Electron Tunneling Processes in Si/SiO_2 Systems

by

Farhan Rana

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of

Master of Science

and

Bachelor of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1997

ARCHIVES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
MAR 06 1997
LIBRARIES

© Farhan Rana, MCMXCVII. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly paper and electronic copies of this thesis document in whole or in part, and to grant others the right to do so.

Author ....
Department of Electrical Engineering and Computer Science
September 30, 1996

Certified by ....
Dimitri A. Antoniadis
Professor EECS
Thesis Supervisor

Accepted by ....
Frederic. R. Morgenthaler
Chairman, Departmental Committee on Graduate Theses

Electron Tunneling Processes in Si/SiO_2 Systems

by

Farhan Rana

Submitted to the Department of Electrical Engineering and Computer Science
on September 30, 1996, in partial fulfillment of the
requirements for the degrees of
Master of Science
and
Bachelor of Science in Electrical Engineering and
Computer Science

Abstract

Ultra dense electrical memories can be built in which the memory devices utilize direct electron tunneling through very thin Silicon oxides. To understand the behavior of such electron devices we provide a comprehensive theoretical framework for modeling electron tunneling processes in Si/SiO_2 systems, and back it up with experimental evidence. We have developed fully self-consistent quantum mechanical models to calculate tunneling currents through sub-50Å oxides. Our theoretical model is in remarkably good agreement with experimental data. We have also formulated a multiband crystalline WKB approximation that takes into account the correct energy dispersion relation in the mid-gap region of SiO_2. We have also developed a novel method to study the dynamic response of two dimensional electron gas in MOS Field Effect devices to external time-dependent perturbations, and we use it to study the dynamic image potential problem in tunneling. We have also modeled the behavior of electrons in quantum dots coupled to the channel of MOS devices. Such structures have the potential for being used in high density multi-state memories. We provide a theoretical formulation to model the time independent and time dependent behavior of such devices.

Thesis Supervisor: Dimitri A. Antoniadis
Title: Professor EECS

Acknowledgments

I am deeply indebted to Sandip Tiwari for his years of support and guidance. His unique intuition and uncanny insight, both in theoretical and experimental sciences, has been continual source of motivation for me. His remarkable personality, with its very diverse elements of intelligence, humility, aggressiveness, and humor, made my stay at IBM intellectually rewarding, challenging and also very pleasant at the same time. I am also very grateful to Mari, whose wonderful dinners made me feel at home while being at IBM, and of course to Nachiketa and Kunal, who always reminded me of my own younger brothers.

I would like to express my gratitude to Max Fishetti and Paul Solomon for the stimulating discussions and to Steve Laux for teaching me the basics of numerical calculations. I am also thankful to Doug Buchanan, without whose experimental data the theory developed in this thesis could not have been verified. I am also grateful to the staff at IBM including Hussain Hanafi, Khalid Ismail, Kevin Chan, and Way Chan for helping me out at various occasions.

I would also like to thank my MIT thesis advisor Prof. Dimitri Antoniadis for providing important guidance and advice, and my undergraduate academic advisor Prof. Jesus Del Alamo for his valuable guidance and for being always extremely helpful. I am also grateful to Prof. Qing Hu for being very patient with me while I worked on this thesis.

I am specially thankful to my friends for making my life at MIT very enjoyable. With my very old friends Asad Naqvi, Jalal Khan, Agha Mirza, Ahmad Shah, Asim Khwaja, and Yassir Elley I share some of my happiest times and many indelible memories. Their support and friendship has been invaluable and will always be treasured by me. I also wish to express my gratitude to Farhan Bhai and Fuazia Appa for their kind and generous hospitality. I am grateful to my friends in PAKSMIT and MITMSA with whom I spent many pleasant moments. There are so many of them and all so important that I apologize to them for not being able to mention their names here. I am also grateful to my new friends Bin, Christina, Ilya, Arif, Gert, and

Eric for their pleasant and enjoyable company during the last year.

Most of all I am grateful to my family whose unconditional support have made it possible for me to study at MIT. Their love, their sacrifices, and their trust has been a constant driving force in my life. My feelings and sense of gratitude for them cannot be encompassed just by words.

Farhan Rana.

Contents

1 Introduction **12**

1.1 Background 12

1.2 Motivation 14

1.3 Research Presented in this Thesis 17

1.3.1 Physics of Tunneling througn Ultra Thin Oxides 17

1.3.2 Tunneling Processes, and Charge Statistics and Fluctuations in Quantum Dots Coupled to FETs 18

1.4 Conclusion 19

2 Semi-classical Theory of Tunneling Through Oxides **20**

2.1 Introduction 20

2.2 The Physics of the Semi-classical Model 21

2.2.1 Solution of the Poisson Equation 23

2.2.2 Calculation of the Transmission Probability 26

2.2.3 Calculation of the Tunneling Currents 28

2.3 Discussion 30

2.3.1 Conservation of Transverse Kinetic Momentum 30

2.3.2 The Oxide Effective Mass and the Question of the Image Force 32

2.3.3 Quantization of Electron Motion in Accumulation and Inversion Layers 32

2.4 Conclusion 33

3 Self-Consistent Quantum Mechanical Model for Tunneling through

Oxides 34
3.1 Foreword 34
3.2 Self-Consistent Solution of Poisson and Schrodinger Equations 35
3.3 Electron Tunneling from Bound States 39
3.3.1 Calculation of Lifetimes from Path Integral Formulation . . . 40
3.3.2 Calculation of Current from Quasi-Bound States with Non-equilibrium Green Function Technique 45
3.4 Calculation of Tunneling Currents in MOS Devices 50
3.5 Conclusion 52

4 Numerical Results and Comparison with Experiments 54
4.1 Foreword 54
4.2 Semi-Classical and Self-Consistent Results for Inversion Layers 55
4.3 Semi-Classical and Self-Consistent Results for Accumulation Layers and Comparison with Experimental Data 61
4.3.1 Theoretical Results for Accumulation Layers 61
4.3.2 Comparisons With Experimental Data 64
4.4 Conclusion 70

5 Advanced Issues in Physics of Electron Tunneling through Oxides : The Mid-Gap Energy Dispersion Relation in SiO_2 and the Effect of Image Forces 72
5.1 Introduction 72
5.2 Energy Dispersion Relation in Mid-Gap Region of SiO_2 and the Crystalline WKB Approximation 73
5.3 The Dynamic Image Force Problem in Tunneling 80
5.3.1 Background 80
5.3.2 Dynamical Image Potential Problem in MOS Structures and the Dynamic Response of 2-DEG 82
5.3.3 On the Neglect of Image Potential Corrections in Calculating Transmission Probabilities 96

6 Modeling of Electronic Processes in Quantum Dots Coupled to FET'S 100

6.1 Introduction 100

6.2 On the Nature of Coulomb Energy 102

6.3 Quantum Kinetic Equations For Modelling Tuuneling Processes in a Quantum Dot Coupled to an Inversion Layer 104

6.4 Carrier Statistics and Fluctuations Inside the Dot 110

6.4.1 A Two Level Model 112

6.5 Quantized Threshold Voltage Shifts 113

6.6 Channel Conductance Fluctuations 114

6.7 Calculation of Coupling Constants 116

6.8 Numerical Results 118

6.8.1 Numerical Results for the Steady State 118

6.8.2 Numerical Results for the Time Dependent Case 120

6.9 Conclusion 124

7 Conclusion 126

List of Figures

1-1 Planar memory cell 14
1-2 Vertical memory cell 15
1-3 Quantum dot memory cell 16

2-1 Electron Tunneling from Accumulation Layer 22
2-2 Electron Tunneling from Inversion Layer 23
2-3 Region in which the Poisson equation is solved 24
2-4 Region in which the Schrodinger equation is solved to get the transmission probabilities 27
2-5 Carrier Pockets in the First Brillouin Zone of Silicon 29

3-1 Region in which self-consistent solution is sought for accumulation layers 36
3-2 quasi-bound state decaying into the gate 41
3-3 Potential for calculating the coupling constants 49

4-1 Tunneling currents from inversion layer in 15, 20, 25, 30, and 35Å oxide N-channel MOS devices. The substrate doping is $10^{17}/cm^3$. Potential drop in n^+ Poly-Si gate is ignored. 56
4-2 Energies of the first five subbands measured w.r.t. the bottom of the conduction band edge at the Si/SiO_2 interface. The two different sets of curves are for two different pockets in which the electron effective mass perpendicular to the interface is either m_l^{si} or m_t^{si}. The device has an oxide thickness of 15Å and substrate doping of $10^{17}/cm^3$. . . . 57

4-3 The potential drop in the substrate as a function of gate voltage. The device has an oxide thickness of 15Å and substrate doping of $10^{17}/\text{cm}^3$. 58

4-4 The potential drop in the oxide as a function of gate voltage. The device has an oxide thickness of 15Å and substrate doping of $10^{17}/\text{cm}^3$. 59

4-5 The lifetimes of electrons in 15, 20, 25, 30, and 35Å oxide MOS devices. For each oxide thickness, the two sets of curves are for the lowest two subbands corresponding to effective masses m_l^{si} and m_t^{si}. The top most set of curves is for the thickest 35Å oxide. The substrate doping is $10^{17}/\text{cm}^3$ in each case. 60

4-6 Capacitance of 15, 20, 25, 30, and 35Å MOS devices calculated from the self-consistent and the semi-classical model. The substrate doping in each case was $10^{17}/\text{cm}^3$. 61

4-7 The relative error in extracting oxide thicknesses from electrical measurements performed in the strong inversion regime. 62

4-8 Self-consistent charge density for a 15Å MOS device at a gate voltage of 3.0 Volts. 63

4-9 Tunneling currents from accumulation layer in 15, 20, 25, 30, and 35Å oxide n-MOS capacitors. The substrate doping is $10^{17}/\text{cm}^3$. Potential drop in n^+ Poly-Si gate is ignored. 64

4-10 Capacitance of 15, 20, 25, 30, and 35Å n-MOS capacitors calculated from the self-consistent and the semi-classical model. The substrate doping in each case is $10^{17}/\text{cm}^3$. 65

4-11 The relative error in extracting oxide thicknesses from electrical measurements performed in the strong accumulation regime. 66

4-12 Energies of the first five subbands measured w.r.t. the bottom of the conduction band edge at the Si/SiO_2 interface. The two different sets of curves are for two different pockets in which the electron effective mass perpendicular to the interface is either m_l^{si} or m_t^{si}. The device has an oxide thickness of 15Å and substrate doping of $10^{17}/\text{cm}^3$. . . . 67

4-13 The potential drop in the substrate as a function of gate voltage. The device has an oxide thickness of $15\AA$ and substrate doping of $10^{17}/cm^3$. 68
4-14 A typical set of experimental data on tunneling currents from accumulation layers in 14, 20, 23, 27, and $35\AA$ oxides. All devices had a n-type substrate with doping $10^{17}/cm^3$ and a n^+ Poly-Si gate with doping $5 \times 10^{19}/cm^3$. 69
4-15 Tunneling currents from accumulation layers calculated from the semi-classical and the self-consistent model for 15, 20, 25, 30, and $35\AA$ oxides compared with the experimental measurements for 14, 20, 23, 27, and $35\AA$ oxides. All devices had a n-type substrate with doping $10^{17}/cm^3$ and a n^+ Poly-Si gate with doping $5 \times 10^{19}/cm^3$. 70

5-1 Energy band diagram of SiO_2. 78
5-2 Energy dispersion relation in the band gap of SiO_2. The zero of energy is the conduction band of SiO_2. 79
5-3 MOS structure with Aluminum gate. 82
5-4 Dynamical screening charge density of a 2-DEG shown as a function of time and radial distance from the position of the external point charge. $z_o = 12.5\AA$ and $t_{ox} = 30\AA$ 93
5-5 Dynamical image potential felt by an external point charge shown as a function of time. $z_o = 12.5\AA$ and $t_{ox} = 30\AA$ 94
5-6 Dynamical image potential felt by an external point charge shown as a function of time. $z_o = 6\AA$ and $t_{ox} = 15\AA$ 95
5-7 δ_{ox} needed to give the same results for tunneling currents as would a 0.3 Volts reduction in barrier height. 98

6-1 Quantum dot memory cell using Si nano crystals 101
6-2 Single quantum dot memory cell 101
6-3 Quantum dot coupled to a 2-DEG 105
6-4 Mean number of electrons in the dot as a function of gate voltage. . . 120
6-5 Threshold voltage shift ΔV_T as a function of gate voltage 121

6-6 Variance of electron number in the dot as a function of gate voltage . 121

6-7 Mean number of electrons in the dot as a function of time on application of a 3.0 Volt pulse at the gate. 123

6-8 Shift in the threshold voltage of the device as a function of time. . . . 124

6-9 Mean number of electrons in the dot as a function of time on application of a -3.0 Volt pulse at the gate. The four curves are for different initial number of electrons in the dot. 125

Chapter 1

Introduction

1.1 Background

Three different types of semiconductor MOS memories are currently in vogue :

- Dynamic Random Access memories (DRAM), that typically read and write in sub-100 nano seconds time but need to be refreshed in time period of seconds.
- Static Random Access memories (SRAM), that read and write in time period of nano seconds, and do not need to be refreshed.
- Non-volatile memories (NV-RAM), for example E^2PROMs, that also read in sub-100 nano seconds but write in milli seconds and do not need to be refreshed.

SRAMs are used in computation for local fast memories, such as that required for caches. Each SRAM cell usually consists of multiple transistors in flip flop configuration, thus consuming a substantial amount of power and space. This is a major limitation in portable low power applications.

NV-RAMs are used where data is stored for large periods of time and small writing times is not a requirement, such as in programmable logic chips. Each NV-RAM cell may consist of one or more transistors among which at least one transistor is capable of storing charge in a floating gate embedded in the gate oxide. The floating gate is charged by electrons tunneling from the substrate. As a result of the large potential

barrier (3eV) due to conduction band offset between silicon and silicon dioxide, the charge in the floating gate can stay for as long as 10 years. However, to improve charge retention thick oxides (few hundred angstroms) are used. Writing is done at high gate voltages (15-50 volts) with large electric fields in the oxide. Tunneling of electrons takes place in the Fowler-Nordheim regime. In the presence of large electric fields, electrons gain large kinetic energies and these energetic electrons are the primary cause of generating defects in the oxide [29]. This hot carrier induced damage limits the number of times the device can be made to go through write and erase cycles (typically 10^5 or 10^6). This together with the slow write times do not allow the use of this device in the temporary storage of information during computation.

DRAMs are almost always the choice where memory is needed for local temporary storage of information during computation. Each DRAM cell consists typically of a single transistor and a capacitor. DRAMs have provided the highest densities because of the relatively small cell size. With continuing increase in densities, the need to refresh in times smaller than seconds, and reading and writing a large number of bits, DRAM has continued to increase the power consumption, and its beginning to become prohibitive for portable applications. For example, the amount of power consumed by 16 mega bytes, employing 4 mega bit DRAMs, during intensive use is about 12 Watts, and in stand-by condition is about 150 milli Watts [30]. Typical portable battery energy is about 50 Watt-hour/kg. Thus, a 100g battery, with current state of technology, would last half an hour of intensive use using 16 mega bytes of memory.

A sub-100 nano second memory with significantly low power consumption and high density is needed to meet the increasing demands of memory intensive computing. The development of giga bit levels of integration in a chip occupying area of 1 cm^2 requires a bit footprint of less than 0.1 μm^2. With conventional silicon technology, the construction of such a small single cell containing a pass transistor and a capacitor, poses a daunting challenge. Even if this were possible, another challenge arises from the sub-threshold current of pass-transistors of gate lengths less than 1000$\AA$, which result in loss of capacitor charge, decrease in refresh times and, therefore, increase in power dissipation.

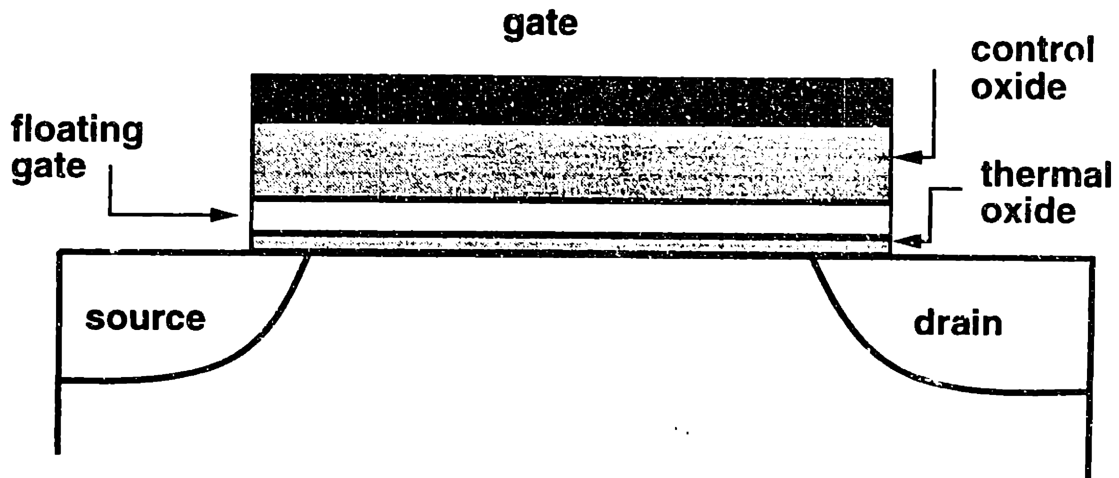

Figure 1-1: Planar memory cell

It seems that large improvements in current memory device technology may require revolutionary changes. In order to meet the challenges of the giga bit generation some novel memory device structures and architectures are being explored at IBM. These devices provide the motivation behind the research being carried out in the present work.

1.2 Motivation

A major limitation of DRAM in power comes about because of the leakage of stored charge and the need to refresh. Density limitations in DRAM and SRAM come about because of multiple elements employed in a single cell. NV-RAMs have long retention times but suffer from long write times. A way to attack all these problems is to use novel single transistor memory cells as shown in figure (1-1) and figure (1-2). The device in figure (1-1) is almost like a E^2PROM but it utilizes ultra thin gate oxide, typically between 15$\mathring{A}$ and 25$\mathring{A}$. Charge can be placed in the floating gate via direct tunnel injection of electrons from the substrate. The advantage of this over conventional E^2PROMs is two fold :

a) Current densities through such thin oxides are large reaching typically around 1-10 A/cm^2 for 15$\mathring{A}$ oxides with gate voltages between 1-2 volts. Thus to put a charge of 10^{12} electrons/cm^2 in the floating gate would require a write time of less than 100

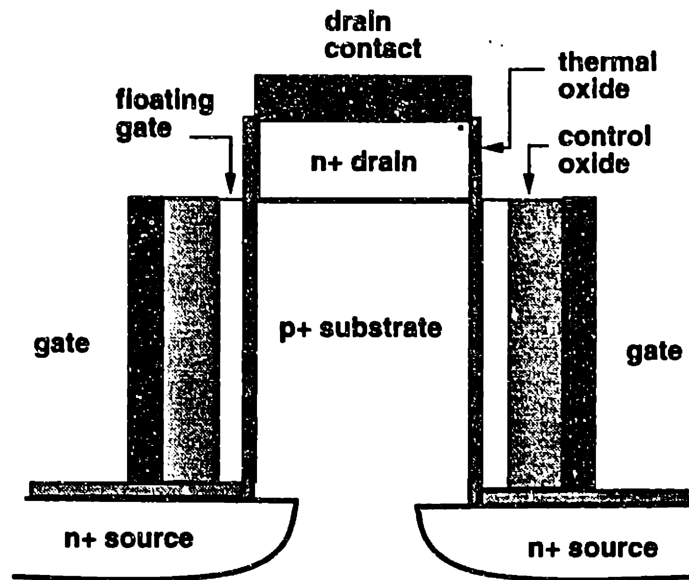

Figure 1-2: Vertical memory cell

nano seconds at fairly low gate voltages. This speed is almost comparable to that of DRAM.

b) Current injection takes place via direct tunnel injection (as opposed to Fowler-Nordheim injection). Electrons thus do not acquire sufficient kinetic energy while traversing the oxide. It has been experimentally observed that two most common type of hot carrier related oxide degradation mechanisms are trap creation by release of mobile hydrogen, and generation of trapped holes via impact ionization [29]. The former process requires electrons' energies over 2eV, and the later requires energies over 9eV. Electrons undergoing direct tunneling are therefore not expected to contribute significantly to such degradation mechanisms. Preliminary experimental results on thin oxide memory devices show that they can be cycled (through write and erase operations) more than 10^{10} times without any significant oxide degradation [31].

The advantage over DRAM lies in the use of just a single transistor to store a bit without the need of a capacitor which usually occupies a large area. In addition these devices have large retention times (form few tens of seconds to minutes) [31]. Without the need of frequent refresh, these devices are expected to consume far less power than conventional DRAM. With short writing and erasing times, it is expected

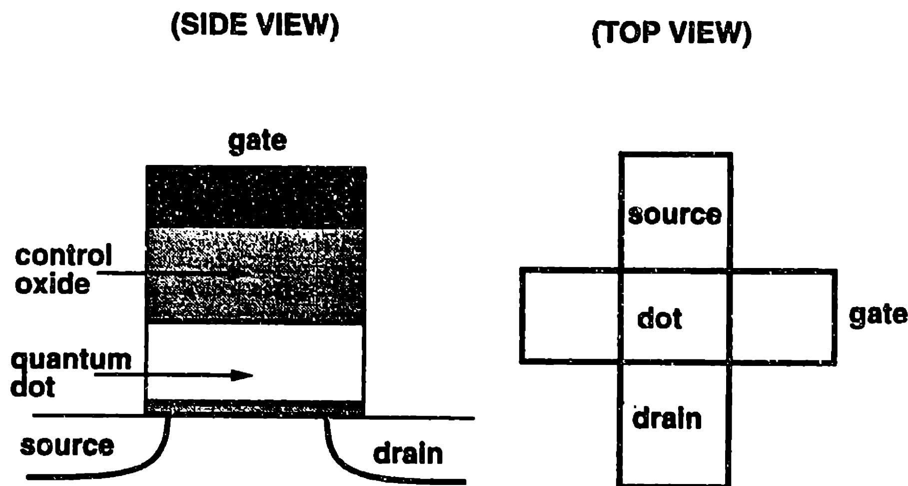

Figure 1-3: Quantum dot memory cell

that these DTRAMs (direct tunneling RAMs) may be able to compete with DRAMs in the temporary storage of information during computation.

Extensions of the above idea can be used in more useful designs. Two such designs will be described here. Figure (1-2) shows a vertical DTRAM. The channel is vertical with a poly-silicon gate surrounding it. This structure can significantly reduce the area required per transistor, making this device an ideal candidate for the giga bit generation. Also the subthreshold characteristics are expected to be better than planar designs, especially at very short channel lengths (less than 0.25 μm). Another very useful design is shown in figure (1-3). In this device the floating gate is in the form of a small quantum box (or dot), coupled to the substrate via a thin thermal oxide. The dimensions of the box are small enough so that effects associated with coulomb energy are significant. Specifically, it is expected that the number of electrons in the box will go up with an applied positive bias on the gate in discrete jumps (electron number quantization). The shifts in threshold voltage, on applying a voltage pulse on the gate, is also expected to have discrete values, depending upon the magnitude and duration of the pulse. This device offers the unique possibility of a multi-state memory cell consisting of only a single device.

1.3 Research Presented in this Thesis

The operation of DTRAM cells relies heavily on the physical and electronic properties of thermal oxides. To understand the behavior of DTRAM devices, it is necessary to study these properties, especially those of ultra thin oxides. The aim of this thesis research was to study the physics of quantum mechanical tunneling in ultra thin oxides, develop theoretical models for DTRAM structures, and to verify these models with experimental data. In modeling DTRAMs we have also focused upon developing the physics associated with coupling quantum dots with the channel of FETs, since this has never been studied before. Below we give a brief summary of the problems we have addressed in this work.

1.3.1 Physics of Tunneling through Ultra Thin Oxides

The physics of tunneling through oxides has been studied previously by many authors [32]. However, with recent developments in condensed matter physics, the availability of fast computers for large calculations and the need to predict accurately tunneling currents in ultra thin oxides it is necessary to develop more accurate models for tunneling processes. There are many questions related to electron tunneling through thin oxides that we feel are not answered satisfactorily in the existing literature. Some of these are :

Self-consistency : The tunneling currents are strongly depended upon the quantity of electron charge present in accumulation or inversion layer, and also the distribution of these electrons in energy. We know that self-consistent solutions generally give different results for these quantities than the semi-classical solutions. Therefore, it is necessary to calculate tunneling currents using fully self-consistent models. In chapter three we present a fully self-consistent model for calculating tunneling currents.

Transmission probability vs life-times : In present literature all tunneling current models use the concept of transmission probabilities. We feel that this

concept is meaningless for quasi bound quantum states that are present in the accumulation and inversion layers. For these states a more useful concept is that of lifetimes. In chapter three we develop a fully quantum mechanical model to describe tunneling from quasi-bound states.

Mid gap dispersion relations in SiO_2 : Exact energy dispersion relation of electrons tunneling through very thin oxides is presently unknown. In chapter five we present a multiband crystalline WKB approximation to model the mid gap dispersion relation of SiO_2 in the presence of an electric field. Our model is adequate to describe electron tunneling in the direct regime.

Image force corrections : Whether image force corrections should enter the calculation of tunneling currents has been a debatable issue among physicists. The fundamental question is whether electrons inside the electrodes can respond quickly enough to the field created by a tunneling electron. We feel that the answer depends upon the nature of the electrodes. In chapter five we model the dynamic response of a two dimensional electron gas to external charge perturbations. We show that the dominant contribution to image force comes from the static dielectric constant mismatch between Si and SiO_2, and also from the metallic gate electrode, and not from the 2-DEG. The Mechanism of image force in MOS devices is explained in detail in this chapter.

1.3.2 Tunneling Processes, and Charge Statistics and Fluctuations in Quantum Dots Coupled to FETs

Understanding the operation of quantum dot memories and designing efficient and useful structures necessitates a good knowledge of tunneling processes, and charge statistics and fluctuations in these devices. In this context, we have explored the following :

Tunneling processes, carrier statistics and fluctuations : In order to determine the speed of these devices accurate calculation of rates of charging and dis-

charging of the quantum box are needed. In chapter six we develop appropriate analytical tools to study the dynamics of tunneling processes in quantum dots coupled to the channel of MOS devices. Carrier statistics and fluctuations in the quantum dot are also studied in chapter six.

Effects on conductivity of the channel : In chapter six we also show that charge fluctuation in single dot devices induce conductivity fluctuations of the inversion layer underneath. This offers a possibility of experimentally determining the time scale of charge fluctuations in the quantum dots by looking at the time scale of conductivity fluctuations of the channel. The physics associated with these fluctuations is developed in chapter six.

In chapters four and six we discuss the numerical results obtained from the various analytical models developed and compare these results with experimental data.

1.4 Conclusion

The new DTRAM memory structures developed at IBM offer possibilities for novel memories for the giga bit generation and at the same time provide an opportunity to study physics associated with electron tunneling processes in very thin oxides and also explore the behavior of electrons in nano structures. The research done in this thesis may also prove useful for the operation of sub 0.1 micron MOS devices which employ very thin gate oxides to control short channel effects. In addition, another object of this thesis would be to develop theoretical models that may be tested with experiments and then integrated into DAMOCLES, IBM's quantum monte carlo device simulator.

Chapter 2

Semi-classical Theory of Tunneling Through Oxides

2.1 Introduction

The semi-classical theory of tunneling through Silicon dioxide (SiO_2) films has a long history (see [2]). The study of tunneling in solids began with work in late twenties on field emission from metal surfaces [3] and in the thirties on metal semiconductor contacts [1]. First theoretical formulations of tunneling in thin insulating films were published in fifties [4, 5]. A number of experimental results were published in the sixties. The first widely accepted results on tunneling through SiO_2 films were published in late sixties by Lenzlinger and Snow [6], who showed Fowler-Nordheim tunneling through relatively thick ($\sim 1000\mathring{A}$) oxides used at that time in MOS devices. Since then a large number of papers have been published which present theoretical and experimental work on tunneling through both thin and thick oxides. We can cite only a few of them here [2, 7, 8, 9, 10]. Over the years the semi-classical theory of electron tunneling in thin oxides has evolved considerably. Many variants of the theory have been presented in literature. With the exception of few, these variants mostly differ from each other depending upon the nature of approximations made to get simple analytic expressions. In this chapter we will review the physics associated the semi-classical model of electron tunneling. We will present a semi-classical model in which

no analytical approximations will be made. The final results will be obtained through numerical means. In subsequent chapters we will compare these results with those obtained from a fully self-consistent quantum mechanical model. We will mostly be concerned with the case when the electrons tunnel from the conduction band of Silicon into the gate electrode, rather than vice versa, since this is more relevant to the memory devices being studied at IBM. However, the opposite case when the electrons tunnel from the gate electrode into the conduction band of Silicon can also be handled equally well by our theory. Also since in all the experimental work done at IBM, the devices had Poly-silicon gates, we will develop the theory for only such devices. Besides, devices with Poly-silicon gates are technologically more relevant.

2.2 The Physics of the Semi-classical Model

The semi-classical model of electron tunneling in MOS structures makes the following main assumptions :

1. The occupation statistics for electrons in Silicon are described by the Fermi-Dirac statistics.

2. The quantization of electron motion perpendicular to the Si-SiO_2 interface in accumulation and inversion layers is ignored.

3. The flux of electrons through the oxide is determined by a transmission probability, which may be calculated by solving the Schrodinger's equation in the effective mass approximation.

The two common and most important cases are when the electrons from the conduction band in Silicon tunnel from either the accumulation layer or the inversion layer into the gate electrode. Figures (2-1) and (2-2) show these processes in the energy band diagrams. It may happen that when the oxide is relatively thick ($> 75\AA$), and a high bias is applied across the MOS structure, tunneling electrons may enter the conduction band of the oxide. When this happens, tunneling is said to enter the 'Fowler-Nordheim' regime. In devices with thin oxides ($< 50\AA$), tunneling electrons

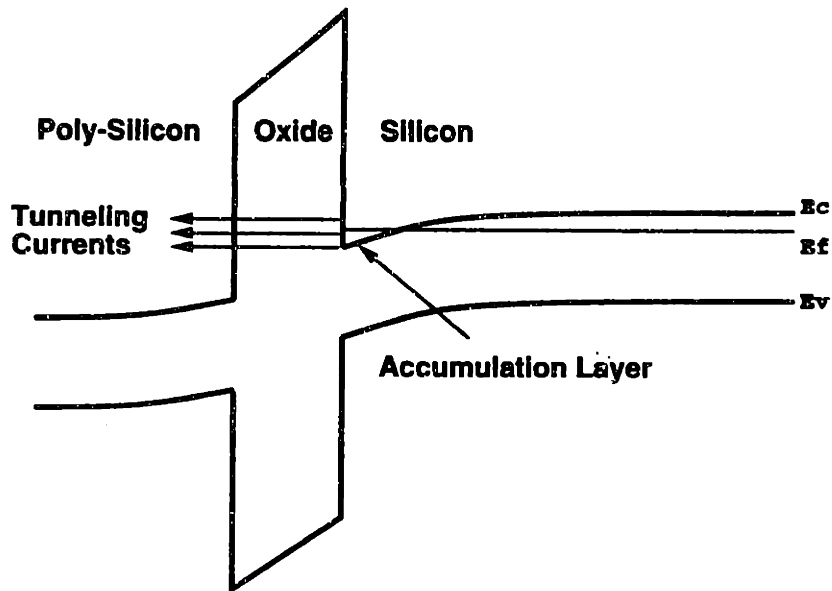

Figure 2-1: Electron Tunneling from Accumulation Layer

usually never enter the conduction band of the oxide even for high gate biases all the way up to the point where the oxide breaks down. Tunneling in this regime is called 'direct'. Both direct and Fowler-Nordheim tunneling can be treated on equal footing in the semi-classical model.

Semi-classical calculations for tunneling currents are done in three steps :

Step I : For a given value of gate voltage, the Poisson equation is solved for the entire structure shown in figure (2-3). This gives the potential drop in the Silicon substrate, in the oxide and in the gate Poly-Silicon.

Step II : For the oxide electric field determined in step I, Schrodinger equation is solved in the effective mass approximation in each region to obtain the transmission probability.

Step III : Transmission probability found in step II is used to calculate the tunneling currents.

