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

ABSTRACT

DEGRADATION AND BREAKDOWN IN ULTRA-THIN SILICON OXIDES

by
Christopher Franck

Emerging trends in the semiconductor device industry call for detailed knowledge of the properties of devices whose dimensions are small enough to exploit Quantum Mechanical effects. This thesis presents a complete picture of oxide degradation in MOS direct tunnel diodes (t_{ox}<3.4 nm). It is demonstrated that for structures fabricated at different facilities and stressed with either gate or substrate injection, a universal degradation mode is revealed which is manifested as the build up of positive charge in the oxide. The data gathered demonstrates that the positive charging phenomena in sub-3.5 nm oxides is independent of oxide thickness, and is characterized by a voltage threshold and two-regime temperature dependence. Further, the catastrophic failure, or breakdown, of these oxides was studied and the strongest evidence to date is presented which links the positive charging phenomena to the oxide breakdown. This thesis concludes with the presentation of a novel device design which can exploit the properties of degradation and breakdown, in thin oxides, to achieve an EEPROM memory cell of superior endurance.

DEGRADATION AND BREAKDOWN IN ULTRA-THIN SILICON OXIDES

by
Christopher Franck

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 Applied Physics

Department of Applied Physics

May 1996

APPROVAL PAGE

DEGRADATION AND BREAKDOWN IN ULTRA-THIN SILICON OXIDES

Christopher Franck

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

Dr. Ken Chin, Committee Member — Date
Professor of Physics, NJIT

Dr. Nuggehalli Ravindra, Committee Member — Date
Associate Professor of Physics, NJIT

BIOGRAPHICAL SKETCH

Author: Christopher Franck

Degree: Master of Science

Date: May 1996

Undergraduate and Graduate Education:

- Master of Science in Applied Physics
  New Jersey Institute of Technology, Newark, NJ, 1996

- Bachelor of Science in Electrical Engineering
  New Jersey Institute of Technology, Newark, NJ, 1994

Major: Applied Physics

"The truth is out there..."

ACKNOWLEDGEMENT

I would like to thank all of the faculty members in the Physics and Electrical Engineering departments at NJIT, especially Dr. Hensel and Dr. Whitman, for their inspiring instruction and discussions with me. Special thanks are also in order for my thesis advisor Dr. Kenneth Farmer for his patience, guidance, and instruction during the development of this thesis.

Also, I would like to thank Dr. Doug Buchanan at IBM, Yorktown Heights, and Dr. Jimmie Wortman and Brian Hornung at North Carolina State University for their generous contributions of experimental devices.

This work was funded in part by grants from Research Corporation and National Science Foundation (ECS-9313937).

TABLE OF CONTENTS

Chapter		Page
1	MOTIVATION FOR BREAKDOWN STUDIES	1
2	OXIDE DEGRADATION	4
	2.1 MOS Capacitors	4
	2.2 Silicon Oxide Degradation in MOS Capacitor	6
	2.3 Previously Reported Properties of Silicon Oxide Charging	7
	2.4 Properties of Silicon Oxide Charging Being Reported in This Work	11
	2.4.1 Experimental Devices	13
	2.4.2 Weak Fluence Dependance for Positive Charging	13
	2.4.3 Temperature Dependance	16
	2.5 Conclusion of Oxide Degradation Studies	19
3	SILICON OXIDE BREAKDOWN	21
	3.1 Introduction	21
	3.2 Intrinsic and Defect-Related Breakdown	22
	3.3 Results of Breakdown Studies on Experimental Devices	23
4	RESONANT TUNNELING	31
	4.1 Introduction	31
	4.2 Qualitative Description Of The Resonant Tunneling Phenomena	31
	4.3 Calculation Of Transmission Coefficient For Double Barrier Potential	34

TABLE OF CONTENTS
(Continued)

Chapter	Page
5 OPERATING PRINCIPLES OF RESONANT TUNNELING ELECTRICALLY ERASABLE PROGRAMMABLE READ ONLY MEMORY	47
5.1 Introduction	47
5.2 Device Function	49
5.2.1 Critical Design Parameters	49
5.2.2 Device Function	49
5.3 Inversion of Substrate at Quiescence	52
REFERENCES	56

LIST OF TABLES

Table		Page
2-1	Initial values of current density for some of the data in Fig. 2-7	19
4-1	Comparison of resonant state calculations based on WKB approximation, and exact Quantum Well and Double Barrier calculations	40

LIST OF FIGURES

Figure		Page
2-1	(a) Simplified diagram of the physical structure of a MOS capacitor (b) Simplified energy band diagram of a MOS capacitor	4
2-2	Energy band diagrams for a MOS capacitor under both (a) low and (b) high bias. The numbered arrows indicate the various electron transport mechanisms present	5
2-3	(a) Negative charge generation within oxide leading to a decrease in tunnel current with time and (b) positive charge generation leading to an increase in tunnel current	9
2-4	Typical plot of fractional change in current density versus time	10
2-5	Fractional change in current density for a set of thin oxide devices stressed at several voltages for 1000 seconds. Notice the voltage threshold for the onset of positive charging	12
2-6	Current density versus voltage for experimental devices	14
2-7	Fractional change in current density versus gate voltage after 1000 seconds of stress on experimental devices	15
2-8	Temperature dependance of positive charging, defined as $\Delta J/J_0$ measured at 1000 seconds for 2.8 nm oxides, versus positive gate bias	17
2-9	Temperature dependance of positive charging for three oxide thicknesses after 1000 seconds of stress	18
3-1	Time dependent Weibull breakdown distribution for an 11 nm oxide measured at various fields [28]	23
3-2	Normalized distribution of breakdown times for 110 250x250 micron devices	24
3-3	Normalized distribution of breakdown times for 40 12x12 micron devices	25
3-4	Weibull plot of time to breakdown data in Fig. 3-2	25

LIST OF FIGURES
(Continued)

Figure		Page
3-5	Average time to breakdown for 250x250 micron devices versus gate voltage	27
3-6	Maximum time to breakdown for 12x12 micron devivces versus gate voltage	29
3-7	Fractional change in current density versus gate voltage after 1000 seconds of stress on experimental devices	30
4-1	Double barrier structure formed by a $SiO_2/Si/SiO_2$ sandwich	32
4-2	Double barrier profiles used to calculate transmission probability	35
4-3	Transmission probability versus electron incident energy for double barrier and single barrier with an oxide thickness of twice the double barrier oxides. The solid line represents the double barrier	37
4-4	Double barrier structure under negative gate bias	42
4-5	Ideal current density versus voltage characteristics for an RT diode. The dashed curve is the result using Eq. 4-8 for T(E) in Eq. 4-2. The solid curve is the result using Eq. 4-7	46
5-1	Schematic diagram of EEPROM memory cell	47
5-2	Schematic diagram of RTEEPROM	49
5-3	Band diagrams of Double Barrier depicted in various states	51
5-4	Depletion layer width and maximum depletion layer width versus dopant concentration	56

CHAPTER 1

MOTIVATION FOR BREAKDOWN STUDIES

It is the continuing goal of the semiconductor device industry to produce components that are smaller, faster, more reliable, and less expensive than the existing technology [1]. In keeping with this trend a new generation of semiconductor devices is emerging called Quantum Effect devices. This technology appears to be the most promising with respect to meeting the goals of the industry [1], as well as creating the possibility for novel devices. At this time, however, most Quantum Effect structures have been studied in Gallium Arsenide and therefore offer little to an industry that is dominated by the less costly silicon processing technology [2]. Hence, there is a need to study the properties of structures whose dimensions exploit Quantum Effects in silicon.

This thesis represents the preliminary work for just such a study. Concepts are explored leading to a novel approach to the design of Floating Gate Electrically Erasable Programmable Read Only Memory cells, or FG EEPROM cells. A design concept is presented which may be used to construct a FG EEPROM cell with superior endurance as compared with conventional technology. The endurance of EEPROM cells is based on the number of times they can be cycled from a logical "1" state into a logical "0" state and then back into the "1" state. Memory arrays currently available on the market can offer at least 10^5 cycles [3-4], and recently it has been reported [5] that 10^9 cycles has been achieved in the laboratory using a modified structure. It is expected that ultimately a device which exploits Quantum Effects coupled with recently observed properties of degradation and breakdown in ultra-thin

silicon oxides [6], will produce memory cells that can be cycled, effectively, an infinite number of times.

Preceding the discussion of the proposed memory cell design, the physical properties of thin oxides on which the gain in endurance is based will be presented. Those properties are (1) the voltage threshold [6-8] for the onset of positive charge generation (discussed in chapter 2), and (2) the linking of oxide breakdown to the generation of positive charge (to be demonstrated in chapter 3). This latter property of silicon oxide breakdown is necessary because the proposed device design directly exploits the voltage threshold. Electrons which tunnel through thin oxides with a bias below 3 volts for an oxide on a P-type silicon, or 1 volt on N-type silicon, will not lead to the generation of positive charge within the oxide. Hence, if oxide breakdown is linked to positive charge generation then one would not expect eventual oxide breakdown to occur in devices where there is no build up of positive charge in the oxide. A memory structure that exploits this property must (1) be limited to a bias below the threshold voltage while still being recyclable, and (2) display charge retention. One way this could be achieved is by the use of Resonant Tunneling to transport electrons into and out of the Floating Gate of the memory cell. This aspect of the design proposal will be elaborated upon in chapter 4.

This thesis is mainly concerned with the degradation modes in ultra-thin silicon oxides and the link between breakdown and positive charge generation. It is important these issues are well understood before one can have confidence in a device design that promises such an enormous gain in endurance. The efforts of this author were

mostly focused on creating a reliable picture of ultra-thin oxide degradation and breakdown so as to provide a foundation for the study of enhanced FG EEPROM cell endurance through the use of Resonant Tunneling.

CHAPTER 2

OXIDE DEGRADATION

2.1 MOS Capacitors

To study the degradation of Silicon Dioxide (SiO_2) insulating layers one constructs a Metal Oxide Semiconductor (MOS) capacitor, Fig. 2-1, and biases the device with a DC voltage. The capacitor is held at the DC bias level while the tunnel current through the oxide is recorded. MOS capacitors fall into two categories depending on their oxide thickness (t_{ox} in Fig. 2-1). They are either thick oxide or thin oxide structures, and are differentiated based on the electron transport modes that can make a significant contribution to the tunnel current. For thick oxides (t_{ox}> 3.5 nm), the electron transport from the gate into the substrate is mainly due to Fowler-Nordheim (FN) tunneling [9-11]. This is not the only electron transport mechanism in these

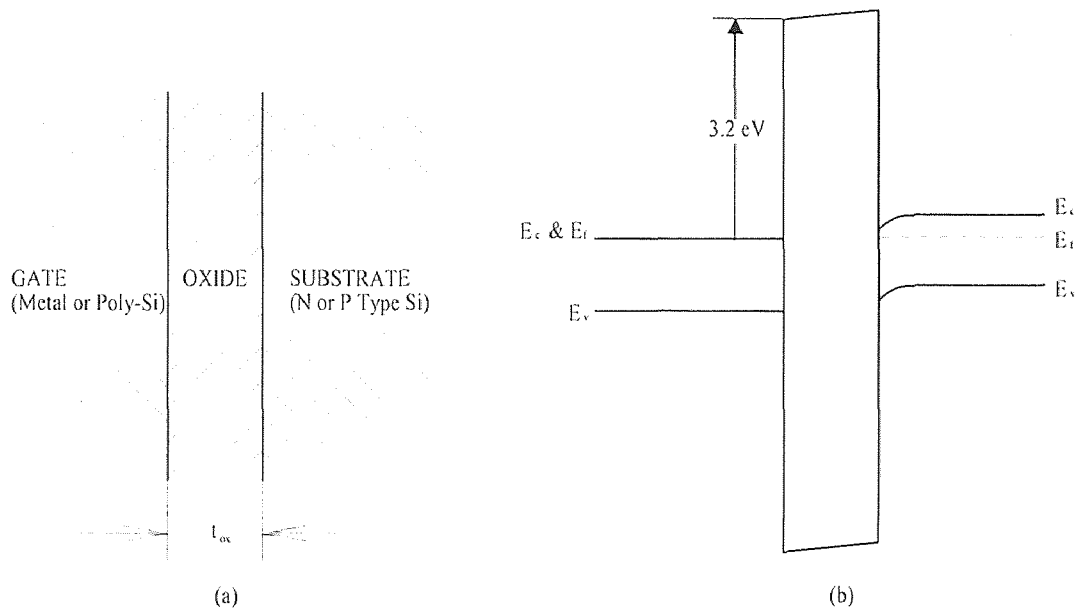

Fig. 2-1: (a) Simplified diagram of the physical structure of a MOS capacitor. (b) Simplified energy band diagram of a MOS capacitor.

structures, however, and in actuality one can identify at least 7 processes [9] in total (see Fig. 2-2). The situation is somewhat different for thin oxides, especially in the low bias regime. When oxides are thin enough ($t_{ox} \leq 3.5$ nm), under low bias the tunneling of electrons directly from the gate into the substrate dominates the current. Accordingly, this electron transport mechanism is called Direct Tunneling (DT). Also, in DT oxides inelastic tunneling processes such as inelastic trap assisted tunneling may be present. In fact, inelastic trap assisted tunneling may dominate the current-voltage characteristics in DT oxides [24].

Fig. 2-2 shows energy band diagrams for both the low and high bias regimes [9]. The capacitor being depicted has an arbitrary oxide thickness and could therefore be either a thin or thick oxide device. DT can just as well take place in the

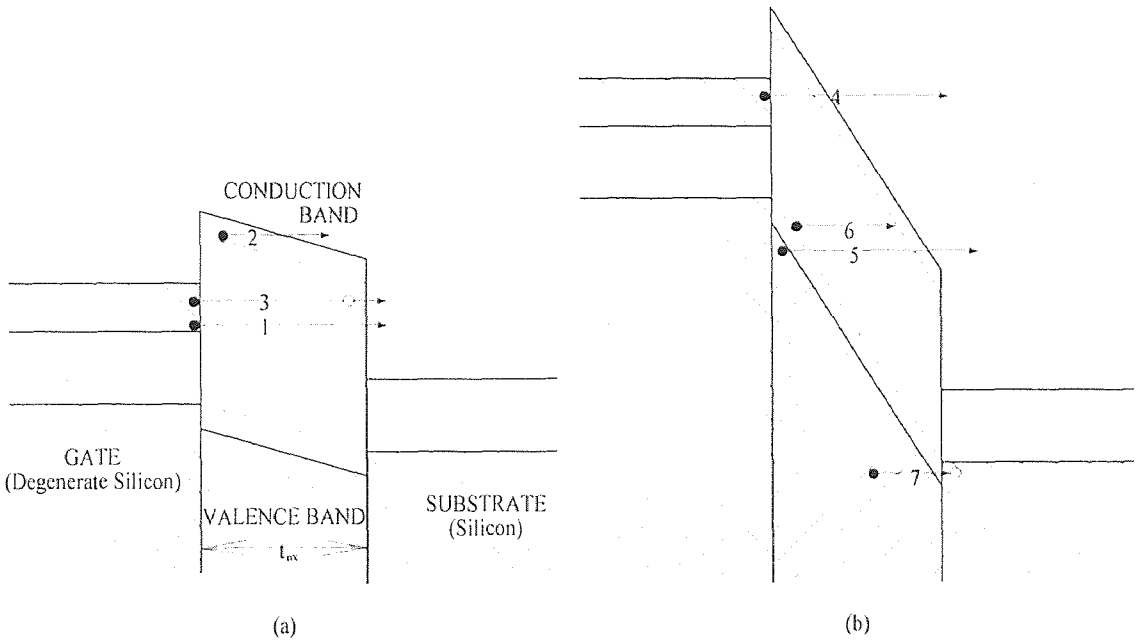

Fig. 2-2: Energy band diagrams for a MOS capacitor under both (a) low and (b) high bias. The numbered arrows indicate the various electron transport mechanisms present.

thick oxide device, however, when t_{ox} is greater than 3.5 nm the probability that an electron tunnels directly into the substrate from the gate electrode is vanishingly small. In Fig. 2-2a three electron tunneling processes are identified. The process represented by arrow 1 is DT. Process 2 is the ionization of impurity states. The dashed line just below the conduction band edge represents a discrete energy level introduced by the impurities present in the oxide. When a bias is applied across the oxide, electrons can tunnel from the impurity state into the conduction band of the oxide. Process 3 is elastic trap-assisted tunneling. In Fig. 2-2b four additional electron transport mechanisms arise due to the high bias applied across the oxide. Process 1 is essentially replaced by process 4, which is the tunneling of electrons from the gate electrode into the conduction band of the oxide (Fowler-Nordheim tunneling). Process 2 is still present. Process 5 is Zener tunneling, which is the tunneling of electrons in the valence band of the oxide into the conduction band of the oxide. Process 6 is intraimpurity tunneling. In this process electrons in the lower impurity state (represented by the dashed line just above the valence band) tunnel into the upper impurity state, and can then enter the oxide conduction band. The last process, number 7 in Fig. 2-2b, is usually referred to as hole tunneling.

