oxide surface. Thus we account for the depth dependence by explicitly adding an exponential time dependence to the first-order kinetics relationship. As seen in Figure 1, the agreement between the model and the data is quite good. As will be shown in the full paper, the model provides physically meaningful parameter values.

The relaxation effect in these devices is demonstrated in Figure 2 which compares the degradation of a fresh device at V_g = 2.1 V with that of the same device measured at V_g = 2.1 V after a brief period biased at V_g = 0 V. For clarity, only the fits to the degradation data are shown. The time at 0 V, t_{sit}, ranges from 2 sec to 12 hr for four different devices. Notice that after 12 hours, nearly complete relaxation occurs.

The dynamics of the relaxation can be studied by analyzing the small differences between the curves in each section of Figure 2. This comparison is made in Figure 3 which plots each of the four fit parameters versus t_{sit}. Notice that for the fresh devices, the parameters do not vary with t_{sit}, showing that the initial degradation of each device is similar, as expected. The following observations can be made concerning the previously stressed devices: 1) The parameter n_{max} is essentialy unchanged from that for the initial degradation period, implying that the second degradation period is mainly a recovery of the original number of defects which relaxed during the interval t_{sit}. 2) The parameter n_{min} is already less than 25% of n_{max} after only 2 s of relaxation, and decreases to zero after several hours. 3) The defect generation time constant τ increases with increasing t_{sit}, while the prefactor K (proportional to defect cross-section and charge carrier density) decreases. Both parameters approach their original value for fresh devices, indicating that during relaxation the shallowest defects relax first.

Additional information can be learned by investigating the bias dependence of the relaxation. Figure 4 shows the fractional increase in current at V_g = 2.1 V measured after 2 hr relaxation periods at various bias levels ranging from 0.0 to 1.8 V. The voltage dependence of the parameters extracted from these data is shown in Figure 5. The most important observation from these data is that the parameter n_{min} clearly increases with increasing V_{sit}, especially above $V_g \sim 1.0$ V. If the degradation occurs mainly above a threshold voltage of $V_g \sim 1.0$ V, as has been reported for

substrate injection[5,6], then one would expect to observe nearly complete relaxation below the threshold (i.e., n_{min} = 0), and relaxation of a decreasing number of defects at increasing voltages above 1.0 V (i.e., n_{min} increases with increasing V_g). The n_{min} data in Figure 5 are consistent with this interpretation.

References

1. Y. Taur et al., IEDM Tech. Dig. (1992) 35.6.1.
2. H. Mamose, M. Ono, T. Yoshitomi, T. Ohguro, S. Nakamura, M. Saito and H. Iwai, IEEE Trans. Electron Dev. 43 (1996), 1233.
3. K. R. Farmer, M. O. Andersson and O. Engström, Appl. Phys. Lett. 60 (1992) 730.
4. P. Lundgren, M. O. Andersson, K. R. Farmer and O. Engström, Microelectronic Engineering 28 (1995) 67.
5. K. R. Farmer, C. Debauche, A Giordano, P. Lundgren, M. O. Andersson and D. A. Buchanan, Appl. Surf. Sci. (1996), in press.
6. P. Lundgren, M. O. Andersson, K. R. Farmer and O. Engström, J. Noncryst. Solids 187 (1995) 140.
7. K. R. Farmer, P. Lundgren, M. O. Andersson and O. Engström, Proc. MRS Symp. 284 (1993) 229.
8. H. Preier, Appl. Phys. Lett. 10 (1967) 361.

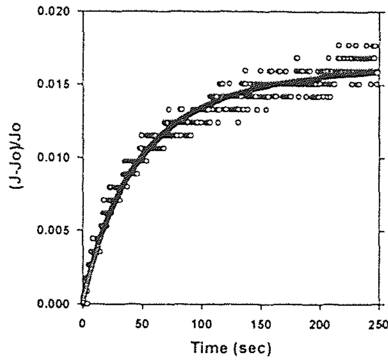


Figure 1. Fractional increase in current at V_g = 2.1 V versus time. The thick line through the data is the result of the fit described in the text, with n_{max} = 0.0187, n_{min} = 0, K = 0.0194 sec^{-1} and τ = 108 sec.