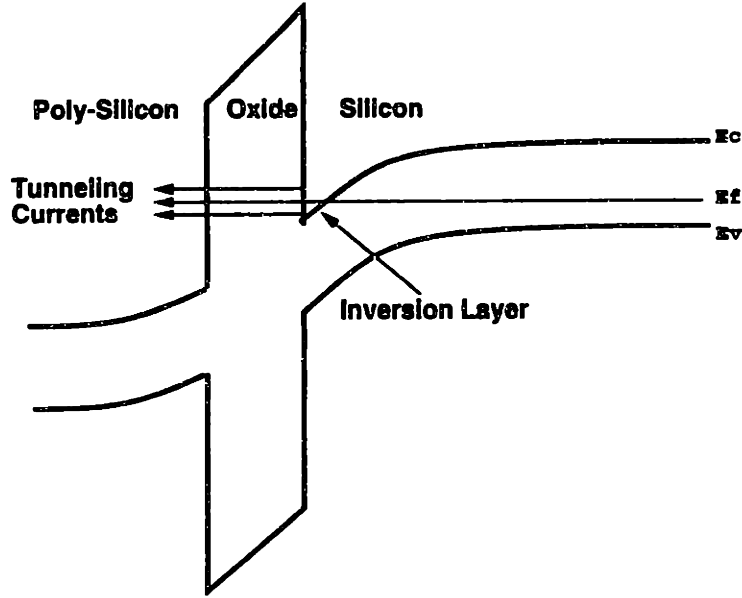

Figure 2-2: Electron Tunneling from Inversion Layer

2.2.1 Solution of the Poisson Equation

Usually analytical analysis of MOS devices is done using the Maxwell-Boltzman occupation statistics for electrons [11]. Even numerical simulators, like IBM's FIELDAY, avoids using Fermi-Dirac statistics since this makes the program take longer to converge. However, since we will be primarily interested in tunneling currents from heavily accumulated or inverted layers where Maxwell-Boltzman statistics do not apply, it is necessary that we use Fermi-Dirac statistics. Thus the Poisson equation needs to be solved using Fermi-Dirac statistics. This can accomplished numerically using the finite difference or the finite element method.

The calculation begins by guessing an approximate form for the potential $\phi_{guess}(x)$ in the region $-t_{ox} \leq x \leq L$, where L is the thickness of the substrate and t_{ox} is the thickness of the oxide. The form of the potential is usually taken to be

$$\phi_{guess}(x) = A\exp(-\frac{x}{\alpha l_D}) + B\exp(-\frac{x}{\beta l_D}) \quad \text{for} \quad 0 \leq x \leq L \tag{2.1}$$

$$\phi_{guess}(x) = \phi(0) - F_{ox}x \quad \text{for} \quad -t_{ox} \leq x \leq 0 \tag{2.2}$$

F_{ox} is the field strength in the oxide and is related to the parameters A, B, α, and β

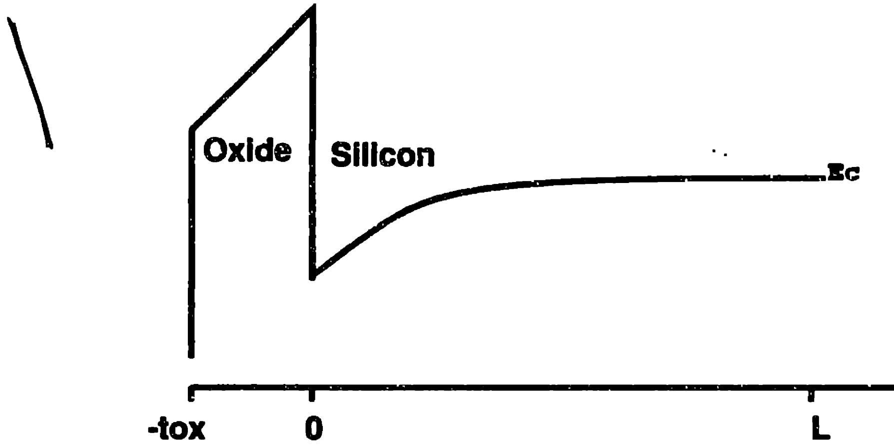

Figure 2-3: Region in which the Poisson equation is solved

by

$$F_{ox} = -\frac{\epsilon_{si}}{\epsilon_{ox}} \frac{\partial \phi_{guess}(x = 0^+)}{\partial x} = \frac{\epsilon_{si}}{\epsilon_{ox}} \frac{(A\alpha + B\beta)}{l_D} \tag{2.3}$$

The values of these parameters are chosen empirically to give the best results and fastest convergence. l_D is the debye length ($\sqrt{e^2 N_D / \epsilon_{si} kT}$) of bulk Silicon. ϵ_{si} and ϵ_{ox} are the static dielectric constants of Silcon and Silicon dioxide repectively. N_D is the doping in the Silicon substrate. Local charge density $\rho(x)$ can be written as $\rho(x) = e(N_D + p(x) - n(x) - N_A)$. Electron density $n(x)$ and hole density $p(x)$ can be written as [12]

$$n(x) = Nc(T)\ F_{1/2}(\frac{Ef - Ec(x)}{KT}) \tag{2.4}$$

$$p(x) = Nv(T)\ F_{1/2}(\frac{Ev(x) - Ef}{KT}) \tag{2.5}$$

where $Ec(x) = Eco - e\phi_{guess}(x)$ and $Ev(x) = Evo - e\phi_{guess}(x)$, and Eco and Evo are the conduction band and valence band edges deep in the substrate where $\rho(x) = 0$. $\rho(x)$ thus obtained can be used to solve the Poisson equation

$$\frac{\partial}{\partial x} \frac{\epsilon(x) \partial \phi(x)}{\partial x} = -\rho(x) \tag{2.6}$$

with boundary conditions

$$\phi(x=-t_{ox}) = V_{gate} - V_{flat-band}, \qquad \phi(x=L)=0, \text{ and}$$

$$\epsilon_{si}\frac{\partial\phi(x=0^+)}{\partial x} = \epsilon_{ox}\frac{\partial\phi(x=0^-)}{\partial x} \tag{2.7}$$

to obtain a new potential $\phi(x)$. The procedure can be iterated until $\phi_{guess}(x)$ and $\phi(x)$ converge. However, a much better approach is to assume that

$$\phi(x) = \phi_{guess} + \delta\phi(x) \tag{2.8}$$

Since $\rho(x)$ is also a functional of $\phi(x)$ we may write to first order in $\delta\phi(x)$

$$\rho[\phi(x)] = \rho[\phi_{guess}(x)] + \frac{\delta\rho(x)}{\delta\phi(x)}\delta\phi(x) \tag{2.9}$$

where $\frac{\delta\rho(x)}{\delta\phi(x)}$ is

$$\frac{\delta\rho(x)}{\delta\phi(x)} = -\frac{e^2}{KT}(Nc\ F'_{1/2}(\frac{Ef - Eco + e\phi_{guess}(x)}{KT}) - Nv\ F'_{1/2}(\frac{Evo - e\phi_{guess}(x) - Ef}{KT})) \tag{2.10}$$

The equation that needs to be solved for $\delta\phi(x)$ is then

$$\frac{\partial}{\partial x}\frac{\epsilon(x)\partial\delta\phi(x)}{\partial x} + (\frac{\delta\rho(x)}{\delta\phi(x)})\delta\phi(x) = 0 \tag{2.11}$$

Also $\delta\phi(x)$ satisfies the same boundary conditions as $\phi(x)$ except that $\delta\phi(x=-t_{ox}) = 0$. Equation (2.11) for $\delta\phi(x)$ can be solved through finite difference method which yields the following equation

$$\frac{\delta\phi(x_{i+1}) - \delta\phi(x_i)}{\epsilon^{-1}_{i+\frac{1}{2}}} - \frac{\delta\phi(x_i) - \delta\phi(x_{i-1})}{\epsilon^{-1}_{i-\frac{1}{2}}} + \Delta^2(\frac{\delta\rho(x_i)}{\delta\phi(x_i)})\delta\phi(x_i) = 0 \tag{2.12}$$

Δ is the length of an element of the mesh. Since we are solving in one dimension only, it is convenient to use a mesh of uniform size. The finite difference method

yields a tridiagonal matrix which can easily be solved by forward elimination and back substitution [13]. Once the equation is solved, a new guess for the potential $\phi(x)$ is made

$$\phi_{new-guess}(x) = \phi_{old-guess} + \delta\phi(x) \tag{2.13}$$

and the procedure is iterated until $\delta\phi(x)$ becomes less than 5×10^{-5} volts for all x. This procedure is usually found to converge in less than ten iterations.

2.2.2 Calculation of the Transmission Probability

The semi-classical model assumes that in the effective mass approximation the wavefunction of an electron in the substrate and the gate electrode can be described by a plane wave. The transmission probability is then defined as the ratio of electron flux transmitted through the SiO_2 barrier to that incident upon the barrier. Figure (2-4) shows the approximate potential profile used in the semi-classical picture to solve the Schrodinger equation

$$-\frac{\hbar^2}{2}\frac{\partial}{\partial x}\frac{1}{m_{eff}(x)}\frac{\partial\psi(x)}{\partial x} + V(x)\,\psi(x) = E\,\psi(x) \tag{2.14}$$

The potential $V(x)$ in each region shown in figure (2-4) is

$$V(x) = 0 \quad \text{for} \quad -\infty \leq x \leq 0$$

$$V(x) = e\phi_B - e|F_{ox}|x \quad \text{for} \quad 0 \leq x \leq t_{ox}$$

$$V(x) = e\phi_B - e|F_{ox}|t_{ox} \quad \text{for} \quad t_{ox} \leq x \leq \infty$$

ϕ_B is the conduction band discontinuity between Silicon and Silicon dioxide. The corresponding solutions of the Schrodinger equation in each region are

$$\psi(x) = e^{ik_i x} + r\, e^{-ik_i x} \quad \text{for} \quad -\infty \leq x \leq 0$$

$$\psi(x) = C\, Ai(y(x)) + D\, Bi(y(x)) \quad \text{for} \quad 0 \leq x \leq t_{ox}$$

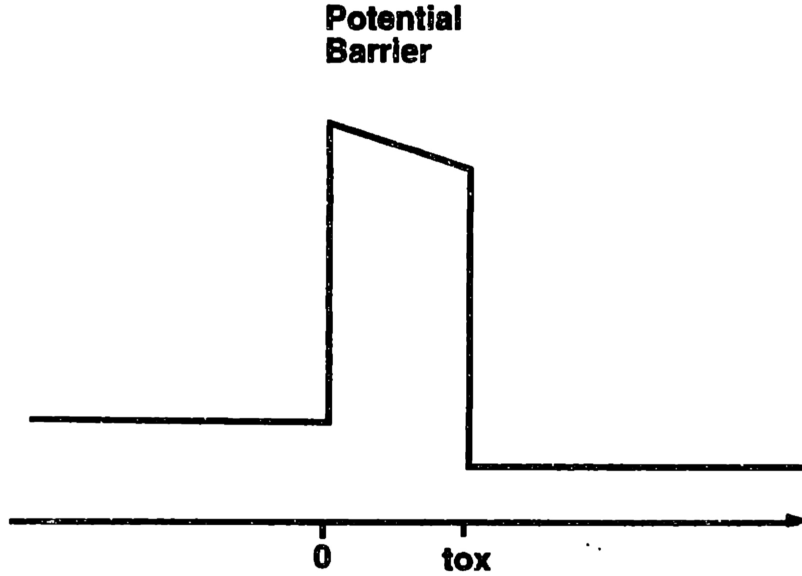

Figure 2-4: Region in which the Schrodinger equation is solved to get the transmission probabilities

$$\psi(x) = t\ e^{ik_f(x-t_{ox})} \quad \text{for} \quad t_{ox} \leq x \leq \infty$$

where

$$y(x) = \sqrt[3]{\frac{2m^{ox}e|F_{ox}|}{\hbar^2}}\left(\frac{e\phi_B - E}{e|F_{ox}|} - x\right)$$

In writing the solutions above, we have made an implicit assumption that the transverse kinetic energy of electrons is conserved while tunneling. Therefore the problem is reduced to the single dimension perpendicular to the Si/SiO_2 interface. With this assumption we can also write

$$k_i = \sqrt{2m_{xx}^{si}E}/\hbar, \qquad k_f = \sqrt{2m_{xx}^{si}(E + e|F_{ox}|t_{ox})}/\hbar$$

The boundary conditions on the electron wavefunction are

$$\psi(x = 0^-) = \psi(x = 0^+), \qquad \psi(x = t_{ox}^-) = \psi(x = t_{ox}^+)$$

and

$$\frac{1}{m_{xx}^{si}}\frac{\partial\psi(x = 0^-)}{\partial x} = \frac{1}{m^{ox}}\frac{\partial\psi(x = 0^+)}{\partial x}, \qquad \frac{1}{m^{ox}}\frac{\partial\psi(x = t_{ox}^-)}{\partial x} = \frac{1}{m_{xx}^{si}}\frac{\partial\psi(x = t_{ox}^+)}{\partial x}$$

These boundary conditions ensure the continuity of the wavefunction and the probability current across the Si/SiO_2 and the SiO_2/Poly-Si interfaces. Using the above boundary conditions we can solve for the transmission amplitude t. The transmission probability T is then defined as

$$T = \frac{k_f}{k_i}|t|^2$$

and is found to be

$$T = \frac{\frac{4}{\pi^2}\alpha^2 k_i\, k_f}{(Q-P)^2 + (R-S)^2} \tag{2.15}$$

where α, β, Q, P, R, and S stand for

$$\alpha = m^{ox}/m^{si}_{xx}$$

$$\beta = -\sqrt[3]{\frac{2m^{ox}e|F_{ox}|}{\hbar^2}}$$

$$Q = \beta Bi'(y(x=t_{ox}))Ai'(y(x=0)) + \frac{\alpha^2 k_i k_f}{\beta} Bi(y(x=t_{ox}))Ai(y(x=0))$$

$$P = \beta Bi'(y(x=0))Ai'(y(x=t_{ox})) + \frac{\alpha^2 k_i k_f}{\beta} Bi(y(x=0))Ai(y(x=t_{ox}))$$

$$R = \alpha k_i Bi'(y(x=t_{ox}))Ai(y(x=0)) - \alpha k_f Bi(y(x=t_{ox}))Ai'(y(x=0))$$

$$S = \alpha k_i Bi(y(x=0))Ai'(y(x=t_{ox})) - \alpha k_f Bi'(y(x=0))Ai'(y(x=t_{ox}))$$

Note that the transmission probability is only a function of the energy of the incident electrons perpendicular to the Si/SiO_2 interface. This is a result of our assumption that transverse component of the kinetic energy is conserved during tunneling. We will say more about this assumption later.

2.2.3 Calculation of the Tunneling Currents

Finally the tunneling current can be written as

$$J = 2e \int \frac{\hbar k_x}{m^{si}_{xx}} T(E_x)\ (\ f_D(E_{||} + E_x - Ef_L) - f_D(E_{||} + E_x - Ef_L)\ )\ \frac{d^3\vec{k}}{(2\pi)^3} \tag{2.16}$$

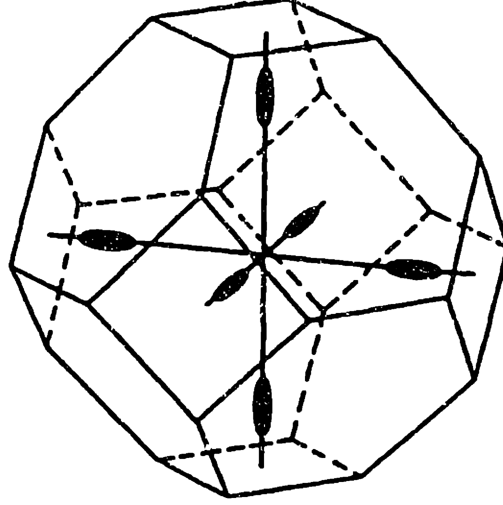

Figure 2-5: Carrier Pockets in the First Brillouin Zone of Silicon

Integration over k_x variable is restricted to the positive half of k_x space. After integrating over the transverse channels this reduces to

$$J = \frac{e\sqrt{m_{yy}^{si} m_{zz}^{si}} KT}{2\pi^2\hbar^3} \int_0^\infty T(E_x) \log \frac{1+\exp(\frac{Ef_L - E_x}{KT})}{1+\exp(\frac{Ef_R - E_x}{KT})} dE_x \tag{2.17}$$

In Silicon there are six carrier pockets in the first Brillouin zone, and assuming a [100] Surface orientation, we may write the total tunneling current as

$$\begin{aligned} J_{total} &= \frac{2e\sqrt{m_t^{si} m_l^{si}} KT}{\pi^2\hbar^3} \int_0^\infty T^t(E_x) \log \frac{1+\exp(\frac{Ef_L - E_x}{KT})}{1+\exp(\frac{Ef_R - E_x}{KT})} dE_x \quad + \\ & \frac{e m_t^{si} KT}{\pi^2\hbar^3} \int_0^\infty T^l(E_x) \log \frac{1+\exp(\frac{Ef_L - E_x}{KT})}{1+\exp(\frac{Ef_R - E_x}{KT})} dE_x \end{aligned} \tag{2.18}$$

In above expression the superscript t and l on $T(E_x)$ mean that the effective mass used in calculating the transmission probability is m_t^{si} and m_l^{si} respectively. Ef_L and Ef_R are the fermi levels on the left (substrate) and right (gate) sides respectively. If a positive bias V_{gate} is applied to the gate electrode then the splitting of the fermi levels is $Ef_L - Ef_R = eV_{gate}$. Note that fermi levels in the substrate and the gate are measured with respect to the conduction band edge right at the interface with SiO_2.

The integrals in the equation (2.18) can be done numerically to get the magnitude of tunneling current as a function of the gate bias.

Results of the calculations described in this chapter will not be presented here but in later chapters when we compare these results with those obtained from the fully self-consistent quantum mechanical model. We will now discuss the shortcomings of the semi-classical model, as presented in this chapter, and justify the need for a more accurate model.

2.3 Discussion

Although the semi-classical model that we have presented above does not suffer from analytical approximations, it makes a number of physical assumptions that are hard to justify. We will discuss the nature of these assumptions in this section.

2.3.1 Conservation of Transverse Kinetic Momentum

In deriving the expression for the transmission probability we made an assumption that the transverse kinetic energy of electrons is conserved while tunneling. In Silicon the six carrier pockets are ellipsoids located along the six equivalent $< 100 >$ axis (Γ to X direction). The centers of these ellipsoids are at a distance of roughly three-fourths of the Γ-X distance. However, in Silicon dioxide the conduction band minimum occurs at the zone center (Γ point). The question arises whether the transverse crystal momentum (i.e. momentum measured w.r.t. the zone center) is conserved, or the transverse kinetic momentum (momentum measured w.r.t. the minimum of conduction band edge) is conserved, or the transverse kinetic energy is conserved in tunneling. This question has been a puzzle for a long time (see [8, 9]) and till now no satisfactory answer is available. However, experimental measurements seem to suggest that crystal momentum is not conserved while tunneling. The argument is as follows. Suppose that transverse crystal momentum is conserved. Consider the six carrier pockets of Silicon shown in figure (2-5), and suppose that the direction perpendicular to the Si/SiO_2 interface is the positive x direction. Electrons form the

carrier pocket [010] can tunnel provided the total energy inside the oxide is the same as the initial total energy. Thus we may write

$$Ec_{si} + \frac{(\hbar k_x)^2}{2m_t^{si}} + \frac{(\hbar k_y)^2}{2m_t^{si}} + \frac{(\hbar k_z)^2}{2m_l^{si}} = Ec_{ox} - \frac{(\hbar \kappa)^2}{2m^{ox}} + \frac{(\hbar(K_y + k_y))^2}{2m^{ox}} + \frac{(\hbar k_z)^2}{2m^{ox}}$$

where K_y ($\simeq 0.75\frac{\pi}{a}$) is the distance of the pocket at [010] from the zone center, and κ is the decay constant of the wavefunction inside the oxide. Assuming for a moment that the effective masses in Si and SiO_2 are roughly equal, we may write for κ

$$\begin{aligned} \frac{(\hbar \kappa)^2}{2m^{ox}} &\simeq (Ec_{ox} - Ec_{si}) + \frac{(\hbar K_y)^2}{2m^{ox}} - E_x^{si} \\ &\simeq e\phi_B + \frac{(\hbar K_y)^2}{2m^{ox}} - E_x^{si} \end{aligned} \tag{2.19}$$

This means that the effective barrier height for electrons has been increased by about $\frac{(\hbar K_y)^2}{2m^{ox}}$, which is around 3eV. Such a large increase should greatly decrease the transmission probability for all electrons in the four carrier pockets located in directions parallel to the interface, and should result in very little tunneling current from these carrier pockets. Such large reductions in tunneling currents are not observed in experiments. Tunneling measurements done for Silicon surfaces with $< 111 >$ and $< 110 >$ orientations also do not support the conservation of transverse crystal momentum [8, 9]. The excellent agreement between our calculations and experimental data also does not support this idea.

However, there also does not seem to be any evidence to support whether transverse kinetic energy or transverse kinetic momentum is conserved during tunneling. For simplicity, we have assumed throughout this thesis that transverse kinetic energy is conserved. This assumption reduces the three dimensional problem to a single dimensional one. Since on average the difference between the effective mass of electrons in Si is almost that in SiO_2, we expect that the assumption of conservation of transverse kinetic momentum should not give results that are much different than ours.

An exact theory explaining the behavior of electron wavefunction near the Si/SiO_2

is difficult, if not completely impossible, given the amorphous nature of the oxide. The local order in amorphous thermal oxides resembles that of the crystalline α-quartz [20]. Some physics learned from the structure of α-quartz may be used to understand the nature of amorphous oxides, but in the end one must rely on experimental results.

2.3.2 The Oxide Effective Mass and the Question of the Image Force

All calculations done in this thesis assume an oxide effective mass of $0.5m_o$ and a value of 3.15eV for ϕ_B (the discontinuity between the Si and SiO_2 conduction bands). We have to justify that the energy dispersion relation in the mid-gap region of SiO_2 can be modeled by assuming an effective mass of $0.5m_o$. Also we have ignored the barrier height reduction effects caused by the image force. We also need to justify this assumption. We will discuss both these issues more carefully in later chapters.

2.3.3 Quantization of Electron Motion in Accumulation and Inversion Layers

The semi-classical model ignores the quantization of electron motion perpendicular to the Si/SiO_2 interface. The tunneling current is extremely sensitive to the distribution of incident electrons in energy and the electric field strength in the oxide. From the self-consistent studies carried out in references [14, 15, 16, 17], it is obvious that the distribution of carriers in energy as predicted by the semi-classical model is very different from that obtained from self-consistent calculations, and the distribution of applied bias between the substrate and the oxide also turns out to be very different. Therefore it is necessary that experimental data be compared with results obtained from self-consistent calculations rather than those deduced from the semi-classical model. Also results from self-consistent calculations published recently show oxide tunneling currents to be roughly two orders of magnitude larger than predicted by the semi-classical model [18, 19]. This provides additional motivation to develop a more accurate quantum mechanical model for tunneling currents. In the next chapter

we will present the details of our self-consistent quantum mechanical model.

2.4 Conclusion

In this chapter we have described in detail the semi-classical model for calculating tunneling currents. We have also discussed the short comings of the model and the need to develop a more sophisticated model. In addition, we have also discussed the assumptions that have been made in deriving the semi-classical model and the need to justify these assumptions. This justification will be presented in later chapters.

Chapter 3

Self-Consistent Quantum Mechanical Model for Tunneling through Oxides

3.1 Foreword

In the previous chapter we described a semi-classical model for calculating tunneling currents through oxides. We also mentioned the shortcomings associated with the semi-classical formulation and justified the need to develop a more rigorous self-consistent model. The purpose of this chapter is to describe the details of a self-consistent model which is suitable for calculating tunneling currents from both accumulation and inversion layers. We will also introduce the concept of electron 'lifetime' to calculate tunneling currents from quasi-bound eigenstates, and describe various means to calculate these lifetimes.

The calculation of tunneling currents in the self-consistent model proceeds in three steps :

Step I : For a given value of gate voltage, Poisson and Schrodinger equations are solved self-consistently in the entire region of interest to obtain the energies of all the quasi-bound states and the potential drops in the substrate, oxide and

the gate.

Step II : Using the values of oxide electric field and eigenenergies found in step I, lifetimes of electrons in each subband are calculated.

Step III : Lifetimes calculated in step II are used to calculate contributions to oxide tunneling current from each subband.

Details of each step will now be described.

3.2 Self-Consistent Solution of Poisson and Schrodinger Equations

Self-consistent modeling of inversion layers has received considerable attention in recent years (see [14, 15, 16]) as a consequence of its importance in transport in MOS devices. However, self-consistent modeling of accumulation layers has received much less attention. We are aware of only two publications [18, 19] that have presented self-consistent results for accumulation layers. However, the method described in these references used the rather tedious 'shooting' methods (see for example [13]) to calculate eigenenergies of various subbands. In this section we will describe a method that works equally well for both inversion and accumulation layers.

The Poisson and Schrodinger equations are solved self-consistently for the structure shown in figure (3-1) for both accumulation and inversion layers. Self-consistent calculations of accumulation layers are more complex than those of inversion layers due to the absence of a separation region between the bulk extended states and the quasi-bound states near the interface. In case of inversion layers this separation is provided by the depletion region. The solution to this problem is to treat both extended and bound states equally in the case of accumulation layers. The calculation starts by using a guess for electrostatic potential in the region $-t_{ox} \leq x \leq L$. In case of accumulation layers, a suitable form for this potential is

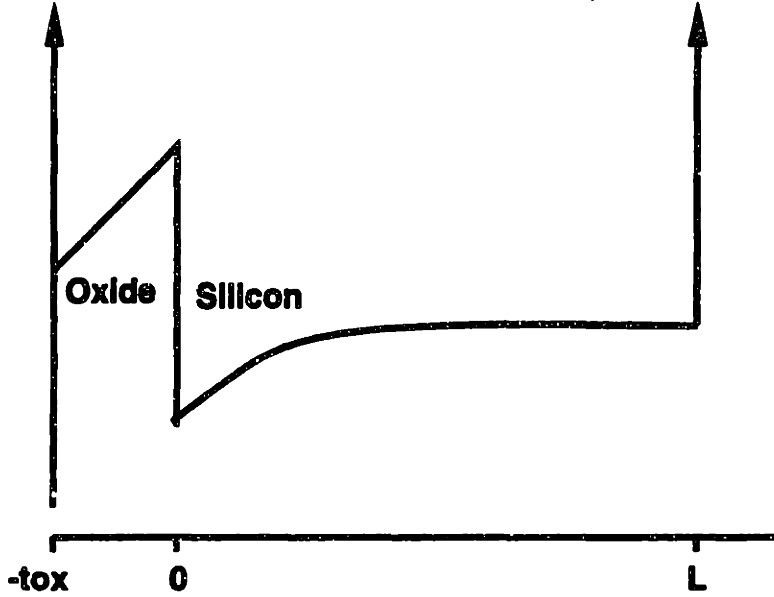

Figure 3-1: Region in which self-consistent solution is sought for accumulation layers

$$\phi_{guess}(x) = A \exp\left(-\frac{x}{\alpha l_D}\right) + B \exp\left(-\frac{x}{\beta l_D}\right) \quad \text{for} \quad 0 \leq x \leq L \tag{3.1}$$

$$\phi_{guess}(x) = \phi(0) - F_{ox} x \quad \text{for} \quad -t_{ox} \leq x \leq 0 \tag{3.2}$$

where F_{ox} is related to the empirical parameters A, B, α, and β as described in equation (2.3). For inversion layers near threshold a suitable form for the potential is

$$\phi_{guess}(x) = 0 \quad \text{for} \quad x_{depl} \leq x \leq L \tag{3.3}$$

$$\phi_{guess}(x) = \frac{eN_A(x - x_{depl})^2}{2\epsilon_{si}} \quad \text{for} \quad 0 \leq x \leq x_{depl} \tag{3.4}$$

$$\phi_{guess}(x) = \phi(0) - F_{ox} x \quad \text{for} \quad -t_{ox} \leq x \leq 0 \tag{3.5}$$

x_{depl} is the length of the depletion region. Schrodinger equation in the effective mass approximation is solved in the region $-t_{ox} \leq x \leq L$ using finite element method. The discretized version of the equation for the n'th eigenstate is

$$-\frac{\hbar^2}{2}\frac{\psi_n(x_{i+1}) - \psi_n(x_i)}{m_{i+\frac{1}{2}}} - \frac{\psi_n(x_i) - \psi_n(x_{i-1})}{m_{i-\frac{1}{2}}} - e\Delta^2 \phi_{guess}(x_i)\psi_n(x_i) = \Delta^2\, E_n\, \psi_n(x_i) \tag{3.6}$$

The boundary conditions on the electron wavefunction are

$$\psi(x=-t_{ox}) = \psi_n(x=L) = 0, \qquad \psi_n(x=-0^-) = \psi_n(x=0^+) \quad \text{and}$$

$$\frac{1}{m^{si}}\frac{\partial\psi_n(x=0^+)}{\partial x} = \frac{1}{m^{ox}}\frac{\partial\psi_n(x=0^-)}{\partial x} \tag{3.7}$$

Equation (3.6) results in a tridiagonal matrix. The lowest few eigenvalues of a symmetric tridiagonal matrix can easily be found to very high accuracy using sturm sequencing and bisection. Search for eigenvalues is done only uptill energies which are not so high to have a negligible occupancy. The upper limit of the search interval is usually fixed at about 0.8eV above the fermi level. The eigenvector corresponding to each eigenvalue can be found by performing inverse iteration. The numerical methods mentioned here are commonly used for the solution of matrix eigensystems and will, therefore, not be described here in detail (see for example [13]).

Once the eigenenergies and the wavefunctions have been obtained the charge density (assuming a $< 100 >$ Si surface orientation) can be calculated from the expression

$$\begin{aligned}
\rho(x) = \quad & e(N_D - N_A) + Nv(T)\, F_{1/2}\left(\frac{Evo - e\phi_{guess}(x) - Ef_L}{KT}\right) \\
- \quad & \frac{4e\sqrt{m_t^{si} m_l^{si}}KT}{\pi\hbar^2}\sum_n \log\left(1+\exp\left(\frac{Ef_L - E_n^t}{KT}\right)\right) \mid \psi_n^t(x) \mid^2 \\
- \quad & \frac{2em_t^{si}KT}{\pi\hbar^2}\sum_n \log\left(1+\exp\left(\frac{Ef_L - E_n^l}{KT}\right)\right) \mid \psi_n^l(x) \mid^2
\end{aligned} \tag{3.8}$$

The second term on the left hand side is the density of hole charge in the semi-classical approximation. The last two terms are the electron charge densities. The superscript $t(l)$ on electron wavefunctions and eigenenergies indicate that the electrons belong to the carrier pocket in which the electron mass perpendicular to the interface is $m_t^{si}(m_l^{si})$. The above expression for the charge density is valid for accumulation as well as inversion layers.

Charge density obtained from equation (3.8) is used to solve the Poisson equation using the finite element method with triangular basis functions [13] which results in

the following equation

$$\frac{\phi(x_{i+1}) - \phi(x_i)}{\epsilon^{-1}_{i+\frac{1}{2}}} - \frac{\phi(x_i) - \phi(x_{i-1})}{\epsilon^{-1}_{i-\frac{1}{2}}} = \Delta^2 \rho(x_i) \tag{3.9}$$

The boundary conditions on the potential are

$$\phi(x = -t_{ox}) = V_{gate} - V_{flat-band}, \quad \phi(x = L) = 0, \text{ and}$$

$$\epsilon_{si}\frac{\partial\phi(x = 0^+)}{\partial x} = \epsilon_{ox}\frac{\partial\phi(x = 0^-)}{\partial x} \tag{3.10}$$

The potential obtained from Poisson equation ($\phi(x)$) is compared with the potential used in the Schrodinger equation ($\phi_{guess}(x)$). If $|\phi(x) - \phi_{guess}(x)| < 5 \times 10^{-5}$eV for all x then $\phi(x)$ is the self-consistent solution. Otherwise a new guess for the potential is made as follows

$$\phi_{new-guess}(x) = \phi_{guess}(x) + r(x)(\phi(x) - \phi_{guess}(x)) \tag{3.11}$$

The function r(x) is such that $0 < |r(x)| < 1$ for all x, and its exact form is chosen empirically to speed up convergence.

In the self-consistent calculations for accumulation layers the boundary condition, $\psi_n(x = L) = 0$, on the electron wavefunctions produces a net positive charge density near $x \approx L$. This artificial pile up of charge is ignored and this does not have any effect on the results as long as the length, L, of the entire substrate region is chosen long enough so that there exists a neutral region between the charge density in the accumulation layer and the artificial pile up of charge near $x \approx L$.