2.2 Oxide Degradation in MOS Capacitors

At the outset it might be advantageous to define what is meant by "degradation" in silicon oxides, at least within the context of this work. In fact, it may very well be convention to regard the build up of either positive or negative charge within a fresh oxide that has undergone an electrical stress to be the measure of oxide degradation.

Within the context of this thesis, this is the measure of oxide degradation. The build up of negative charge within a SiO_2 insulating layer can be attributed to two factors both related to electron trapping [7, 12-16]. One process is the trapping of electrons into sites that are created as a natural by-product of the processing stage. The other kind of trapping occurs due to trap sites that are actually generated by the FN stress imposed on the device. Degradation also occurs in thin SiO_2 insulating layers being maintained in the DT current mode, and it is different in character as compared with thick oxides. Degradation in thin oxides is manifested as an increase in the tunnel current with time as, the device is maintained at fixed low voltage DC bias. This current versus time behavior has been attributed to increased trap assisted tunneling as well as, effectively, the build up of positive charge within the oxide [6]. 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 [6]. Since this thesis is only concerned with thin oxides, however, the positive charging that occurs in thin oxides where DT is measurable will be the focus of the remaining discussion.

2.3 Previously Reported Properties of Oxide Charging

Because positive and negative charging in SiO_2 insulating layers arise from two different tunnel mechanisms, there is no reason to believe they would display the same characteristics. Studies have revealed, however, that they do share one interesting property (electron kinetic energy/voltage threshold for the onset of degradation), and to be demonstrated in what follows, they differ critically in another (weak fluence dependance for positive charging induced by DT stress). Before going into these

properties, the methods used to measure oxide charging will be briefly reviewed.

Positive or negative charging in silicon oxides is manifested as a change in the tunnel current (or current density) with time (see Fig. 2-3). One may also look at the evolution of the current (or current density) with respect to electron fluence; the measure of the total amount of charge per square centimeter that has passed through the device. Electron fluence can be seen on the top of the plots in Fig. 2-3. Another way to study the charging of an oxide is to look at the fractional change in current (or current density) as a function of time or fluence. This is given by the equation,

$$\frac{\Delta J}{J_0} = \frac{J - J_0}{J_0} \tag{2-1}$$

Where J is the current density. A typical plot of the fractional change in current density is given in Fig. 2-4.

As was stated at the beginning of the section, there is one property of degradation similar to thin as well as thick oxides. For both thick and thin oxides there exists an observed threshold for the onset of degradation. In thick oxides, where electrons tunnel into the oxide conduction band, it has been observed that trap generation does not occur unless the electrons gain about 2 eV from the electric field in the oxide [7]. The situation is somewhat different for thin oxides. Positive charging does not occur in DT devices unless the bias across the oxide exceeds a threshold voltage. As will be discussed in what follows, the threshold is different depending on whether electron injection is occurring from the gate to the P-type Si substrate (-3 V), or from an N-type substrate to gate (+1 V). Also, the voltage threshold in DT devices is independent of the oxide thickness. This is strikingly

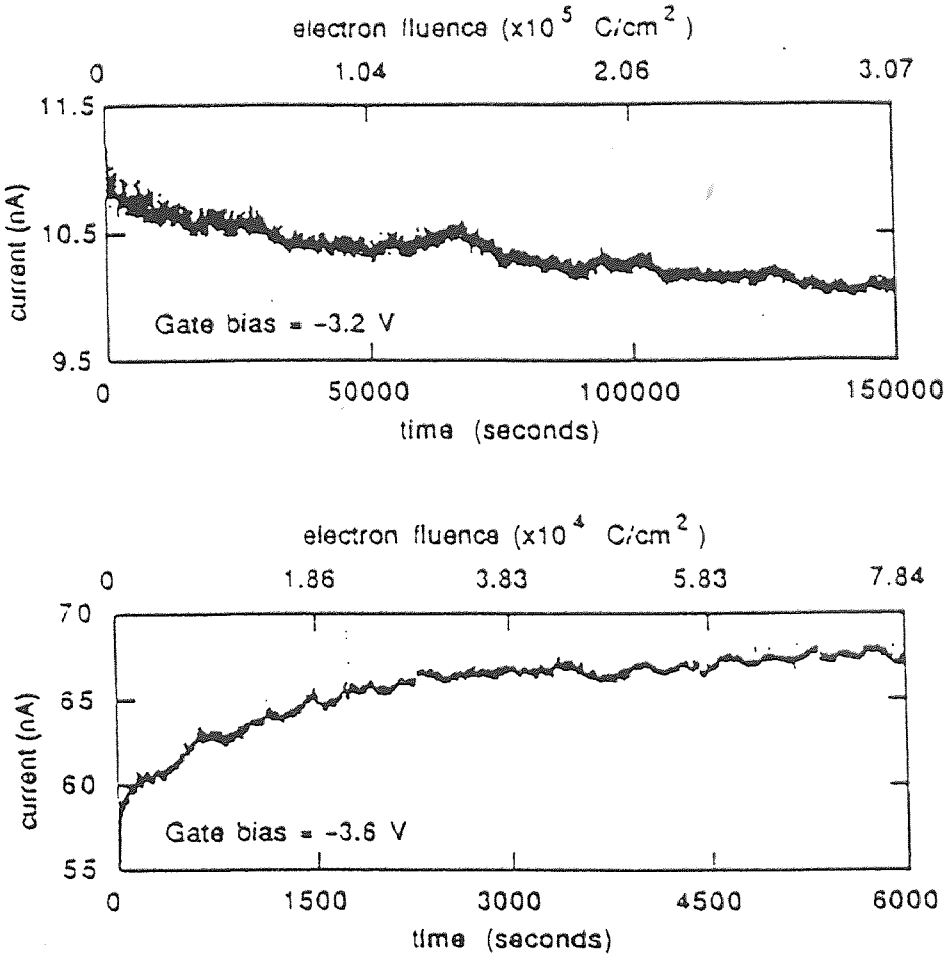

Fig. 2-3: (a) Negative charge generation within oxide leading to a decrease in tunnel current with time and (b) positive charge generation leading to an increase in tunnel current.

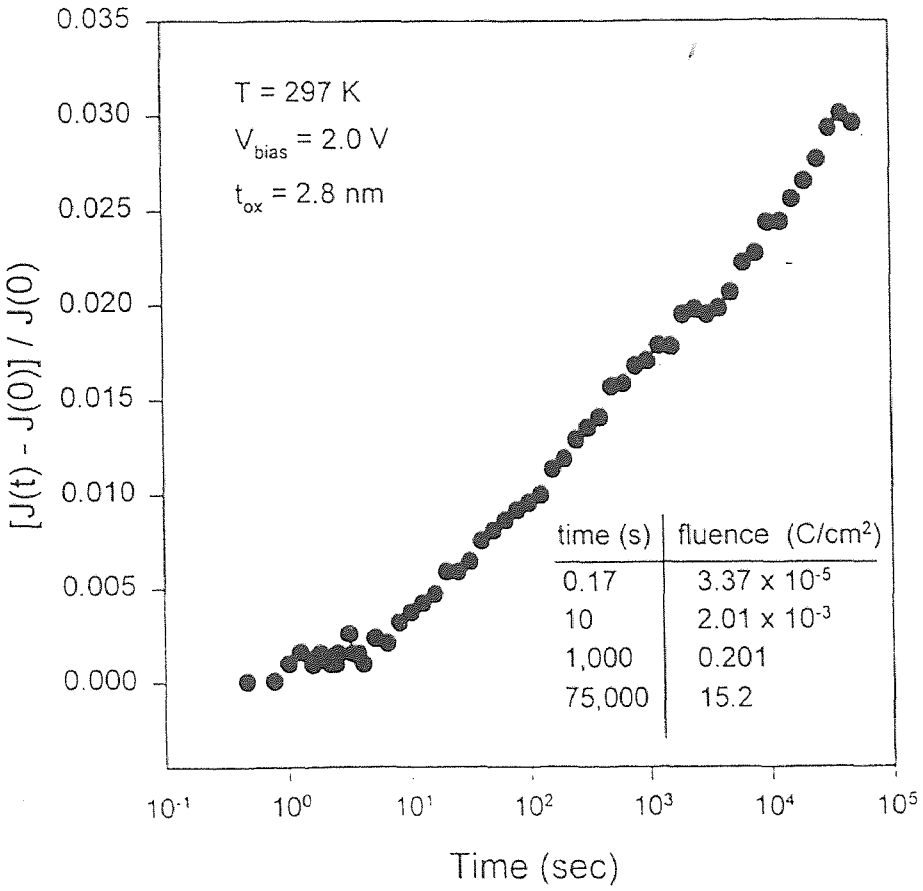

Fig. 2-4: Typical plot of fractional change in current density versus time.

different from the thick oxide threshold which depends on the oxide thickness for the electron to gain the threshold kinetic energy from the electric field. These differences clearly imply that the threshold for degradation in thick oxides depends on the field energy, and the threshold for thin oxides only depends on the voltage across the oxide independent of the oxide thickness [6].

Fig. 2-5 is a plot of experimental data which demonstrates the oxide thickness independence of the threshold voltage for the onset of positive charging in thin oxides. The plots show the fractional change in current density as a function of applied bias after the devices were stressed for 1000 s.

2.4 Properties Being Reported in This Work

Positive charge generation in thin oxides has been reported for electron injection from the gate, formed from either aluminum or polycrystalline silicon, into the substrate. In depth studies have been conducted using P-type silicon substrates, preliminary work has been reported for N-type substrates. During this study, as stated in the introduction, a complete picture of degradation in thin oxides is desirable. Thus, injection from the N-type substrate was examined in as much detail as previously reported for injection into P-type substrates, and it was found to exhibit the same charging behavior. It is therefore concluded that positive charging in devices where direct tunneling is measurable is a universal effect. It is reasonable to believe that this positive charge generation is the primary bias related instability in DT devices on the basis of this study.

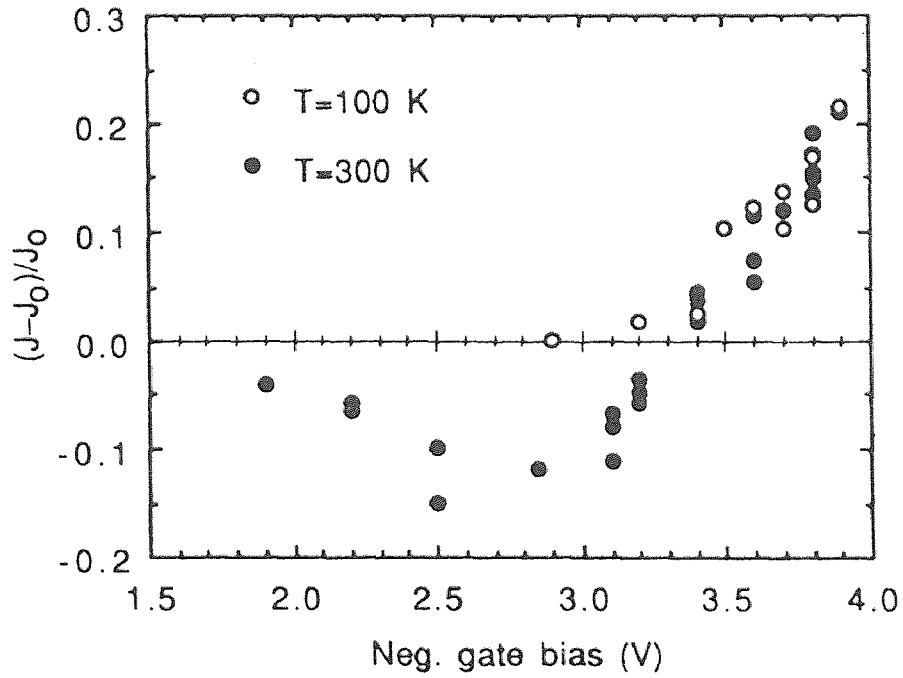

Fig. 2-5: Fractional change in current density for a set of thin oxide devices stressed at several voltages for 1000 seconds. Notice the voltage threshold for the onset of positive charging.

2.4.1 Experimental Devices

The devices used during this study were fabricated on the silicon line at the Yorktown Heights, New York IBM facility. Active tunnel oxides for four different nominal thicknesses were formed. The thicknesses are 2.0, 2.4, 2.8, and 3.4 nm. The tunnel oxides were formed at 700°C in dry O_2 in windows that were opened in the field oxide. The field oxide was grown on <100> oriented, ~0.005-0.02 ohm-cm, n-type silicon. Window sizes ranging 1 μm^2 to 1 mm^2 were made available on each chip, however, measurements were mainly performed on 250 $\mu m \times 250 \mu m$ for this work. In Fig. 2-6 on a plot of the current density versus positive gate voltage is provided for each of the four oxide thicknesses. As can been seen in the figure the current density increases by a factor of about 100 for each 0.4 nm decrease in oxide thickness. This scaling agrees with fundamental tunneling theory [18].

2.4.2 Weak Fluence Dependance for Positive Charging

The experimental results that will now be provided reveal a strikingly different character for oxide degradation in thin oxides as compared with thick oxides. The specific feature of degradation that is being compared is the electron fluence dependance for degradation. It has been demonstrated in the FN regime for oxide thicknesses from 4.5 to 5.5 nm that trap generation tends to scale with electron fluence [17]. This does not appear to be the situation for thin oxides in the DT regime, and at best there exists only a weak dependance on electron fluence for positive charging.