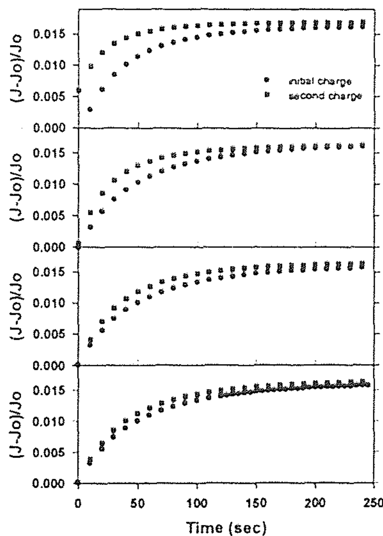


Figure 2. Initial and second degration curves measured on four different devices. The relaxation period between the two curves during which the device is biased at V_g = 0 V is 2 sec, 30 min, 2 hr, and 12 hr for figures a, b, c and d, respectively.

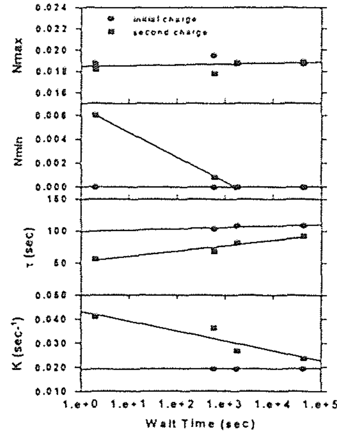


Figure 3. Relaxation time dependence of the fit parameters resulting from the data in Figure 2.

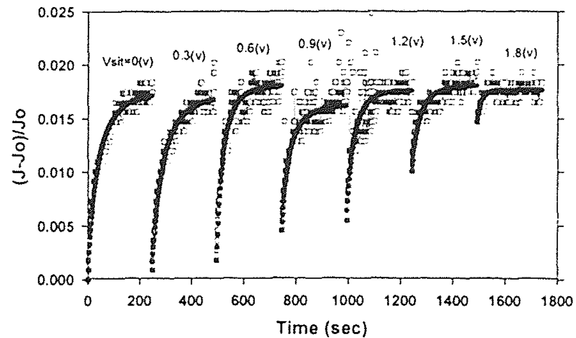


Figure 4. Fractional increase in current at V_g = 2.1 V versus time, measured after 2 hour relaxation periods at various bias levels ranging from 0.0 to 1.8 V.

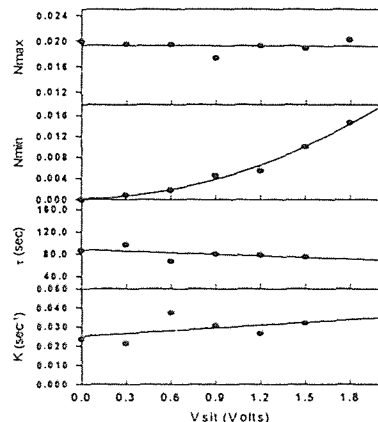


Figure 5. Relaxation voltage dependence of the fit parameters resulting from the data in Figure 4.

Defect Instability in Ultra-Thin Oxides on Silicon

L. Xie[a], K. R. Farmer[a] and D. A. Buchanan[b]

[a]Department of Physics, New Jersey Institute of Technology,
University Heights, Newark, NJ 07102, USA

[b]IBM Research Division, T. J. Watson Research Center,
P.O. Box 218, Yorktown Heights, NY 10598, USA

The instability of defects created in ultra-thin insulator, metal-oxide-silicon devices biased in the direct tunnel regime is investigated. For the case of electron injection from the silicon substrate, nearly complete defect relaxation is observed after the bias is removed, allowing the possibility of re-generating the defects. Modeling the defect generation process and examining differences between initial and subsequent degradation periods lead to an improved picture of both the relaxation process and the nature of the involved defects.