Finally, the potential drop in the n^+ Poly-silicon gate is calculated using the semi-classical formula with the oxide electric field as the boundary condition

$$V_{poly} = \frac{(\epsilon_{ox}F_{ox})^2}{2\epsilon_{si}eN_D^+} \tag{3.12}$$

where N_D^+ is the doping density in the n^+ Poly-silicon gate. The correct value of the

gate bias V_{gate} becomes (compare with equation (3.10))

$$V_{gate} - V_{flat-band} = V_{poly} + \phi(x = -t_{ox}) \tag{3.13}$$

Note that $\phi(x = -t_{ox})$ is just the potential drop in the oxide and in the substrate. If doping in the gate is more than $5 \times 10^{20}/cm^3$ then V_{poly} is usually negligible.

The method for obtaining self-consistent solutions described in this section works well for both accumulation and inversion layers and is fairly robust. We have tested our self-consistent solver for both accumulation and inversion layers at 300°K and at 77°K. The number of iterations required to obtain convergence depends upon the gate bias, and is usually between twenty to hundred. In the next section we will describe various methods to calculate lifetimes for the quasi-bound states obtained from the self-consistent solution.

3.3 Electron Tunneling from Bound States

The problem of calculating tunneling currents from quasi-bound states in MOS devices has been to the best our knowledge addressed in two publications [18, 8]. We believe that methods described in these references are rather awkward and even hard to justify. For quasi-bound states the concept of transmission probability is no longer meaningful. Transmission probability is defined as the ratio of transmitted to incident flux. In case of tunneling from quasi-bound states, there is no incident flux. However, the lifetime of quasi-bound states can be used to calculate tunneling currents. The concept of lifetime is useful for a decaying state if the following conditions are satisfied :

a) It has a lifetime much longer compared to $\epsilon/\hbar$, where ϵ is the energy of the state.

b) The state is coupled very weakly to other states (and hence its long lifetime)

c) The state is decaying into a continuum, or a very large number of final states so that there is no chance of coherent recurrence.

All the above conditions are satisfied for quasi-bound states in the accumulation and inversion layers. The high (~3.15eV) potential barrier provided by the SiO_2 ensures that the quasi-bound states have a long lifetime, and the very large number of final states in the gate electrode into which these quasi-bound states decay, implies that there is almost zero probability for coherent regeneration of the initial state.

In this section we will describe two different methods to calculate lifetimes of quasi-bound states, and then using non-equilibrium Green function technique we will show how tunneling currents may be calculated from quasi-bound states.

3.3.1 Calculation of Lifetimes from Path Integral Formulation

Consider the quasi-bound state shown in figure (3-2). Suppose an electron is placed in this state at time $t = 0$. The wavefunction of the electron at time $t = 0$ is $\psi(x, t = 0)$. If $\psi(x, t)$ were an exact eigenstate of the system, then we would expect that at a later time t the wavefunction would be

$$\psi(x,t) = e^{-\frac{i}{\hbar}E\,t}\ \psi(x,t=0)$$

But as result of the finite thickness of the SiO_2 barrier the state gets coupled to the states in the gate electrode, and it is therefore no longer an exact energy eigenstate. If this coupling is weak, the time development of the state is expected to be of the form

$$\psi(x,t) = e^{-\frac{i}{\hbar}E\,t}\ e^{-\frac{t}{2\tau}}\psi(x,t=0)$$

and therefore the probability $P(t)$ that the electron will be found in the quasi-bound state at a later time decays exponentially as

$$P(t) = \int dx\ |\psi(x,t)|^2 = e^{-\frac{t}{\tau}}$$

Since the coupling with the states in the gate is weak, we can calculate the value of

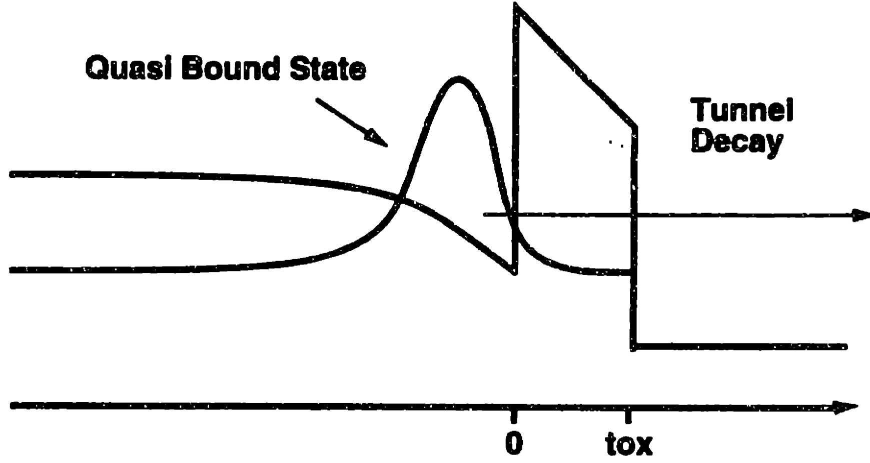

Figure 3-2: quasi-bound state decaying into the gate

the lifetime τ perturbatively.

We define a resolvent $G(\omega)$ as [21]

$$G(\omega) = tr\left(\frac{1}{\omega - \frac{\hat{H}}{\hbar} + i\eta}\right) \tag{3.14}$$

Thus the eigenenergies of the hamiltonian $\hat{H}$ are the poles of $G(\omega)$. We may also write $G(\omega)$ as

$$G\omega) = -i\int_0^\infty dT\, e^{i\omega T}\int dx\ < x|e^{-\frac{i}{\hbar}\hat{H}T}|x > \tag{3.15}$$

$$G(\omega) = -i\int_0^\infty dT\, e^{i\omega T}\int dx_o \int_{x(0)=x_o}^{x(T)=x_o} D[x(t)]e^{\frac{i}{\hbar}S[x(t)]} \tag{3.16}$$

where

$$S[x(t)] = \int_0^T dt\ \left(\frac{m}{2}\dot{x}(t)^2 - V(x(t))\right) \tag{3.17}$$

is the classical action and the path integral in equation (3.16) is over all trajectories satisfying $x(T) = x(0) = x_o$. We can expand the action around the classical path

$x_{cl}(t)$ which satisfies the Euler-Lagrange equation

$$m\ddot{x}_{cl}(t) = -V'(x_{cl}(t)) \tag{3.18}$$

with boundary conditions

$$x_{cl}(T) = x_{cl}(0) = x_o \tag{3.19}$$

and get

$$\int_{x(0)=x_o}^{x(T)=x_o} D[x(t)] e^{\frac{i}{\hbar}S[x(t)]} = \sqrt{\frac{i}{2\pi\hbar}\frac{\partial^2 S_{cl}}{\partial x_o \partial x_o}}\, e^{\frac{i}{\hbar}S_{cl}[x_{cl}(t)]} = A\, e^{\frac{i}{\hbar}S_{cl}[x_{cl}(t)]} \tag{3.20}$$

The above expression is exact for potentials linear and quadratic in x. Note that action $S_{cl}[x(t)]$ is only a function of x_o and T. We may define momentum p and a constant of classical motion ϵ associated with the classical path from the relations

$$p[x_{cl}(t)] = m\dot{x}_{cl}(t) \tag{3.21}$$

$$\epsilon = \frac{m}{2}\dot{x}_{cl}(t)^2 + V(x_{cl}(t)) \tag{3.22}$$

The time period T of the classical trajectory is then

$$T = \oint_{x_o}^{x_o} dx\, \sqrt{\frac{m}{2(\epsilon - V(x))}} = \oint dx\, \frac{m}{p(x)} \tag{3.23}$$

The classical action can also be written as [22]

$$S_{cl}(x_o, T) = \oint_{x_o}^{x_o} p(x)\, dx \ -\ \epsilon(x_o, T)T \tag{3.24}$$

The expression for the resolvent becomes

$$G(\omega) = -i\int_0^\infty dT\, e^{i\omega T} \int dx_o\, A\, e^{\frac{i}{\hbar}S_{cl}[x_{cl}(t)]} \tag{3.25}$$

The integral over x_o can be performed in the stationary phase approximation which

means that from all possible $S_{cl}(x_o, T)$, corresponding to different classical trajectories, keep only those for which $\frac{\delta S_{cl}(x_o,T)}{\delta x_o} = 0$. This condition gives

$$\frac{\delta S_{cl}(x_o,T)}{\delta x_o} = \frac{\delta S_{cl}(x_o,T)}{\delta x_{cl}(T)} + \frac{\delta S_{cl}(x_o,T)}{\delta x_{cl}(0)} = p[x_{cl}(T)] - p[x_{cl}(0)] = 0 \tag{3.26}$$

Thus both the initial and final co-ordinates and momenta are supposed to be equal. In other words only those classical paths which have periodic trajectories are allowed. We may write

$$G(\omega) = -i \int_0^\infty dT \, A \, e^{i\omega T + \frac{i}{\hbar} S_{cl}(T)} \tag{3.27}$$

We can again use stationary phase approximation and keep only those terms for which

$$\omega = -\frac{1}{\hbar}\frac{\delta S_{cl}(T)}{\delta T} = \frac{\epsilon}{\hbar} \tag{3.28}$$

In general there will be many classical periodic paths with different time periods T_n such that the energy associated with these paths equals $\hbar\omega$. Therefore the final expression for $G(\omega)$ will just be the sum over all these paths

$$G(\omega) = -i \sum_n A_n \, e^{i\omega T_n + S^n_{cl}(T)} = -i \sum_n A_n \, e^{i \oint p(x)dx} \tag{3.29}$$

When the contour integral $\oint p(x)dx$ in the above expression passes through a classically forbidden region where $\epsilon < V(x)$ then p(x) becomes imaginary and therefore the integral $\oint p(x)\, dx$ may have imaginary parts. In performing summation over all the paths in (3.29) the path dependence of the prefactor can be neglected if one observes the following rule [22] : Whenever a path bounces back from a classical turning point into a classically forbidden region A gets a factor $\frac{i}{2}$, and whenever a path bounces back from a classical turning point into a classically allowed region A gets a factor $-i$.

For the problem shown in figure (3-2) we may define the W_{si} and W_{ox} as

$$W_{si} = 2\int_{x_{turning}}^{0} dx \frac{\sqrt{2m^{si}(\hbar\omega - V(x))}}{\hbar} \tag{3.30}$$

$$W_{ox} = 2\int_{0}^{t_{ox}} dx \frac{\sqrt{2m^{ox}(V(x) - \hbar\omega)}}{\hbar} \tag{3.31}$$

Now we can write the summation in equation (3.29) as sum over all trajectories beginning in either Si or SiO_2 and containing all combinations of cycles in each region

$$\begin{aligned} G(\omega) \;\propto\; & \sum_{n=1}^{n=\infty}\left(-ie^{iW_{si}}\left(-i+\frac{i}{2}e^{-W_{ox}}+(\frac{i}{2})^3e^{-W_{ox}}+....\right)\right)^n + \\ & \sum_{n=1}^{n=\infty}\left(\frac{i}{2}e^{-W_{ox}}\left(\frac{i}{2}+-ie^{iW_{si}}+(-i)^3e^{i3W_{si}}+....\right)\right)^n \end{aligned} \tag{3.32}$$

which gives

$$G(\omega) \propto \frac{(1-0.25e^{-W_{ox}})e^{iW_{si}} + 0.25(1-e^{iW_{si}})e^{-W_{ox}}}{(1+0.25e^{-W_{ox}}) + (1-0.25e^{-W_{ox}})e^{iW_{si}}} \tag{3.33}$$

The real part of the pole of $G(\omega)$ can be found by neglecting terms containing $e^{-W_{si}}$ in the denominator giving

$$1 + e^{iW_{si}} = 0 \tag{3.34}$$

which implies

$$W_{si} = 2\oint_{x_{turning}}^{0} p(x)dx = 2\pi(n+\frac{1}{2}) \qquad n = 0,1,2... \tag{3.35}$$

Equation (3.35) is the just the Bohr-Sommerfield quantization condition, implying that poles in ω plane occur at discrete frequencies ω_n. These discrete frequencies correspond to the discrete spectrum of the quasi-bound states. To take into account the correction terms due to $e^{-W_{ox}}$ we can expand W_{si} as

$$W_{si}(\omega) = W_{si}(\omega_n) + \frac{\partial W_{si}(\omega_n)}{\partial\omega}\Delta\omega_n \tag{3.36}$$

Using equations (3.36), (3.23), (3.30), and (3.33) the value of $\Delta\omega$ is found to be

$$\Delta\omega_n = -\frac{i}{\hbar}\frac{e^{-W_{ox}(\omega_n)}}{2T(\omega_n)} \tag{3.37}$$

where $T(\omega_n)$ is the time period for classical motion in the n'th quasi-bound state. Thus it is now obvious that the energy of the quasi-bound state has acquired an imaginary part and the lifetime of the n'th state is therefore

$$\frac{1}{\tau(\epsilon_n)} = \frac{e^{-W_{ox}(\frac{\epsilon_n}{\hbar})}}{T(\frac{\epsilon_n}{\hbar})} = \frac{e^{-2\int_0^{t_{ox}} dx \frac{\sqrt{2m^{ox}(V(x)-\epsilon_n)}}{\hbar}}}{\oint_{x_o}^{x_o} dx \sqrt{\frac{m}{2(\epsilon_n - V(x))}}} \tag{3.38}$$

We can write this in more suggestive form as

$$\frac{1}{\tau_n} = f_n \times |t(\epsilon_n)|^2 \tag{3.39}$$

Where f is the classical frequency of oscillation of a confined particle and $|t(\epsilon_n)|^2$ is the transmission probability in the WKB approximation. Equation (3.39) is the main result of this section.

3.3.2 Calculation of Current from Quasi-Bound States with Non-equilibrium Green Function Technique

In the last section we derived a formula for lifetime of quasi-bound states using the path integral expansion of the resolvent operator. In this section we will use another technique to get the same result and also calculate the total tunneling current when the fermi level in the substrate is shifted higher than that in the gate by the application of a positive bias on the gate. We again consider the situation depicted in figure (3-2). Using the tunneling hamiltonian formalism [24, 25] we can write the hamiltonian for the system as

$$H = H_L + H_R + H_{L,R} \tag{3.40}$$

where

$$
\begin{aligned}
H_L &= \sum_n (\epsilon_n + eV)\ a_n^\dagger a_n \\
H_R &= \sum_m \epsilon_m\ c_m^\dagger c_m \\
H_{L,R} &= \sum_{n,m} T_{mn}\ c_m^\dagger a_n + \text{c.c.}
\end{aligned}
\tag{3.41}
$$

The indices n and m label the states on the left and right side of the barrier repectively. As shown in figure (3-2) the states on the left side of the barrier are the quasi-bound states.

The total current going from the substrate (left side) into the gate (right side) is

$$
\begin{aligned}
J_{L\to R} &= -e\frac{dN_L(t)}{dt} = -e\frac{i}{\hbar}\sum_{m,n}\left(T_{mn} < c_m^\dagger(t)a_n(t) > -T_{nm} < a_n^\dagger(t)c_m(t) >\right) \\
&= -\frac{e}{\hbar}\sum_{m,n}\left(T_{mn}G_{LR}^{-+}(n,t;m,t) - T_{nm}G_{RL}^{-+}(m,t;n,t)\right)
\end{aligned}
\tag{3.42}
$$

where $G_{LR}^{-+}(n,t;m,t) = i < c_m^\dagger(t)a_n(t) >$ and $G_{RL}^{-+}(m,t;n,t) = i < a_n^\dagger(t)c_m(t) >$. The order of the subscripts 'L' and 'R' on the Green function indicate the regions to which the states whose indices appear inside the brackets belong. In order to calculate the current we need analytical expressions for the various Green functions appearing in equation (3.42). These Green functions can be calculated using non-equilibrium perturbation technique developed by Keldysh [26].

By doing perturbation expansion in the hamiltonians $H_{L,R}$ and $H_{R,L}$ on the Keldysh contour [23] in complex time it can be shown that [27]

$$
\begin{aligned}
G_{LR}^{-+}(n,t;m,t) = \sum_{n'm'} \frac{T_{n'm'}}{\hbar} \int_{-\infty}^{\infty} dt_1 \Big(& g_{LL}^{-+}(n,t;n',t_1)\ g_{RR}^{a}(m',t_1;m,t) + \\
& g_{LL}^{r}(n,t;n',t_1)\ g_{RR}^{-+}(m',t_1;m,t)\Big)
\end{aligned}
\tag{3.43}
$$

$$
\begin{aligned}
G_{RL}^{-+}(k,t;n,t) = \sum_{n'm'} \frac{T_{m'n'}}{\hbar} \int_{-\infty}^{\infty} dt_1 \Big(& g_{RR}^{r}(m,t;m',t_1)\ g_{LL}^{-+}(n',t_1;n,t) + \\
& g_{RR}^{-+}(m,t;m',t_1)\ g_{LL}^{a}(n',t_1;n,t)\Big)
\end{aligned}
\tag{3.44}
$$

g_{LL} and g_{RR} represent Green functions of the left and right side in the uncoupled system, respectively. These are as follows [28]

$$g_{LL}^{-+}(n', t_1; n, t_2) = i f_D(\epsilon_n - Ef_L)\ e^{-\frac{i}{\hbar}\epsilon_n(t_1 - t_2)} \delta_{n',n} \tag{3.45}$$

$$g_{LL}^{r}(n', t_1; n, t_2) = -i\, \theta(t_1 - t_2)\, e^{-\frac{i}{\hbar}\epsilon_n(t_1 - t_2)}\ \delta_{n',n} \tag{3.46}$$

$$g_{LL}^{r}(n', t_1; n, t_2) = i\, \theta(t_2 - t_1)\, e^{-\frac{i}{\hbar}\epsilon_n(t_1 - t_2)}\ \delta_{n',n} \tag{3.47}$$

Green functions for the right side can be obtained by substituting R for L and m for n in the above equations. Using equations (3.43) and (3.44) and the expressions for Green functions given above, we can calculate tunneling current from equation (3.42) which comes out to be

$$J_{L\to R} = \frac{2\pi e}{\hbar} \sum_{n,m} |T_{mn}|^2\ (f_D(\epsilon_n - Ef_L)\ -\ f_D(\epsilon_m - Ef_R))\ \delta(\epsilon_n - \epsilon_m) \tag{3.48}$$

If we define the lifetime of the n'th state on the left side as

$$\frac{1}{\tau_n} = \frac{2\pi}{\hbar} \sum_m |T_{mn}|^2 \delta(\epsilon_n - \epsilon_m) \tag{3.49}$$

then in terms of these lifetimes, the tunneling current from the quasi-bound states on the left side can be written as

$$J_{L\to R} = e \sum_n \frac{1}{\tau_n}\ (f_D(\epsilon_n - Ef_L)\ -\ f_D(\epsilon_n - Ef_R)) \tag{3.50}$$

Above equation is the central result of this section. In the next section we will describe how to calculate tunneling currents from MOS inversion and accumulation layers using expression (3.50).

We have defined the lifetime of a quasi-bound states as follows

$$\frac{1}{\tau_n} = \frac{2\pi}{\hbar} \sum_m |T_{mn}|^2 \delta(\epsilon_n - \epsilon_m) \tag{3.51}$$

Now we will make a connection between lifetime as defined above and its definition given in equation (3.39)

$$\frac{1}{\tau_n} = f_n \times |t(\epsilon_n)|^2 \tag{3.52}$$

The coupling constants T_{mn} used in the tunneling hamiltonian are given by the relation [24, 25]

$$T_{mn} = -\frac{\hbar^2}{2m^{ox}} \int \left(\psi_m^* \vec{\nabla}\psi_n \; - \; \psi_n \vec{\nabla}\psi_m^*\right) \cdot d\vec{S} \tag{3.53}$$

The above integral is over a surface lying inside the barrier and separating the left and right hand sides. The left side states ψ_n are defined for the hamiltonian in which the barrier extends all the way upto $+\infty$. Similarly the right side states ψ_m are defined for the hamiltonian in which the barrier extends all the way upto $-\infty$. If the eigenstates also have transverse components we can choose a composite labelling scheme $\{n, \vec{k_{||}}\}$ and $\{m, \vec{q_{||}}\}$ for the left and right side states respectively. In case of MOS devices the eigenstates are of the form

$$\psi_{n,\vec{k}_{||}} = \chi_n(x) e^{i\vec{k}_{||}\cdot\vec{r}} \tag{3.54}$$

$$\psi_{m,\vec{q}_{||}} = \chi_m(x) e^{i\vec{q}_{||}\cdot\vec{r}} \tag{3.55}$$

The coupling constant therefore becomes

$$T_{\{n,\vec{k}_{||}\}\{m,\vec{q}_{||}\}} = -\frac{\hbar^2}{2m^{ox}} \left(\chi_m^* \frac{\partial \chi_n}{\partial x} \; - \; \chi_n^* \frac{\partial \chi_m}{\partial x}\right) \delta_{\vec{k}_{||},\vec{q}_{||}} \tag{3.56}$$

To get a simple analytical expression for lifetime we suppose that quasi-bound state in the left side is confined in a square well potential of length l, and the state on the right side is confined in a macroscopically large box of length L as shown in figure (3-3). The analytical form of the left and right eigenstates of this system are

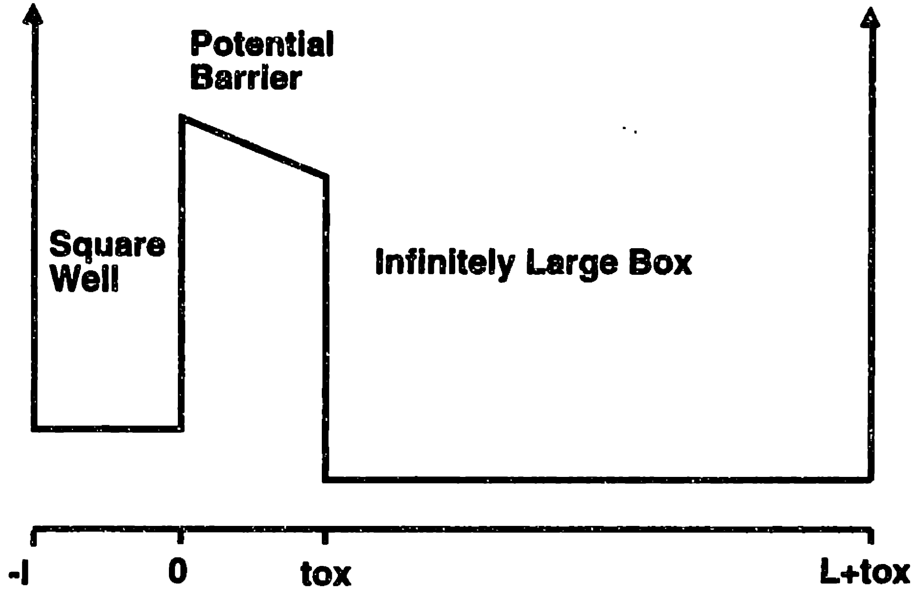

Figure 3-3: Potential for calculating the coupling constants

$$\chi_{left}(x) \approx A \sin k_i(x+l)\theta(-x) + B\frac{e^{-\int_0^x \kappa(x)dx}}{\sqrt{(x)}}\theta(x) \tag{3.57}$$

$$\chi_{right}(x) \approx C \sin k_f(x - t_{ox} - L)\theta(x - t_{ox}) + B\frac{e^{-\int_{t_{ox}}^x \kappa(x)dx}}{\sqrt{(x)}}\theta(t_{ox} - x) \tag{3.58}$$

where A,B,C, and D are appropriate normalization constants. Using these eigenstates in equation (3.56) and using equation (3.51) we get for the lifetime

$$\frac{1}{\tau_{k_i}} = \frac{\hbar k_i}{2lm^{si}}\left(\frac{16(m^{ox}m^{si})^2 k_i k_f \kappa(0)\kappa(t_{ox})e^{-2\int_0^{t_{ox}} \kappa(x)dx}}{((\kappa(0)m^{si})^2 + (k_i m^{ox})^2)((\kappa(t_{ox})m^{si})^2 + (k_f m^{ox})^2)}\right) \tag{3.59}$$

The part outside the paranthesis in the above expression is the classical oscillation frequency of a particle in the quasi-bound state and the part inside the paranthesis is just the WKB transmission probability of particle with initial momentum k_i and final momentum k_f. Thus we have reproduced the results derived before by a different method, and this time we even have the correct pre-exponential factor in the WKB approximation.

Generalizing from what we have learned we propose the following ansatz for the lifetime of a quasi-bound state of energy ϵ_n which is valid even in the Fowler-Nordheim tunneling regime

$$\frac{1}{\tau_n} = \frac{T(\epsilon_n)}{2\int_{x_n}^{0} dx \sqrt{\frac{m^{si}}{2(E_n - V(x))}}} \tag{3.60}$$

where $T(\epsilon_n)$ is the transmission probability calculated using the Airy functions, x_n is the classical turning point of the state and the expression in the denominator is just the time period for one complete classical round trip. In this thesis we have used the above formula in numerical calculations.

3.4 Calculation of Tunneling Currents in MOS Devices

In this section we will derive the expression for tunneling currents from accumulation and inversion layers in MOS devices. In the last section we obtained the following formula for tunneling currents from quasi-bound states

$$J_{L\to R} = e \sum_{n} \frac{1}{\tau_n} \left(f_D(\epsilon_n - Ef_L) \; - \; f_D(\epsilon_n - Ef_R)\right) \tag{3.61}$$

Taking into account the transverse components of wavefunctions we can write this as

$$J_{L\to R} = e \sum_{n,\vec{k}_{||}} \frac{1}{\tau_n} \left(f_D(\epsilon_n + \epsilon_{\vec{k}_{||}} - Ef_L) \; - \; f_D(\epsilon_n + \epsilon_{\vec{k}_{||}} - Ef_R)\right) \tag{3.62}$$

Note that according to our assumptions the lifetime is independent of the transverse part of the wavefunction. Now assuming a $< 100 >$ surface orientation for the Silicon substrate we can perform the summation over the transverse channels in the above

expression and get

$$
\begin{aligned}
J_{L\rightarrow R} &= \frac{4e\sqrt{m_t^{si}m_l^{si}}KT}{\pi\hbar^2}\sum_n \log\left(\frac{1+\exp\dfrac{Ef_L-\epsilon_n^t}{KT}}{1+\exp\dfrac{Ef_R-\epsilon_n^t}{KT}}\right)\frac{1}{\tau_n(\epsilon_n^t)} \\
&+ \frac{2em_t^{si}KT}{\pi\hbar^2}\sum_n \log\left(\frac{1+\exp\dfrac{Ef_L-\epsilon_n^l}{KT}}{1+\exp\dfrac{Ef_R-\epsilon_n^l}{KT}}\right)\frac{1}{\tau_n(\epsilon_n^l)} \qquad (3.63)
\end{aligned}
$$

where, as before, the superscript $t(l)$ on the energies mean that the state belongs to the pocket in which the effective mass in direction perpendicular to the Si/SiO_2 interface is $m_t^{si}(m_l^{si})$ respectively. In some cases the tunneling currents may have a contribution from extended states as well. This contribution may be calculated using semi-classical methods already described in chapter two. Usually, this contribution is negligible.

In deriving equation (3.63) we have made an important and subtle approximation. We have assumed that the electrons in the quasi-bound states are in equilibrium among themselves and, therefore, their distribution in energy is described by a quasi-fermi level Ef_L. This assumption needs some justification. In case of tunneling from inversion layers the electrons are injected into the inversion layer from the source and drain ends of the device. In practice, the source and drain ends of the device are tied together at the same potential with respect to the gate for making tunneling current measurements. Electrons injected from the drain and source travel under the gate and after some time manage to escape into the gate by tunneling through the oxide. Tunneling currents from very thin oxides can be very large - reaching upto 10 A/cm^2 for 15$\mathring{A}$ oxides. Therefore, at such large current densities the quasi-fermi level in the source and drain regions is not expected to be the same as in the channel under the gate. Infact the distribution function of electrons in the channel may not even be describable in terms of a Fermi-Dirac distribution function with an appropriate quasi-fermi level. We therefore expect our numerical results for tunneling currents to be accurate only at relatively small current densities (< 0.1A/cm^2). In case of

accumulation layers the electrons in the channel come from the substrate. What we have said above for the case of inversion layers applies equally well for the present case also. The quasi-fermi level for electrons in the substrate may not be the same quasi-fermi level as that for the electrons in the bound states in the channel. And for very high current densities the distribution function of bound electrons may not be a Fermi-Dirac distribution. However, the situation is not so bleak. In the next chapter we will show that the lifetime of electrons in MOS quasi-bound states come out to be very large - much larger than the energy relaxation times. These lifetimes range from micro-seconds to seconds. The energy relaxation times are of the order of pico-seconds. Thus we may expect that electrons in the channel are in equilibrium among themselves. This means that the distribution function of electrons in the channel is indeed Fermi-Dirac. Long lifetimes also suggest that the sources (e.g. source and drain in case of inversion layers and substrate in case of accumulation layers) can supply the replacements for electrons that tunnel out much faster than the rate at which electrons tunnel out. Therefore, as a reasonably good approximation we may even take the channel electrons to be in equilibrium with their sources.

3.5 Conclusion

In this chapter we have presented a fully self-consistent quantum mechanical model to describe tunneling from quasi-bound states in accumulation and inversion layers of MOS devices. This model overcomes the problems associated with the semi-classical model by including the effects associated with quantization of the electron motion in direction perpendicular to the Si/SiO_2 interface. We have also presented two different methods to calculate lifetimes of quasi-bound states and showed that both methods give approximately the same results. However our approach was limited to the WKB approximation. Based upon the intuition gained from these methods we proposed an ansatz for the lifetime of quasi-bound states that is valid even when the tunneling takes place in the Fowler-Nordheim regime.

In the last section we used the non-equilibrium Green function technique to cal-

culate tunneling currents from quasi-bound states under non-equilibrium conditions. In the next chapter we will present the results of numerical calculations and also compare them with the results obtained from the semi-classical methods. We will also compare the calculations with actual experimental data.

Chapter 4

Numerical Results and Comparison with Experiments

4.1 Foreword

In this chapter we will present the results of numerical calculations carried out for the semi-classical model and self-consistent quantum mechanical model to compute tunneling currents. We will focus upon tunneling transport through very thin ($\sim 15 - 35 \AA$) oxides. In such thin oxides, tunneling occurs only in the direct regime, since thin oxides usually break down at large gate voltages much before the Fowler-Nordheim regime is reached. We will also compare our calculations with experimental data and show an excellent agreement between theory and experiment. We will also present self-consistent results for various important device parameters related to tunneling transport through thin oxides. We start by showing the theoretical results for the case of inversion layers.

4.2 Semi-Classical and Self-Consistent Results for Inversion Layers

In this section the numerical results obtained from the semi-classical and the self-consistent models will be presented for the case of tunneling from inversion layers. We have carried out these calculations at room temperature for an N-channel device with a substrate doping of $10^{17}/cm^3$. The surface orientation of the substrate is assumed to be $< 100 >$. The energy dispersion relation in the oxide is taken to be isotropic and parabolic with an effective mass of $m^{ox} = 0.5m_o$. The conduction band discontinuity between Si and SiO_2 is assumed to be 3.15 eV. The barrier reduction effects caused by image forces have been neglected. Justification for these assumptions will be provided in the next chapter. The potential drop in gate n^+-Poly-Si will usually be neglected, unless stated otherwise. As stated earlier, this becomes significant only when the doping in the gate is less than $5 \times 10^{19}/cm^3$. Besides, there is no important physics associated with it, other than the fact that its inclusion is important when comparisons with actual experimental data are made.