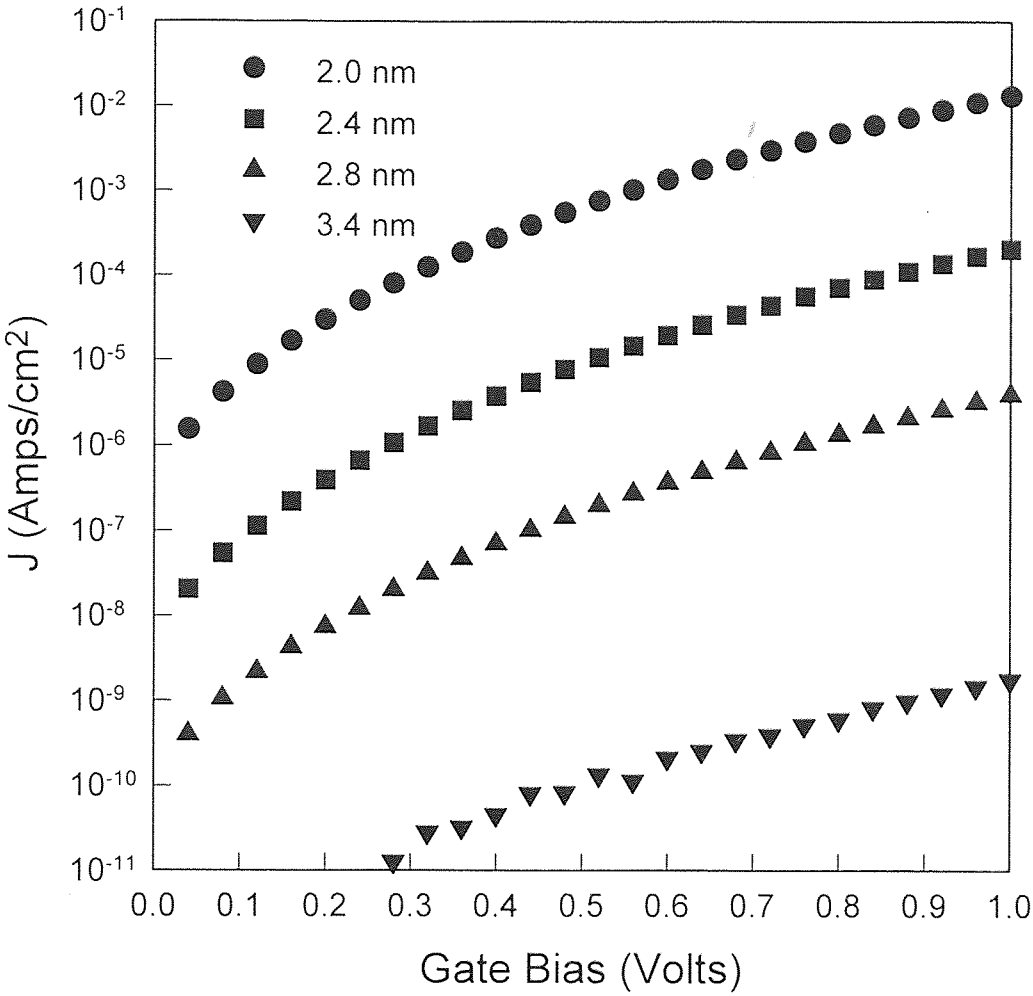

Fig. 2-6: Current density versus voltage for experimental devices.

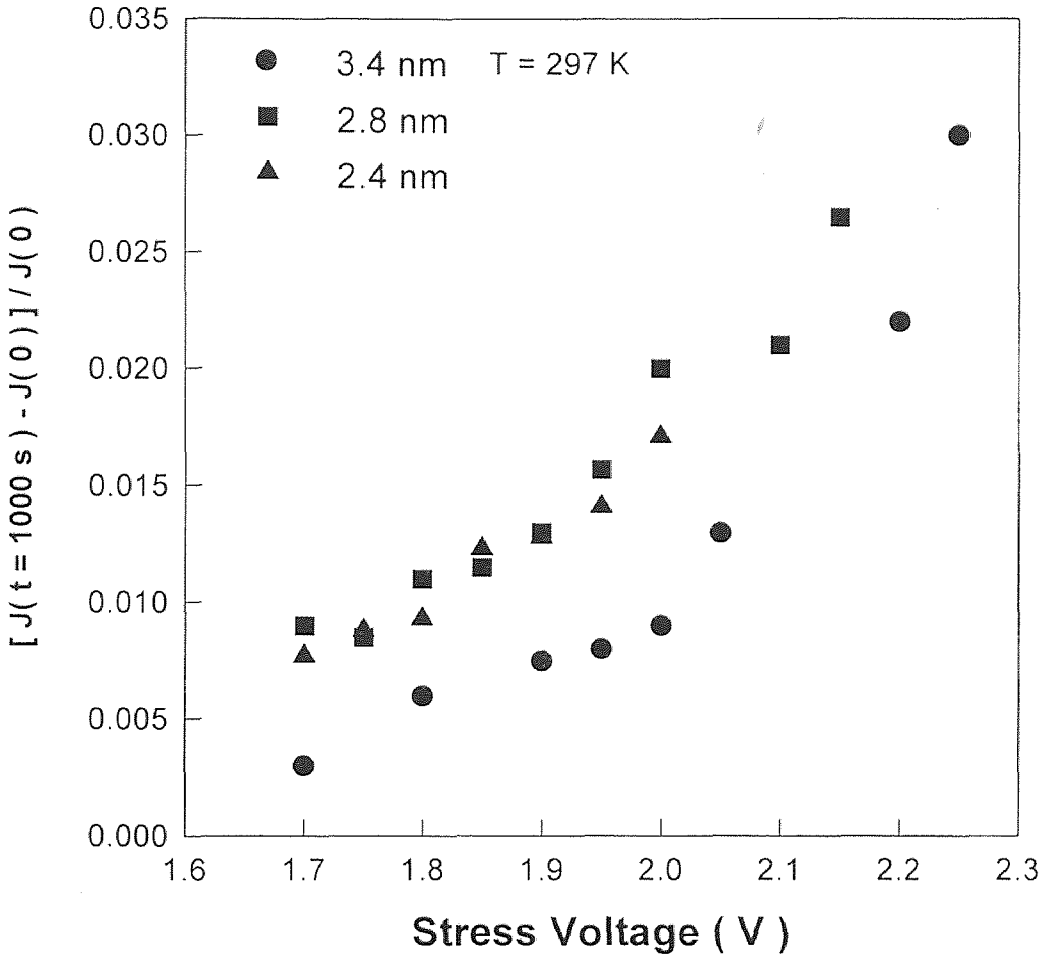

Fig. 2-7: Fractional change in current density versus gate voltage after 1000 seconds of stress on experimental devices.

Devices with three separate oxide thicknesses were studied, with the results plotted in Fig. 2-7. Fig. 2-7 represents the fractional change in current, after 1000 seconds, as a function of applied bias. Data was gathered for 2.4, 2.8, and 3.4 nm oxide devices. To determine the role played by the electron fluence in these curves, Table 2-1 provides J_0 for each device at two bias levels, 1.7 and 2.0 volts. In Table 2-1 it can be seen that J_0, and hence the electron fluence, differs by five orders of magnitude. Returning to Fig. 2-7, however, the fractional change in current density differs by no more than a factor of 2. Even more compelling, from the plot in Fig. 2-7, is the fact that the amount of charge generated in the 2.4 and 2.8 nm oxides appears to be identical. In Table 2-1, however, J_0 for these two oxides differs by two orders of magnitude at one bias level and one order of magnitude at another. Clearly, if the generation of positive charge were dependent on electron fluence, then one would not expect the same amount of charge generation in two devices with current densities separated by orders of magnitude.

2.4.3 Temperature Dependance of Positive Charging

For gate injection it has been shown that positive charge generation is dependent on temperature above ~150 K; below this temperature it appears to be relatively temperature independent [25]. Fig. 2-8 and Fig. 2-9 show plots of the data for the temperature dependence of positive charging. Fig. 2-8 is a plot of fractional change in current after 1000 seconds of stress versus gate voltage for 2.8 nm devices. Four plots are given for temperatures of 77 K, 150 K, 240 K, and 300 K. It can be seen that the charge generation rate decreases strongly with temperature. Fig. 2-9 is an Arrhenius

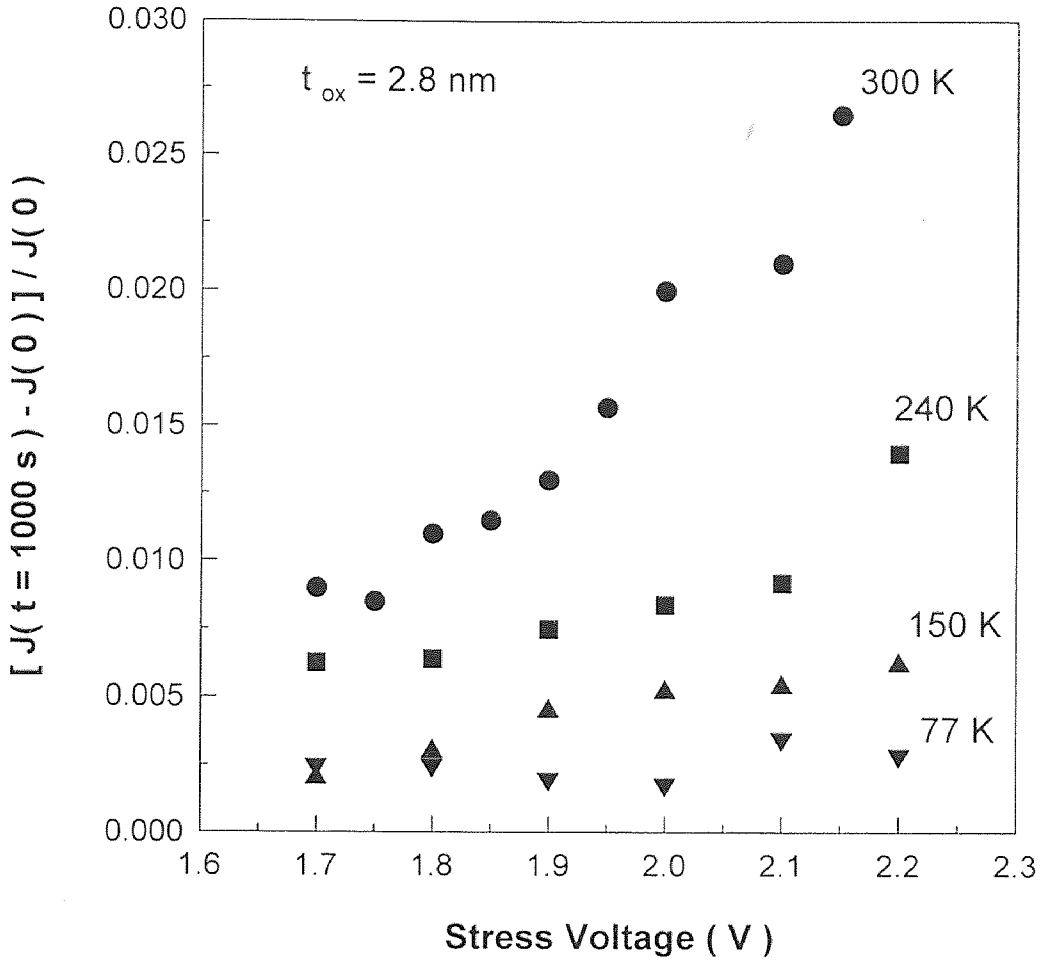

Fig. 2-8: Temperature dependance of positive charging, defined as $\Delta J/J_0$ measured at 1000 seconds for 2.8 nm oxides, versus positive gate bias.

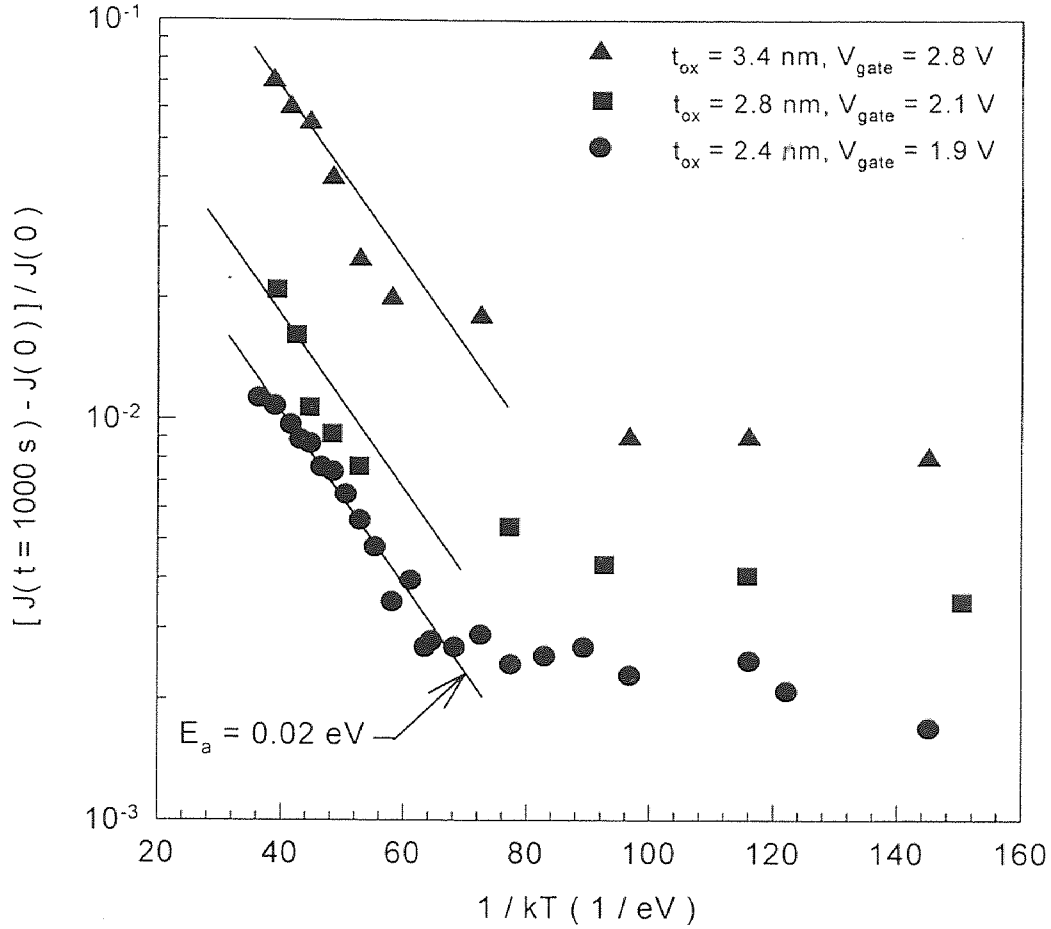

Fig. 2-9: Temperature dependance of positive charging for three oxide thicknesses after 1000 seconds of stress.

Table 2-1: Initial values of current density for some of the data in Fig. 2-7.

t_{ox} (nm)	J_0 (A/cm^2) V_{bias} = 1.7 V	J_0 (A/cm^2) V_{bias} = 2.0 V
3.4 nm	2×10^{-8}	6.4×10^{-8}
2.8 nm	6.4×10^{-5}	1.6×10^{-4}
2.4 nm	3.2×10^{-3}	8.0×10^{-3}

plot of the charge generation measured at a gate voltage of V_G=2.1 V after 1000 seconds. Again, similar to injection from the gate, DT substrate injection shows a strong temperature dependance above ~150 K, and weak temperature dependance below this value. Curve fitting suggests that at low temperatures charge generation may be proportional to T^2, but more measurements are required to confirm this. The behavior illustrated in Fig. 2-9 is similar to that reported for both interface state generation [26] and trap creation [8] in thicker oxides which undergo FN degradation. This suggests the possibility that while the DT defect generation mode is distinctly different from the FN mode, the nature of the defects created may be similar in both cases.

2.5 Conclusion of Oxide Degradation Studies

The properties of DT degradation presented for substrate injection complete the picture of degradation in thin oxides where DT transport of electrons is measurable. Devices have now been studied with Al and polycrystalline silicon gates, fabricated at different

facilities, and stressed by either gate or substrate injection. The data that has been gathered reveals the following universal characteristics of DT degradation:

- Weak fluence dependance in devices where DT current is measurable.
- Threshold voltage for the onset of degradation.
- Apparent thermal activation above ~150 K, and weak temperature dependance below ~150 K.

Having studied the properties of degradation in thin oxides, the phenomenon of oxide breakdown can now be addressed. It still remains to link positive charge generation to the ultimate failure of the oxide. This will be the subject of the next chapter.