1. INTRODUCTION

The importance of studying defect dynamics in ultra-thin oxides on silicon has increased recently due to reports of logic elements being fabricated using sub-3.5 nm gate oxides.[1,2] In such thin dielectric layers, significant direct tunneling (DT) through the trapezoidal oxide barrier occurs, and with it, a new defect creation mode, characterized by an increasing current over time at constant voltage.[3] It has been demonstrated that this DT degradation is a universal defect creation mode, found in devices fabricated at different facilities and undergoing either substrate or gate electron injection.[4,5] One recent study has focused on the instability of these defects, reporting a partial relaxation of the degradation for the case of injection from the gate into p-type silicon substrates.[6] In the present study we extend the investigation of the defect instability to the case of injection from n-type silicon substrates into degenerate n-type polycrystalline silicon gates. Unlike the situation involving p-type silicon and gate injection, we find that in the n-type system with substrate injection, nearly complete relaxation can occur at 300 K, allowing the possibility of re-generating the defects. By modeling the defect generation process and examining differences between initial and subsequent defect creation periods, we are able to obtain an improved picture of both the relaxation process and the nature of the involved defects. Besides this increased understanding of the defects, an additional important contribution of this work is the demonstration that a straightforward first-order kinetics process may underlie the degradation. This is surprising since it has been shown previously that the dominant electron tunnel current does not control the rate of degradation.[3,5] The success of our model suggests that a conduction process other than this dominant direct tunnel current component may drive the degradation.

2. DEVICE DESCRIPTION

The devices used in this study have been fabricated on the silicon line at IBM, Yorktown Heights. Active tunnel oxides of four different nominal thicknesses, 2.0, 2.4, 2.8 and 3.4 nm are formed at 700 °C in dry O_2 in windows opened in a field oxide grown on <100> oriented, 0.005-0.02 ohm-cm, n-type silicon wafers. After the gate oxidation, the active oxide is immediately covered by polycrystalline silicon deposition. Although each chip contains window areas ranging from 1 μm^2 to 1 mm^2, measurements for this work have been carried out using mainly 250 μm × 250 μm devices. Tunnel oxide thicknesses have been estimated

electrically from capacitance-voltage characteristics. For all of the different oxides, devices exhibit current-voltage (I-V) characteristics which scale well both with device area, and with oxide thickness in good agreement with first principles tunneling theory.[7] For the measurements reported in this work, care has been taken to select devices for each thickness which display almost identical initial I-V characteristics.

3. RESULTS AND DISCUSSION

The direct tunnel degradation effect is illustrated in Figure 1 for a 2.8 nm device biased at constant gate voltage, V_g = 2.1 V. The gradually increasing current over time has previously been shown to correspond both to an increase in positive charge trapped in the ultra-thin oxide which increases the tunnel probability for DT electrons[6], and to the growth of a defect-assisted conduction path through the oxide.[8] If the magnitude of the current enhancement arising from newly created individual defects is randomly distributed in time about some mean value, then because a large number of defects is created in our large area device, the fractional increase in the current can be used as a measure of the number of these defects, with each defect assumed to contribute the average current enhancement. Thus in Figure 1, we take the vertical axis, $(J-J_o)/J_o$, to be a measure of the number of defects created during the degradation period.

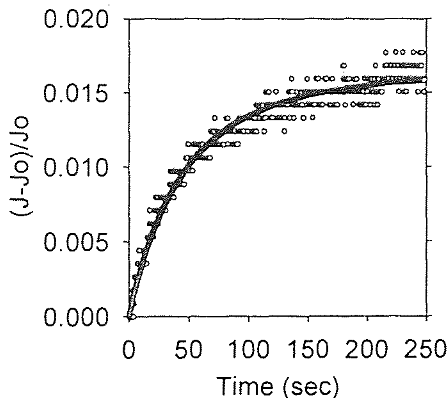


Figure 1. Fractional increase in current at V_g = 2.1 V versus time. The thick line through the data is the result of the fit described in the text, with n_{max} = 0.0187, n_{min} = 0, K = 0.0194 sec^{-1} and τ = 108 sec.

The thick line passing through the data in Figure 1 is the result of a fit to a modified first-order kinetics model. In this model we make the reasonable assumption that the defect creation probability is exponentially dependent on the depth of the defect precursor in the oxide, implying that tunneling is an important aspect of the defect creation mechanism. Such an argument has been made previously in connection with oxide defects, for example for interpreting conductance measurements.[9] The exponential dependence on depth means that the defect creation rate will have an exponential time dependence, *i.e.*, the precursors closer to the oxide surface will be converted to defects exponentially faster than those located farther from the oxide surface. Thus we account for the depth dependence by explicitly adding an exponential time dependence to the first-order kinetics relationship. Hence, the defect creation rate is given by:

$$dn/dt = n_h v \sigma_o e^{-t/\tau}(n_{max} - n) = K e^{-t/\tau}(n_{max} - n) \quad (1)$$