Figure (4-1) shows the calculated results for tunneling currents from inversion layers obtained from the semi-classical and the self-consistent models for 15, 20, 25, 30, and 35$\AA$ oxide N-channel MOS devices. The most interesting thing is the fact that both the self-consistent model and the semi-classical model predict the same magnitude of tunneling current for all oxide thicknesses over almost the entire range of gate voltages shown in the figure. This observation deserves some explanation. Figure (4-2) shows the energies of first five subbands for both the possible values of effective masses (m_l^{si} and m_t^{si}) for a 15$\AA$ oxide as a function of gate voltage. Figure (4-3) shows the potential drop in the substrate calculated from the self-consistent and semi-classical models for a 15$\AA$ oxide as a function of gate voltage. In the semi-classical model the electron motion perpendicular to the Si/SiO_2 is not quantized. Therefore most of the electrons occupy states with very little energy perpendicular to the interface. However, in the self-consistent model electron energies perpendicular to the interface are quantized. Therefore, on average, the electrons have higher

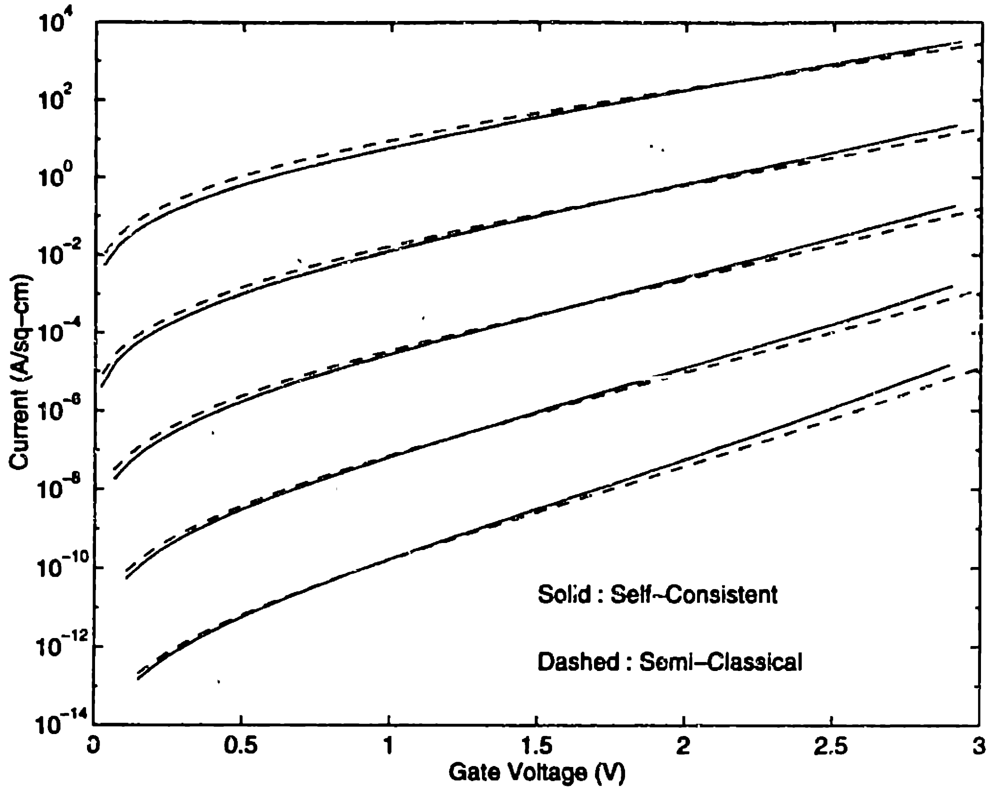

Figure 4-1: Tunneling currents from inversion layer in 15, 20, 25, 30, and 35$\AA$ oxide N-channel MOS devices. The substrate doping is $10^{17}/cm^3$. Potential drop in n^+ Poly-Si gate is ignored.

energies perpendicular to the interface than as predicted by the semi-classical model. Since tunneling rates increase exponentially with energy, we might expect the self-consistent model to predict higher tunneling currents than those calculated from the semi-classical model. However, figure (4-3) shows that the self-consistent model gives a larger potential drop in the substrate than that calculated semi-classically. This is a well known fact [15]. Consequently, the oxide potential drop, and therefore the oxide electric field, will be smaller in the self-consistent result. This is shown explicitly in figure (4-4).

Tunneling rates also increase exponentially with increase in oxide electric field. Therefore, the higher oxide electric field in the semi-classical result will cause tunneling currents to be higher in the semi-classical solution.

Now at very low gate voltages (less than 0.25 volts) it is reasonable to expect that the difference in the magnitude of tunneling currents predicted by the semi-classical

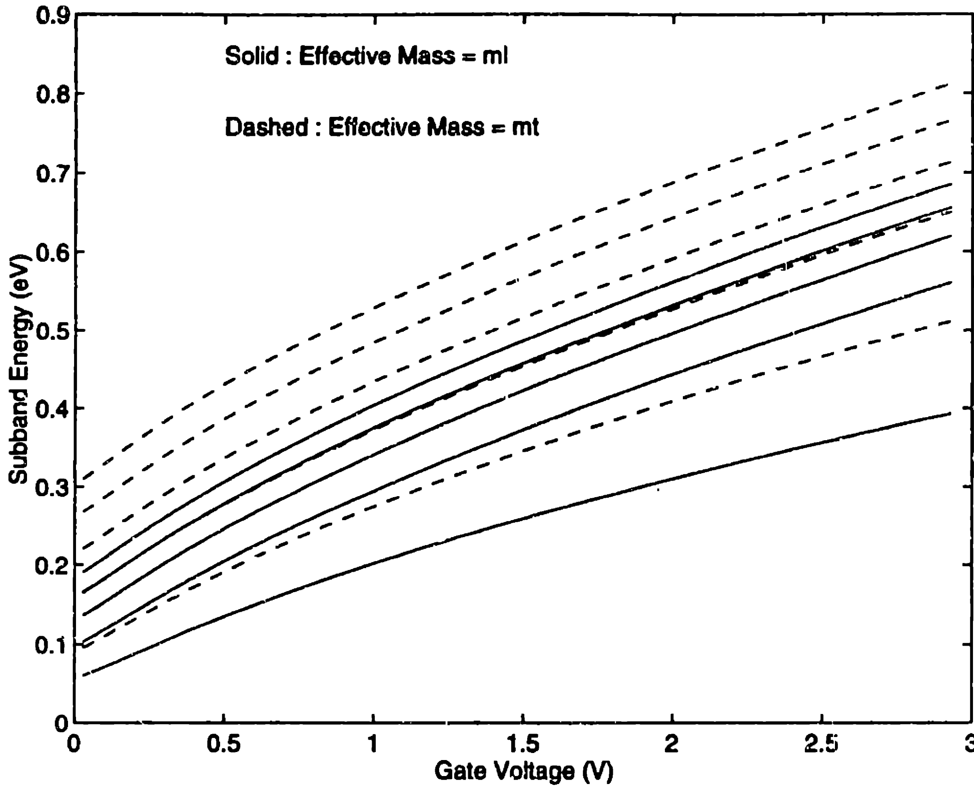

Figure 4-2: Energies of the first five subbands measured w.r.t. the bottom of the conduction band edge at the Si/SiO_2 interface. The two different sets of curves are for two different pockets in which the electron effective mass perpendicular to the interface is either m_l^{si} or m_t^{si}. The device has an oxide thickness of 15$\AA$ and substrate doping of $10^{17}/\mathrm{cm}^3$.

and the self-consistent model will be small since quantum mechanical effects are small (in other words electron density and subband energies are both small), and therefore the electron gas must show semi-classical behavior. This is visible in the fact that potential drop in the substrate and in the oxide are almost the same at low gate voltages in both the semi-classical and the self-consistent solutions. Therefore, it is not surprising that tunneling currents predicted by the two models agree at low gate voltages. As the gate voltage is increased the subband energies increase (see figure (4-2)). Quantization effects become more prominent, but at the same time, and perhaps as a result, the difference in the oxide electric field strengths predicted by the semi-classical and the self-consistent model also increases. The net result of these two effects is that tunneling currents calculated from the semi-classical and the self-consistent models come out fortuitously to be approximately the same over the entire

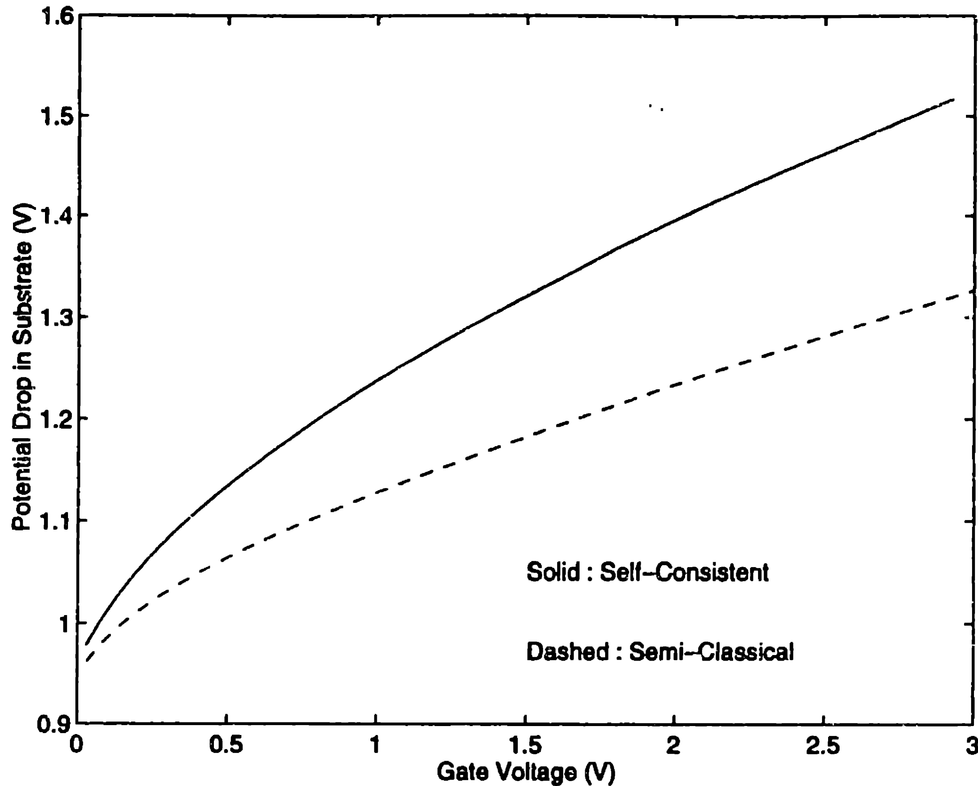

Figure 4-3: The potential drop in the substrate as a function of gate voltage. The device has an oxide thickness of 15$\AA$ and substrate doping of $10^{17}/\text{cm}^3$.

range of gate voltages from 0 to 3 volts. We have not yet investigated theoretically whether this agreement by chance remains at higher gate voltages. However, in actual experiments the oxides usually break down around 3 volts.

The lifetime of electrons in the lowest two subbands for various oxide thicknesses are shown in figure (4-5). The two lowest subbands are associated with the two different effective masses (m_l^{si} and m_t^{si}) of electrons in Silicon. Interestingly, the lifetimes can vary from a few nano seconds to almost a year ($\sim 10^7$s) depending upon the oxide thickness and the applied gate bias. But even for the thinnest 15$\AA$ oxide the lifetimes are larger than nano seconds. Thus the assumption made in the self-consistent model that the electron distribution function for the quasi-bound states be describable by a Fermi-Dirac function with an appropriate quasi-fermi level is justified. This is because the energy relaxation times at room temperature are of the order of 10^{-13} seconds, and the electrons are expected to reach local equilibrium among themselves at rates much faster than tunneling rates. Even at liquid Helium

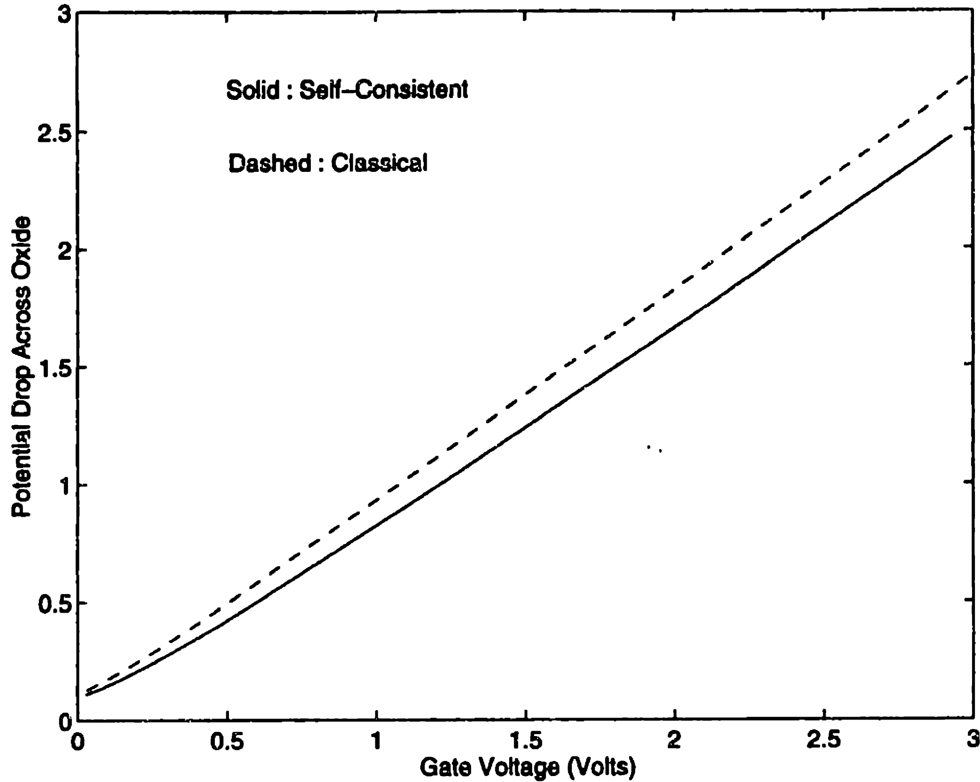

Figure 4-4: The potential drop in the oxide as a function of gate voltage. The device has an oxide thickness of 15Å and substrate doping of $10^{17}/\mathrm{cm}^3$.

temperatures, where relaxation times are a few pico seconds, local equilibrium seems a robust assumption.

Finally, we show the inversion layer capacitance calculated from the self-consistent and the semi-classical model. Since oxide thicknesses are measured electrically using capacitance spectroscopy, it is important to know how much error is incurred in such measurements if the quantum effects associated with the inversion layer charge distribution are neglected. Figure (4-6) shows the capacitances calculated from the self-consistent and the semi-classical model for various oxide thicknesses. Figure (4-7) shows the relative error in extracting oxide thickness from electrical methods if the simple formula

$$t_{ox} = \epsilon_{ox}/C \tag{4.1}$$

is used to determine thickness from the electrically measured capacitance C in strong inversion. The origin of this error lies in the fact that charge distribution in the inver-

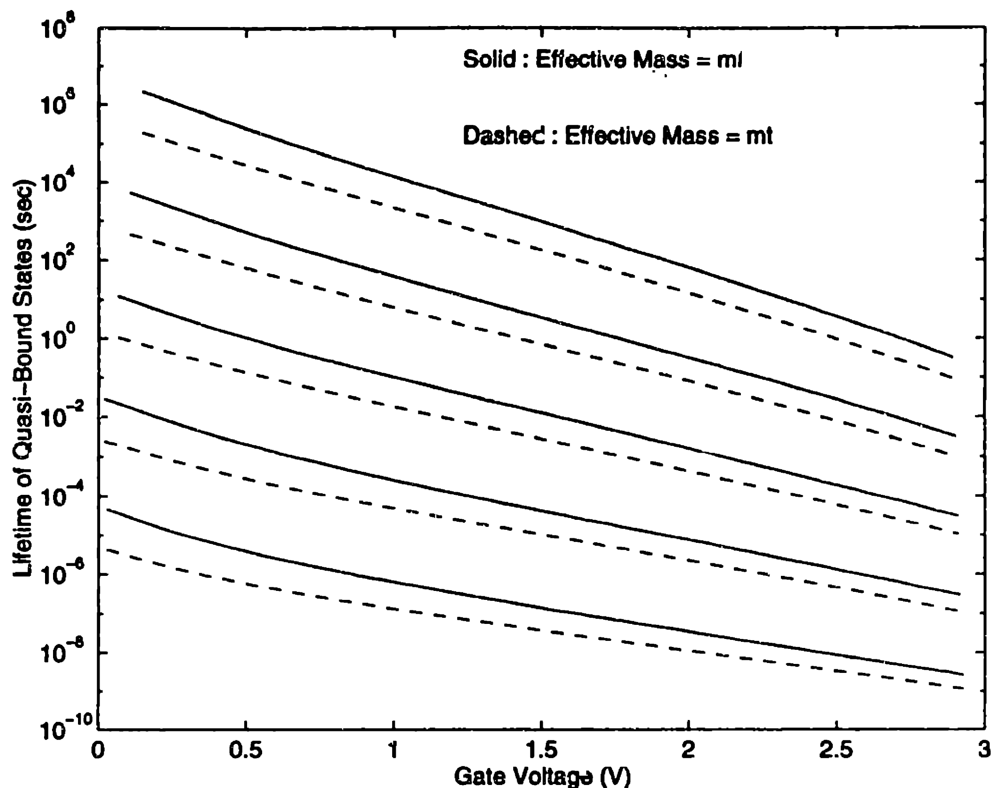

Figure 4-5: The lifetimes of electrons in 15, 20, 25, 30, and 35Å oxide MOS devices. For each oxide thickness, the two sets of curves are for the lowest two subbands corresponding to effective masses m_l^{si} and m_t^{si}. The top most set of curves is for the thickest 35Å oxide. The substrate doping is $10^{17}/cm^3$ in each case.

sion layer calculated self-consistently does not have its peak at the Si/SiO_2 interface but at a distance of few (3-5) Angstroms away from it. This is shown in figure (4-8) which shows the self-consistent charge density profile for a 15Å MOS device. Note that we have not modeled the contribution to the error caused by the depletion region in the Poly-Si gate. Depending upon the doping in the gate, this contribution may be large.

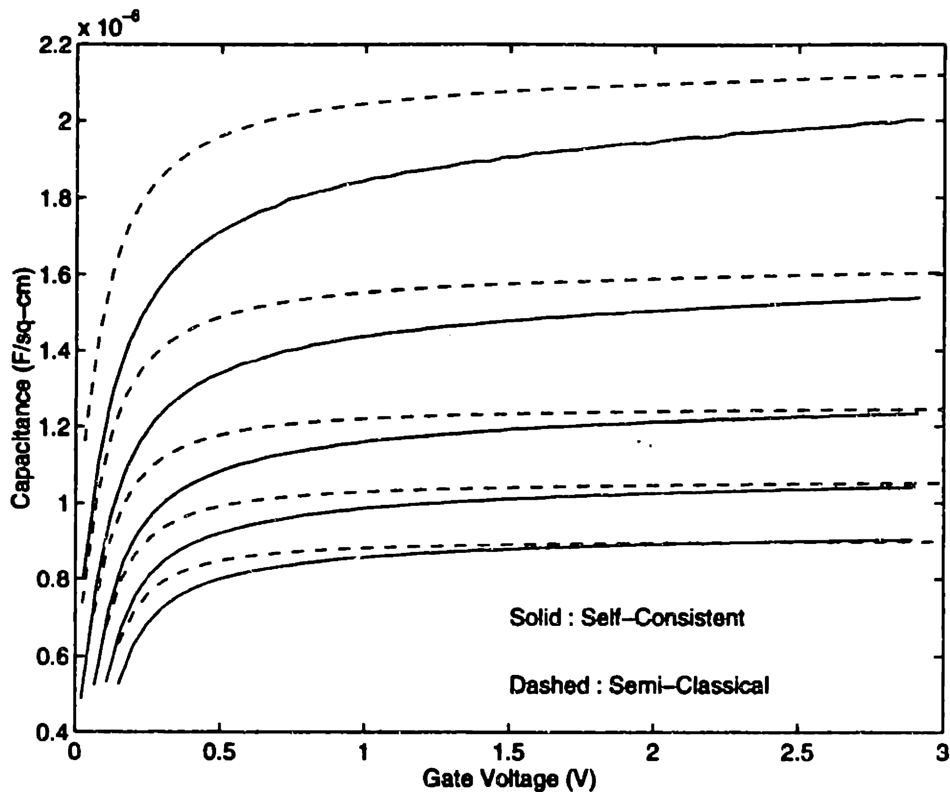

Figure 4-6: Capacitance of 15, 20, 25, 30, and 35Å MOS devices calculated from the self-consistent and the semi-classical model. The substrate doping in each case was $10^{17}/\text{cm}^3$.

4.3 Semi-Classical and Self-Consistent Results for Accumulation Layers and Comparison with Experimental Data

4.3.1 Theoretical Results for Accumulation Layers

In the last section we presented results of theoretical calculations for the case of inversion layers. The primary goal of this section is to compare theoretical calculations with experimental data. However, the experimental measurements were made on MOS capacitors with n-doped substrates. The tunneling in these devices was therefore from accumulation layers instead of from inversion layers. The motivation of performing tunneling measurements from accumulation layers instead of inversion layers is twofold

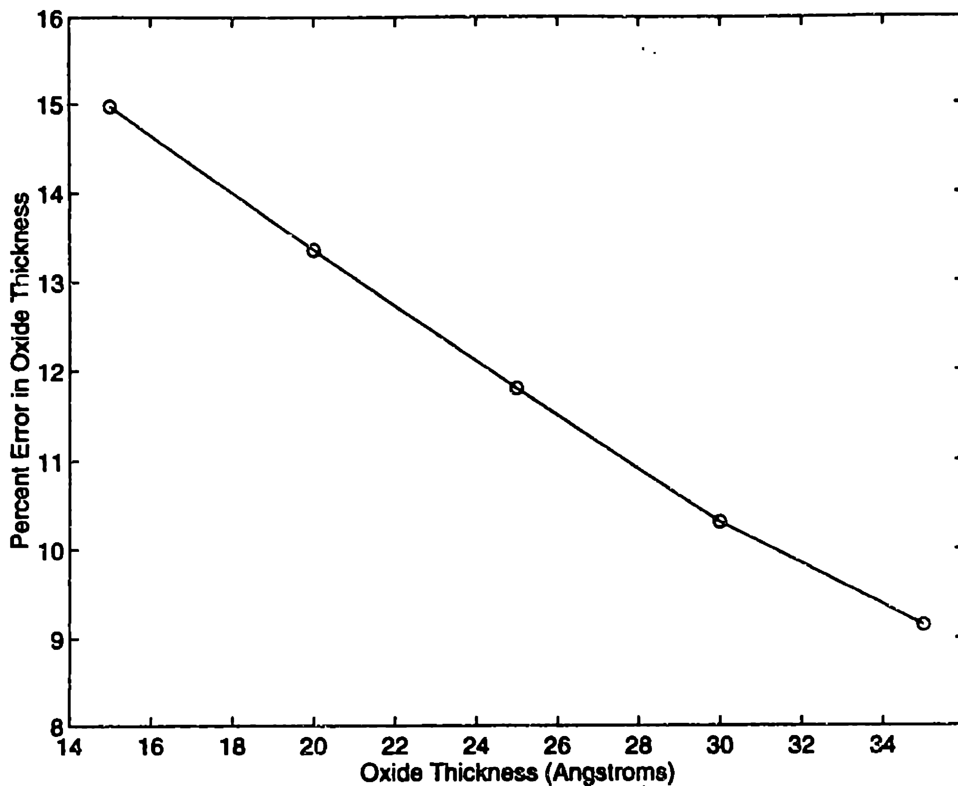

Figure 4-7: The relative error in extracting oxide thicknesses from electrical measurements performed in the strong inversion regime.

- It is much easier and also it takes much less time to fabricate n-MOS capacitors which can be used to study tunneling currents from accumulation layers than make n-channel FET's which are necessary to study tunneling currents from inversion layers.

- The lateral transport in the channel also needs to be carefully modeled for a complete description of tunneling from inversion layers. Such modeling adds unnecessary complications and increases the number of unknown parameters in the theory.

Since experimental data presented in this thesis has been obtained from n-MOS capacitors, we have also carried out self-consistent and semi-classical calculations for tunneling currents from accumulation layers. Results of these calculations will now be presented.

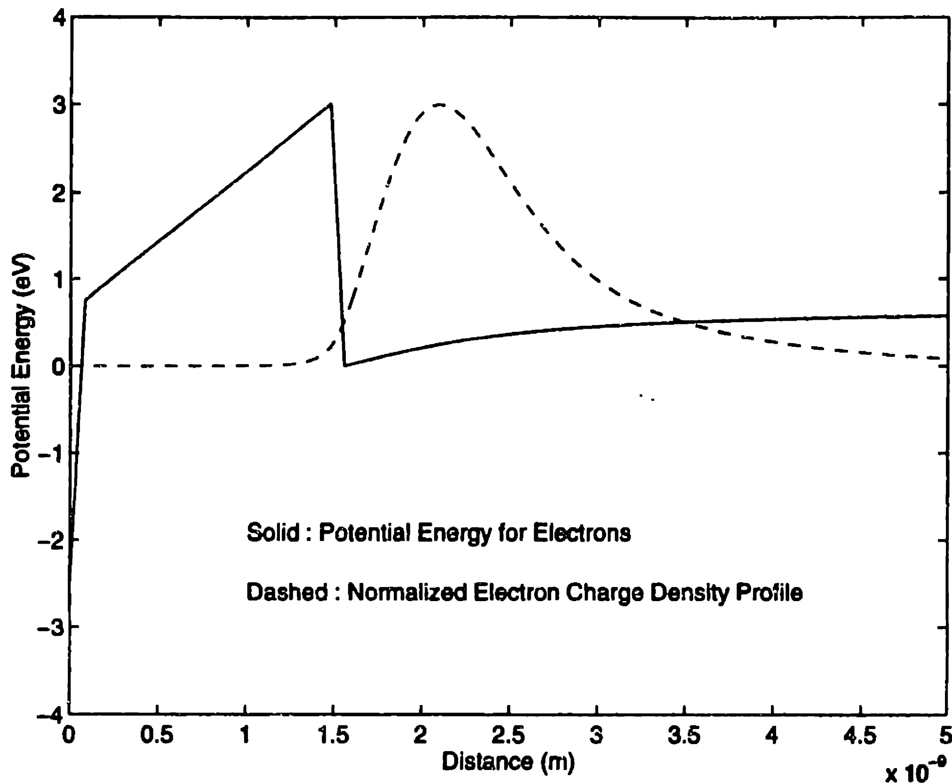

Figure 4-8: Self-consistent charge density for a 15Å MOS device at a gate voltage of 3.0 Volts.

Figure (4-9) shows the calculated tunneling currents from accumulation layers for various oxide thicknesses. Just like in the case of inversion layers, the tunneling currents predicted by the semi-classical and the self-consistent model come out to be almost the same for all values of gate voltage from 0 to 3 volts. Figure (4-10) shows the capacitance calculated from the semi-classical and self-consistent models, and figure (4-11) shows the relative error in extracting oxide thicknesses from electrical methods if expression (4.1) is used to calculate oxide thickness. As before, in all calculations potential drop in the n+ Poly-Si gate has been ignored.

For completeness, we also present here the plots of the energies of first five sub-bands for both the possible values of effective masses (m_l^{si} and m_t^{si}) for a 15Å oxide as a function of gate voltage (figure (4-12)), and the potential drop in the substrate calculated from the self-consistent and the semi-classical model for a 15Å oxide as a function of gate voltage (figure (4-13)). These plots show trends similar to those observed in case of the inversion layers and these have already been discussed in detail

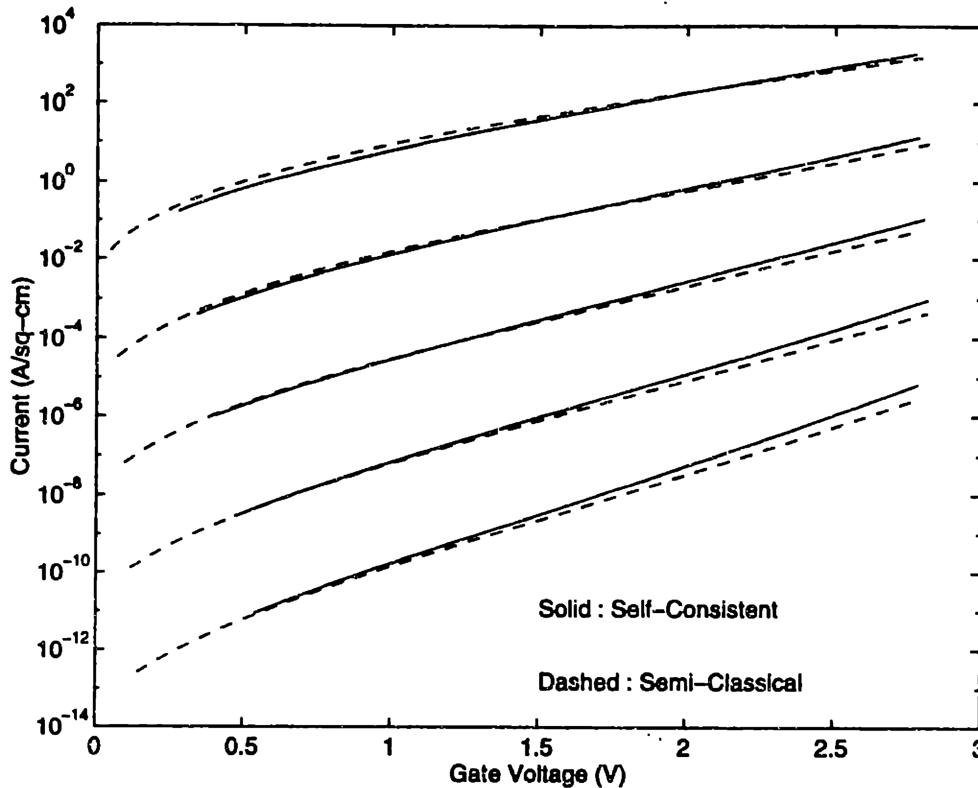

Figure 4-9: Tunneling currents from accumulation layer in 15, 20, 25, 30, and 35Å oxide n-MOS capacitors. The substrate doping is $10^{17}/cm^3$. Potential drop in n^+ Poly-Si gate is ignored.

in the previous section.

4.3.2 Comparisons With Experimental Data

Before presenting a comparison of the theoretically calculated currents with experimental measurements, we would like to mention a few things about the experimental measurement of oxide thicknesses. In practice it is difficult to grow thin oxide of a desired thickness. Usually the grown oxide thickness comes out to be within an error of 2-3Å of the desired value. Oxides thicknesses can be measured experimentally using a variety of methods such as ellipsometery, capacitance spectroscopy, x-ray photoemission e.t.c.. The interested reader is referred to a recent review article in reference [33]. The thicknesses measured by different methods do not come out to be the same, and indeed they should not because each of these measurements is sensitive to a different physical property. For example, whereas capacitance measurements are sensitive to

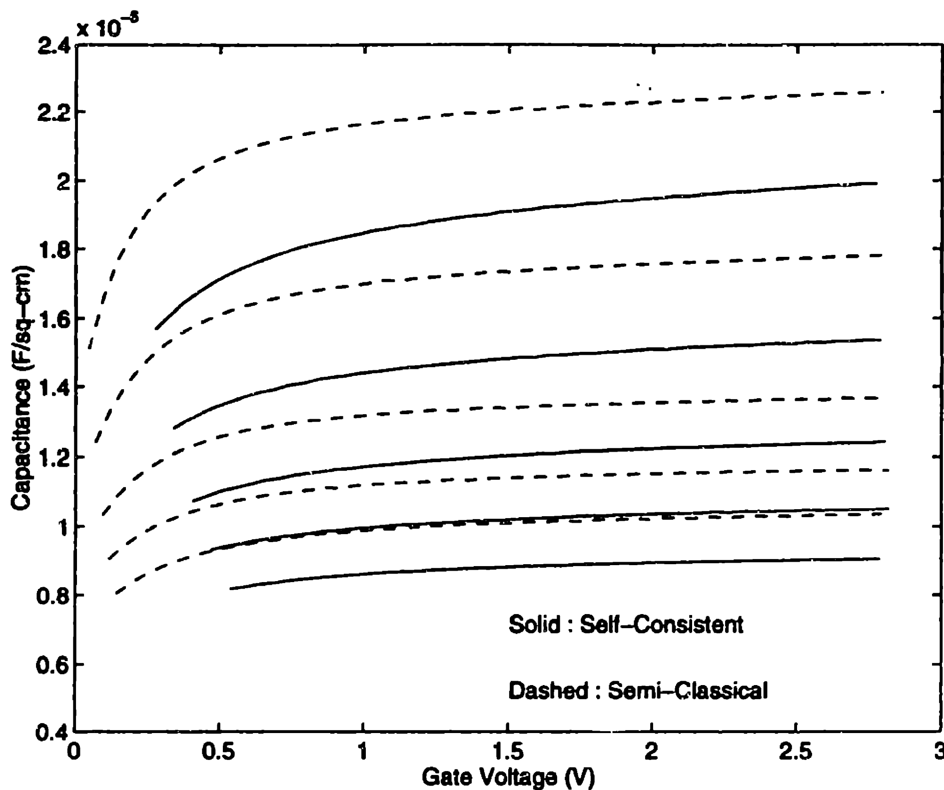

Figure 4-10: Capacitance of 15, 20, 25, 30, and 35Å n-MOS capacitors calculated from the self-consistent and the semi-classical model. The substrate doping in each case is $10^{17}/cm^3$.

the position of the peak of the electron charge density in the channel, ellipsometric measurements detect the average position of the plane where the dielectric constant changes from 3.9 (bulk SiO_2 value) to 11.7 (bulk Si value). Given this ambiguity in the measurement process, the question that arises is which oxide thickness should be used in the theoretical formulas for calculating tunneling rates. At present we do not have a reliable answer to this question. However, we believe that thicknesses measured by ellipsometer may be used as a best first guess. The reason for this is that the ellipsometer allows one to determine the plane at which the material properties of the medium change sufficiently so that the effective macroscopic dielectric constant of the medium also switches values. It is reasonable to assume that this plane may also be the plane at which the energy dispersion relation of electrons change sufficiently so that electronic wavefunctions become decaying exponentials from plane waves. All oxide thicknesses quoted in this thesis have been measured by ellipsometer.