CHAPTER 3

SILICON OXIDE BREAKDOWN

3.1 Introduction

The universal positive charge generation described in the previous chapter seems to be the primary bias-related instability in below 3.5 nm oxides. By itself, it is an important phenomenon to study since it is the signature of changing oxide characteristics over time. As presented in the last chapter, the detailed study of the positive charging provides insight into the microscopic mechanism underlying the degradation. Perhaps a more important question, beyond the study of defect mechanisms, would be, is there a relationship between the stress-related degradation and the ultimate catastrophic failure of the oxide? In this chapter the possibility is explored. A study of the breakdown of hundreds of DT devices under different bias conditions is presented. A distinction between "defect-related breakdown" and so-called "intrinsic breakdown," is identified, and strong evidence is presented linking the intrinsic breakdown to the build-up of the positive charge. In previous studies on thicker oxides this relationship has never been demonstrated. In fact, in oxides over 10.0 nm thick it has been concluded that positive charge generation is not related to the ultimate failure of the device [27]. In thinner oxides, 5.0 nm thick, where the electron injection mechanism is still FN tunneling, a correlation between stress-induced positive charge and breakdown has been inferred recently, but not demonstrated conclusively [28]. The work being presented in this chapter gives the first indication of a relationship between the positive charging and breakdown in the DT thickness regime, and gives the strongest evidence to date for such a relationship in any thickness regime.

3.2 Intrinsic and Defect-Related Breakdown

The IBM chip set containing the test DT devices has a large number of 250 x 250 micron capacitors. Most of the work presented in the previous chapter was conducted using this size of device. The chip set also contains small area capacitors down to 1 x 1 micron. In the breakdown studies both the large area devices and a group of ~12 x 12 micron devices were used.

By studying breakdown in devices of different area it is possible to distinguish two breakdown regimes: defect-related and intrinsic [29]. Breakdown that occurs at a weak spot in the oxide is generally termed defect related breakdown. If the dielectric strength of these weak spot defects is distributed, and breakdown occurs at the weakest spot, then the average breakdown voltage, or the average time to breakdown at a given voltage, will be lower in devices which contain more defects. Assuming that a single type of defect is responsible for the weak spot, and the defects are distributed randomly over the surface, then the distribution of breakdown times at a given field is characterized by a Poisson distribution with a single time constant. This is given by,

$$Probability = \frac{\mu^t}{t!} e^{-\mu} \qquad (3\text{-}1)$$

where μ is the time constant. Assuming a constant defect density that is independent of the device area, and owing to the probabilistic nature of breakdown, the distribution of breakdown times is expected to scale with device area by a factor $\ln(A_1/A_2)$ where A_1 and A_2 are the respective areas, and the time axis is increased for decreasing areas [29]. On a Weibull plot, plotting $\ln(-\ln(1-F))$ versus $\log(t_{bd})$, where t_{bd} is the measured time to breakdown and F is the fraction of samples that have broken down

after a time less than t_{bd}, defect-related breakdown gives rise to a shallow slope. In contrast, so-called intrinsic breakdown gives rise to a very steep slope. The steep slope indicates that beyond a threshold stress time, essentially all devices are prone to experience catastrophic failure, hence the term intrinsic breakdown. Fig. 3-1 illustrates the two breakdown modes in a thick oxide. This data was taken from Wolters, et al [28].

3.3 Results of Breakdown Studies on Experimental Devices

The IBM samples mainly fail by a defect related mechanism. This is confirmed in Fig. 3-2 and Fig. 3-3 for a 2.8 nm oxide. Fig. 3-2 shows the distribution of breakdown times for 110 250 x 250 micron devices stressed at 2.4 V, where the data is normalized using the number of breakdowns divided by half the number of devices.

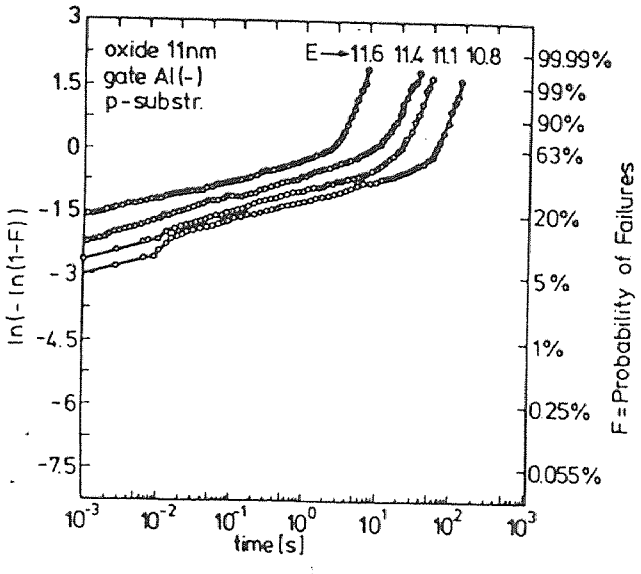

Fig. 3-1: Time dependent Weibull breakdown distribution for an 11 nm oxide measured at various fields [28].

The data can be fit reasonably well using the sum of two Poisson distributions, assuming that half of the devices broke down with the shorter time constant, the other half broke down with the longer time constant. Fig. 3-3 shows the normalized distribution of breakdown times for 40 ~12 x 12 micron devices, also stressed at 2.4 V. Notice that the time axis is expanded by a factor of six compared to the time axis for the larger area devices. On this scale, despite the small amount of data, the distribution of breakdown times is remarkably similar to that for the larger area devices, except scaled by the factor of 6 in time. The natural log of the ratio of the areas, ln(650,000/144) is approximately 6, which leads to the conclusion that defect-related breakdown is mainly occurring. The data of Fig. 3-2 is replotted in Fig. 3-4 on a Weibull plot. Except for the very longest lived devices, this plot has a shallow slope, consistent with defect related breakdown. The two peaks in Fig. 3-2 indicate that two

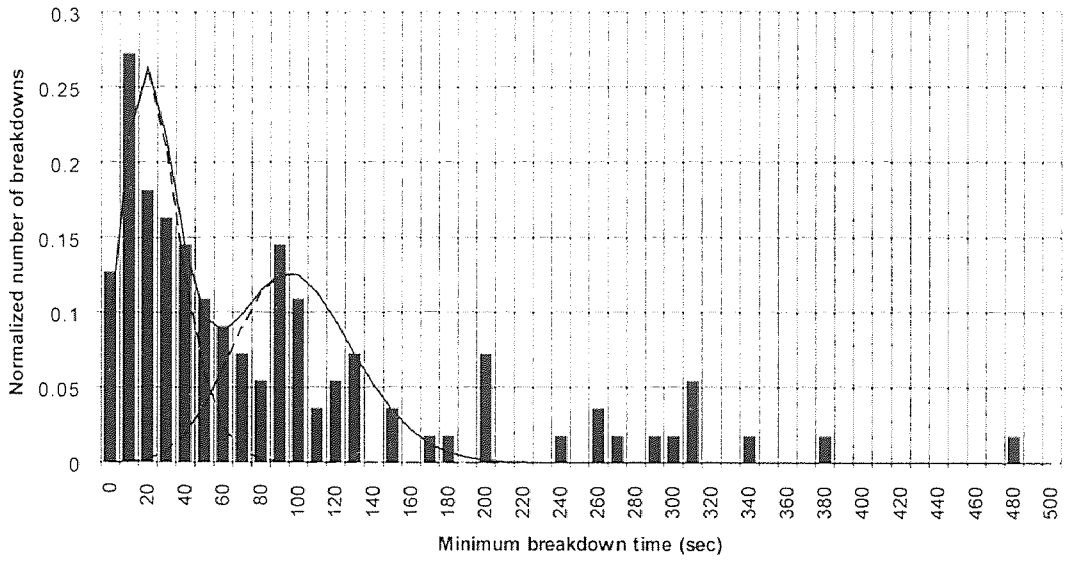

Fig. 3-2: Normalized distribution of breakdown times for 110 250x250 micron devices.

Fig. 3-3: Normalized distribution of breakdown times for 40 12x12 micron devices.

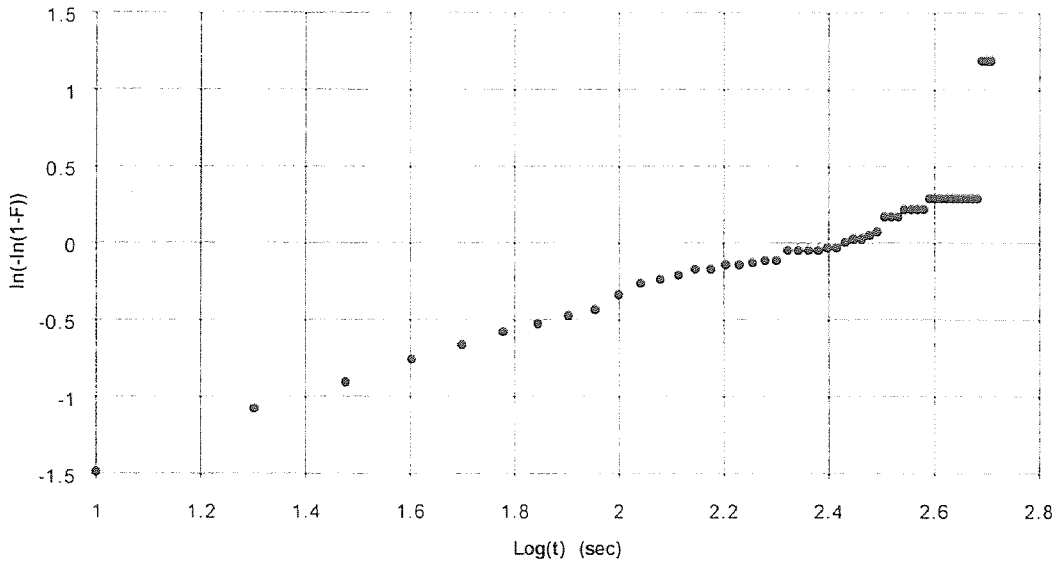

Fig. 3-4: Weibull plot of time to breakdown data in Fig. 3-2.

different types of defects may be responsible for the observed oxide failures, each defect type claiming roughly half of the devices. This suggests that in these DT oxides, some sort of defect breakdown spectroscopy may be possible. However, such defects generally are to be avoided, and in optimized fabrication processes, at least in thicker oxides, defect-related breakdown can be almost entirely avoided.

Fig. 3-5 shows the average time to breakdown of 250 x 250 micron devices as a function of applied gate voltage for four different oxide thicknesses. The time to breakdown curves are roughly parallel, shifted to lower voltages for the thinner oxides. This is identical to the findings of Schuegraf and Hu, who claimed to be studying intrinsic breakdown [30]. It is certain that the data presented in Fig. 3-5 are not due to intrinsic breakdown, but rather to defect-related breakdown.

It is possible to study *intrinsic* breakdown in the experimental devices, but to do so requires looking at the longest lived devices at the highest voltage levels possible. Fig. 3-6 shows the longest of five breakdown times measured at various bias levels using 12 x 12 micron devices. Four different device thicknesses were studied. These breakdowns were measured at very high voltage levels compared to those for Fig. 3-5. In general it is found that no 250 x 250 micron device could survive at this high bias, and roughly half of the 12 x 12 micron devices failed immediately after the high bias was applied. The other half of the smaller area devices lasted for times ranging from 100 to 10,000 seconds depending on the applied bias and oxide thickness. Because of the distinctly long lifetime at high bias, the failure is attributed to intrinsic breakdown. Notice that there is a significant difference in the thickness dependence of defect-related breakdown shown in Fig. 3-5 and the thickness

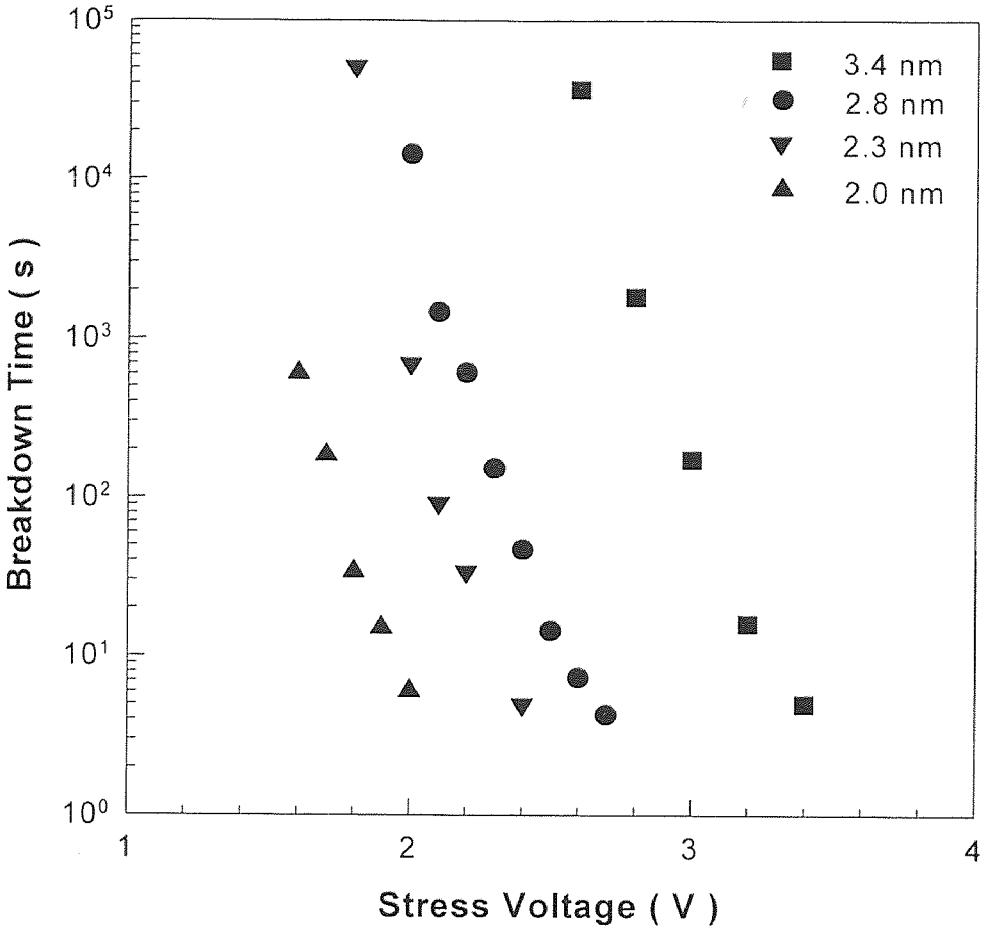

Fig. 3-5: Average time to breakdown for 250x250 micron devices versus gate voltage.

dependence of the intrinsic breakdown shown in Fig. 3-6. In particular, there is no parallel shift in the time-to-breakdown voltage dependence for intrinsic failure of the 2.0 to 2.8 nm oxides in Fig. 3-6. In other words, the time to intrinsic breakdown of the 2.0 to 2.8 nm oxides at a given bias voltage is roughly independent of oxide thickness, while the time to breakdown of the 3.4 nm oxides is somewhat higher. This behavior is strikingly similar to the thickness dependence of the build-up of positive charge. Fig. 3-7 repeats the data from chapter 2 which shows that the build-up of positive charge is relatively independent of oxide thicknesses of 2.8 nm and below, and somewhat decreased for the 3.4 nm oxides. The correlation between thickness dependence of the build-up of positive charge and the time to intrinsic breakdown suggests that the positive charge may be at the root of intrinsic oxide failure. In this interpretation, positive charge evolves over time as described in chapter 2, independent of oxide thickness, until a critical amount of charge is present. Above this critical threshold, intrinsic breakdown is inevitable. If this interpretation is correct, then the use of charge to breakdown, Q_{BD}, is completely inappropriate as a characterization parameter for predicting the lifetime of thin oxides in the DT regime. Q depends strongly on the oxide thickness, but intrinsic breakdown in DT oxides is an effect which appears merely to be a function of time, independent of thickness.