In this equation, n_h is the density of charge carriers which control the degradation, written suggestively with the subscript "h" to imply that these carriers may be holes, v is the charge carrier thermal velocity, and $\sigma_o e^{-t/\tau}$ is the capture cross-section for these charges, including the explicit dependence on the oxide depth by way of an exponential time dependence with time constant τ. The parameter n_{max} is the total number of defect precursors and n is the number of precursors transformed to active defects. For simplicity, we combine the constant prefactors into one parameter K. The solution to this equation can be written:

$$n(t) = n_{max} - (n_{max} - n_{min})\exp[K\tau(e^{-t/\tau} - 1)] \quad (2)$$

Here, n_{min} is the number of defects which are already present at the start of a degradation period, assumed to be zero for the initial stressing of a device, but depending on the extent of relaxation, not necessarily zero for subsequent stressing periods. As seen in Figure 1, the agreement between the fit and the data is quite good. Furthermore, the fit is consistent with physically meaningful values for the parameters n_h, v and σ_o. We interpret the values of v and n_h by assuming that v is the drift velocity of the charge carriers responsible for the degradation.

In this interpretation, the current component of these charge carriers is given by $J_h = n_h q v$, where q is the absolute charge of an electron. Thus a reasonable value of the cross-section, for example $\sigma_0 \sim 10^{-16}$ cm^2, gives a charge carrier current density of $J_h \sim 10^{-12}$ A/cm^2. This current density is very low suggesting that roughly one defect is created per injected charge carrier.

The relaxation effect in these devices is demonstrated in Figure 2 which compares the degradation of a fresh device at $V_g = 2.1$ V with that of the same device measured at $V_g = 2.1$ V after a brief period biased at $V_g = 0$ V. For clarity, only the fits to the degradation data are shown. The time at 0 V, t_{sit}, ranges from 2 sec to 12 hr for four different devices. Notice that after 12 hours, nearly complete relaxation occurs. In other words, the degradation curve measured on a fresh device is nearly the same as the curve measured on the same device after the degradation followed by a 12 hour sit at $V_g = 0$ V.

The dynamics of the relaxation can be studied by analyzing the small differences between the curves in each section of Figure 2. This comparison is made in Figure 3 which plots each of the four fit parameters versus t_{sit}. Notice that for the fresh devices, the parameters do not vary with t_{sit}, showing that the initial degradation of each device is

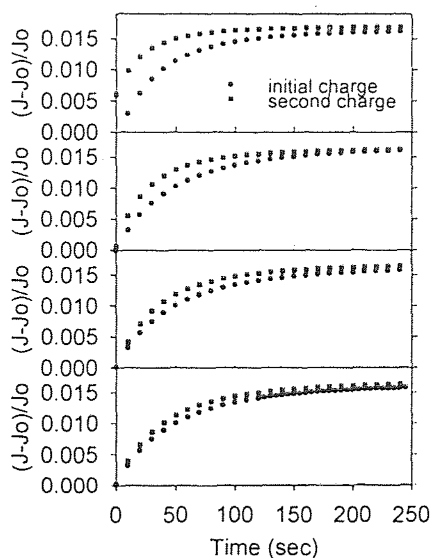


Figure 2. Initial and second degradation curves measured on four different devices. The relaxation period between the two curves during which the device is biased at $V_g = 0$ V is 2 sec, 30 min, 2 hr, and 12 hr for the top to bottom plots, respectively.

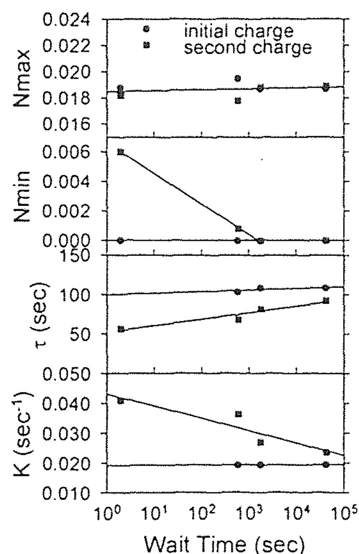


Figure 3. Relaxation time dependence of the fit parameters resulting from the data in Figure 2.