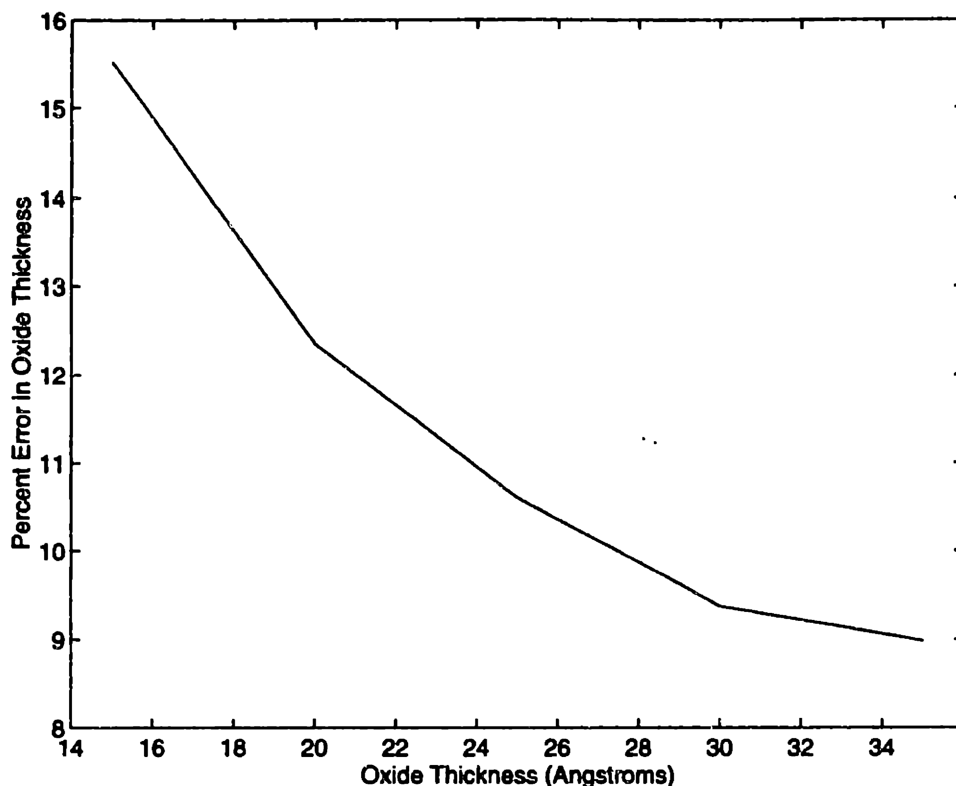

Figure 4-11: The relative error in extracting oxide thicknesses from electrical measurements performed in the strong accumulation regime.

All experimental measurements were made on MOS capacitors with n-doped Si substrates having a uniform doping concentration of $10^{17}/\mathrm{cm}^3$ and a $< 100 >$ surface orientation. The devices had n^+ Poly-Si gates with dopings ranging from $5 \times 10^{19} - 5 \times 10^{20}/\mathrm{cm}^3$. Tunneling current measurements are usually made in two different ways :

1. Tunneling currents can be measured by applying a linear voltage staircase across the gate and the substrate. Such a ramp is characterized by a delay time and hold time. Delay time is the time during which no current measurement is made after each small increment in voltage. Hold time is the time during which voltage is kept constant while several current measurements are made. To avoid errors caused by displacement (or capacitive) currents, hold times must be chosen much larger than the *RC* times associated with the device. Such measurements can easily be made by a HP4145 Semiconductor Parameter Analyzer.

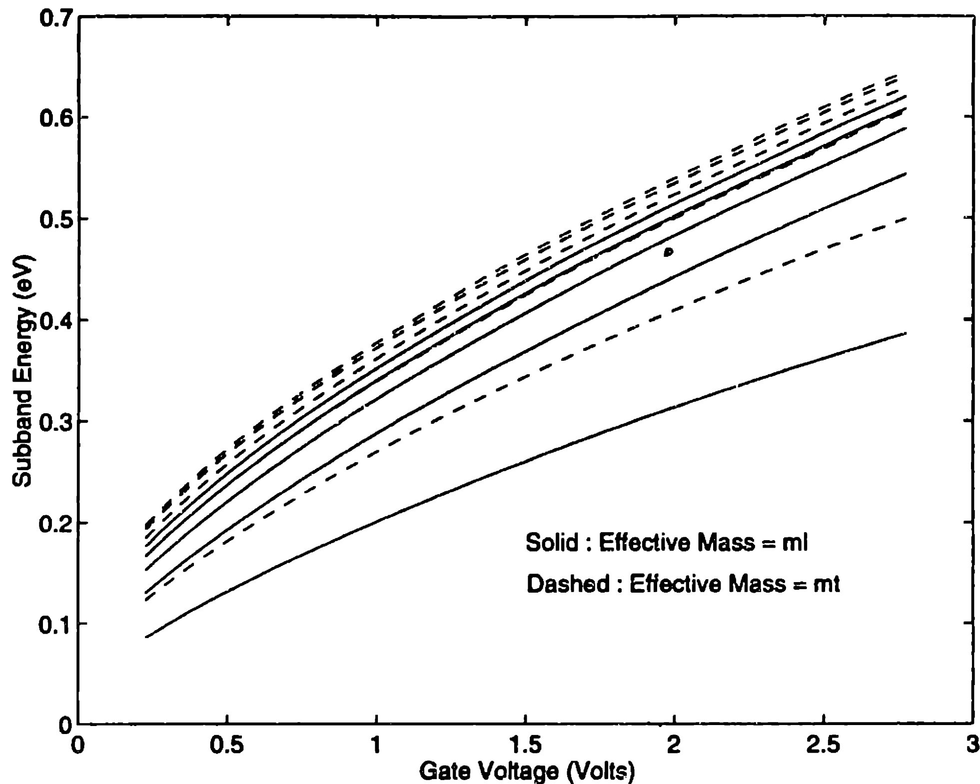

Figure 4-12: Energies of the first five subbands measured w.r.t. the bottom of the conduction band edge at the Si/SiO_2 interface. The two different sets of curves are for two different pockets in which the electron effective mass perpendicular to the interface is either m_l^{si} or m_t^{si}. The device has an oxide thickness of 15Å and substrate doping of $10^{17}/cm^3$.

2. A much better way of measuring tunneling currents through oxides is to apply short voltage pulses, and find the total charge that flows in the external circuit by integrating the current. Contributions from displacement currents, being equal and of opposite sign at the rising and falling edges of the voltage pulse, cancel out. In addition, by using very short pulses (a few micro seconds long), one can avoid sending large quantities of charge through the oxide. Transport of charge through the oxide causes oxide degradation, leading to breakdown [29]. Oxides measured with pulsed gate voltages show the same magnitude of tunneling currents as those which are measured with ramped gate voltages. However, they exhibit, on average, higher breakdown voltages than those oxides which are measured with ramped gate voltages [34].

The experimental data given here were obtained by the pulsed voltage method.

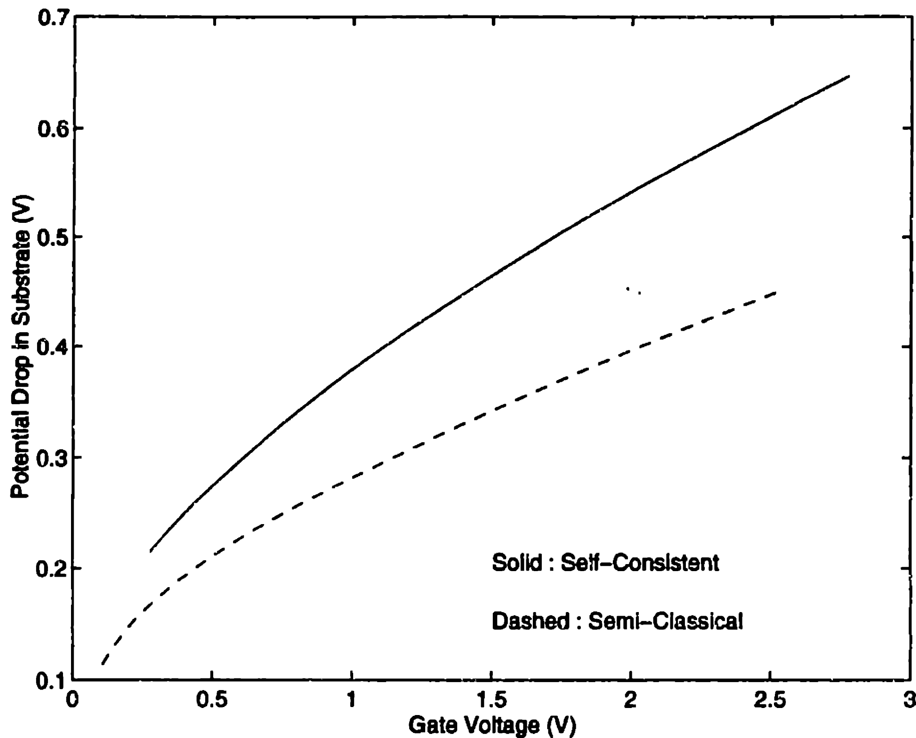

Figure 4-13: The potential drop in the substrate as a function of gate voltage. The device has an oxide thickness of 15Å and substrate doping of $10^{17}/\text{cm}^3$.

Figure (4-14) shows a typical set of measurements on a large number of devices with oxide thicknesses of 14, 20, 23, 27 and 35Å. From the figure one can easily observe the large amount of variability associated with the oxides' characteristics. For example, the 20Å devices show breakdown voltages ranging from 2.2 volts to 4.2 volts, even though all the devices went through the same fabrication sequence. The variability in the magnitude of tunneling currents from different oxides of the same thickness can also be as large as an order of magnitude. This much variability represents the current state of the art in oxides grown at IBM (NY). We believe that the cause of these variations may be the defects in the thermally grown oxides. Such defects are usually associated with non-crystalline materials [35]. Given such a large amount of variability in oxides' characteristics, we feel that it is absurd to use experimental data on tunneling currents to extract physical parameters such as barrier height and oxide effective mass. Indeed, such efforts have been made in literature [7, 10]. They are, perhaps, more useful in case of thicker oxides than for the case of very thin oxides. In

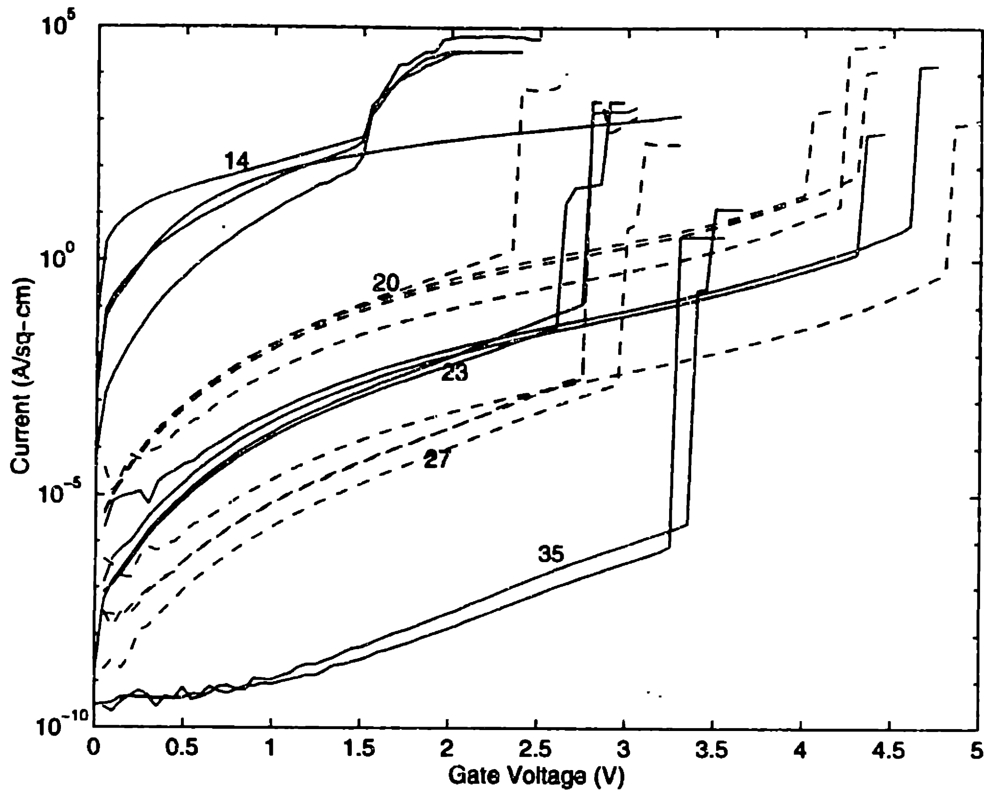

Figure 4-14: A typical set of experimental data on tunneling currents from accumulation layers in 14, 20, 23, 27, and 35$\mathring{A}$ oxides. All devices had a n-type substrate with doping $10^{17}/\mathrm{cm}^3$ and a n^+ Poly-Si gate with doping $5 \times 10^{19}/\mathrm{cm}^3$.

this thesis the main aim has been to use theory without any fitting parameters and find out how far one can go with it in modeling the tunneling characteristics of thin oxides.

Figure (4-15) shows the tunneling currents calculated from the semi-classical and the self-consistent model plotted with the experimentally measured data. The experimental curves were picked from the data shown in figure (4-14). Those curves were picked which showed a relatively high breakdown voltage, and also represented a mean of the tunneling currents measured from devices of a particular oxide thickness. The agreement between theory and experiment is excellent within the error limits imposed by the uncertainty in the measured thicknesses of oxides. This uncertainty, inherent in ellipsometric measurements, is expected to be $\sim 1\mathring{A}$.

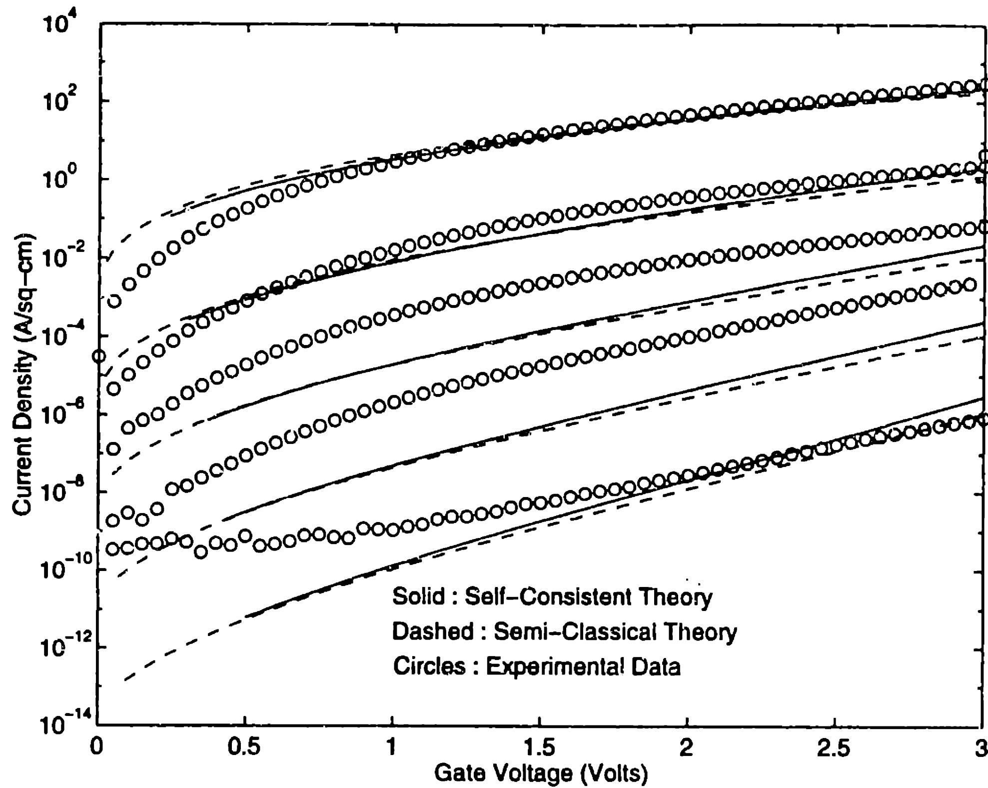

Figure 4-15: Tunneling currents from accumulation layers calculated from the semi-classical and the self-consistent model for 15, 20, 25, 30, and 35Å oxides compared with the experimental measurements for 14, 20, 23, 27, and 35Å oxides. All devices had a n-type substrate with doping $10^{17}/cm^3$ and a n^+ Poly-Si gate with doping $5 \times 10^{19}/cm^3$.

4.4 Conclusion

In this chapter we presented the results of numerical calculations based upon the semi-classical model described in chapter two and the quantum mechanical self-consistent model developed in chapter three. We showed that despite completely different physics associated with the two models, the magnitude of tunneling currents predicted by the two models come out to be almost the same. This was true for tunneling from accumulation as well as inversion layers.

We also presented experimental data, which in our opinion, was typical of a large number of measurements performed on many different devices. The experimental data showed large variations in magnitude of tunneling currents measured from different devices having the same oxide thickness and fabricated via the same process.

Finally we showed that within the error limits present due to the uncertainty in the measured thicknesses of the oxides by ellipsometer, the agreement between theory and experiment was very good. Such a good agreement was surprising. It still remains a mystery to us whether it was accidental or that crystalline effective mass theory really captures the physics associated with tunneling in thin amorphous oxides.

In the next chapter we wish to examine in more detail two assumptions we made in our theoretical models. We assumed that the transport in the mid-gap region of SiO_2 could be described by an effective mass of $m^{ox}(= 0.5m_o)$, and that barrier reduction effects due to image forces can be neglected. We will justify these assumptions in the next chapter and also shed some more light on the physics of tunneling through thin oxides.

Chapter 5

Advanced Issues in Physics of Electron Tunneling through Oxides : The Mid-Gap Energy Dispersion Relation in SiO_2 and the Effect of Image Forces

5.1 Introduction

The theoretical models we presented in chapters two and three to describe tunneling in thin oxides made the following two assumptions :

1. The energy dispersion relation in the mid-gap region of SiO_2 can be described by an effective mass $m^{ox} = 0.5m_o$.

2. The barrier reduction effects due to image forces may be neglected.

In this chapter we wish to justify the first assumption and also explore the physics associated with the dynamical nature of image forces. We start by discussing the issues related to the energy dispersion relation in the mid-gap region of SiO_2.

5.2 Energy Dispersion Relation in Mid-Gap Region of SiO_2 and the Crystalline WKB Approximation

In this section we plan to derive a version of effective mass theory valid for the case of tunnel transport when the tunneling electron's energy does not fall in any one of the allowed bands. In the usual effective mass theory the electron wavefunction in the n'th band is [36]

$$\psi(\vec{r},t) = f(\vec{r},t)\phi_{n,\vec{k}}(\vec{r}) \tag{5.1}$$

where the slowly varying envelope function $f(\vec{r},t)$ satisfies the effective mass equation

$$i\hbar\frac{\partial f(\vec{r},t)}{\partial t} = \left(E_n(\vec{k} - i\vec{\nabla}_r) + V_{ext}(\vec{r})\right) f(\vec{r},t) \tag{5.2}$$

The Bloch functions $\phi_{n,k}(\vec{r})$ satisfy the completeness relation

$$\sum_n \phi^*_{n,\vec{k}}(\vec{r})\phi_{n,\vec{k}}(\vec{r}') = \delta^3(\vec{r} - \vec{r}') \tag{5.3}$$

for each $\vec{k}$ in the first Brillouin zone. Therefore, we can in principle write the wavefunction describing a tunneling electron as

$$\phi(\vec{r},t) = f(\vec{r})\sum_n c_n(t)\phi_{n,\vec{k}}(\vec{r}) \tag{5.4}$$

Let the hamiltonian be written as

$$H = H_o + V_{ext}(\vec{r}) \tag{5.5}$$

Where

$$H_o = -\frac{\hbar^2}{2m_o}\vec{\nabla}^2_r + V_{lattice}(\vec{r}) \tag{5.6}$$

Substituting (5.4) in the Schrodinger equation

$$i\hbar\frac{\partial\psi(\vec{r},t)}{\partial t} = H\ \psi(\vec{r},t) \tag{5.7}$$

with $f(\vec{r}) = e^{\frac{i}{\hbar}\int \vec{Q}(\vec{r}).d\vec{r}}$, we get

$$i\hbar\sum_n \dot{c}_n(t)\phi_{n,\vec{k}}(\vec{r}) = \sum_n c_n(t)\left(\frac{(-i\hbar\vec{\nabla}_r + \vec{Q}(\vec{r}))^2}{2m_o} + V_{lattice}(\vec{r}) + V_{ext}(\vec{r})\right)\phi_{n,\vec{k}}(\vec{r}) \tag{5.8}$$

For time independent solutions with energy E we put

$$c_n(t) = c_n e^{-i\frac{Et}{\hbar}} \tag{5.9}$$

Using the decomposition [36]

$$\phi_{n,\vec{k}}(\vec{r}) = \frac{e^{i\vec{k}.\vec{r}}}{\sqrt{Vol}}u_{n,\vec{k}}(\vec{r}) \tag{5.10}$$

we can write (5.8) as

$$\sum_n c_n\left(E_n(\vec{k}) + \frac{\vec{Q}^2(\vec{r})}{2m_o} + V_{ext}(\vec{r}) - E + \frac{\vec{Q}(\vec{r})}{m_o}.(\frac{\hbar}{i}\vec{\nabla}_r + \hbar\vec{k})\right)u_{n,\vec{k}}(r) = 0 \tag{5.11}$$

Multiplying from the left by $u^*_{m,\vec{k}}(r)$ and integrating over the unit cell located in the vicinity of $\vec{r}$ we get the matrix equation

$$c_m\left(E_m(\vec{k}) + \frac{\vec{Q}^2(\vec{r})}{2m_o} + V_{ext}(\vec{r}) - E\right) = -\sum_n c_n\ \vec{Q}(\vec{r}).\int d^3\Omega\ u^*_{m,\vec{k}}(\vec{r})\frac{(-i\hbar\vec{\nabla}_r + \hbar\vec{k})}{m_o}u_{n,\vec{k}}(r) \tag{5.12}$$

From elementary band theory we know that [36]

$$\frac{1}{\hbar}\vec{\nabla}_{\vec{k}}E_n(\vec{k}) = \int d^3\vec{r}\ u^*_{n,\vec{k}}(\vec{r})\frac{(-i\hbar\vec{\nabla}_r + \hbar\vec{k})}{m_o}u_{n,\vec{k}}(r) = v_n(\vec{k}) \tag{5.13}$$

Time reversal symmetry implies that [37]

$$E_n(-\vec{k}) = E_n(\vec{k}) \quad \text{and} \quad v_n(-\vec{k}) = -v_n(\vec{k}) \tag{5.14}$$

which means that

$$v_n(\vec{k} = 0) = 0 \tag{5.15}$$

Therefore if expansion in equation (5.4) is made in terms of Bloch functions at $\vec{k} = 0$ (or around any point in Brillouin zone where the bands have extrema), only terms for which $n \neq m$ will appear on the L.H.S. of equation (5.12). Equation (5.12) is very general. Suppose the tunneling electron has an energy such that its wavefunction is describable predominantly in terms of Bloch functions belonging to only two bands A and B. Such will be the case if the electron has energy in the mid-gap region where its wavefunction may be expected to be made up predominantly of Bloch functions belonging to the conduction band and the valence band edges. For just two bands, equation (5.12) can be written in matrix form as

$$\begin{pmatrix} E_A + \frac{\vec{Q}^2(\vec{r})}{2m_o} + V_{ext}(\vec{r}) - E & \vec{Q}(\vec{r}).\vec{K}_{BA} \\ \vec{Q}(\vec{r}).\vec{K}^*_{BA} & E_B + \frac{\vec{Q}^2(\vec{r})}{2m_o} + V_{ext}(\vec{r}) - E \end{pmatrix} \begin{pmatrix} c_A \\ c_B \end{pmatrix} = 0 \tag{5.16}$$

Where $E_{A/B} = E_{A/B}(\vec{k} = 0)$, and matrix element $\vec{K}_{BA}$ is

$$\vec{K}_{BA} = \int d^3\vec{r}\, u^*_{B,\vec{k}}(\vec{r}) \frac{(-i\hbar\vec{\nabla}_r)}{m_o} u_{A,\vec{k}}(r) \tag{5.17}$$

From 2-band $\vec{k}.\vec{p}$ theory we know that effective masses for bands A and B at $\vec{k} = 0$ are given by the relation [36]

$$\frac{1}{m_{ij}^{\substack{A\\B}}} = \frac{\delta_{ij}}{m_o} \pm \frac{1}{(E_{\substack{A\\B}} - E_{\substack{B\\A}})} \left(K_i^{AB} K_j^{BA} + K_j^{AB} K_i^{BA}\right) \tag{5.18}$$

This implies that

$$\frac{1}{m_{ij}^A} + \frac{1}{m_{ij}^B} = \frac{2\delta_{ij}}{m_o} \tag{5.19}$$

Above equation must hold at least approximately otherwise a two band model is not a good approximation. In order to solve (5.16) we put the determinant of the matrix on the left hand side equal to zero and using (5.18) we get

$$\left(E_A + \frac{\vec{Q}^2(\vec{r})}{2m_o} + V_{ext}(\vec{r}) - E\right)\left(E_B + \frac{\vec{Q}^2(\vec{r})}{2m_o} + V_{ext}(\vec{r}) - E\right)$$
$$= Q_i(\vec{r})Q_j(\vec{r})\left(\frac{1}{m_{ij}^A} - \frac{1}{m_{ij}^B}\right)\frac{(E_A - E_B)}{4} \tag{5.20}$$

Einstein summation convention is used in the above expression. We now specialize to the case of SiO_2 in which the conduction and valence bands at $\vec{k} = 0$ are istropic with masses $m^A = m^c$ and $m^B = m^v$. The band gap is $(E_A - E_B) = (E_c - E_v) = E_g$. Solving (5.20) we get

$$E = V_{ext}(\vec{r}) + \frac{(E_c + E_v)}{2} + \frac{\vec{Q}^2(\vec{r})}{2m_o} \pm \sqrt{\frac{E_g^2}{4} + \vec{Q}^2(\vec{r})\left(\frac{1}{m^c} - \frac{1}{m_v}\right)\frac{E_g}{4}} \tag{5.21}$$

Using plus sign in the above equation gives the dispersion relation close to the conduction band for small $\vec{Q}(\vec{r})$ and choosing negative sign gives the dispersion relation close to the valence band. Since tunneling transport in SiO_2 usually occurs closer to the conduction band edge than the valence band edge we will choose the plus sign. Since $\vec{Q}(\vec{r})$ has components parallel and perpendicular to the Si/SiO_2 interface we can write

$$\vec{Q}^2(\vec{r}) = \vec{Q}_{||}^2(\vec{r}) + \vec{Q}_{\perp}^2(\vec{r}) \tag{5.22}$$

Now we need to make some approximations. We need to constrain $\vec{Q}_{||}(\vec{r})$ somehow. In chapter two we pointed out that when an electron tunnels from Si through SiO_2 one has two choices available - either conserve transverse kinetic energy or conserve transverse kinetic momentum (note that transverse direction in chapter two was the direction prallel to the Si/SiO_2 interface). We also discussed that in case of tunneling in Si/SiO_2 systems both choices are expected to give similar results for tunneling

currents since on average the effective mass in Si is almost equal to that in SiO_2. Whatever the choice made, $\vec{Q}_{\|}(\vec{r})$ will be much smaller than $\vec{Q}_{\perp}(\vec{r})$. This will be verified explicitly. We anticipate that the following inequality will always hold in all practical cases

$$\frac{\vec{Q}_{\|}^2(\vec{r})}{2m_o} << \frac{\vec{Q}_{\perp}^2(\vec{r})}{2m_o} < E_g \tag{5.23}$$

Therefore, using (5.18) we may expand (5.21) for small $\vec{Q}_{\|}(\vec{r})$ as

$$E = V_{ext}(\vec{r}) + \frac{(E_c + E_v)}{2} + \frac{\vec{Q}_{\|}^2(\vec{r})}{2m^c} + \frac{\vec{Q}_{\perp}^2(\vec{r})}{2m_o} + \sqrt{\frac{E_g^2}{4} + \vec{Q}_{\perp}^2(\vec{r})\left(\frac{1}{m^c} - \frac{1}{m_v}\right)\frac{E_g}{4}} \tag{5.24}$$

The energy of the tunneling electron in Si must be the same as that in SiO_2. If the initial energy of electron in Si is $E = E_c^{si} + E_{\|} + E_{\perp}$, then using (5.24) and assuming conservation of transverse kinetic energy (as was assumed in chapter two and three) we get

$$E_c^{si} + E_{\perp} = V_{ext}(\vec{r}) + \frac{(E_c + E_v)}{2} + \frac{\vec{Q}_{\perp}^2(\vec{r})}{2m_o} + \sqrt{\frac{E_g^2}{4} + \vec{Q}_{\perp}^2(\vec{r})\left(\frac{1}{m^c} - \frac{1}{m_v}\right)\frac{E_g}{4}} \tag{5.25}$$

Note that the superscript si on E_c implies the conduction band edge of Si as opposed to the conduction band edge of SiO_2.

Suppose that $\vec{Q}_{\perp}(\vec{r})$ is small. If we expand the square root in (5.25) for small $\vec{Q}_{\perp}(\vec{r})$ we get

$$-\frac{\vec{Q}_{\perp}^2(\vec{r})}{2m^c} = (e\phi_o + V_{ext}(\vec{r}) - E_{\perp}) \tag{5.26}$$

where $e\phi_o$ is the barrier height (i.e. $E_c - E_c^{si} = e\phi_o$). For $V_{ext}(\vec{r}) = 0$ and electron energies less than the barrier height, $\vec{Q}_{\perp}(\vec{r})$ can only take imaginary values, resulting in a decaying wavefunction inside the oxide. For such decaying wavefunctions we may write

$$Q_{\perp}(\vec{r}) = i\sqrt{2m^c(e\phi_o + V_{ext}(\vec{r}) - E_{\perp})} \tag{5.27}$$

We have therefore verified that at least for small $\vec{Q}_{\perp}(\vec{r})$ the dispersion relation in the band gap of SiO_2 may be described with an effective mass equal to the conduction band effective mass. However in real situations $\vec{Q}_{\perp}(\vec{r})$ is not so small. Solving for

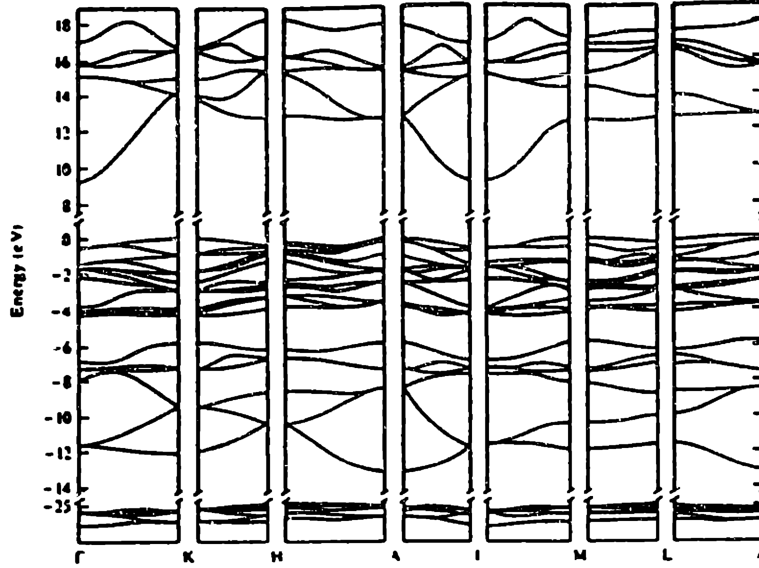

Figure 5-1: Energy band diagram of SiO_2.