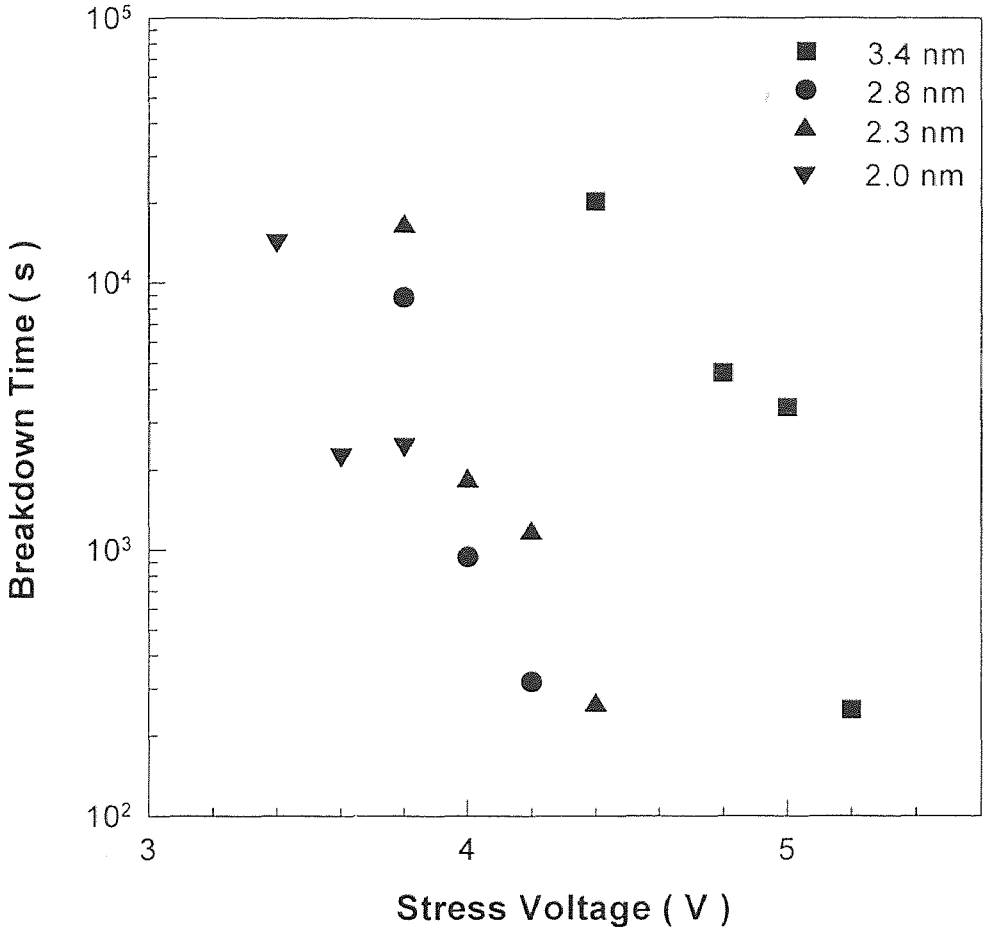

Fig. 3-6: Maximum time to breakdown for 12x12 micron devices versus gate voltage.

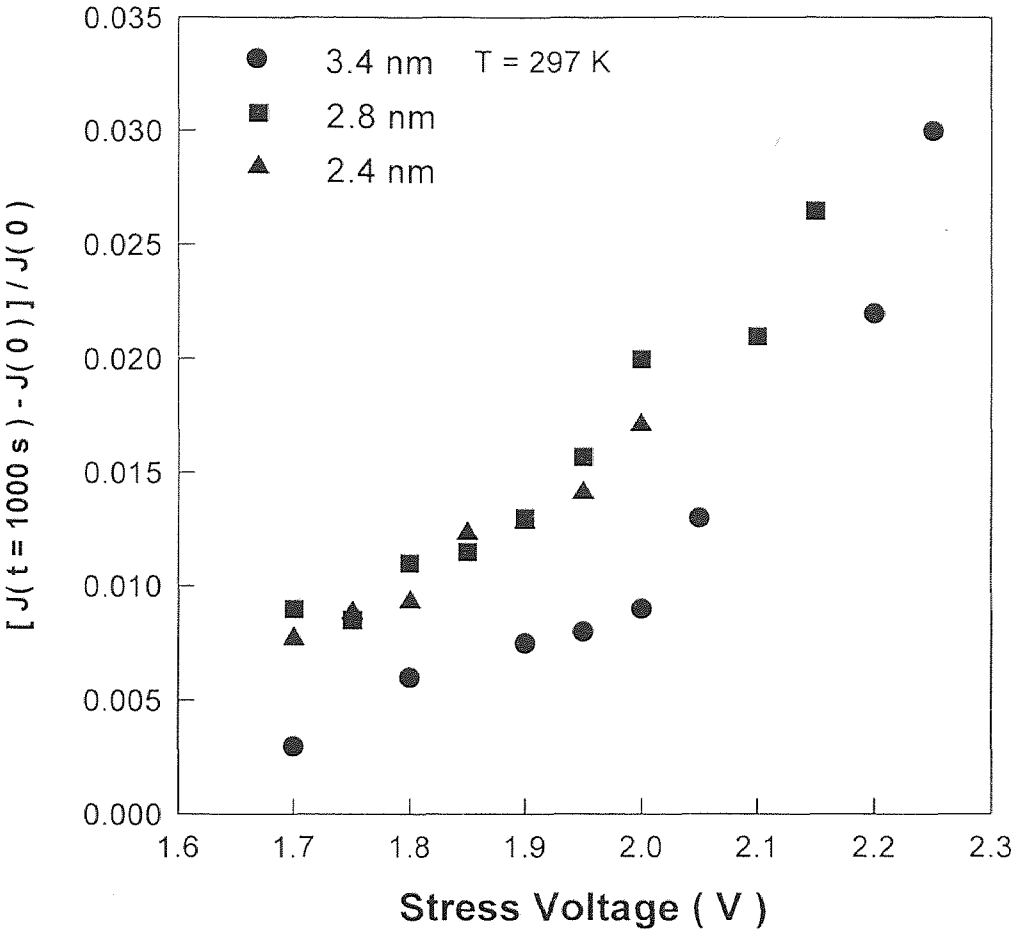

Fig. 3-7: Fractional change in current density versus gate voltage after 1000 seconds of stress on experimental devices.

CHAPTER 4

RESONANT TUNNELING

4.1 Introduction

In this chapter a theoretical description of the Resonant Tunneling (RT) phenomena will be presented, culminating with a calculation of the ideal current density profile that can be expected from a RT diode. The model to be presented incorporates simplifications that will lead to quantitative discrepancies when compared with actual device measurements. For example, it has been found that the electron-phonon interaction is one of the most important processes to include in any model that will achieve accurate quantitative results [19]. Also, the assumption of uniform oxide thickness is another source of quantitative errors [18]. The analysis that will follow is not meant to provide an accurate quantitative solution, rather, it is meant to illuminate the critical features of Resonant Tunneling as they apply to the memory cell design proposal mentioned in chapter 1.

4.2 Qualitative Description of the Resonant Tunneling Phenomena

In chapter 2 the topic of tunneling through a SiO_2 insulating layer was discussed. The energy band diagrams for a single layer of silicon dioxide were given in Fig. 2-1 and Fig. 2-2. These diagrams depict a single potential energy barrier encountered by an electron traveling from the gate electrode towards the substrate. If a thin layer of silicon is sandwiched between two thin oxide layers, a double barrier (DB) potential energy profile will result (Fig. 4-1).

The tunneling of electrons through a DB can lead to resonances in the

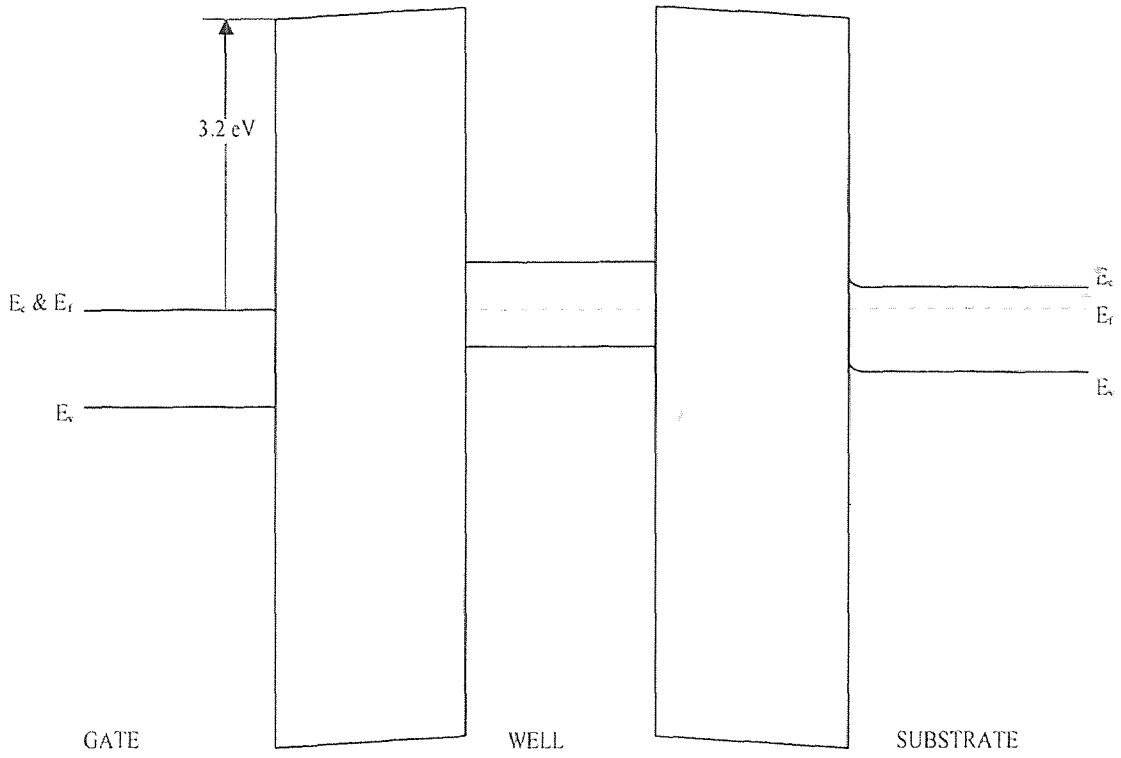

Fig. 4-1: Double barrier structure formed by a $SiO_2/Si/SiO_2$ sandwich.

transmission spectrum. By resonances it is meant that sudden, sharp peaks close to unity in the transmission probability can occur for electrons of specific energies. This is in contrast to tunneling through a single barrier where the probability for tunneling can be expressed by,

$$T = e^{-\frac{4\pi}{h}\int_a^b dx\sqrt{V(x)-E}} \qquad (4\text{-}1)$$

where V(x) is a well behaved function, usually a constant or just linear in x. Eq. 4-1 is the usual expression given by the WKB approximation. Notice in Eq. 4-1 that the probability of tunneling through a single barrier rises exponentially and exhibits no sudden peaks as E is varied from zero. Yet, if a DB structure is formed from two

oxides, separated by a thin layer of silicon, with the same t_{ox} as the oxide in Eq. 4-1, there will be distinct energy levels at which an electron will pass through both oxides layers completely unattenuated. This is the phenomena of Resonant Tunneling.

Resonant Tunneling occurs due to the wave nature attributed to the electron. If the oxide layers are thin then electrons impinging upon the first barrier (on the left in Fig. 4-1), will tunnel directly through the barrier into the silicon in the center of the double barrier structure (to be referred to as the silicon well from this point on). The electron waves will then strike the second barrier and be reflected as well as transmitted. The reflections will be directed back towards the first barrier and be reflected again. As more electrons pass through the first barrier they can constructively interfere with the reflections already bouncing between the two barriers, provided they are phase synchronized. Phase synchronization is disrupted whenever an electron scatters. The repeated reflection of phase synchronized electron waves results whenever a bias is placed across the entire DB structure and will cause standing waves of large amplitude to build up inside the silicon well at distinct energy levels. The resulting probability for an electron to tunnel through both barriers at these energy levels will be close to unity. At other energy levels the electron waves will not be in phase and therefore destructively interfere. The energy levels of high transmission probability will be referred to as resonant states.

In the preceding discussion it was specified that both the silicon well and the oxide barrier layers need to be thin. Clearly, if the oxides are not thin enough to allow DT then one would not expect the build up of large standing waves in the silicon well. But there is a more important need for thin oxide layers and silicon well. The build up

of standing waves in the silicon well depends on the phase coherence of the incoming and reflecting waves, however, electrons in solids will only travel a certain distance before they are scattered. The scattering of an electron changes its momentum and therefore disrupts its phase. The average distance an electron will travel before being scattered is referred to as its mean-free path. Hence, the entire DB structure must be smaller than the mean-free path of the electron. When an electron tunnels without being scattered it is referred to as ballistic tunneling. It has been observed that ballistic tunneling can take place in oxides that are as large as 10 nm [23].

The discussion so far has only provided a qualitative model for the resonant tunneling phenomena. The probability of tunneling still needs to be calculated and this will be the subject of the next section.

4.3 Calculation of Transmission Coefficient for Double Barrier Potential

Before one can compute the theoretical current density versus voltage (JV) curve for a tunnel diode, the tunneling probability as a function of electron energy must be found. This is required by the current density integral which takes on the form [9],

$$J = \int_0^\infty N(E)T(E)dE \tag{4-2a}$$

$$N(E) = \frac{4\pi m_s q k_B T}{h^3} \ln \frac{1 + e^{\beta(E_f - E)}}{1 + e^{\beta(E_f - E - qV)}} \tag{4-2b}$$

where N(E) is referred to as the supply function since it incorporates the Fermi function and density of states, m_s is the electron effective mass in the metal/silicon

electrode, q is the electron charge magnitude, k_B is Boltzman's constant, T is temperature, h is Planck's constant, V is the bias applied across the structure, $\beta=1/k_BT$, E_f is the Fermi-level in silicon, and lastly T(E) is the probability of tunneling as a function of electron energy. The evaluation of this integral will be the subject of the next section, the focus for the remainder of this section will be on finding an approximate expression for T(E).