similar, as expected. The following observations can be made concerning the previously stressed devices: 1) The parameter n_{max} is essentially unchanged from that for the initial degradation period, implying that the second degradation period is mainly a recovery of the original number of defects which relaxed during the interval t_{sit}. (A slight amount of additional new defect creation does occur during the recovery period, as will be shown below in the discussion of Figure 5.) 2) The parameter n_{min} is already less than 35% of n_{max} after only 2 s of relaxation, and decreases to zero after several hours. 3) The time constant τ increases with increasing t_{sit}, while the prefactor K decreases with increasing t_{sit}. Both parameters approach their original value for fresh devices, indicating that during relaxation the shallowest defects relax first. It is important to note that neither τ nor K reach their original values even after a 12 hour sit at $V_g = 0$ V, but n_{min} reaches its original value sooner than this, showing that relaxation involves more than just elimination of defects as sources of increased tunnel current. Evidently additional structural relaxation occurs long after the defect is eliminated as a current enhancing source, placing the relaxed defect precursors physically deeper into the oxide.

Additional information can be learned by investigating the bias dependence of the relaxation. Figure 4 shows the fractional increase in current at

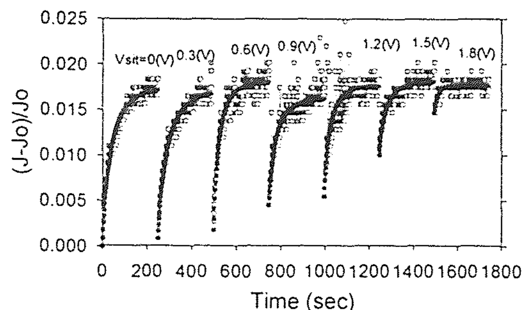


Figure 4. Fractional increase in current at V_g = 2.1 V versus time, measured after 2 hour relaxation periods at various bias levels ranging from 0.0 to 1.8 V.

V_g = 2.1 V measured after 2 hour relaxation periods at various bias levels ranging from 0.0 to 1.8 V. The voltage dependence of the parameters extracted from these data is shown in Figure 5. The following observations can be made: 1) There is a slight increase in n_{max} with increasing V_{sit} indicating that a small amount of defect creation occurs during each of the recovery periods. 2) The parameters τ and K do not change significantly with V_{sit} indicating that after the initial degradation period, the subsequent regeneration periods sample roughly the same distribution of defect precursors. 3) The parameter n_{min} clearly increases with increasing V_{sit}, especially above $V_g \sim 1.0$ V. If the degradation occurs mainly above a threshold voltage of $V_g \sim 1.0$ V, as has been reported for substrate injection[5,6], then one would expect to observe nearly complete relaxation below the threshold (i.e., n_{min} = 0), and relaxation of a decreasing number of defects at increasing voltages above 1.0 V (i.e., n_{min} increases with increasing V_g). The n_{min} data in Figure 5 are consistent with this interpretation.

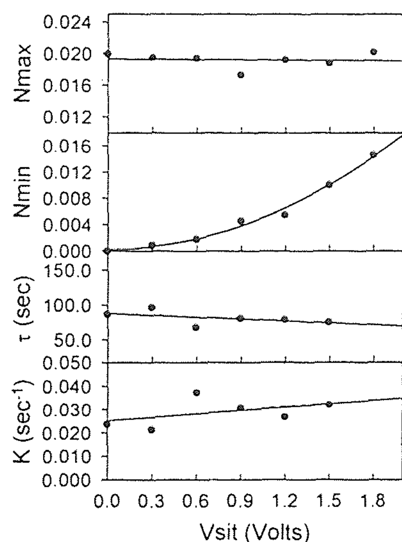


Figure 5. Relaxation voltage dependence of the fit parameters resulting from the data in Figure 4.

4. CONCLUSION

In conclusion, we have investigated defect instability in ultra-thin oxides stressed by injection from n-type silicon substrates into degenerate n-type polycrystalline silicon gates. In this system nearly complete relaxation occurs, allowing the possibility of regenerating the defects. By modeling the defect generation process and examining differences between original and subsequent defect creation periods, we have obtained an improved picture of the relaxation process and the nature of the involved defects. Also, we have demonstrated that a straightforward first-order kinetics process may underlie the degradation. The success of our model suggests that a conduction process other than the dominant direct tunnel current component may drive the degradation with high efficiency. This work was supported in part by National Science Foundation awards No. ECS-9530984 and ECS-9624798.