$\vec{Q}_\perp(\vec{r})$ from (5.25) gives

$$Q_\perp(\vec{r}) = i\ Imag\left(\sqrt{-\frac{B}{2} - \frac{\sqrt{B^2 - 4C}}{2}}\right) = iK(\vec{r}) \tag{5.28}$$

where

$$B = 4m_o\left(e\phi_o + V_{ext}(\vec{r}) - E_\perp - \frac{E_g}{2}\right) - \left(\frac{1}{m^c} - \frac{1}{m^v}\right) m_o^2 E_g \tag{5.29}$$

$$C = 4m_o^2\left(e\phi_o + V_{ext}(\vec{r}) - E_\perp - E_g\right)\left(e\phi_o + V_{ext}(\vec{r}) - E_\perp\right) \tag{5.30}$$

Equation (5.28) gives the dispersion relation throughout the band gap of SiO_2 and is the central result of this section. From (5.4) we may write the approximate form of the wavefunction of a tunneling electron as

$$\phi(\vec{r}, t) = e^{-\frac{1}{\hbar}\int K(\vec{r})\, dr_\perp}\, e^{i\frac{1}{\hbar}\int \vec{Q}_{||}(\vec{r}).d\vec{r}_{||}} \sum_n c_n(t)\phi_{n,\vec{k}}(\vec{r}) \tag{5.31}$$

The above equation shows that transmission probability will be roughly equal to $e^{-2\frac{1}{\hbar}\int K(\vec{r})\, dr_\perp}$. Thus we have shown how the usual WKB approximation may be generalized to include the full mid-gap dispersion relation.

Figure (5-1) shows the energy band diagram of SiO_2 [38]. The conduction band minima occurs at the Γ point. The conduction band is isotropic at $\vec{k} = 0$ with an effective mass $m^c = 0.5m_o$ [38]. The valence band is almost flat at $\vec{k} = 0$ and the

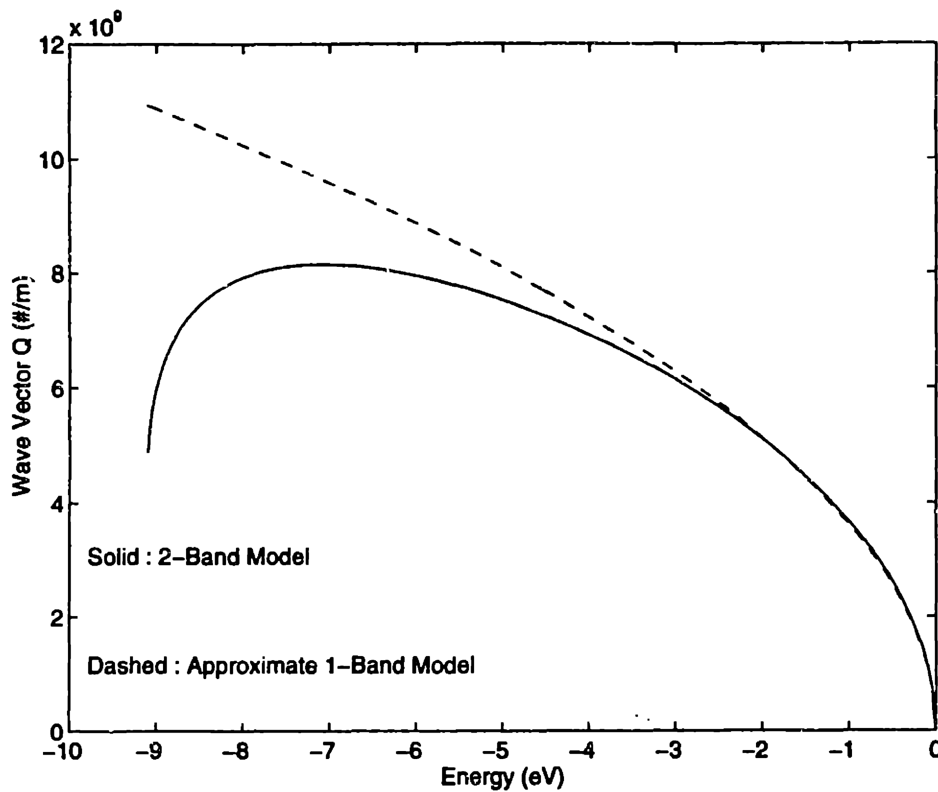

Figure 5-2: Energy dispersion relation in the band gap of SiO_2. The zero of energy is the conduction band of SiO_2.

expected valence band effective mass from the energy band diagram shown in figure (5-1) is $\sim 5m_o$. The condition (5.19) for the applicability of a two band model is approximately satisfied. In figure (5-2) we have plotted the dispersion relations given by the approximate expression (5.27) and by (5.28) for all values of $E_\perp$ that fall within the band gap of SiO_2 in the case where $V_{ext}(\vec{r}) = 0$. Note that in the figure the zero of $E_\perp$ is the conduction band of SiO_2. Interestingly, the approximate relation (5.27) gives almost the correct dispersion relation for all energies lying in the band gap of SiO_2 below its conduction band till about 3eV. This is due to the rather flat nature of the SiO_2 valence band and also due to the large band gap of SiO_2. Beyond 3eV, the more accurate expression (5.28) must be used. Since conduction band of Si is lower than the conduction band of SiO_2 by about 3.15eV, the approximate relation (5.27) will always hold for electrons tunneling from the conduction band of Si. Thus the assumption made in chapter two and three that the mid gap dispersion in SiO_2 may be described by a single effective mass of $0.5m_o$ is justified, and dispersion relation

(5.27) may be used when tunneling electrons originate from the conduction band of Si. However, as is clear from figure (5-2), hole tunneling currents through SiO_2 cannot be described by the dispersion relation (5.27), and expression (5.28) must be used instead.

5.3 The Dynamic Image Force Problem in Tunneling

5.3.1 Background

The problem of dynamic image potential experienced by tunneling electrons is rather an old one. A large number of papers have appeared in literature in the last few decades [39, 40, 41, 42, 43, 44, 45, 46, 47, 48] which deal with dynamical image potentials in case of metallic electrodes. In almost all these papers the relevant physics of the problem is cast in the form of interaction of tunneling electrons with the surface plasmon modes of the metallic electrodes. The tunneling electron polarizes these surface modes. This polarization produces a surface charge density on the metallic electrodes. This surface charge density in turn produces the image potential. The dynamics of the image potential then depends upon how quickly these surface modes respond to the potential produced by a moving electron. As may be expected, the ability of these surface plasmon modes to *track* the motion of a tunneling electron depends on the velocity components of the tunneling electron, both parallel and perpendicular to the surface of the metallic electrode. We will not describe the details of the physics here, but only mention a few relevant points. Interested reader is referred to reference [44] for details. The dispersion relation of surface modes (in case of single electrode only) is given approximately by the relation

$$\epsilon_m(\vec{k},\omega) + \epsilon_b(\vec{k},\omega) = 0 \tag{5.32}$$

where $\epsilon_m(\vec{k},\omega)$ and $\epsilon_b(\vec{k},\omega)$ are the dielectric constants of the metallic electrode and the region outside the electrode, respectively. If one is interested in the surface plasmon modes, then the metallic dielectric constant $\epsilon_m(\vec{k},\omega)$ may be approximated by

$$\epsilon_m(\vec{k},\omega) = \epsilon_m - \epsilon_o \frac{\omega_p^2}{\omega^2} \tag{5.33}$$

where the bulk plasma frequency ω_p is

$$\omega_p^2 = \frac{ne^2}{\epsilon_o m^*} \tag{5.34}$$

and ϵ_m is the contribution to the dielectric constant by the valence and core electrons. For an insulator surrounding the electrode with a constant dielectric constant ϵ_b, the surface plasmon dispersion relation becomes

$$\omega^2 = \frac{ne^2}{2(\epsilon_m + \epsilon_b)m^*} = \omega_s^2 \tag{5.35}$$

Above equation shows that the dispersion relation is $\vec{k}$-independent.

It has been shown in references [39, 40, 41, 42, 43, 44, 45, 46, 47, 48] that the surface plasmon modes can efficiently screen the potential of a tunneling electron and, therefore, provide dynamical image potential if the following conditions are satisfied :

1. The time taken by the tunneling electron to cross the barrier region is much larger than $\frac{1}{\omega_s}$. This is because the time taken by the surface plasmon modes to respond is of the order of $\frac{1}{\omega_s}$.
2. For the case of an electron moving with a velocity $v_{||}$ parallel to the surface of the electrode at a distance d from it, the surface modes can provide dynamical image potential provided

$$\frac{d}{v_{||}} << \frac{1}{\omega_s} \tag{5.36}$$

Since plasma frequency for most metals is fairly large, it is expected that most metals provide efficient dynamical screening for tunneling electrons.

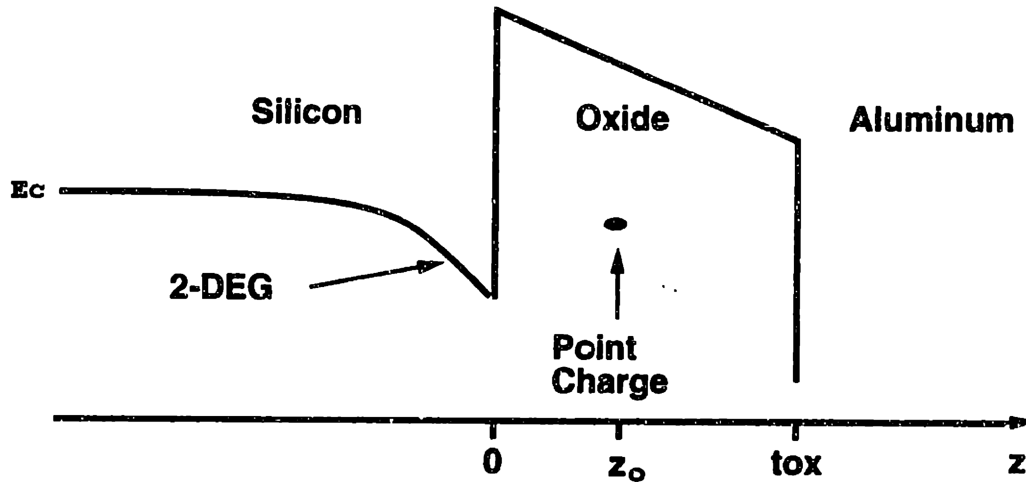

Figure 5-3: MOS structure with Aluminum gate.

However, in case of tunneling in MOS devices, where electrons tunnel from the accumulation or inversion layers, the situation is very different from that in case of metallic electrodes. The response of a two dimensional electron gas cannot be described in terms of surface plasmons. A new theoretical approach is required to study the dynamical image potential problem. In the next section we develop such an approach for the case of a MOS structure.

5.3.2 Dynamical Image Potential Problem in MOS Structures and the Dynamic Response of 2-DEG

Consider the MOS structure shown in figure (5-3). The electrons tunnel from a two dimensional electron gas (either accumulated or inverted layer of electrons), tunnel through the SiO_2 potential barrier, and enter the metallic gate electrode (which might be Aluminum). The response of each material present in the structure (i.e. Si, SiO_2 and Aluminum) to external potential disturbances can be described in terms of its dielectric response function. We discuss the case of Si first.

The full longitudinal dielectric function of bulk Si without local field corrections

can be written as [52]

$$\epsilon_{si}(\vec{k},\omega) = \epsilon_o - \frac{2e^2}{k^2V}\sum_{\substack{l,m\\ \vec{p},\sigma}}\Bigg(\left|< l;\vec{p}+\vec{k}+\vec{G}|e^{-i\vec{G}.\vec{r}}|m;\vec{p}>\right|^2 f_D(E_m(\vec{p}))\,\frac{E_l(\vec{p}+\vec{k}+\vec{G})-E_m(\vec{p})}{(\hbar\omega)^2-(E_l(\vec{p}+\vec{k}+\vec{G})-E_m(\vec{p}))^2}\Bigg) \quad (5.37)$$

where

$$< l;\vec{p}+\vec{k}+\vec{G}|e^{-i\vec{G}.\vec{r}}|m;\vec{p}>= \int \frac{d^3\vec{r}}{V}\, u^*_{l,\vec{p}+\vec{k}+\vec{G}}(\vec{r})e^{-i\vec{G}.\vec{r}}u_{m,\vec{p}}(\vec{r}) \quad (5.38)$$

and $\vec{p}$ is a vector in the first Brillouin zone. For a given $\vec{p}$ and $\vec{k}$, $\vec{G}$ is a unique reciprocal lattice vector such that $\vec{p}+\vec{k}+\vec{G}$ is in the first Brillouin zone. The summation in above expression is over all $\vec{p}$ in the first Brillouin zone, and over all energy bands labeled by the indices l and m, and over the two components of spin ($\sigma = \pm\frac{1}{2}$) of electrons. The summation over m is restricted to all filled bands (because of the Fermi-Dirac factor $f_D(E_m(\vec{p}))$. Suppose that the conduction band of Si and all bands above it are empty, and all bands below it are completely filled (as is the case in undoped Si at low temperatures). Using the f-sum rule for solids

$$\sum_{\substack{l,m\\ \vec{p},\sigma}}\Bigg(\left|< l;\vec{p}+\vec{k}+\vec{G}|e^{-i\vec{G}.\vec{r}}|m;\vec{p}>\right|^2 \Big(E_l(\vec{p}+\vec{k}+\vec{G})-E_m(\vec{p})\Big) f_D(E_m(\vec{p}))\Bigg) = \frac{\hbar^2k^2}{2m_o}\sum_{\substack{m\\ \vec{p},\sigma}} f_D(E_m(\vec{p})) \quad (5.39)$$

we get for the dielectric function

$$\epsilon_{si}(\omega) \approx \epsilon_o - \frac{e^2}{mV}\sum_{\substack{m\\ \vec{p},\sigma}} \frac{f_D(E_m(\vec{p}))}{(\hbar\omega)^2-Eg_m^2} \quad (5.40)$$

where Eg_m is the average gap between the m'th filled band and the conduction band. The above expression has contributions from the core electrons and the valence electrons. If we are interested in contribution to $\epsilon_{si}(\omega)$ just from the valence electrons

we can restrict the summation over m in (5.40) to valence band only which gives

$$\epsilon_{si}(\omega) = \epsilon_o \left(1 - \frac{(\hbar\omega_v)^2}{(\hbar\omega)^2 - Eg^2}\right) \tag{5.41}$$

where ω_v is given by

$$\omega_v = \sqrt{\frac{n_v e^2}{\epsilon_o m}} \tag{5.42}$$

E_g is the band gap of Si, and n_v is the number density of valence electrons. From the form of the dielectric function in (5.41) we can extract the following information

1. For $\omega << E_g$, $\epsilon_{si}(\omega)$ is independent of frequency.

2. We can find the time during which the valence electrons provide screening by finding the frequency of the interband plasma modes by solving the equation $\epsilon_{si}(\omega_p) = 0$. This gives

$$\omega_p^2 = (\frac{Eg}{\hbar})^2 + \omega_v^2 \tag{5.43}$$

For semiconductors like Si and Ge, ω_p^2 is extremely large (around $\sim 10^{16}$ /sec). Such a large plasma frequency means that valence electrons can screen time varying external potentials extremely quickly. Thus for all practical purposes, the expression (5.41) can be replaced by its static limit. One can carry out a similar analysis by keeping contributions from the core electrons as well. The final result is that screening provided by core electrons and valence electrons as a result of interband transitions is extremely fast, and one may use the static limit of (5.40) which is

$$\epsilon_{si} \approx \epsilon_o + \frac{e^2}{mV} \sum_{\substack{m \\ \vec{p},\sigma}} \frac{f_D(E_m(\vec{p}))}{Eg_m^2} \tag{5.44}$$

The above analysis was done to carefully separate out the contributions of valence and core electrons to the dielectric response function from those of the conduction electrons. The value of $11.7\epsilon_o$ for ϵ_{si} usually quoted in literature has contributions only from the valence and core electrons. Now we will discuss the contributions from the conduction electrons.

The conduction electrons in accumulation or inversion layer form a two dimensional electron gas. The dielectric response function of a two dimensional electron gas embedded in a medium of relative dielectric constant of unity in the plasmon pole approximation is [49]

$$\epsilon_{si}(\vec{k},\omega) = \epsilon_o \left(\frac{\omega^2 - \omega_p^2(k)(1 + \frac{k}{\kappa})}{\omega^2 - \omega_p^2(k)\frac{k}{\kappa}} \right) \tag{5.45}$$

where

$$\omega_p^2(k) = \frac{ne^2 k}{2\epsilon_o m^*} \tag{5.46}$$

and

$$\kappa = \frac{2\pi\hbar^2\epsilon_o}{e^2 m^*} \tag{5.47}$$

The dispersion relation of plasmons can be found as before by putting $\epsilon(\vec{k},\omega) = 0$, which gives

$$\omega^2(\vec{k}) = \frac{ne^2 k}{2\epsilon_o m^*}(1 + \frac{k}{\kappa}) = \omega_p^2(k)(1 + \frac{k}{\kappa}) \tag{5.48}$$

Notice the important difference between the dispersion relation of surface plasmons in case of metallic electrodes and plasmons in 2-DEG (compare (5.35) with (5.48)). Whereas ω is $\vec{k}$-independent in case of surface plasmons, $\omega(\vec{k})$ goes to zero as the square root of k for plasmons in 2-DEG. This has drastic consequences. If we introduce scattering in our model the plasmons will have a finite lifetime. The plasmons with small wavevector will have a lifetime shorter than their frequency. In other words, they will not exist as elementary excitations anymore. Their weight in the excitation spectrum will be *washed* out. Thus, it is not possible to develop a theory of dynamical image potential in case of an electron interacting with a 2-DEG by reducing the interaction to that between the electron and plasmons, since any realistic system will have finite scattering. For a 3-d electron gas plasmon modes near $\vec{k} = 0$ are particularly important in providing screening [28]. For a 2-d electron gas plasmon modes for small $\vec{k}$ disappear, and are replaced by the slow diffusive modes which provide the screening [50, 51].

It can easily be shown that in the presence of scattering the dielectric response function of a 2-DEG in the plasmon pole approximation becomes

$$\epsilon(\vec{k},\omega) = \epsilon_o \left(\frac{\omega(\omega + \frac{i}{\tau}) - \omega_p^2(k)(1 + \frac{k}{\kappa})}{\omega(\omega + \frac{i}{\tau}) - \omega_p^2(k)\frac{k}{\kappa}} \right) \tag{5.49}$$

Here τ is a phenomenological relaxation rate. The above expression satisfies all the sum rules [28] and therefore conserves particle number. From (5.49) we can find the susceptibility $\chi(\vec{k},\omega)$ defined as

$$\epsilon(\vec{k},\omega) = \epsilon_o - \frac{e^2\chi(\vec{k},\omega)}{2k} \tag{5.50}$$

which gives

$$\chi(\vec{k},\omega) = \frac{\frac{nk^2}{m^*}}{\omega(\omega + \frac{i}{\tau}) - \frac{n\pi\hbar^2k^2}{(m^*)^2}} \tag{5.51}$$

It can easily be shown [14] that charge density $\rho(\vec{k},\omega)$ induced in a 2-DEG in response to a net poential $\phi(\vec{k},\omega)$ is

$$\rho(\vec{k},\omega) = e^2\chi(\vec{k},\omega)\phi(\vec{k},\omega) \tag{5.52}$$

Equations (5.52) and (5.51) can be used to study the dynamical response of 2-DEG to external perturbations. Specializing (5.51) to the case of 2-DEG in Si becomes a little troublesome since the effective mass in Si is not isotropic. The result is that 2-d plasmons in Si have different frequencies depending upon the direction of travel of the plasmons. However, at low temperatures almost all the electrons occupy the two carrier pockets in the first Brillouin zone in which the effective mass in direction perpendicular to the Si/SiO$_2$ interface is m_l^{si}. Therefore, for these electrons effective mass in all directons parallel to the interface is m_t^{si}. Thus, in this special case all we need to do is replace m^* in (5.51) with m_t^{si}. In what follows, we shall restrict ourselves to this simple case.

So far we have described how to model the dielectric response of core, valence and conduction electrons in Si. In SiO_2 the response of core and valence electrons can be modeled by the static dielectric constant ϵ_{ox}. Finally, the screening properties of the Aluminum gate electrode can be described in terms of surface plasmons. Since these surface plasmons are expected to provide efficient dynamical screening, we may assume that the gate electrode acts as a perfect metal.

Below we will study the dynamic response of a 2-DEG embedded in the geometry shown in figure (5-3). Suppose that in the structure shown in figure (5-3) we place a point charge inside the oxide at a distance z_o from the 2-DEG, and switch it *on* at time $t = 0$. We would like to see how fast the charge gets screened and how its image potential evolves in time. Since the problem has cylindrical symmetry, it is useful to work with quantities that have are fourier transformed with respect to the co-ordinate variables parallel to the Si/SiO_2 interface and also with respect to time. We can write the potentials in Si and in oxide as follows

$$\phi_{ox}(\vec{k}, z, \omega) = \phi_{ext}(\vec{k}, z, \omega) + A(\vec{k}, \omega)e^{-kz} + B(\vec{k}, \omega)e^{kz} \tag{5.53}$$

$$\phi_{si}(\vec{k}, z, \omega) = C(\vec{k}, \omega)e^{kz} \tag{5.54}$$

where $\phi_{ext}(\vec{k}, z, \omega)$ is the potential perturbation produced by the external point charge. For a point charge located at z_o and switched *on* at time $t = 0$, $\phi_{ext}(\vec{k}, z, \omega)$ is

$$\phi_{ext}(\vec{k}, z, \omega) = \frac{i}{\omega + i\eta} \frac{e}{2\epsilon_{ox}k} e^{-k|z - z_o|} \tag{5.55}$$

Other useful situations may also be realized in this formalism. For example, for a point charge moving with a velocity v_x parallel to the interface in the x-direction at a distance z_o from it, we get

$$\phi_{ext}(\vec{ext}, z, \omega) = 2\pi\delta(\omega - k_x v_x) \frac{e}{2\epsilon_{ox}k} e^{-k|z - z_o|} \tag{5.56}$$

The solution consists of finding the coefficients A,B and C. Using the following bound-

ary conditions

$$\phi_{ox}(\vec{k}, z = t_{ox}, \omega) = 0 \quad \text{and} \quad \phi_{ox}(\vec{k}, z = 0, \omega) = \phi_{si}(\vec{k}, z = 0, \omega) \tag{5.57}$$

$$\epsilon_{ox}\frac{\partial\phi_{ox}(\vec{k}, z = 0, \omega)}{\partial z} - \epsilon_{si}\frac{\partial\phi_{si}(\vec{k}, z = 0, \omega)}{\partial z} = -\rho(\vec{k}, \omega) = -e^2\chi(\vec{k}, \omega)\phi_{ox}(\vec{k}, z = 0, \omega) \tag{5.58}$$

we find

$$\rho(\vec{k}, \omega) = ef(\vec{k}, \omega)\frac{e^2\chi(\vec{k}, \omega)}{k}\frac{\left(e^{-kz_o} - e^{-(2kt_{ox} - z_o)}\right)}{\left(\epsilon_{ox}(1 + e^{-2kt_{ox}}) + \epsilon_{si}(1 - e^{-2kt_{ox}}) - \frac{e^2\chi(\vec{k}, \omega)}{k}(1 - e^{-2kt_{ox}})\right)} \tag{5.59}$$

where

$$\begin{aligned} f(\vec{k}, \omega) &= \frac{i}{\omega + i\eta} \quad \text{for the stationary external point charge} \\ &= 2\pi\delta(\omega - k_x v_x) \quad \text{for the external point charge} \\ &\quad\ \text{moving with velocity } v_x \end{aligned} \tag{5.60}$$

Equation (5.59) can be inverse fourier transformed to get $\rho(\vec{r}, t)$ which will give us the time and space dependent charge density response of the 2-DEG. Also $\rho(\vec{k}, t)$ may be used to find the coefficients A and B

$$B(\vec{k}, t) = -\frac{\left(\frac{\rho(\vec{k}, t)}{k} + (\epsilon_{si}\sinh kz_o + \epsilon_{ox}\cosh kz_{ox})\frac{ef(\vec{k}, t)}{\epsilon_{ox}k}\right)}{\left(\epsilon_{ox}(1 + e^{-2kt_{ox}}) + \epsilon_{si}(1 - e^{-2kt_{ox}})\right)}e^{-2kt_{ox}} \tag{5.61}$$

and

$$A(\vec{k}, t) = -B(\vec{k}, t)e^{2kt_{ox}} - \frac{ef(\vec{k}, t)}{\epsilon_{ox}k}e^{kz_o} \tag{5.62}$$

where

$$\begin{aligned} f(\vec{k}, t) &= \theta(t) \quad \text{for the stationary external point charge} \\ &= e^{-ik_x v_x t} \quad \text{for the external point charge} \end{aligned}$$

$$\text{moving with velocity } v_x \tag{5.63}$$

Equ ons (5.59), (5.61) and (5.62) constitute a complete solution for the dynamic image potential experienced by the external point charge. Results can be obtained by numerical integration of these equations. However, we present some physical insight into the nature of the problem below.

Consider first the case where the oxide is extremely thick, so that the gate metal electrode is very far away from the external point charge. In this case the gate will provide no screening. Also suppose that there are no conduction electrons in Si. In this simple case screening will be instantaneous and the image potential experienced by the charge at a distance z_o from the Si/SiO_2 interface will be

$$\phi_{image}(t) = -\left(\frac{\epsilon_{si} - \epsilon_{ox}}{\epsilon_{si} + \epsilon_{ox}}\right)\frac{e}{8\pi\epsilon_{ox}z_o} \tag{5.64}$$

The expression in the paranthesis is about 0.5. Now suppose the gate electrode also provides screening but still there are no conduction electrons in Si. The image potential is again instantaneous and is approximately given by

$$\phi_{image}(t) \approx -\left(\frac{\epsilon_{si} - \epsilon_{ox}}{\epsilon_{si} + \epsilon_{ox}}\right)\frac{e}{8\pi\epsilon_{ox}z_o} - \frac{e}{8\pi\epsilon_{ox}(t_{ox} - z_o)} \tag{5.65}$$

The exact magnitude of image potential will be a little less than that given by the above equation since the image charges in Si will induce counter charges in the gate electrode and vice versa. The minimum value of image potential predicted by (5.65) will occur at $z_o = \frac{t_{ox}}{\sqrt{2}+1}$. We can use this to calculate the apparent reduction in peak barrier height $\Delta\phi_o$

$$\Delta\phi_o = \frac{e}{8\pi\epsilon_{ox}t_{ox}}\left(\frac{3}{2} + \sqrt{2}\right) \tag{5.66}$$

For a $30\mathring{A}$ oxide this comes out to be about 0.17 Volts, which is a pretty big number This shows that a substantial amount of screening is provided just by the gate electrode and the mismatch between the dielectric constants ϵ_{si} and ϵ_{ox}. The conduction electrons in the accumulation or inversion layer in Si will provide additional screening,

whose dynamic nature will now be discussed below.

Again suppose that the gate electrode is at infinity and provides no screening. In this case (5.59) may be reduced to

$$\rho(\vec{k},\omega) = ef(\vec{k},\omega)\frac{e^2\chi(\vec{k},\omega)}{k}\frac{e^{-kz_o}}{\left(\epsilon_{ox}+\epsilon_{si}-\frac{e^2\chi(\vec{k},\omega)}{k}\right)} \tag{5.67}$$

Defining an average dielectric constant ϵ_{avg} as

$$\epsilon_{avg} = \frac{\epsilon_{ox}+\epsilon_{si}}{2} \tag{5.68}$$

(5.67) becomes

$$\rho(\vec{k},\omega) = ef(\vec{k},\omega)\frac{\omega_p^2(k)e^{-kz_o}}{\omega(\omega+\frac{i}{\tau})-\omega_p^2(k)(1+\frac{k}{\kappa})} \tag{5.69}$$

where, in comparison to (5.46) and (5.47), $\omega_p(k)$ and κ are

$$\omega_p^2(k) = \frac{ne^2k}{2\epsilon_{avg}m_t^{si}} \tag{5.70}$$

and

$$\kappa = \frac{2\pi\hbar^2\epsilon_{avg}}{e^2m_t^{si}} \tag{5.71}$$

Finally, $\rho(\vec{r},t)$ can be found by inverse fourier transformation which gives

$$\rho(\vec{r},t) = \int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\int\frac{d^2\vec{k}}{(2\pi)^2}\,\rho(\vec{k},\omega)\,e^{i\vec{k}.\vec{r}}e^{-i\omega t} \tag{5.72}$$

First consider the case of an external point charge moving with a velocity v_x parallel to the Si/SiO$_2$ interface. In that case, as shown before, $f(\vec{k},\omega)$, is

$$f(\vec{k},\omega) = 2\pi\delta(\omega-k_xv_x) \tag{5.73}$$

Substituting $f(\vec{k},\omega)$ from above in (5.72) and performing integration over ω we get

$$\rho(\vec{r},t) = \int \frac{d^2\vec{k}}{(2\pi)^2}\, e\, \frac{\omega_p^2(k)e^{-kz_o}e^{i\vec{k}.\vec{r}}e^{-ik_xv_xt}}{k_xv_x(k_xv_x+\frac{i}{\tau}) - \omega_p^2(k)(1+\frac{k}{\kappa})} \tag{5.74}$$

As v_x approaches zero we get the same results as in the case of static screening. If we let v_x increase then the dymnamic screening will deteriorate when the terms in the denominator of the integrand in (5.74) containing v_x become comparable to $\omega_p^2(k)$. The integral has an upper cut off in $\vec{k}$-space at $\approx \frac{1}{z_o}$. So that k_x, k_y and k can have maximum values of $\approx \frac{1}{z_o}$. Using this fact, and assuming that $\frac{v_x}{z_o} << \frac{1}{\tau}$, we can write the condition for the deterioration of dynamic screening as

$$\frac{v_x}{z_o} \approx \omega_p(k = \frac{1}{z_o}) \tag{5.75}$$

The physical interpretation of the above result is simple. The external potential of the point charge disturbs the 2-DEG at a maximum wavevector of $k_{max} = \frac{1}{z_o}$. The plasma modes at that wavevector have a frequancy of $\omega_p(k = \frac{1}{z_o})$. So that the time during which screening occurs is of the order of $\frac{1}{\omega_p(k=\frac{1}{z_o})}$. The radius of the induced charge density is roughly z_o. If the charge has enough parallel velocity to cross a region of this size in time much less than $\frac{1}{\omega_p(k=\frac{1}{z_o})}$ then screening will depreciate. This is exactly what is expressed by the relation (5.75). Notice the strong similarity between this relation and (5.36) for the case of screening by surface plasmons.

On the other hand if $\frac{v_x}{z_o} >> \frac{1}{\tau}$, then screening will start weakening when

$$\frac{v_x}{z_o} \approx \omega_p^2(k = \frac{1}{z_o})\tau \tag{5.76}$$

In this case we see that the time during which screening occurs is of the order of $\frac{1}{\omega_p^2(k=\frac{1}{z_o})\tau}$.

For the case when a stationary external point charge is simply switched on at time

$t = 0$, approximate analysis, similar to what was done above, yields the following two conclusions :

1. When $\frac{1}{\tau} << \omega_p(k = \frac{1}{z_o})$ screening time is of the order of $\frac{1}{\omega_p(k = \frac{1}{z_o})}$.

2. When $\frac{1}{\tau} >> \omega_p(k = \frac{1}{z_o})$ screening time is of the order of $\frac{1}{\omega_p^2(k = \frac{1}{z_o})\tau}$.

Here we also present the results of numerical simulations showing the response of a 2-DEG to a stationary external point charge placed inside the oxide between the Si substrate and the gate electrode and switched *on* at time $t = 0$. We have inverse fourier transformed (5.59) numerically. The results in figure (5-4) are shown for the case when $z_o = 12.5\mathring{A}$, $t_{ox} = 30\mathring{A}$, density of the 2-DEG is $10^{12}/\text{cm}^2$, and $\tau = 10^{-13}$sec. From the figure it can be seen that, as expected, the radius of the screening charge is $\approx z_o = 12.5\mathring{A}$, and the time taken by the 2-DEG to screen the external charge is $\approx \frac{1}{\omega_p(k = \frac{1}{z_o})} = .01 \times 10^{-12}$sec. Note that the peak of the screening density is larger than the number density assumed for the 2-DEG. Since the screening charge is positive, this means the total charge density of the 2-DEG is becoming positive. This is impossible. What is happening is that our linear theory is breaking down, and the external point charge is strong enough to produce a completely depleted region underneath it. Another interesting thing to notice that is that the screening charge overshoots in the beginning and the extra charge later disperses away in the form of cylinderical charge density waves. Figure (5-5) shows the image potential felt by the external charge as a function of time. As expected, the image potential starts off from about 0.14 Volts, which has no contribution from the conduction electrons, and as the conduction electron screening charge density builds up, the image potential also increases. The image potential also overshoots a little, and settles to its final value of about 0.165 Volts in about $.05 \times 10^{-12}$sec, which is about $\frac{5}{\omega_p(k = \frac{1}{z_o})}$. We may also mention here that value of $12.5\mathring{A}$ for z_o was chosen so that z_o is the point where the image potential is minimum, therefore the image potential at this point will allow us to calculate the peak barrier height.