To evaluate the current integral expressed in Eq. 4-2 an approximate expression for T(E) will be used based on the WKB approximation. A complete derivation can be found in Bohm and Merzbacher [20-21]. Fig. 4-2 depicts the DB structure upon which the analysis of T(E) will be based. In Fig. 4-2 the DB is symmetrically centered about the origin of the coordinate system and t_{ox} and w refer to the oxide and silicon well thicknesses, respectively. The structure is treated as unbiased, which introduces the first simplifying step since the resonant states of this structure will be used at all bias levels. The diagram is divided into five regions. Electrons impinge upon the first

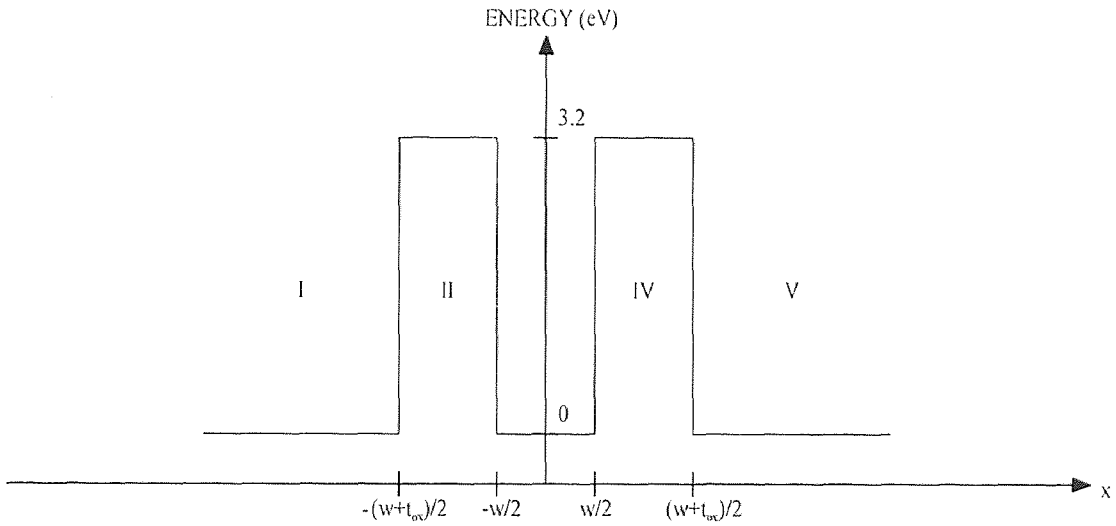

Fig. 4-2: Double barrier profiles used to calculate transmission probability.

barrier traveling from region I towards region II. After tunneling through both barriers the electrons will emerge in region V and continue traveling in the +x direction. The expression for the transmission coefficient, using the WKB approximation, is given by,

$$T=\frac{4}{(4\theta^2+\frac{1}{4\theta^2})^2\cos^2 L+4\sin^2 L} \tag{4-3a}$$

$$\theta=\exp(\int_0^{t_{ox}}\kappa(x)dx) \tag{4-3b}$$

$$L=\int_0^{w}k(x)dx \tag{4-3c}$$

k(x) and κ(x) are trivial for rectangular potentials since they are constant for a given energy (not functions of x), they are defined in the usual way as,

$$k(x)=\frac{2\pi}{h}\sqrt{2m_s E} \tag{4-4a}$$

$$\kappa(x)=\frac{2\pi}{h}\sqrt{2m_{ox}(V_0-E)} \tag{4-4b}$$

where m_s and m_{ox} represent the electron effective mass in silicon and SiO_2, respectively, and V_0 is the barrier heights. Fig. 4-3 shows a plot of T(E) along with a plot of T(E) for a single barrier with a thickness of $2t_{ox}$. The plot of T(E) for the DB follows T(E) for the single barrier, except for the addition of the resonant peaks and the dips below T(E) for the single barrier, between the resonant peaks. Both of these effects can be attributed to either constructive or destructive interference between electron waves. This means that when many scattering events take place as electrons cross the structure, which will "wash-out" the interference effects, the tunneling

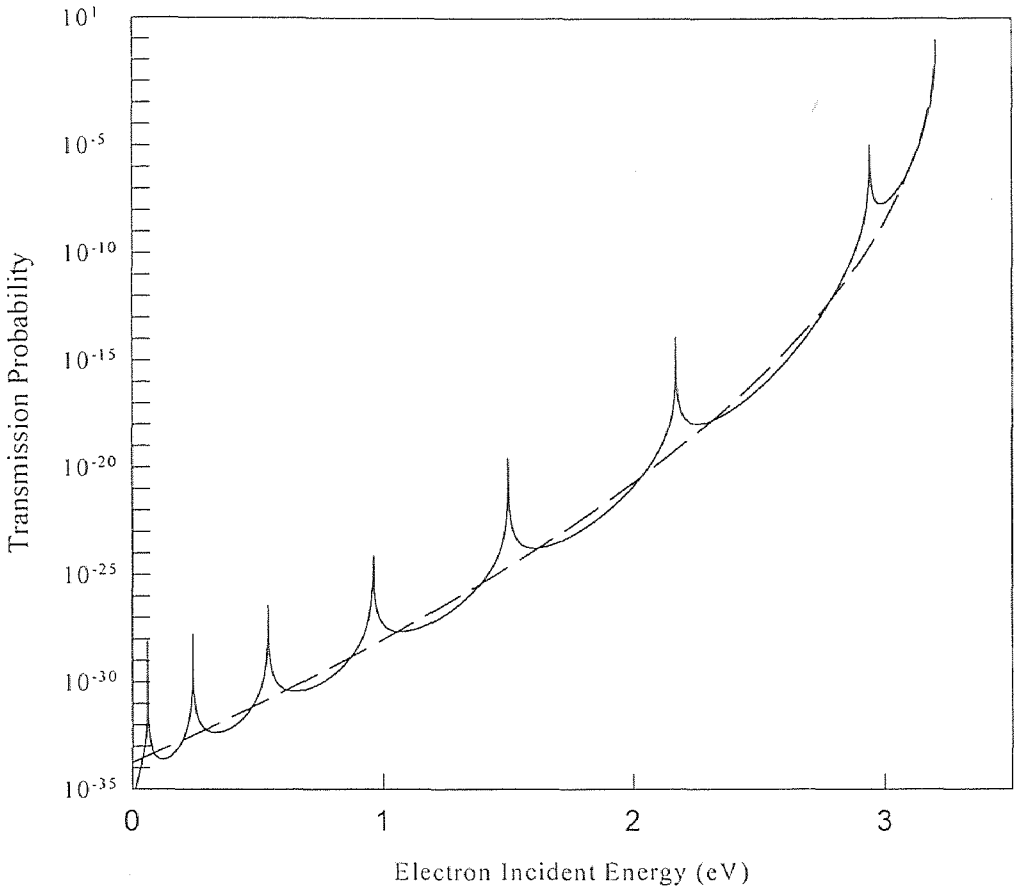

Fig. 4-3: Transmission probability versus electron incident energy for double barrier and single barrier with an oxide thickness of twice the double barrier oxides. The solid line represents the double barrier.

probability will reduce to that of a thin oxide with a thickness of $2t_{ox}$, as one might expect.

The tunneling probability as given by Eq. 4-3 leads to a very simple expression for the resonant states. Since the cos and sin functions in Eq. 4-3 both have the same argument, T(E)=1 when cos(L)=0. Applying this condition, the resonant states are given by,

$$\int_0^w k(x)dx=(2n+1)\frac{\pi}{2} \tag{4-5a}$$

or upon evaluation of the integral,

$$E_n=\frac{(2n+1)^2h^2}{32m_s w^2} \tag{4-5b}$$

Eq. 4-5 is the same expression the WKB approximation gives for the bound states of a potential well, predicting that the resonant states of the DB are the same as the bound states in the equivalent Quantum Well (by equivalent it is meant that the well depth is V_0 and width is w). Since an exact solution of the Quantum Well problem is straightforward (see Gasiorowicz [31]), a comparison between the two will be made. Also, a straightforward exact calculation for the resonant states of the DB structure is given by Roy [10],

$$\tan^{-1}\sqrt{\frac{m_{ox}}{m_s}\frac{V_0-E_n}{E_n}}-\frac{\pi w}{h}\sqrt{2m_sE_n}=-\frac{n\pi}{2} \tag{4-6}$$

Both the Quantum Well solution and Eq. 4-6 lead to transcendental equations that

must be solved numerically. The equations were solved for w = 2 nm, and t_{ox}= 3 nm in the DB case, and are given in table 4-1.

The results in table 4-1 confirm the prediction that the Quantum Well bound states are the same as the DB resonant states, but from the exact calculations, not Eq. 4-5. Even though the WKB resonant states are not very accurate, there is a way to use the results from the exact calculation of the DB resonant states (Eq. 4-6) combined with the WKB approximation.

Bohm demonstrates that for electron energies near the resonant states, Eq. 4-3 reduces to,

$$T(E) \cong \frac{1}{1+\frac{4\pi^2\tau^2}{h^2}(E-E_n)^2\theta^4} \tag{4-7a}$$

where τ is the amount of time an electron travelling classically would require to make two passes through the silicon well. This can be expressed as,

$$\tau = 2w\sqrt{\frac{m_s}{2E}} \tag{4-7b}$$

Using the exactly calculated resonant states and Eq. 4-7 a function can be constructed to give good approximate results with the resonant peaks at the correct energies. When investigating Eq. 4-7 it was found that it gives the same results as Eq. 4-3 within the energy range of interest (to be discussed in the next section). This is not surprising because θ dominates Eq. 4-3. For the purpose of calculating the current density integral, Eq. 4-7 will be used for T(E) and the exact resonant states will be calculated using Eq. 4-6. One difficulty arises, however, when using Eq. 4-7 or Eq. 4-3. Both

Table 4-1: Comparison of resonant state calculations based on WKB approximation, and exact Quantum Well and Double Barrier calculations.

Index (n)	WKB (eV)	Quantum Well (eV)	Double Barrier (eV)
0	0.01	0.06	0.06
1	0.12	0.24	0.24
2	0.40	0.54	0.54
3	0.83	0.97	0.96
4	1.42	1.53	1.50
5	2.16	2.21	2.17
6	3.06	2.98	2.94

equations are very sensitive to the value of E near the resonant states. If the value for E at the resonant state is not made exactly equal to E_n then the result can be orders of magnitude below unity. Because of this another simple approximation of T(E) will be used to evaluate the current integral.

Using a second simple approximation of T(E) also provides an opportunity to determine what the important features of the RT transmission probability curve are. That is, to determine whether it is the peak at unity, or the spread (bandwidth) of the peaks about the resonant states that make the most important contributions to the current integral. This second approximation of T(E) will be given by,

$$T(E)=e^{-\frac{8\pi}{h}\int_0^{t_{ox}}\sqrt{2m_{ox}(V(x)-E)}}+(1-e^{-\frac{8\pi}{h}\int_0^{t_{ox}}\sqrt{2m_{ox}(V(x)-E)}})\sum_n \delta(E-E_n) \qquad (4\text{-}8)$$

where $\delta(E-E_n)$ is the Dirac Delta function. Eq. 4-8 is the equation for tunneling

through a single barrier (see Eq. 4-1) of thickness $2t_{ox}$, replaced by unity at the resonant states.

4.4 Evaluation of Current Density Integral for Double Barrier Potential

To determine if the model of RT presented in the previous section is valid, at least to some degree, one will measure the current density versus voltage on an experimental DB structure. This will require the evaluation of Eq. 4-2 for comparison. The integration of Eq. 4-2 is not trivial and therefore numerical techniques will need to be employed. Moreover, aspects of semiconductor physics need to be considered to determine a valid bias regime. Fig. 4-4 shows a band diagram for an experimental test structure with an P-type silicon substrate, which exaggerates the band bending that will occur at the substrate oxide interface. In Fig. 4-4 the silicon well is depicted as intrinsic, and therefore band bending will occur at its interfaces as well, for simplicity it will be ignored in the following analysis.

When a negative bias is applied to the gate of the device, as is the case in Fig. 4-4, the bands in the substrate will bend up. This is referred to as Accumulation in the P-type substrate [11]. Under this bias condition the device is said to be Tunnel-Limited and the JV curve can be based simply on the evaluation of Eq. 4-2. This is *NOT* true when a positive bias is applied to the gate. Under this bias condition, when the bias is not strong enough to cause Inversion [11], the bands in the silicon substrate are bent down. The hole concentration near the oxide-silicon interface will decrease and a region of fixed donor ions, devoid of charge carriers, will form and extend into the substrate from the interface [11]. This is referred to as Depletion and under this

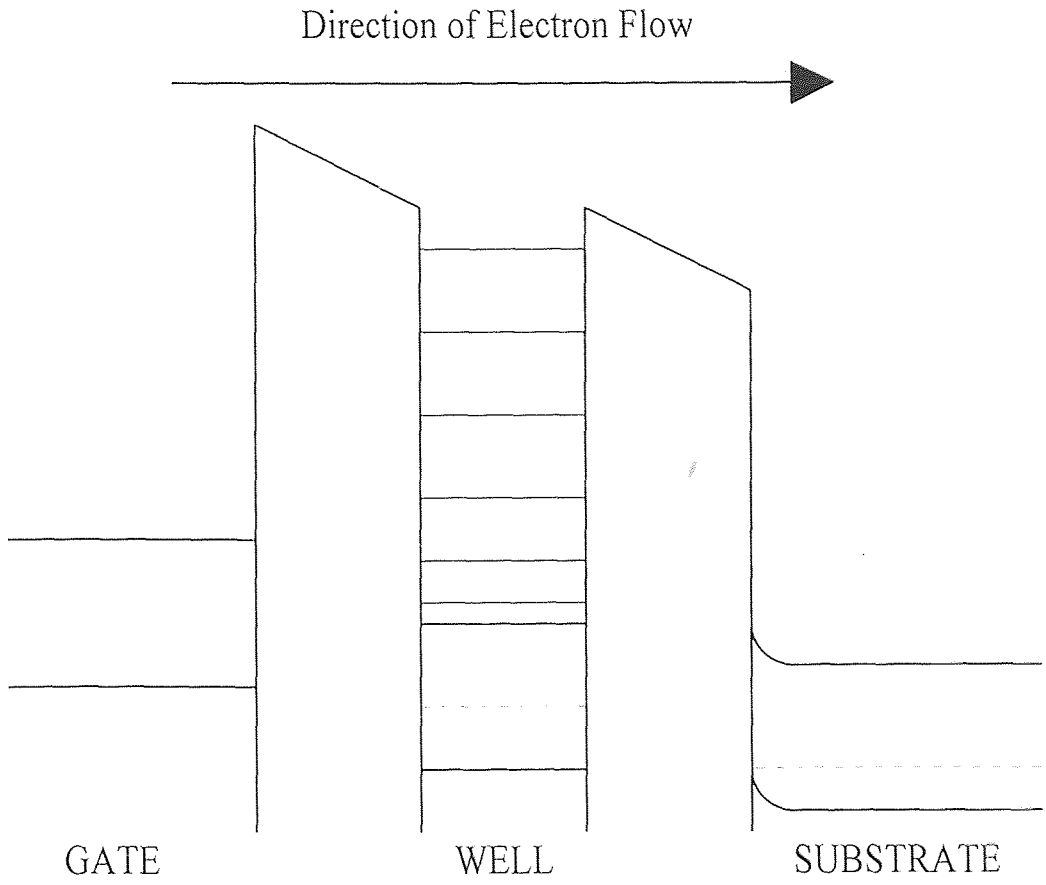

Fig. 4-4: Double barrier structure under negative gate bias.

bias condition the device is said to be Depletion limited. Its JV characteristics will no longer be found simply from Eq. 4-2. For simplicity in the analysis to follow, only the JV characteristics for the Tunnel-Limited case will be calculated, therefore, only a negative gate bias will be considered. Before the entire bias regime is specified, however, factors placing an upper limit of the gate bias must be addressed.

The WKB approximation that has been used to find T(E) (Eq. 4-7) becomes increasingly inaccurate as the bias across the oxide layers is raised. This is not a problem, however, since the purpose of studying the RT diode's characteristics

is for an application restricted to low bias (see chapter 5). It will be assumed that the total voltage across the device divides evenly across each oxide layer and therefore the upper limit on the bias will be placed at -3 volts (-1.5 volts across each oxide). Hence, the entire bias regime has been defined, ie.- 0 to -3 volts. One last factor needs to be considered and this is the band bending induced at the interface of the substrate and oxide layer.