REFERENCES

1. Y. Taur *et al.*, IEDM Tech. Dig. (1992) 35.6.1.
2. H. Mamose, M. Ono, T. Yoshitomi, T. Ohguro, S. Nakamura, M. Saito and H. Iwai, IEEE Trans. Electron Dev. 43 (1996), 1233.
3. K. R. Farmer, M. O. Andersson and O. Engström, Appl. Phys. Lett. 60 (1992) 730.
4. P. Lundgren, M. O. Andersson, K. R. Farmer and O. Engström, Microelectronic Engineering 28 (1995) 67.
5. K. R. Farmer, C. Debauche, A Giordano, P. Lundgren, M. O. Andersson and D. A. Buchanan, Appl. Surf. Sci. 104/105 (1996) 369.
6. P. Lundgren, M. O. Andersson, K. R. Farmer and O. Engström, J. Noncryst. Solids 187 (1995) 140.
7. P. V. Dressendorfer, Ph.D. Thesis, Yale University (1978).
8. K. R. Farmer, P. Lundgren, M. O. Andersson and O. Engström, Proc. MRS Symp. 284 (1993) 229.
9. H. Preier, Appl. Phys. Lett. 10 (1967) 361.

REFERENCES

1. Yuan Taur, S. Cohen,S. Wind, T. Li, C. Hsu et al, "Experimental 0.1μm p-Channel MOSFET with p^{+}-Polysilicon Gate on 35-Å Gate Oxide", *IEEE Electron Letters*, vol. 14, No.6, June 1993.

2. D. J. DiMaria, E. Cartier, and D. Arnold, "Impact iionization, trap creation, degradation, and breakdown in silicon dioxide films on silicon", *J. Appl. Phys.*73 (7), 1 April 1993

3. S. M. Sze, *Physics of Semiconductor Devices*, second edition, Published by: John Wiley & Sons, Singapore.

4. I. Ohdomari, H. Akatsu, Y. Yamakoshi and K. Kishimoto, *J. Appl. Phys.* 62, 3751 (1989)

5. M. V. Fischetti, Z. A. Weinberg, and J. A. Calise, *J. Appl. Phys.* 57, 418 (1985).

6. D. J. Dimaria and J. W. Stasiak, *J. Appl. Phys.* 65,2342 (1989).

7. D. J. Dimaria, *J. Appl. Phys.* 68,5234 (1990).

8. Z. A. Weinberg and T. N. Ngugeu, *J. Appl. Phys.* 61, 1947 (1987).

9. K. R. Farmer, M. O. Andersson and O. Engstrom, *Appl.Phys.Lett.* 58 (23), 10 June (1991).

10. K. R. Farmer, M. O. Andersson and O.Engstrom, *Appl. Phys. Lett.* 60 (6), 10 Feb. 1992.

11. P. Lundgren, M. O. Andersson, K. R. Farmer and O.Engstrom, *Microelectronic Engineering* 28, (1995) 67.

12. P. Lundgren, M. O. Andersson, K. R. Farmer and O. Engstrom, *J. Non-cryst. Solids* 187 (1995) 140.

13. K. R. Farmer, C. Debanche, A. Giordano, P. Lundgren and M. O. Andersson, *Appl. Surf. Sci.* (1996) accepted for publication.

14. K. R. Farmer, P. Lundgren, M. O. Andersson and O. Engstrom, *Mat. Res. Soc. Symp. Proc.* vol.284, 1993 Materials Research Society.

15. M. O. Andersson, K. R. Farmer and O. Engstrom, J. Appl. Phys. 71 1846 (1992).

16. K. R. Farmer, C. P. Debauche, C. Franck, W. Buchwald and D. A. Buchanan, *Manuscript.*

17. P. V. Dressendorfer, Ph.D. Dissertation, Dept. of Electrical Engineering, Yale University, New Haven, CT.

18. E. H. Nicollian, J. R. Brews, *MOS Physics and Technology*, Published by: John Wiley & Sons, New York, Chapter 16, pp 824-825.

19. H. Preier, *Appl. Phys. Lett.* 10 (1967) 361.