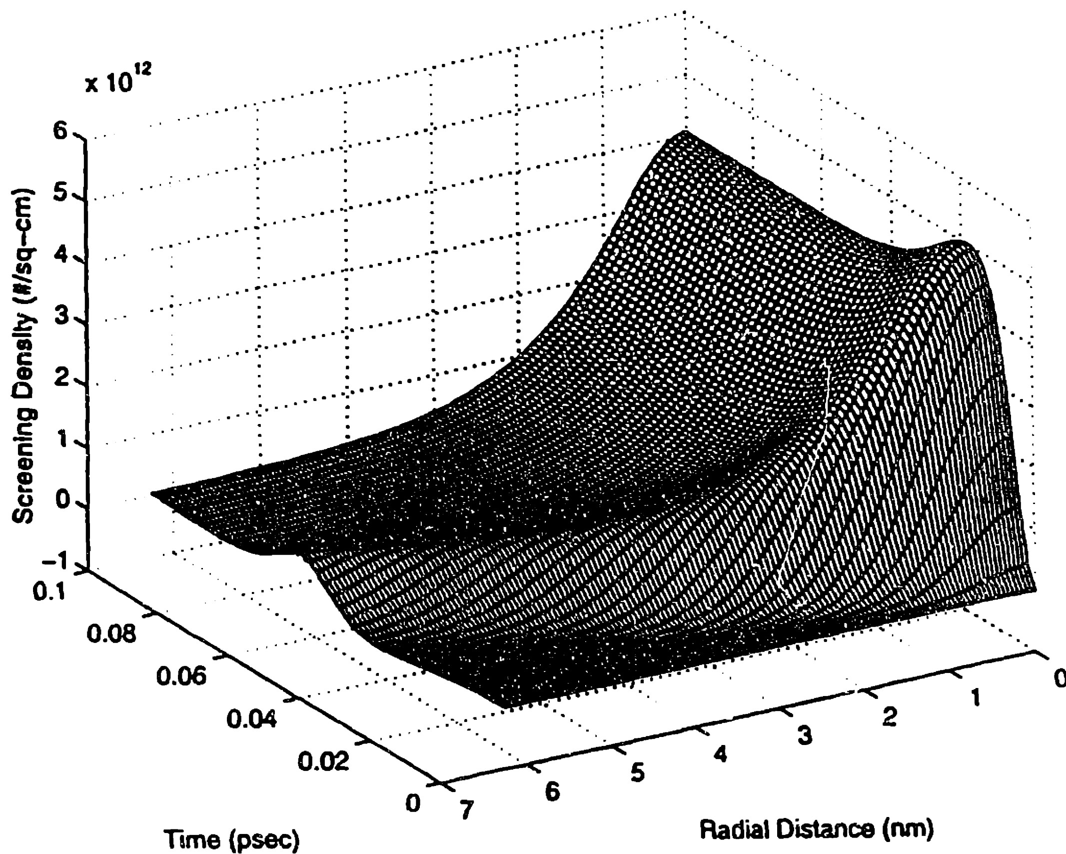

Figure 5-4: Dynamical screening charge density of a 2-DEG shown as a function of time and radial distance from the position of the external point charge. $z_o = 12.5\mathring{A}$ and $t_{ox} = 30\mathring{A}$

Figure (5-6) shows the time dependent image potential felt by the external charge in the case when $z_o = 6\mathring{A}$ and $t_{ox} = 15\mathring{A}$. Notice that because of the proximity of the gate electrode and the Si/SiO_2 interface to the charge, the image potential is much larger.

Having discussed the dynamic response of the 2-DEG to external perturbations, we move on to discuss the effect of image forces in tunneling. From the discussion above, it is obvious that screening of charges which move with a velocity parallel to the interface is a sufficiently complicated process and depends on the parallel velocity of the charge and also on its distance from the interface. The actual tunneling scenario is, however, more complicated. The tunneling electron starts off from the 2-DEG. When

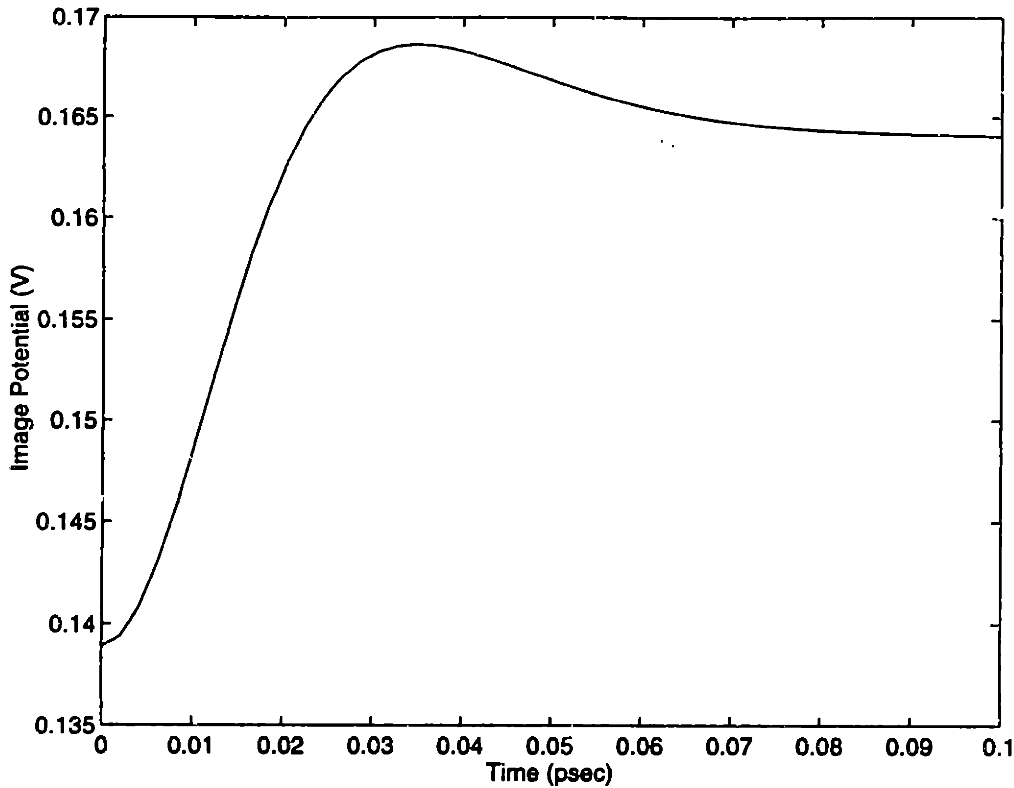

Figure 5-5: Dynamical image potential felt by an external point charge shown as a function of time. $z_o = 12.5\mathring{A}$ and $t_{ox} = 30\mathring{A}$

inside the 2-DEG it has a correlation hole around it (i.e. a region in which the electron density is less than the average density as a result of many-body exchange correlation effects [14]). As the electron leaves the 2-DEG and enters the oxide region, it leaves behind its exchange correlation hole (see [42] for details). The exchange correlation hole has a net positive charge density and it provides a non-local image potential. As the electron moves away from the Si/SiO_2 interface, the exchange correlation hole spreads out and provides the usual local image force. Thus, a complete theory of image potential must be able to provide a complete dynamical description of this process starting from when the electron resides inside the 2-DEG and till it reaches the gate electrode. In such a theory the component of the velocity of the tunneling electron perpendicular to the Si/SiO_2 interface will also play an important role.

A qualitative description of this process may be as follows. As the electron leaves the 2-DEG at the Si/SiO_2 interface, its correlation hole is able to follow its motion parallel to the interface since the electron, being very close to the 2-DEG, is

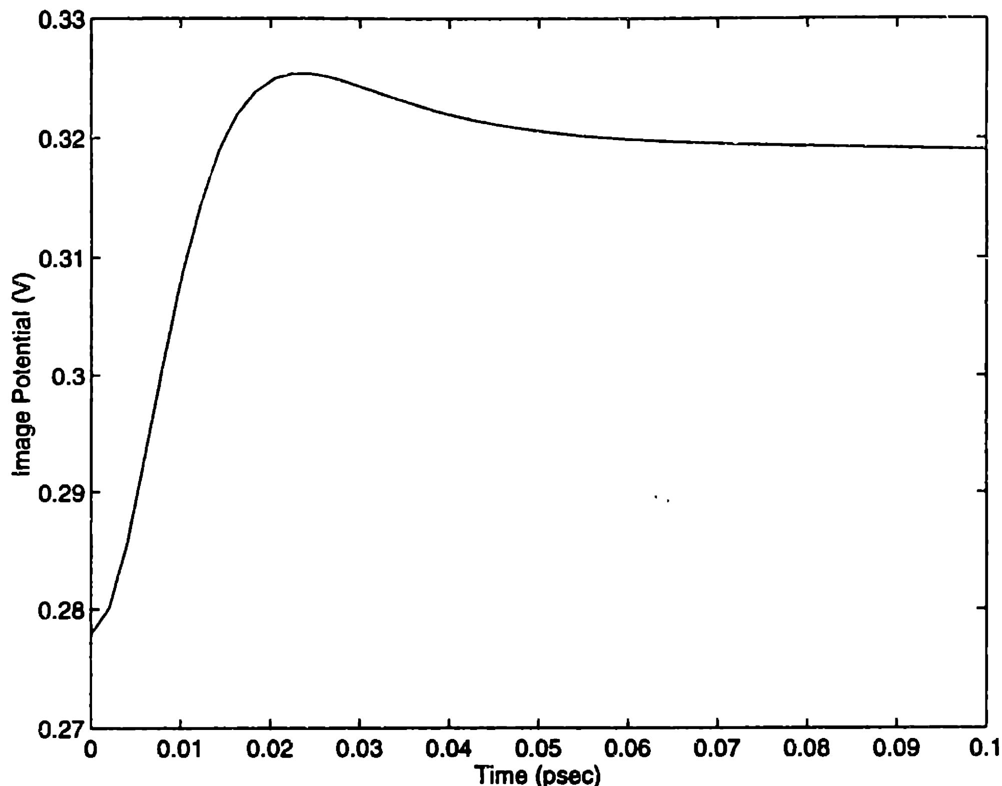

Figure 5-6: Dynamical image potential felt by an external point charge shown as a function of time. $z_o = 6\mathring{A}$ and $t_{ox} = 15\mathring{A}$

able to excite plasma modes of arbitrary high frequency. However, as the electron moves further away from the 2-DEG, the response of the 2-DEG becomes more and more sluggish. We expect the dynamic image potential to depend on the velocity components of the tunneling electron both parallel and perpendicular to the Si/SiO_2 interface. Tunneling currents measured experimentallly have contributions from electrons that have velocity components parallel to the interface ranging from zero to the fermi velocity. This poses a daunting challenge to anybody interested in calculating transmission probabilities since the transmission probability will then also become dependent on the energy of tunneling electrons parallel to the interface. However, things can be simplified. We have showed that for thin oxides almost all the image potential contribution comes from the gate electrode and the dielectric constant mismatch between Si and SiO_2. In the examples discussed above in which an external stationary point charged was placed in the oxide region, only about 18 percent of the total image potential was from the conduction electrons in case of a $30\mathring{A}$ oxide. In

case of a 15Å oxide this contribution was only about 12 percent. Thus, as a first approximation, the contribution of the conduction electrons can even be ignored. This simplifies things a lot.

Our numerical analysis, although far from providing a complete description of dynamic image potential in tunneling, had the following objectives

1. To check our qualitative estimate on the time required for the 2-DEG to respond to perturbations.

2. To provide a comparison between the magnitude of the image potential contributed by the conduction electrons in Si and that contributed by the gate electrode and the dielectric constant mismatch between Si and SiO_2.

We may also mention here that the formalism developed here is capable of handling the complete dynamic image potential problem as described in the paragraph above. In (5.59) if we let z_o go to zero and modify $f(\vec{k},\omega)$ as follows

$$f(\vec{k},\omega) = \left(\frac{i}{\omega - k_x v_x + ikv_z} - \frac{i}{\omega - k_x v_x + ikv_z}\right) \tag{5.77}$$

then these changes describe a situation where an electron is moving with a velocity v_x inside the 2-DEG for all negative times, and for positive times it develops an additional component of velocity v_z perpendicular to the Si/SiO_2 interface and shoots out from the 2-DEG into the oxide towards the gate electrode. We have not carried out numerical simulations with these changes in this thesis.

5.3.3 On the Neglect of Image Potential Corrections in Calculating Transmission Probabilities

Finally, we would like to provide justification for ignoring the contribution of image forces in our calculations of transmission probabilities in chapter two and three. This may perhaps seem a bit strange at first, given that image potential contribution to reduction in peak barrier height may be as large as 0.28 Volts for a 15Å oxide. However, our argument will be presented from an experimental measurement point

of view. Of course, we do not intend to argue that image force is negligible since this is certainly not the case, as already shown above.

The transmission probability for an electron, with energy E, through an oxide of thickness t_{ox} is roughly given by the expression

$$T(E) \approx e^{-2 \int_0^{t_{ox}} dx \frac{\sqrt{2m^{ox}(e\phi_o - eF_{ox}x - E)}}{\hbar}} \tag{5.78}$$

In case of thin oxides, where WKB approximation is quite adequate (since tunneling occurs in the direct regime and never in the Fowler-Nordheim regime), (5.78) holds approximately for even relatively large electric fields ($\sim 10^7$V/cm). In chapters two and three we had $m^{ox} = 0.5m_o$ and $\phi_o = 3.15$ Volts. Since bulk of the electrons lie in the two lowest states whose energies are roughly around 0.2eV, we will assume $E \approx 0.2$eV in (5.78). In actual experimental situations the oxide thickness is known with an uncertainty of about $1 - 2\mathring{A}$. The determination of oxide thicknesses by ellipsometery induces an error which is expected to be at least $1 - 2\mathring{A}$. Infact, it is not even possible to define even theoretically the exact location of the Si/SiO_2 interface with an accuracy greater than this [33].

The magnitude of experimentally measured tunneling currents give a rough estimate of the entire expression in the exponential but does not yield information about each parameter present in the expression. Suppose one is interested in finding out by making tunneling current measurements whether the image force correction to barrier height of 0.3 Volts is present for a $15\mathring{A}$ or not. To be able to extract such information from tunneling current measurements one must know apriori the magnitude of other parameters in the expression in the exponential in (5.78) to a sufficient accuracy. The error δt_{ox} in oxide thickness which will exactly produce the same results for tunneling currents as those produced by an image potential correction of $\delta\phi_o$ to the barrier height can be calculated from (5.78) and comes out to be

$$\delta t_{ox} = \frac{\delta\phi_o}{F_{ox}} \left(\frac{\sqrt{x_o} - \sqrt{x_o - t_{ox}}}{\sqrt{x_o - t_{ox}}} \right) \tag{5.79}$$

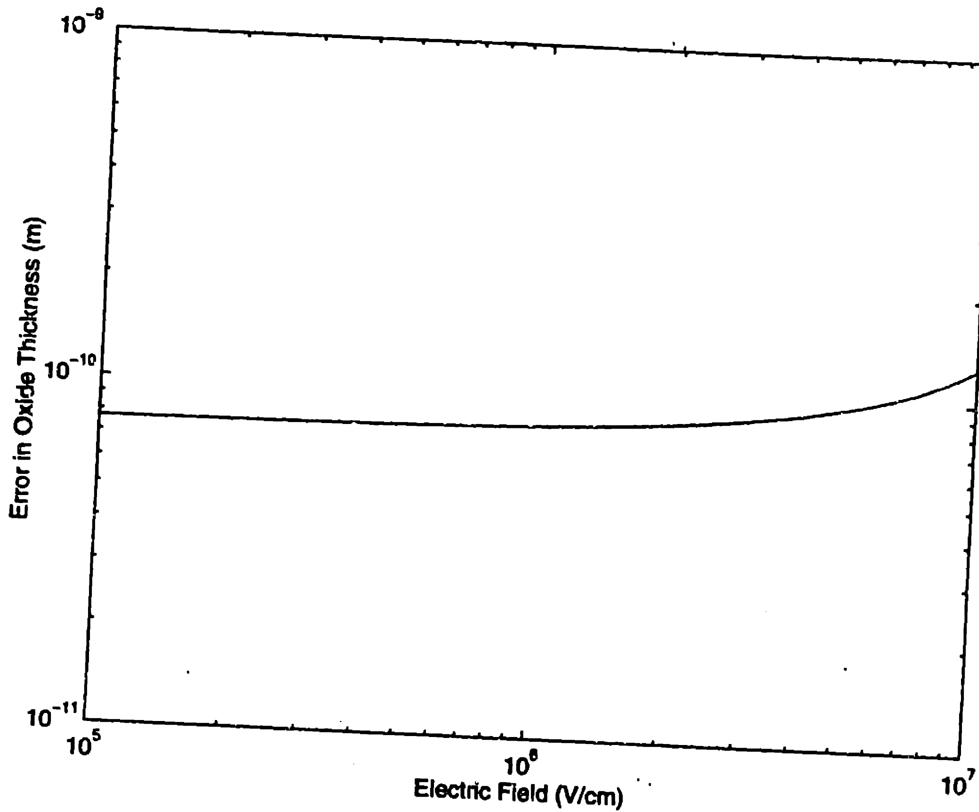

Figure 5-7: δ_{ox} needed to give the same results for tunneling currents as would a 0.3 Volts reduction in barrier height.

where

$$x_o = \frac{e\phi_o - E}{eF_{ox}} \tag{5.80}$$

Above equations imply that roughly

$$\frac{\delta t_{ox}}{t_{ox}} = \frac{\delta\phi_o}{2\phi_o} \tag{5.81}$$

Figure (5-7) shows the values of δt_{ox} needed to give the same results for tunneling currents as would a 0.3 Volts reduction in barrier height for different values of electric field strength. We see that δt_{ox} is much less than 1.5Å for all values of electric field upto 10^7V/cm. But 1.5Å is well within the uncertainty in the knowledge of oxide thickness. Also since δt_{ox} seems largely independent of electric field, we can safely conclude that tunneling current measurements in thin oxides cannot distinguish between effects produced by image potential or those that result from inaccurate knowledge of the oxide thickness. In all the calculations carried out in chapter two

and three we neglected image potential corrections. In comparing our calculations with experimental data we used oxide thicknesses obtained from ellipsometery. If it is the case that ellipsometery yields oxide thicknesses that are always in error by $0.8 - 1.5\mathring{A}$ then just by doing tunneling current measurements it is not possible to confirm experimentally the presence or absence of image potential corrections. We know from chapter four that our theory without image potential corrections agrees very well with experimental data provided we use oxide thicknesses that were determined using ellipsometery. Thus we have two options available to us - either adopt a theory that has image potential corrections and postulate that ellipsometery gives values of oxide thickness that are slightly smaller than actual thicknesses, or adopt a simple theory that neglects image potentials corrections but accepts oxide thicknesses measured using ellipsometery as correct. We have chosen the second option just because it keeps the theoretical model simple. We have also shown that image potential corrections can be fairly large. Clearly more work, both theoretical and experimental, is required here to understand the error induced in ellipsometeric measurements and this may perhaps tie the remaining loose ends in our modeling.

Chapter 6

Modeling of Electronic Processes in Quantum Dots Coupled to FET'S

6.1 Introduction

In chapter one of this thesis we introduced the idea of building few electron memory devices using quantum dots coupled to the channel of MOS FET's. Figures (6-1) and (6-2) show two possible ways of making such devices. Figure(6-1) shows nano crystals, either of Si or Ge, about 50$\mathring{A}$ in radius, embedded in the gate oxide of an MOS transistor. These nano crystals can be deposited by CVD and they subsequently nucleate on the surface of the wafer to form small clusters. Very crudely speaking, these nano crystals behave rather like the floating gate of E^{2}PROMS. They can store electrons and thereby change the threshold voltage of the MOS device. However, the small size of the nano crystals results in large coulomb energies. As a result it is expected that the mean number of electrons in these nano crystals will increase with the magnitude (and/or duration) of the write pulse in discrete steps. The shifts in threshold voltage of the MOS device is, therefore, also expected to go up in discrete steps with the magnitude (and/or duration) of the write pulse. Thus the device has

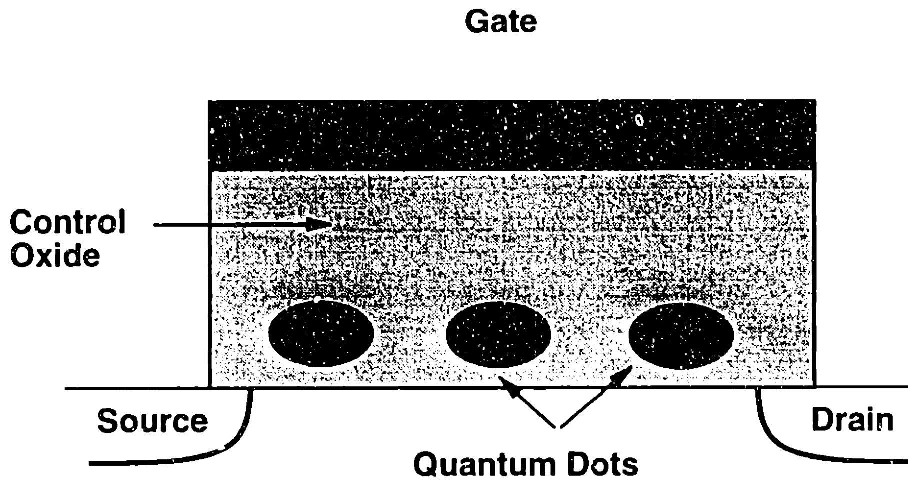

Figure 6-1: Quantum dot memory cell using Si nano crystals

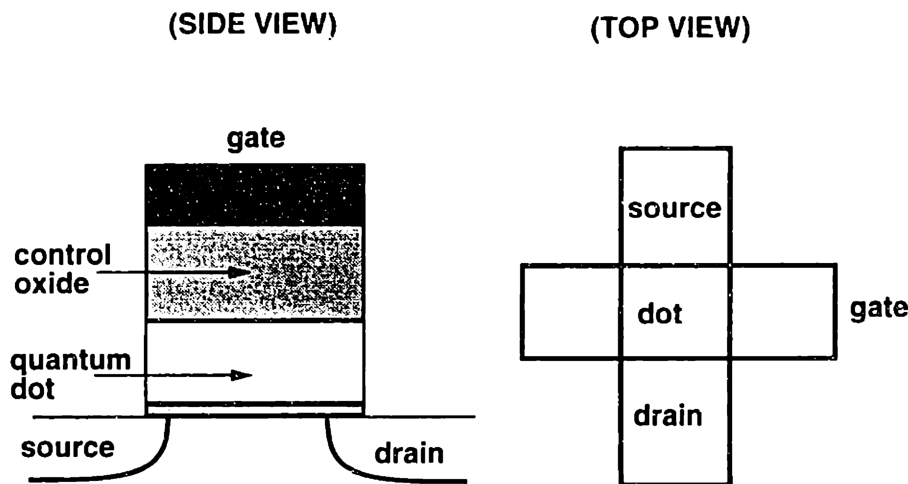

Figure 6-2: Single quantum dot memory cell

a potential for being used as a multistate memory cell. Figure (6-2) shows a single quantum dot coupled to the channel of a MOS device. Such a device can be made by lithography at IBM. Research is currently being done at IBM to make them smaller than 20×20 nm^2.

The purpose of this chapter is to lay down a theoretical framework which can describe electron tunneling processes, and charge statistics and fluctuations in devices containing quantum dots coupled to the channel of MOS devices. We will also be interested in the write times and erase times of these devices and the effect of charge fluctuations in the dots on the channel conductivity. The main object will be to

develop simple models that capture the essential physics and that may be incorporated into a device simulator like IBM's DAMOCLESE.

6.2 On the Nature of Coulomb Energy

Before describing the details of modeling electronic processes in quantum dot memories, we consider it important to spend sometime on the nature of coulomb energy in mesoscopic devices. Although this is an extremely common and well known concept, it is unfortunately poorly understood by many in the field. Consider a system of free electrons and metallic conductors. The total electrostatic energy E_s of the system is

$$E_s = \frac{\epsilon}{2} \int d^3\vec{r}\, \vec{E}(\vec{r}).\vec{E}(\vec{r}) \tag{6.1}$$

Since $\vec{E}(\vec{r}) = -\vec{\nabla}V(\vec{r})$ we may write

$$\begin{aligned} E_s &= -\frac{\epsilon}{2} \int d^3\vec{r}\, \vec{E}(\vec{r}).\vec{\nabla}_r V(\vec{r}) \\ &= -\frac{\epsilon}{2} \int d^3\vec{r}\, \vec{\nabla}.(V(\vec{r})\vec{E}(\vec{r})) \quad + \quad \frac{\epsilon}{2} \int d^3\vec{r}\, V(\vec{r})\vec{\nabla}.\vec{E}(\vec{r}) \\ &= \frac{1}{2}\sum_n V_n Q_n \quad + \quad \frac{1}{2}\int d^3\vec{r}\, V(\vec{r})\rho(\vec{r}) \end{aligned} \tag{6.2}$$

The sum on n is over all the conducting bodies with potentials V_n and total charges Q_n, and the in the last integral $\rho(\vec{r})$ is the volume charge density residing in the space between the metallic conductors. Note that $V(\vec{r})$ satisfies the Poisson equation

$$\nabla^2 V(\vec{r}) = -\frac{\rho(\vec{r})}{\epsilon} \tag{6.3}$$

with the boundary conditions such as to match the potentials V_n on the conductors. The last term in (6.2) is a little problematic. It contains electron self-energy contributions. To see that suppose $\rho(\vec{r}) = \sum_i e_i \delta^3(\vec{r} - \vec{r}_i)$, then assuming there are no

metallic conductors, we have

$$V(\vec{r}) = \frac{1}{4\pi\epsilon} \int d^3\vec{r} \sum_i \frac{e_i}{|\vec{r} - \vec{r}_i|} \tag{6.4}$$

Using above expression for $V(\vec{r})$ in (6.2) gives terms that are infinite. In practice, one is only interested in changes in the total electrostatic energy, so that self-energy terms cancel out. A better way is to write (6.2) as

$$E_s = \frac{1}{2}\sum_n V_n Q_n \quad + \quad \frac{1}{2}\sum_i \int d^3\vec{r}\, V_i(\vec{r})\rho_i(\vec{r}) \tag{6.5}$$

where $\rho_i(\vec{r}) = e_i\delta^3(\vec{r} - \vec{r}_i)$ is the charge density due to the i'th charge and we have defined $V_i(\vec{r})$ as the potential that satisfies the modified poisson equation

$$\nabla^2 V_i(\vec{r}) = -\sum_{j\neq i} \frac{\rho_j(\vec{r})}{\epsilon} \tag{6.6}$$

and satisfies the same boundary conditions as $V(\vec{r})$. By defining $V_i(\vec{r})$ in this way we have eliminated the unphysical self-energy contributions. It is very important to mention here that a charge is only affected by the potential produced by other charges, the image charges on the electrodes of other charges, and its own image charges. Its not affected by the potential it produces itself, and all we have done above is taken this unphysical contribution out from (6.2). We may also mention here that the usual numerical methods to perform self-consistent Hartree calculations (e.g. IBM's SCRAP), which are accurate for structures containing a large number of electrons, will not be accurate for strutures containing a small number of electrons (two or three e.t.c.) since these methods do not take into account the fact that a charge is not affected by its own potential.

If we write $V(\vec{r})$ as

$$V(\vec{r}) = V_{int}(\vec{r}) + V_{ext}(\vec{r}) \tag{6.7}$$

where $V_{int}(\vec{r})$ is the part contributed by $\rho(\vec{r})$ and is given exactly as in (6.4), and $V_{ext}(\vec{r})$ is contributed by charges on the metallic conductors, then it follows from

(6.2) that

$$E_s = (\text{self} - \text{energy terms }) \ + \ \frac{1}{2}\sum_{i<j}\frac{e_i e_j}{4\pi\epsilon|\vec{r_i} - \vec{r_j}|} \ + \ \frac{1}{2}\sum_i e_i V_{ext}(\vec{r_i}) \qquad (6.8)$$

The above equation shows the relationship between the electrostatic interaction between charges and the total coulomb energy of the system. Of course, $V_{ext}(\vec{r})$ will be affected by the charges e_i since it will have contribution from the image charges of all the charges e_i of the system. This makes (6.8) difficult to use in calculations. However, equation (6.2) is ideal for performing numerical calculations. In case there are no metallic conductors in the system (e.g. this will be the case if the gate electrodes in devices shown in figures (6-1) and (6-2) are made of lightly doped Poly-Si and are therefore not treated as perfect metals) the change in electrostatic energy δE_s following any event which causes electrostatic potential and charge density to change from $V^i(\vec{r})$ and $\rho^i(\vec{r})$ to $V^f(\vec{r})$ and $\rho^f(\vec{r})$, respectively, is simply

$$\delta E_s = \frac{1}{2}\int d^3\vec{r}\, V^f(\vec{r})\rho^f(\vec{r}) \ - \ \frac{1}{2}\int d^3\vec{r}\, V^i(\vec{r})\rho^i(\vec{r})) \qquad (6.9)$$

Above equation is suitable for implementation in a numerical simulator like IBM's DAMOCLES. In the presence of metallic electrodes the additional term

$$\frac{1}{2}\sum_n V_n^f Q_n^f \ - \ \frac{1}{2}\sum_n V_n^i Q_n^i \qquad (6.10)$$

must be added to the left hand side of (6.9).

6.3 Quantum Kinetic Equations For Modelling Tunneling Processes in a Quantum Dot Coupled to an Inversion Layer

In this section we will drive a set of quantum kinetic equations in the quantum Markoff approximation [53] which will enable us to model the charging and discharging

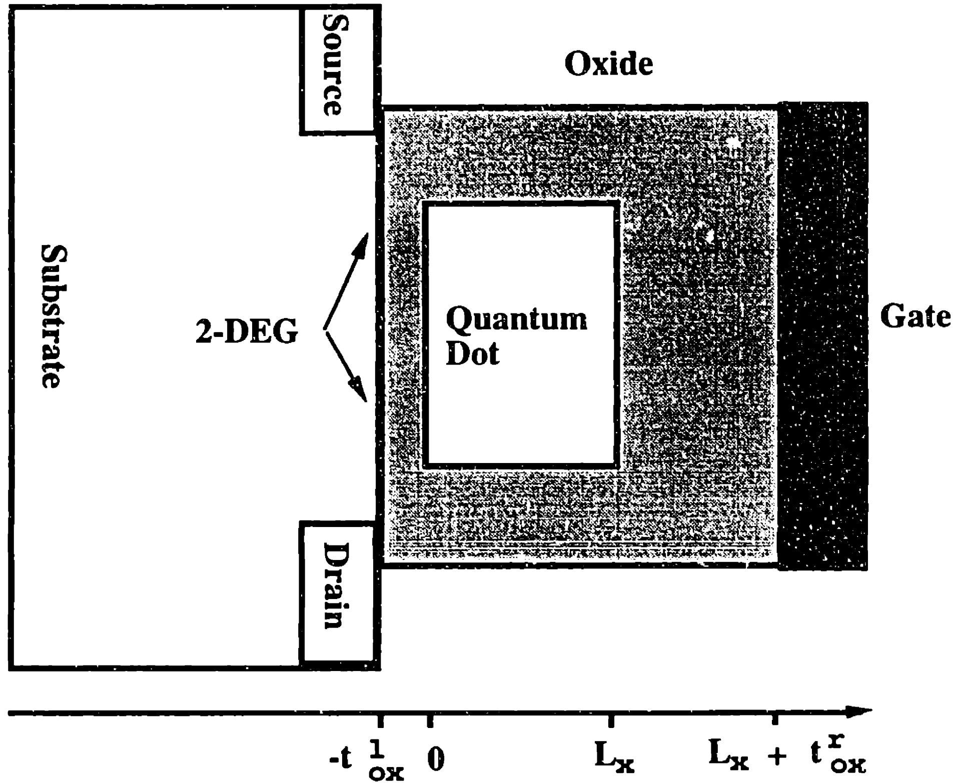

Figure 6-3: Quantum dot coupled to a 2-DEG

processes and carrier statistics and fluctuations in quantum dots which are coupled to a 2-D electron gas.