The result of band bending under a negative gate bias is to effectively lower the Fermi-level with respect to the band edges in the substrate silicon, at the interface. thereby changing the number of electrons tunnelling across the barrier. A very simple approach will be used to account for this during the evaluation of Eq. 4-2. As was previously stated, it will be assumed that half of the total bias across the entire structure appears across each oxide layer. Treating the silicon/oxide/silicon sandwiches as ideal capacitors, the surface charge at the oxide-substrate interface will be calculated using Guass's law,

$$\frac{V}{t_{ox}} = -\frac{\sigma}{\epsilon_{ox}} \tag{4-9}$$

where the left hand side of Eq. 4-9 is the electric-field expressed in terms of the voltage and oxide thickness, σ is the surface charge density at the oxide-substrate interface, and ε_{ox} is the permittivity of SiO_2. The surface charge in the substrate silicon can be related to the band-bending at the surface using,

$$\rho(x) = q[N_D - 2n_i \sinh\frac{k_B T}{q}(\phi_B - \phi(x))] \tag{4-10}$$

which is the usual expression for the charge density within a semiconductor, and ϕ_B

and $\phi(x)$ are defined consistent with the treatment given by Sze [11]. Eq. 4-10 can be used to find the surface charge density at the oxide-substrate interface by assuming, when in Accumulation, all of the holes are concentrated into a very thin layer at the interface. This assumption leads to the convenient integral representation,

$$\sigma = \int_{-\infty}^{\infty} \rho\, \delta(x) dx \tag{4-11}$$

Upon substitution of Eq. 4-10 into Eq. 4-11 the surface charge density becomes,

$$\sigma = q[N_D - 2qn_i \sinh \frac{k_B T}{q}(\phi_B - \phi_s)] \tag{4-12}$$

where $\phi(x)$ is replaced by the constant ϕ_s at the interface only. A value for ϕ_s is found by rearranging Eq. 4-12, which is then used in Eq. 4-2b to account for the effective shift in the Fermi-level. This changes Eq. 4-2b to,

$$N(E) = \frac{4\pi m_s q k_B T}{h^3} \ln \frac{1 + e^{\beta(E_f - E)}}{1 + e^{\beta(E_f + q\phi_s - E - qV)}} \tag{4-13}$$

With a bias regime defined and band bending taken into account, the integral of Eq. 4-2 can be evaluated. It should be pointed out that Eq. 4-2 treats the conduction band of the gate as the energy zero for the entire system. The integral does not have to be taken to infinity since the concentration of electrons, as determined by the fermi function and density of states [22], approaches less than 1000 electrons per cm^3 per eV as E approaches 1.4 eV. It was found that taking the integral to 1 eV proves sufficient. Fig. 4-5 shows the results of the integrations using the two expression for T(E) in Eq. 4-2, at T=4K, in normalized units.

Having presented an analysis of the RT phenomena, device structures can be fabricated to look for the JV characteristics predicted in Fig. 4-5. Because of non-uniform oxide thicknesses and electron-phonon interactions, the ideal curves presented in Fig. 4-5 can never be achieved. However, the significance of Fig. 4-5, with respect to device design, is that a large current may be achievable at low bias. Although, the peaks in a real JV characteristic will not be as many orders of magnitude above the conventional tunnel current as depicted in Fig. 4-5. The significance of the large current, however, is that it will allow fast switching rates at low bias. If this were not the case then defeating the DT degradation would not matter, because practical devices (fast switching times) could not be made. With the combined information that has been presented in chapters 1 through 4, a novel approach to the design of FG EEPROM memory cells will now be presented.

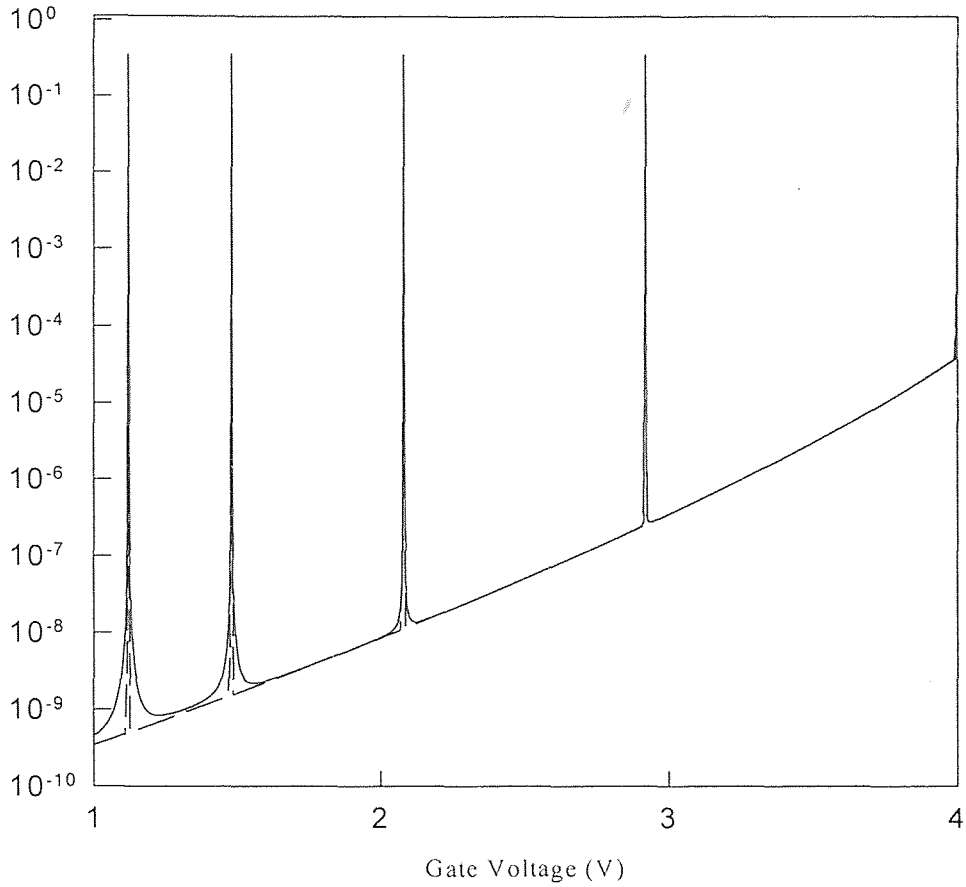

Fig. 4-5: Ideal current density versus voltage characteristics for an RT diode. The dashed curve is the result using Eq. 4-8 for T(E) in Eq. 4-2. The solid curve is the result using Eq. 4-7.

CHAPTER 5

OPERATING PRINCIPLES OF RESONANT TUNNELING ELECTRICALLY ERASABLE PROGRAMMABLE READ ONLY MEMORY CELL

5.1 Introduction

A practical memory cell requires two stable operating states with which it represents a logical "1" and a logical "0". This has been realized in conventional floating gate (FG) electrically erasable programmable read only memory (EEPROM) cells by the use of a FG. Depicted in Fig. 5-1, the FG is an island of metal or polycrystalline silicon electrically isolated from the gate electrode by the field oxide, and electrically isolated

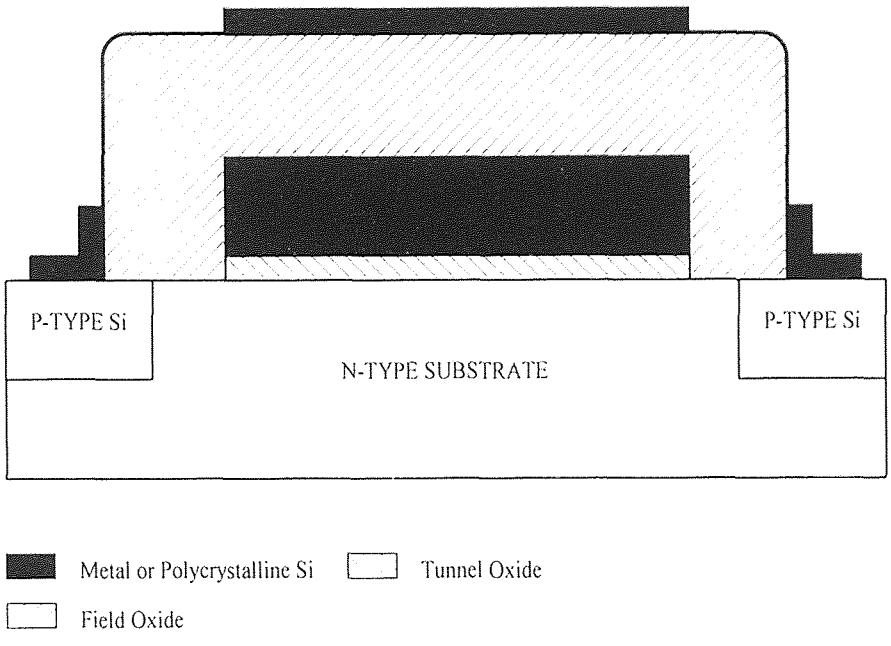

Fig. 5-1: Schematic diagram of EEPROM memory cell.

from the substrate by a tunnel oxide layer. Charge can be transported into and out of the FG by tunneling electrons across the tunnel oxide. The thickness of the tunnel oxide is determined from two requirements: (1) it must be thin enough so that electrons can tunnel to the floating gate at a reasonable bias, and (2) it must be thick enough so that electrons do not "leak" off the floating gate by direct tunneling back through the oxide. The typical tunnel oxide thickness is 5.0 - 10.0 nm, so the conduction mechanism through the oxide is Fowler-Nordheim tunneling. This transport mechanism, however, leads to degradation in the tunnel oxide and therefore limits the number of times the charge state of the FG can be changed.

Since the FG is usually implemented in a modified Metal Oxide Semiconductor Field Effect Transistor (MOSFET), sensing the charge state of the FG is achieved by passing current from the Drain to the Source of the transistor. The charge state of the FG alters the conduction properties of the substrate silicon just below the tunnel oxide layer. For example, if the Drain and Source are formed from P-type silicon and the substrate from N-type, when positive charge is stored on the FG the Drain to Source path is poorly conductive. But, when negative charge is stored on the FG, an inversion layer can be formed just below the tunnel oxide layer and create a highly conductive path between the Drain and Source. In this way the two logical states are achieved.

In what follows the operating principles of a novel approach to the design of FG EEPROM memory cells will be presented. Essentially, the single tunnel oxide layer is replaced by a $SiO_2/Si/SiO_2$ sandwich (double barrier as described in chapter 4). The goal of this approach is to effect the transport of electrons to and from the FG via resonant tunneling. This will keep the required bias, for writing and erasure, below the

threshold for the onset of DT degradation (discussed in earlier chapters). In this way, it may be possible to achieve an EEPROM cell that will not suffer appreciable degradation during its use over any realistic time frame.

5.2 Device Function

5.2.1 Critical Design Parameters

The proposed device structure is depicted in Fig. 5-2. Similar to conventional FG EEPROM design, the Resonant Tunneling EEPROM (RTEEPROM) is essentially a MOSFET with the inclusion of a FG and tunnel oxide layer. The tunnel oxide, however, is separated into two ultra-thin layers (~3nm) by an ultra-thin layer of Silicon (~2nm). The dimensions of these layers are the first critical design parameters since the

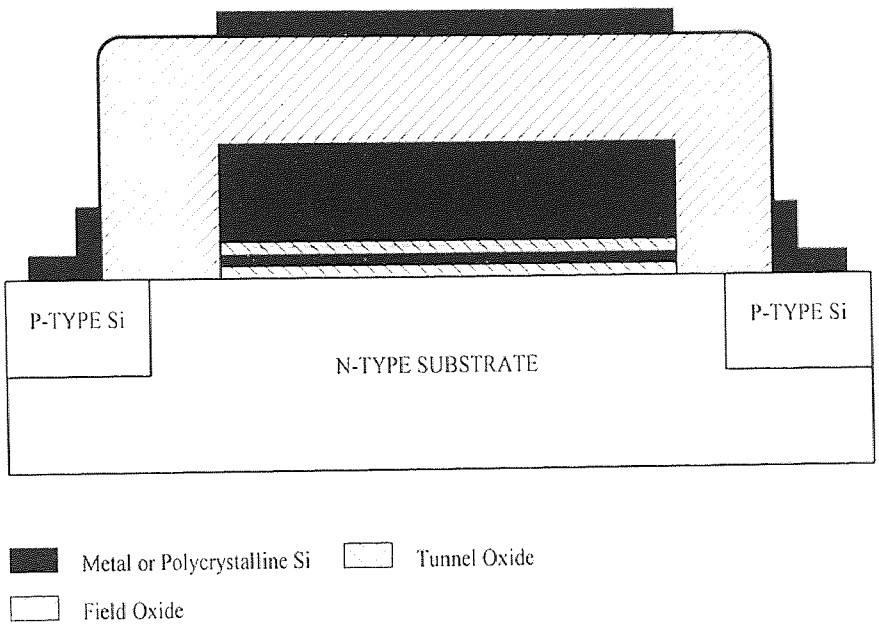

Fig. 5-2: Schematic diagram of RTEEPROM.

quasi-bound states of the silicon well depend on them as discussed in Gasiorowicz [31]. The ground state of the silicon well increases as the silicon well width decreases. In what follows, it will be demonstrated that the ground state of the silicon well determines the amount of charge stored in the FG. An additional critical design parameter is the doping concentration of the substrate. The ability of the charge stored in the FG to push the substrate into inversion is strongly dependent upon the doping concentration of the substrate.

5.2.2 Device Function

Fig. 5-3 shows the band structure of the tunnel layers sandwiched between the FG and the substrate silicon in four different states. It is assumed that the Silicon Well is intrinsic, and the substrate and FG Silicon is N-type. In Fig. 5-3a the equilibrium or "no charge" state is depicted. All the fermi levels are aligned with each other, and a slight amount of band bending occurs in the substrate due to the difference in Work functions between the Silicon Well and substrate. The band bending gives rise to a small amount of charge stored at the interfaces. In Fig. 5-3b a positive voltage (V_a) has been applied to the gate of the transistor. This causes the fermi level of the substrate to fall below the fermi level of the FG, therefore, electrons begin to tunnel into the FG as indicated by the arrow. A negative charge forms in the FG and therefore a positive charge forms in the substrate. As electrons tunnel into the floating gate an internal electric field (E_{int}) develops which opposes the applied field (E_a). Once $E_{int}=E_a$, the Fermi levels again line up and electron flow halts. Fig. 5-3c shows the situation directly after the applied bias is removed. E_{int} now creates a potential

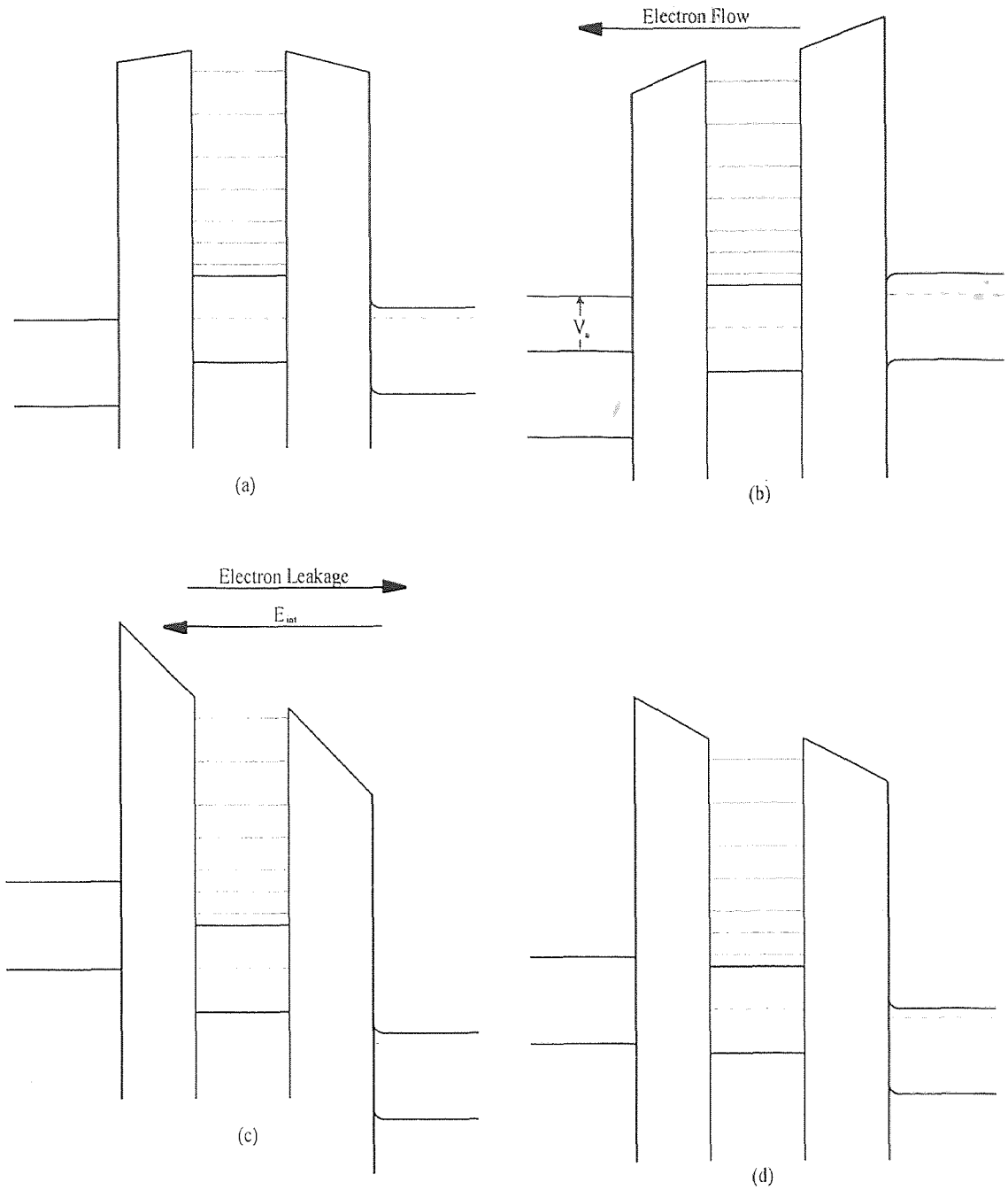

Fig. 5-3: Band diagrams of Double Barrier depicted in various states.

difference across the double barrier and the Fermi level of the FG is above the Fermi level of the substrate. The upward bending of the conduction band in the substrate reflects the fact that positive charge is stored there. Electrons will now tunnel out of the FG, and will do so until the fermi level of the FG lines up with the ground state of the silicon well. This is depicted in Fig. 5-3d and shall be referred to as the Quiescent State. In this state any further leakage of charge from the FG would lower the Fermi level of the FG below the ground state of the silicon well. Since the electron population in the FG essentially vanishes above the Fermi level, charge in the FG would no longer have a path to tunnel through and therefore charge leakage due to resonant tunneling must halt. Since DT from the FG requires passing through the combined thickness of both oxides, the probability for tunneling will be vanishingly small. In this way charge retention on the FG can be achieved, as well as transporting charge at low bias to defeat DT degradation.

5.3 Inversion of Substrate at Quiescence

The Quiescent state is critical to the operation of the device. When the device is in this state the path between the Drain and Source becomes highly conductive. This occurs when the charge on the FG pushes the substrate into inversion. When a positive bias is applied to the gate, electrons will tunnel into the FG. If left undisturbed the FG will remain at the quiescent state with a net negative charge stored in it. The negative charge on the FG will induce a positive charge in the substrate. The positive charge in the substrate can be formed by either the creation of a space charge region, or the combination of a space charge region and a positive surface charge of holes [11]. The

latter case is the inversion phenomena. To ensure that the substrate is in inversion when the device is in the quiescent state, the doping concentration of the substrate must be chosen properly. The following demonstrates how this choice can be made.

Fig. 5-3d shows a diagram of the device in the Quiescent state. Since the voltage drop across the double barrier is 0.12 V, the surface charge in the FG and substrate can be calculated using $\sigma=C_{ox}V_{ox}$, where C_{ox} is the total capacitance of the tunnel layers and V_{ox} is the voltage across the tunnel layers. The total capacitance of the tunnel layers can be found from $C=\varepsilon/t_{ox}$. Using $\varepsilon_{ox}=3.9\varepsilon_0$, $t_{ox}=3$ nm, and treating the tunnel layers as two ideal capacitors in series, the total capacitance of the tunnel layers is $C_{ox}=.006$ F/m^2. This leads to $\sigma=720$ μC/m^2. Since the charge in the FG is negative it will be assumed that all of σ is created by a very thin layer of electrons at the interface of the FG and tunnel oxide. The positive charge in the substrate, however, will need to be a combination of a thin layer of holes and a space charge region.

The space charge region in the substrate forms as a consequence of the negative charge in the FG repelling the electrons in the substrate near the substrate-tunnel oxide interface. Since the positively charged donor ions in the substrate are fixed in position, they form an effective positive surface charge at the interface of the substrate and tunnel oxide. The magnitude of the effective surface charge is given by,

$$\sigma_s=qN_DW \tag{5-3}$$

In Eq. 5-3 W is the width of the space charge region. The space charge region will reach a maximum width, W_m, which is determined by the substrate dopant

concentration. Once W_m is reached the positive surface charge is completed by a thin layer of holes that forms at the interface of the substrate and tunnel oxide.

As is depicted in Fig. 5-3d all of the bands in the substrate, except for E_f, are bent upward as a result of the stored charge in the device. Following the discussion found in Sze [11], the potentials ψ and ψ_B can be defined such that $\psi=E_i-E_{i0}$ and $\psi_B=E_f-E_{i0}$ (refer to [11]). With these definitions the carrier concentration in the substrate is given by,

$$\rho = q[N_D - 2n_i \sinh(\frac{q}{k_B T}(\psi_B - \psi))] \tag{5.4}$$

It can be seen in Eq. 5.4 that as ψ approaches ψ_B the total charge concentration in the substrate approaches N_D. When ψ exceeds ψ_B the total charge concentration in the substrate must be partly due to N_D and partly due to holes, since N_D is only a constant in the equation. Since the holes are mobile it is assumed that they are attracted to the surface. Following Sze [11], the calculations used to find N_D for the substrate will assume that the charge density in the space charge region is given entirely by N_D. With the stated assumptions, N_D for the substrate will now be calculated.

Sze [11] demonstrates that the maximum space charge region width is given by,

$$W_m = \sqrt{\frac{4\varepsilon_s k_B T}{q^2 N_D} \ln(\frac{N_D}{n_i})} \tag{5.6}$$

This means the space charge region will only grow as wide as W_m before inversion sets in. In choosing N_D, therefore, we want W calculated from Eq. 5.3 to be greater than W_m so as to ensure that the substrate goes into inversion. Fig. 5-4 shows a plot of W and W_m versus doping concentration, where W is represented by the dashed line and W_m is represented by the solid line. Clearly, as long as N_D is kept below approximately 10^{22} m^{-3} the substrate will be in inversion.

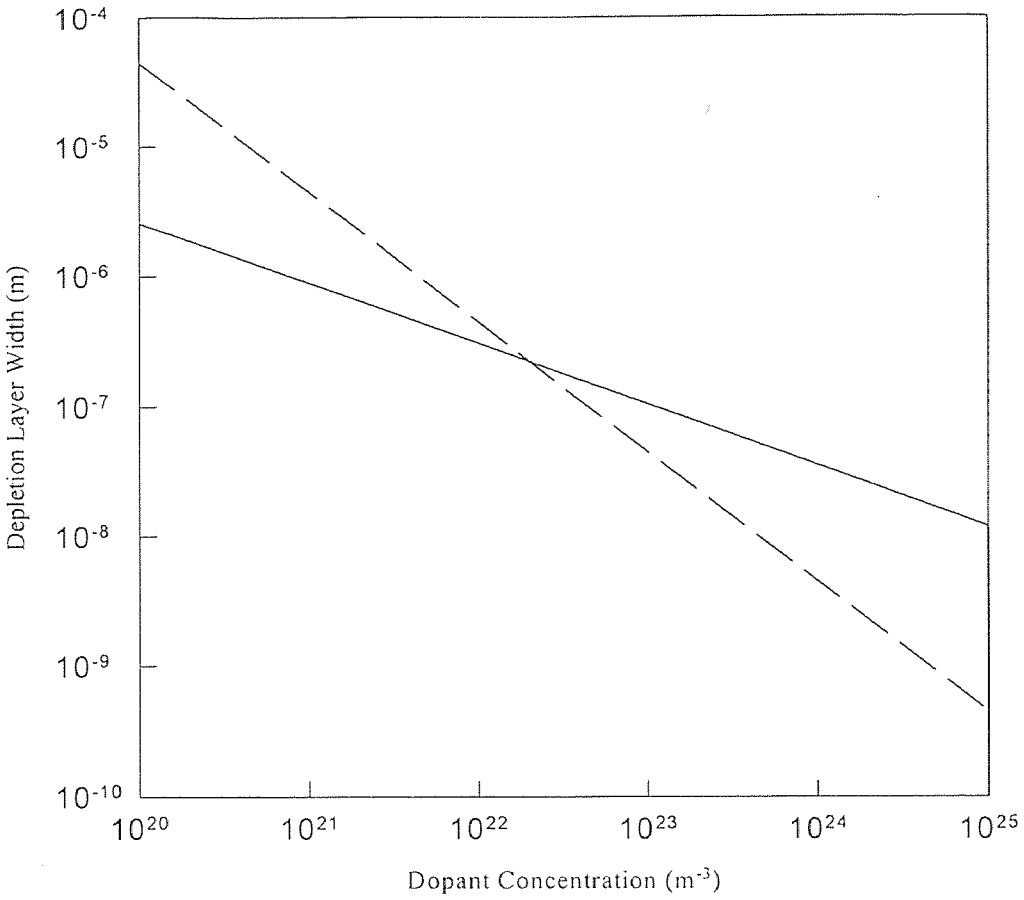

Fig 5-4: Depletion layer width and maximum depletion layer width versus dopant concentration.

REFERENCES

1. R. E. Hummel, *Electronic Properties of Materials*, Springer-Verlag, New York, NY, 1993.

2. R. Tsu, "Silicon-Based Quantum Wells," Nature, vol. 364, pp. 19, July 1993.

3. R. D. Pashley, and S. K. Lai, "Flash Memories: The Best of Two Worlds," IEEE Spectrum, pp. 30-33, December 1989.

4. A. Watts, "Unique Cell Design Gives High EEPROM Endurance," Electronic Engineering, pp. 43-46, July 1992.

5. S. Tiwari, H. Hanafi, A. Hartstein, E. F. Crabbé, and K. Chan, "A Silicon Nano-Memory," IEEE Electron Device Letters, Submitted.

6. K.R. Farmer, M.O. Andersson, and O. Engström, "Tunnel Electron Induced Charge Generation in Very Thin Silicon Oxide Dielectrics," Applied Physics Letters, vol. 58, pp. 2666-2668, June 1991.

7. D. J. DiMaria and J. W. Stasiak, "Trap Creation in Silicon Dioxide Produced by Hot Electrons," Journal of Applied Physics, vol. 65, pp. 2342-2355, March 1989.

8. D. J. DiMaria, "Temperature Dependence of Trap Creation in Silicon Dioxide," Journal of Applied Physics, vol. 68, no. 10, pp. 5234-5245, November 1990.

9. C. B. Duke, *Tunneling in Solids*, Academic Press, New York, NY, 1969.

10. D.K. Roy, *Quantum Mechanical Tunnelling and its Applications*, World Scientific, Philadelphia, PA, 1986.

11. S. Sze, *Physics of Semiconductor Devices*, Wiley Interscience, New York, NY, 1981.

12. B. Euzent, N. Boruta , J. Lee, and C. Jenq, "Reliability Aspects of a Floating Gate E^2PROM," IEEE Proceedings on Semiconductor Device Reliability, pp. 10-16, 1981.

13. J. S. Witters, G. Groeseneken, and H. Maes, "Degradation of Tunnel-Oxide Floating-Gate EEPROM Devices and the Correlation with High Field-Current-Induced Degradation of Thin Gate Oxides," IEEE Transactions on Electron Devices, vol. 36, no. 9, pp. 1663-1681, September 1989.

14. K. R. Farmer, M. O. Andersson, and O. Engström, "Negative Charging in Ultrathin Metal-Oxide-Silicon Tunnel Diodes," Journal of Applied Physics, vol. 71, no. 4, pp. 1846-1852, February 1992.

15. S. Aritome, R. Shirota, G. Hemink, T. Endoh, and F. Masuoka, "Reliability Issues of Flash Memory Cells," Proceedings of The IEEE, vol. 81, no. 5, pp. 776-787, May 1993.

16. C. Papadas, G. Ghibaudo, C. Monserié, and G. Pananakakis, "Reliability Issues of Silicon Dielectric Structures - Application to FLOTOX EEPROM Cells," Microelectronic Reliability, vol. 33, no. 10/11, pp. 1867-1908, 1993.

17. D. A. Buchanan, D. J. DiMaria, C. A. Chang, and Y. Taur, "Defect Generation in 3.5 nm Silicon Dioxide Films," Applied Physics Letters, vol. 65, pp. 1820-1822, October 1994.

18. P. V. Dressendorfer, "Interface and Electron Tunneling Properties of Thin Oxides on Silicon," Ph.D. Dissertation, Yale University, Hartford Connecticut, 1978.

19. N. S. Wingreen, K. W. Jacobsen, and J. W. Wilkins, "Resonant Tunneling with Electron-Phonon Interaction: An Exactly Solvable Model," Physical Review Letters, vol. 61, no. 11, pp. 1296-1299, September 1988.

20. D. Bohm, *Quantum Theory*, Prentice-Hall Inc., Englewood Cliffs, NJ, 1951.

21. E. Merzbacher, *Quantum Mechanics,* John Wiley & Sons Inc., New York, NY, 1970.

22. R. F. Pierret, *Modular Series on Solid State Devices: Advanced Semiconductor Fundamentals*, Addison-Wesley Publishing Company, Reading Massachusetts, 1987.

23. J. Maserjian and N. Zamani, "Behavior of the Si/SiO_2 interface observed by Fowler-Nordheim Tunneling," Journal of Applied Physics, vol. 53, no. 1, pp. 559-567, January 1982.

24. K. R. Farmer, P. Lundgren, M. O. Andersson, and O. Engstrom, "Injection Related Damage in Metal-Oxide-Silicon Tunnel Diodes," Materials Research Society Symposium Proceedings, vol. 284, pp. 229-235, December 1992.

25. K. R. Farmer, M. O. Andersson, and O. Engström, "Time-Dependent Positive Charge Generation in Very Thin Silicon Oxide Dielectrics," Applied Physics Letters, vol. 60, no. 6, pp. 730-732, February 1992.

26. D. Vuillaume, "Nature of the Defects Generated by Electric Field Stress at the Si-SiO_2 Interface," Applied Physics Letters vol. 58, no. 5, pp. 490-492, February 1991.

27. Z. A. Weinberg, "High-field positive-charge generation and its relation to breakdown in a-SiO_2," *The Physics and Technology of Amorphous SiO_2*, pp. 427, Plenun Publishing Corporation, New York, NY, 1987.

28. H. Yamada and T. Makino, "Correlations Between Stress-Induced Positive Charges and Time-Dependent Dielectric Breakdown in Ultrathin Silicon Oxide Films," Applied Physics Letters, vol. 59, no. 17, pp. 2159, October 1991.

29. D. R. Wolters and J. J. Van Der Schoot, "Dielectric Breakdown in MOS Devices Part 1: Defect-Related and Intrinsic Breakdown," Philips Journal of Research, vol. 40, no. 3, pp. 115-136, 1985.

30. K. Schuegraf and C. Hu, "Reliability of thin SiO_2," Semiconductor Science and Technology, vol. 9, pp. 989-1004, November 1993.

31. Gasiorowicz, *Quantum Physics*, John Wiley and Son Inc., NewYork, NY, 1974.