Consider the structure shown in figure (6-3). Suppose an electron makes a transition from an energy level ϵ_n in the 2-DEG to a level ϵ_m in the dot, increasing the number of electrons inside the dot by one. During this transition the voltage source does some work, δW. Let the change in electrostatic energy of the system (which may be calculated from any one of the formulas given in the previous section) following this transition be δE_s. Since the work done by the voltage source will result only in increasing the energy of electrons (assuming there is no dissipation) we may write the condition of conservation of energy as

$$\{\text{initial energy of system}\} + \{\text{work done by bettery }\} = \{\text{final energy of system }\} \tag{6.11}$$

Assuming that energies of all other electrons do not change before and after this

transition, we may write

$$\epsilon_n + \delta W = \epsilon_m + \delta E_s \tag{6.12}$$

It can easily be shown by simple calculations that work done by the battery is simply the potential difference between the dot and the 2-DEG before the transition took place and assuming there were no other electrons in the dot prior to this transition. δE_s is, however, more difficult to calculate. If we assume a simple two capacitor model for our structure, then using the formulas given in the previous section, it can be shown that E_s depends only on the total number of electrons in the dot and is given by the expression

$$E_s = \frac{(Ne)^2}{2C} \tag{6.13}$$

Where

$$\frac{1}{C} = \frac{1}{C_l} + \frac{1}{C_r} \tag{6.14}$$

and

$$C_l \sim \frac{\epsilon_{ox} A}{(t^l_{ox} + \frac{d}{2}\frac{\epsilon_{ox}}{\epsilon_{si}})} \tag{6.15}$$

$$C_r \sim \frac{\epsilon_{ox} A}{(t^r_{ox} + \frac{d}{2}\frac{\epsilon_{ox}}{\epsilon_{si}})} \tag{6.16}$$

δE_s, for the case when the number of electrons inside the dot increases from N by one, becomes

$$\delta E_s = \frac{Ne^2}{C} + \frac{e^2}{2C} \tag{6.17}$$

We expect that the above model for electrostaic energy is a little too simplified for the devices under consideration and more careful calculations using equations given in the previous secton are needed to get more accurate results.

The hamiltonian of the entire system can be written as

$$H = H_{2-DEG} + H_{DOT} + H_T \tag{6.18}$$

where

$$H_{2-DEG} = \sum_n (\epsilon_n + eV) a_n^\dagger a_n \tag{6.19}$$

$$H_{DOT} = \sum_m \epsilon_m b_m^\dagger b_m + E_s(N) \tag{6.20}$$

$$H_T = \sum_{n,m} T_{nm} a_n^\dagger b_m \quad + \quad c.c \tag{6.21}$$

V is the potential difference between the 2-DEG and the dot, assuming there are no electrons in the dot. For simplicity, we have assumed that E_s depends only on the total number of electrons N inside the dot.

Let at time t_o the system be described by a density matrix $\hat{P}$. We wish to study its time development for times greater than t_o. In the Heisenberg representation we have the equation of motion,

$$i\hbar \frac{\partial \hat{P}_H(t)}{\partial t} = [H, \hat{P}_H(t)] \tag{6.22}$$

We may write $H = H_o + H_T$, where $H_o = H_{2-DEG} + H_{DOT}$. If we define a density matrix $\hat{P}_I(t)$ in the interaction representation as

$$\hat{P}_I(t) = e^{\frac{i}{\hbar}\int_{t_o}^{t} H_o\, dt'} \hat{P}_H(t) e^{-\frac{i}{\hbar}\int_{t_o}^{t} H_o\, dt'} \tag{6.23}$$

we get the following equation of motion for $\hat{P}_I(t)$

$$i\hbar \frac{\partial \hat{P}_I(t)}{\partial t} = [H_T(t), \hat{P}_I(t)] \tag{6.24}$$

where

$$H_T(t) = e^{\frac{i}{\hbar}\int_{t_o}^{t} H_o\, dt'} H_T e^{-\frac{i}{\hbar}\int_{t_o}^{t} H_o\, dt'} \tag{6.25}$$

Solution of (6.24) to first order in $H_T(t)$ is

$$\hat{P}_I(t) = \hat{P}_I(t_o) - \frac{i}{\hbar}\int_{t_o}^{t} dt' \, [H_T(t'), \hat{P}_I(t')] \tag{6.26}$$

Substituting (6.26) in (6.24) yields

$$\frac{\partial \hat{P}_I(t)}{\partial t} = -\frac{i}{\hbar}[H_T(t), \hat{P}_I(t_o)] + \left(\frac{i}{\hbar}\right)^2 \int_{t_o}^{t} dt' \; [H_I(t), [H_I(t'), \hat{P}_I(t')]] \tag{6.27}$$

In the quantum Markoff approximation the density operator inside the integral, $\hat{P}_I(t')$, is replaced by the density operator $\hat{P}_I(t)$ at time t. This approximation is valid provided the coherence times are very small compared to the time taken by $\hat{P}_I(t)$ to change significantly. For further justification of this assumption see [53]. We will work in the qantum Markoff approximation.

Let the many body state of the entire system be described by the quantum state $|\{n_n\}, \{n_m\} >$. This state is characterized by a particular configuration of occupation numbers $\{n_n\}$ and $\{n_m\}$ of the 2-DEG and the dot. Let $p(\{n_m\})(t)$ be the probability that the dot is described by occupation numbers $\{n_m\}$ at time t. $p(\{n_m\})(t)$ may be calculated from the density matrix $\hat{P}_I(t)$ by taking the trace over the states belonging to the 2-DEG.

$$p(\{n_m\})(t) = \sum_{\{n_n\}} < \{n_n\}, \{n_m\}|\hat{P}_I(t)|\{n_n\}, \{n_m\} > \tag{6.28}$$

$p(\{n_m\})(t)$ contains a lot more information than needed to model the system. Usually, one would only be interested in the probability $p_N(t)$ that the dot contains a total of N electrons at time t. This can be obtained from $p(\{n_m\})(t)$ by summing over all possible configurations with occupation numbers $\{n_m\}$ such that $\sum_m n_m = N$. Thus

$$\begin{aligned} p_N(t) &= \sum_{\{n_m\}} p(\{n_m\})(t) \; \delta_{\sum_m n_m = N} \\ &= \sum_{\{n_m\}} \sum_{\{n_n\}} < \{n_n\}, \{n_m\}|\hat{P}_I(t)|\{n_n\}, \{n_m\} > \; \delta_{\sum_m n_m = N} \end{aligned} \tag{6.29}$$

We can now substitute for $\hat{P}_I(t)$ from (6.27) in the above equation and perform the indicated summations. After a lengthy and tedious algebra we get the following

stochastic equation

$$\begin{aligned}\frac{\partial p_N(t)}{\partial t} &= W_{N+1\to N}p_{N+1}(t) + W_{N-1\to N}p_{N-1}(t) \\ &\quad -W_{N\to N+1}p_N(t) - W_{N\to N-1}p_N(t)\end{aligned} \tag{6.30}$$

$W_{N\to M}$ are the transition rates for going from the state of the dot with N electrons inside it to the state where there are M electrons inside it. These transition rates are given as follows

$$\begin{aligned}W_{N+1\to N} &= \frac{2\pi}{\hbar}\sum_{n,m}|Tnm|^2\,\delta(\epsilon_n + E_s(N+1) + eV - \epsilon_m - E_s(N)) \\ &\times\ (1 - f_D(\epsilon_n - E_f))f_{N+1}(\epsilon_m) \\ W_{N-1\to N} &= \frac{2\pi}{\hbar}\sum_{n,m}|Tnm|^2\,\delta(\epsilon_n + E_s(N-1) + eV - \epsilon_m - E_s(N)) \\ &\times\ f_D(\epsilon_n - E_f)(1 - f_{N-1}(\epsilon_m)) \\ W_{N\to N+1} &= \frac{2\pi}{\hbar}\sum_{n,m}|Tnm|^2\,\delta(\epsilon_n + E_s(N) + eV - \epsilon_m - E_s(N+1)) \\ &\times\ f_D(\epsilon_n - E_f)(1 - f_N(\epsilon_m)) \\ W_{N\to N-1} &= \frac{2\pi}{\hbar}\sum_{n,m}|Tnm|^2\,\delta(\epsilon_n + E_s(N) + eV - \epsilon_m - E_s(N-1)) \\ &\times\ (1 - f_D(\epsilon_n - E_f))f_N(\epsilon_m)\end{aligned} \tag{6.31}$$

f_D is the Fermi-Dirac distribution function for electrons in the 2-DEG. E_f is the fermi level of the 2-DEG. In the above equation $f_N(\epsilon_m)$ is the probability that the state with energy ϵ_m in the dot is occupied given that there are total of N electrons in the dot. Since the lifetimes of electron in the dots are expected to be much larger than the relaxation times, we may compute $f_N(\epsilon_m)$ for a canonical ensemble with N electrons. Note, that since the system is not attached to a particle reservoir, usual

Fermi-Dirac statistics are not applicable. Thus we may write

$$f_N(\epsilon_m) = \frac{\sum_{\{n_{m'}\}} e^{-\frac{1}{KT}\sum_{m'}\epsilon_{m'}n_{m'}} (\delta_{\sum_{m'} n_{m'}=N})(\delta_{n_m=1})}{\sum_{\{n_{m'}\}} e^{-\frac{1}{KT}\sum_{m'}\epsilon_{m'}n_{m'}} (\delta_{\sum_{m'} n_{m'}=N})} \tag{6.32}$$

For a dot with few electrons, $f_N(\epsilon_m)$ can be calculated numerically.

Equations (6.30), (6.31), and (6.32) comprise a complete set of equations needed to describe the behavior of electrons in quantum dots coupled to 2-DEG's.

6.4 Carrier Statistics and Fluctuations Inside the Dot

Consider equation (6.30). We can write it as

$$\frac{\partial \vec{P}(t)}{\partial t} = W.\vec{P}(t) \tag{6.33}$$

Where the vector $\vec{P}(t)$ is $[p_0(t), p_1(t), ...p_{N_o}(t)]$, and W is the transition matrix of dimension $N_o \times N_o$. We have turncated the matrix W so that the maximum number of allowed electrons inside the dot is N_o. This is done for computational ease. The matrix W is expected to have all positive eigenvalues and the solution will therefore reach a stationary value $\vec{P}^s$. This can be found by setting $\frac{\partial \vec{P}(t)}{\partial t} = 0$. The stationary probabilities for $N \neq 0$ are found to be

$$p_N^s = \frac{\prod_{K=1}^{N} \frac{W_{K-1\to K}}{W_{K\to K-1}}}{1 + \sum_{Q=1}^{N_o-1} \prod_{K=1}^{Q} \frac{W_{K-1\to K}}{W_{K\to K-1}}} \tag{6.34}$$

and

$$p_0^s = \frac{1}{1 + \sum_{Q=1}^{N_o-1} \prod_{K=1}^{Q} \frac{W_{K-1\to K}}{W_{K\to K-1}}} \tag{6.35}$$

From knowledge of the probabilities p_N^s, one can find the mean number of electrons and the the variance in the number distribution by the expressions

$$m_N =< N >= \sum_{N=0}^{N_o} N\, p_N^s \tag{6.36}$$

$$\sigma_N^2 =< N^2 > - < N >^2 = \sum_{N=0}^{N_o} N^2\, p_N^s - \left(\sum_{N=0}^{N_o} N\, p_N^s \right)^2 \tag{6.37}$$

The spectrum of number fluctuations $S_n(\omega)$ inside the dot can be found from the expression

$$S_N(\omega; t_o) = \frac{1}{2} \int_0^\infty d(t - t_o)\, e^{i\omega(t - t_o)} \left(< N(t)N(t_o) + N(t_o)N(t) > -2 < N(t_o) >^2 \right) \tag{6.38}$$

where

$$< N(t)N(t_o) >= \sum_{N=0,M=0}^{N_o,N_o} N\, M\; p(N, t|M, t_o)\, p(M, t_o) \tag{6.39}$$

$p(M, t_o)$ is the probability that the dot has M electrons at time t_o and $p(N, t|M, t_o)$ is the conditional probability that the dot will have N electrons at time t given that it had M electrons at time t_o. It seems that that $S_n(\omega; t_o)$ depends on the artificial time parameter t_o. This is, however, not the case as we will show here. We expect all the averages to have time translational invariance, i.e.

$$< N(t)N(t_o) >=< N(t + t')N(t_o + t') > \tag{6.40}$$

and if we choose t' big enough so that $p(M, t_o + t') = p_M^s$ we get

$$< N(t)N(t_o) >= \sum_{N=0,M=0}^{N_o,N_o} N\, M\; p(N, t|M, t_o)\, p_M^s \tag{6.41}$$

Time translational invariance implies $p(N, t|M, t_o)$ depends upon $t - t_o$. Therefore

$p(N,t|M,t_o) = p(N,t|M,0)$. Finally we can write $S_N(\omega)$ as

$$S_N(\omega) = \frac{1}{2}\int_{-\infty}^{\infty} dt\, e^{i\omega t} \left(\sum_{N=0,M=0}^{N_o,N_o} N\, M\, p(N,t|M,0)\, p^s_M\, \theta(t) + \sum_{N=0,M=0}^{N_o,N_o} N\, M\, p(N,0|M,t)\, p^s_M\, \theta(-t) - 2(\sum_{N=0}^{N_o} N\, p^s_N)^2 \right) \quad (6.42)$$

The quantities $p(N,t|M,0)$ and $p(N,0|M,t)$ can easily be found by solving equation (6.30) as will be shown below for a simple case.

6.4.1 A Two Level Model

Here we will present a simplified two state model of a quantum dot. Suppose we apply a certain gate voltage to the single quantum dot device shown in figure (6-2) such that the stationary probabilities p^s_N are all almost zero for all $N < K$ and for all $N > K+1$. Thus at that gate voltage the dot has a high probability for containing either K or $K+1$ electrons. Such a situation arises in actual devices at that gate voltage at which the mean number of electrons inside the dot is $\sim K + \frac{1}{2}$. At this gate voltage the dot may modeled as a two level system with the two states being those in which the number of electrons inside the dot is either K or $K+1$. We may write the stochastic equations that follow from (6.30) for this case as

$$\frac{\partial}{\partial t}\begin{bmatrix} p_K(t) \\ p_{K+1}(t) \end{bmatrix} = \begin{bmatrix} -u & v \\ u & -v \end{bmatrix}\begin{bmatrix} p_K(t) \\ p_{K+1}(t) \end{bmatrix} \quad (6.43)$$

where $u = W_{K\to K+1}$ and $v = W_{K+1\to K}$. The matrix can be solved by diagonalization giving

$$p(K,t|N,0) = \frac{v}{u+v} + e^{-(u+v)t}\left(\frac{u}{u+v}\delta_{N,K} - \frac{v}{u+v}\delta_{N,K+1}\right) \quad (6.44)$$

$$p(K+1,t|N,0) = \frac{u}{u+v} - e^{-(u+v)t}\left(\frac{u}{u+v}\delta_{N,K} - \frac{v}{u+v}\delta_{N,K+1}\right) \quad (6.45)$$

As $t \to \infty$

$$p(K, t|N, 0) \to p^s_K = \frac{v}{u+v} \tag{6.46}$$

$$p(K+1, t|N, 0) \to p^s_{K+1} = \frac{u}{u+v} \tag{6.47}$$

Also

$$m_N =< N >= (K)\frac{v}{u+v} + (K+1)\frac{u}{u+v} = K + \frac{u}{u+v} \tag{6.48}$$

For $u \approx v$, $m_N \approx K + \frac{1}{2}$. The variance in electron number is

$$\sigma^2_N =< N^2 > - < N >^2 = \frac{uv}{(u+v)^2} \tag{6.49}$$

For $u \approx v$ the variance in electron number is about $\frac{1}{2}$. The results in (6.44), (6.45), (6.46) and (6.47) can be plugged in (6.42) to give the spectrum of number fluctuations which comes out to be

$$S_N(\omega) = \frac{u+v}{\omega^2 + (u+v)^2}\sigma^2_N \tag{6.50}$$

The number fluctuation spectrum gives information about the rates at which number fluctuations take place. We see that for our two level model $S_N(\omega)$ is a lorentzian with a width equal to the sum of the two important transition rates characterizing the system.

6.5 Quantized Threshold Voltage Shifts

The storage of electrons inside the quantum dots will result in shifts in the threshold voltage of the MOS devices shown in figures (6-1) and (6-2). For the case of the single dot device shown in figure (6-2) the threshold voltage shift $\Delta V_T(N)$ when the dot contains N electrons may be given by the approximate expression

$$\Delta V_T(N) = \frac{Ne}{A\epsilon_{ox}}\left(\frac{d\epsilon_{ox}}{2\epsilon_{si}} + t^r_{ox}\right) \tag{6.51}$$

A is the area of the dot, and d and t_{ox}^r are as shown in figure (6-3). For the nano crystal device shown in figure (6-1) the threshold voltage shift can be written as

$$\Delta V_T(N) = \frac{n_d N e}{\epsilon_{ox}} \left(\frac{d\epsilon_{ox}}{2\epsilon_{si}} + t_{ox}^r \right) \tag{6.52}$$

where n_d is the number density of nano crystals per unit area and N is the number of electrons per dot. From the analysis carried out in the last section it is obvious that if a voltage pulse is applied to the gate of the device then the mean number of electrons m_N inside the dots at the end of the pulse will depend on both the magnitude and duration of the applied pulse. As a result of the coulomb energy associated with charging the dots we expect the mean number of electrons in the dots to increase in discrete steps with the magnitude of the gate voltage pulse. It follows that the threshold voltage shifts will also increase in discrete steps with the magnitude of the gate voltage pulse. Thus these devices offer the novel possibility of realizing multi-state memory devices using just a single transistor.

6.6 Channel Conductance Fluctuations

In the last section we showed the relationship between the number of electrons inside the dots and the threshold voltage shift of the device. For a MOS device the channel current for small values of V_{DS} is given approximately by the relation [11]

$$I_{DS} = K(V_{GS} - V_T)V_{DS} \tag{6.53}$$

Therefore the channel conductance G becomes

$$G = K(V_{GS} - V_T) \tag{6.54}$$

The mean value of drain current $m_{I_{DS}}$ and mean value of channel conductance m_G will be

$$m_{I_{DS}} = m_G V_{DS} \tag{6.55}$$

$$m_G = \langle G \rangle = K(V_{GS} - \langle V_T \rangle) \tag{6.56}$$

From (6.36), (6.51) and (6.52)

$$\langle V_T \rangle = V_{TO} + \frac{m_N e}{A\epsilon_{ox}} \left(\frac{d\epsilon_{ox}}{2\epsilon_{si}} + t^r_{ox} \right) \tag{6.57}$$

for a single dot device and

$$\langle V_T \rangle = V_{TO} + \frac{n_d m_N e}{\epsilon_{ox}} \left(\frac{d\epsilon_{ox}}{2\epsilon_{si}} + t^r_{ox} \right) \tag{6.58}$$

for a nano crystal device. The variance of fluctuations in current $\sigma_{I_{DS}}$ and conductance σ_G can be also be written as a function of variance in fluctuations of the threshold voltage σ_{V_T}

$$\sigma^2_{I_{DS}} = \sigma^2_G \, V^2_{DS} \tag{6.59}$$

$$\sigma^2_G = K^2 \, \sigma^2_{V_T} \tag{6.60}$$

From (6.37), (6.51) and (6.52) we have

$$\sigma^2_{V_T} = \left(\frac{\sigma_N e}{A\epsilon_{ox}} \right)^2 \left(\frac{d\epsilon_{ox}}{2\epsilon_{si}} + t^r_{ox} \right)^2 \tag{6.61}$$

for a single dot device and

$$\sigma^2_{V_T} = \left(\frac{n_d \sigma_N e}{\epsilon_{ox}} \right)^2 \left(\frac{d\epsilon_{ox}}{2\epsilon_{si}} + t^r_{ox} \right)^2 \tag{6.62}$$

for a nano crystal device. The experimentally relevant quantity is the spectrum of current fluctuations $S_{I_{DS}}(\omega)$. This can also be related to the spectrum of conductance fluctuations $S_G(\omega)$ and spectrum of threshold voltage fluctuations $S_{V_T}(\omega)$ as follows

$$S_{I_{DS}}(\omega) = S_G(\omega) \, V^2_{DS} \tag{6.63}$$

$$S_G(\omega) = K^2 \, S_{V_T}(\omega) \tag{6.64}$$

Again, we can relate $S_{V_T}(\omega)$ to $S_N(\omega)$ given in equation (6.42)

$$S_{V_T}(\omega) = \left(\frac{S_N(\omega)e}{A\epsilon_{ox}}\right)^2 \left(\frac{d\epsilon_{ox}}{2\epsilon_{si}} + t^r_{ox}\right)^2 \tag{6.65}$$

for a single dot device and

$$S_{V_T}(\omega) = \left(\frac{n_d S_N(\omega)e}{\epsilon_{ox}}\right)^2 \left(\frac{d\epsilon_{ox}}{2\epsilon_{si}} + t^r_{ox}\right)^2 \tag{6.66}$$

for a nano crystal device. The spectrum of current fluctuations is thus related to the spectrum of number fluctuations. Current fluctuation spectrum can be measured experimentally at low temperatures. Such measurements can provide valuable information about the rates of charging and discharging these devices as already shown by the two level model of the dot discussed in the last section.

6.7 Calculation of Coupling Constants

We have defined the transition rates W in terms of the coupling constants T_{mn} between state ψ_m inside the dot with state ψ_n inside the 2-DEG. These coupling constants were introduced in the hamiltonian. Here we will describe briefly how to calculate these coupling constants. We will explicitly focus upon the single dot device shown in figure (6-2). As shown in chapter three, T_{mn} is given by the expression

$$T_{mn} = -\frac{\hbar^2}{2m^{ox}} \int \left(\psi^*_m(\vec{r})\vec{\nabla}\psi_n(\vec{r}) - \psi_n(\vec{r})\vec{\nabla}\psi^*_m(\vec{r})\right) .d\vec{S} \tag{6.67}$$

The above integral is taken over a surface lying in the middle of the injecting oxide. We assume the co-ordinate axis to be oriented as in figure (6-3). The states in the 2-DEG are labeled by the components of momenta parallel to the interface (which are q_y and q_z) and the energy ϵ_n of the subband. Thus inside the oxide region we may write the wavefunction of an electron belonging to the 2-DEG in the WKB approximation

as

$$\psi_{n,q_y,q_z}(\vec{r}) \approx \sqrt{\frac{2}{A\,L_w}}\,e^{i(q_y y + q_z z)}\,\frac{\sqrt{\frac{\beta(-t^l_{ox})}{\beta(x)}}\,q_x\,m^{ox}\,e^{-\int_{-t^l_{ox}}^{x}\beta(x')\,dx'}}{\sqrt{(m_x^{si}\beta(-t^l_{ox}))^2 + (m^{ox}q_x)^2}} \tag{6.68}$$

In the above expression

$$q_x = \frac{\sqrt{2m_x^{si}\epsilon_n}}{\hbar} \tag{6.69}$$

$$m_x^{si} = \text{either } m_t^{si} \text{ or } m_l^{si} \text{ assuming a } <100> \text{ crystal orientation} \tag{6.70}$$

$$L_w = \text{approximate width of the well confining the } 2-\text{DEG} \tag{6.71}$$

$$A = \text{arbitrary area of a region in which the wavefunctions are normalized} \tag{6.72}$$

$$\beta(x) = \frac{\sqrt{2m_x^{si}(e\phi_o - eF_{ox}x - \epsilon_n}}{\hbar} \tag{6.73}$$

$$F_{ox} = \text{electric field inside the oxide} \tag{6.74}$$

As long as electron tunneling occurs in the direct regime, WKB approximation is an excellent approximation.

The states inside the dot may be computed accurately in the Hartree approximation. However, we expect that for the purposes of calculating coupling constants choosing aproximate wavefunctions of an infinite 3-D quantum box will make little difference in the final result. Therefore to keep analysis simple we choose the following form for the wavefunction of the state of the dot inside the oxide region

$$\psi_{k_x,k_y,k_z}(\vec{r}) \approx \sqrt{\frac{8}{L_y L_z L_x}}\,\sin(k_y y)\sin(k_z z)\,\frac{\sqrt{\frac{\alpha(0)}{\alpha(x)}}\,k_x\,m^{ox}\,e^{\int_0^x \alpha(x')\,dx'}}{\sqrt{(m_x^{si}\alpha(0))^2 + (m^{ox}k_x)^2}} \tag{6.75}$$

where

$$k_x = \frac{\sqrt{2m_x^{si}\epsilon_x}}{\hbar} \tag{6.76}$$

$$L_x, L_y, L_z = \text{x, y, z dimensions of the dot} \tag{6.77}$$

$$\alpha(x) = \frac{\sqrt{2m_x^{si}(e\phi_o - eF_{ox}x - \epsilon_x)}}{\hbar} \tag{6.78}$$

$$\epsilon_x = \text{energy associated with motion in the x} - \text{direction inside the dot} \tag{6.79}$$

The total energy E_{k_x,k_y,k_z} of this state is simply

$$E_{k_x,k_y,k_z} = \frac{\hbar^2 k_x^2}{2m_x^{si}} + \frac{\hbar^2 k_y^2}{2m_y^{si}} + \frac{\hbar^2 k_z^2}{2m_z^{si}} \tag{6.80}$$

Using the wavefunctions (6.68) and (6.75) we can compute the coupling constant $T_{\{n,q_y,q_z\},\{k_x,k_y,k_z\}}$ which comes out to be

$$\begin{aligned} T_{\{n,q_y,q_z\},\{k_x,k_y,k_z\}} &= -\frac{\hbar^2}{2m^{ox}}\sqrt{\frac{16}{AL_wL_xL_yL_z}} \\ &\times \frac{\sqrt{\alpha(0)\beta(0)}\; k_x\, q_x (m^{ox})^2}{\sqrt{(m_x^{si}\beta(-t_{ox}^l))^2 + (m^{ox}q_x)^2}\,\sqrt{(m_x^{si}\alpha(0))^2 + (m^{ox}k_x)^2}} \\ &\times \left(\sqrt{\frac{\alpha(-t_{ox}^l/2)}{\beta(-t_{ox}^l/2)}} + \sqrt{\frac{\beta(-t_{ox}^l/2)}{\alpha(-t_{ox}^l/2)}}\right) \\ &\times e^{\left(\int_0^{-t_{ox}^l/2} \alpha(x')\,dx' + \int_{-t_{ox}^l/2}^{-t_{ox}^l} \beta(x')\,dx'\right)} \\ &\times \left(\int_0^{L_y} dy\, e^{iq_y y} \sin(k_y y)\right)\left(\int_0^{L_z} dz\, e^{iq_z z} \sin(k_z z)\right) \end{aligned} \tag{6.81}$$

Transition rates W can be calculated numerically using this expression for the coupling constant and equations (6.31). Also note that the expression for coupling constant above holds irrespective of whether the state inside the dot is coupled to a state in the 2-DEG (which is confined in one dimension) or a to state in a 3-D continuum. However, the difference comes when the summations are performed in the calculation of transition rates W using equations (6.31).

6.8 Numerical Results

6.8.1 Numerical Results for the Steady State

To test our theoretical model we have carried out calculations for a single quantum dot coupled to an inversion layer as shown in figure(6-3). Our calculations are carried out in the following steps

1. For a given number of electrons N inside the dot and a gate voltage, Poisson equation is solved numerically for the structure shown in figure (6-3) to calculate the potential distribution.

2. Equations (6.31) are used to calculate transition rates $W_{N\to N+1}$ and $W_{N\to N-1}$ using coupling constants given by (6.81).

3. For the same gate voltage the number of electrons N inside the dot is varied from 0 to some upper number $N_o(\sim 3$ or $4)$ and for each value of N steps (1) and (2) above are repeated.

4. Once all the transitions rates have been obtained, equations (6.34), (6.35), (6.36) and (6.37) are used to obtain the stationary probabilities p_N^s, mean number of electrons and the variance in the electron number.

5. The gate voltage is varied and steps (1), (2) and (3) above are repeated for each value of gate voltage.

Here we present the numerical results for dot of dimensions $L_x = 60\mathring{A}$, $L_y = 100\mathring{A}$, and $L_z = 100\mathring{A}$. The thickness t_{ox}^l of the injecting oxide is $15\mathring{A}$, and the thickness t_{ox}^r of the control oxide is $50\mathring{A}$. The substrate doping is $10^{17}/\text{cm}^3$ p-type. The threshold voltage V_{TO} of the device without any electrons inside the dot is about 0.3 Volts. We assumed that both the substrate and the dot have $< 100 >$ crystal orientation. This assumption is not valid in actual devices since the dots are deposited by CVD and therefore have no fixed crystal orientation. Infact, they are perhaps not even single crystal but poly-crystalline.

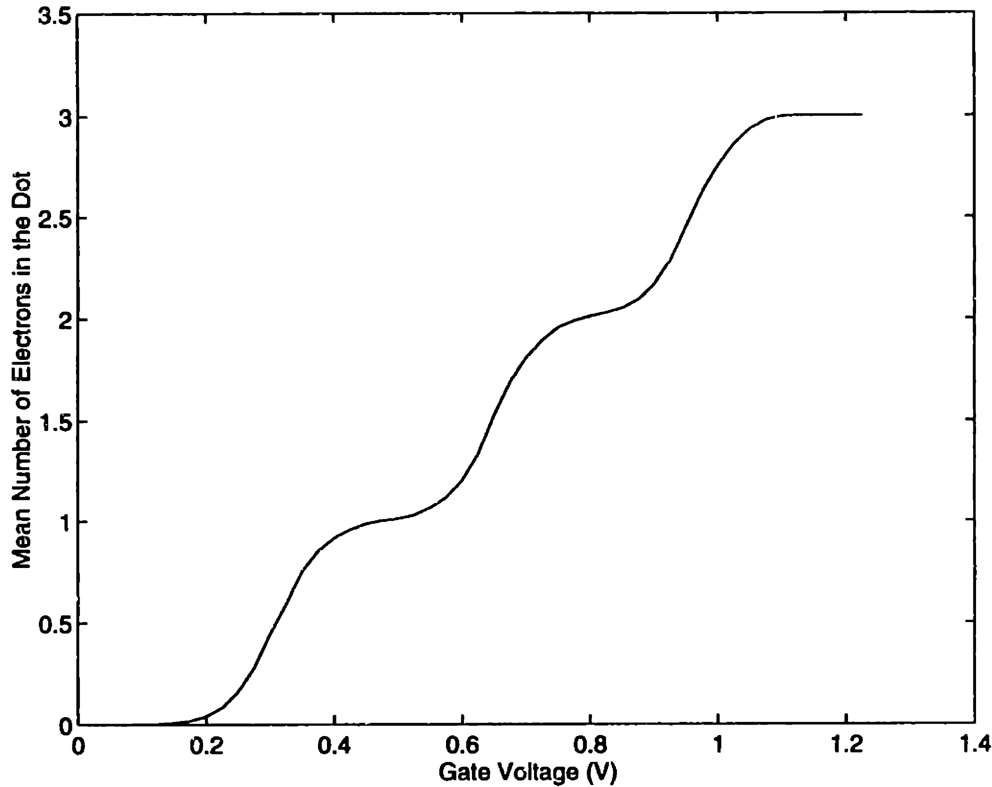

Figure 6-4: Mean number of electrons in the dot as a function of gate voltage.

Figure (6-4) shows the mean number of electrons inside the dot as a function of the gate to substrate voltage. Notice that the first electron does not appear inside the dot until the gate voltage exceeds the threshold voltage V_{TO} and an inversion layer is formed to supply that electron. The presence of one electron inside the dot shifts the threshold voltage by $\Delta V_T(N=1)$ whose value is given in (6.51). For this device $\Delta V_T(N=1)$ comes out to be about 0.3 Volts. Thus to put a second electron inside the dot the gate voltage needs to be increased by at least $\Delta V_T(N=1)$ beyond the voltage needed to put in the first electron, which is approximately V_{TO}. We say 'at least' because the electron already present inside the dot will try to occupy the lowest available states, leaving the higher ones for the second electron. Thus the second electron can be put in at a gate voltage which is little higher than just $V_T + \Delta V_T(N=1)$. Generalizing this we can say that to put the N'th electron in the dot requires a gate voltage of approximately $V_{TO} + \Delta V_T(N-1)$. This trend can be clearly seen in figure (6-4).

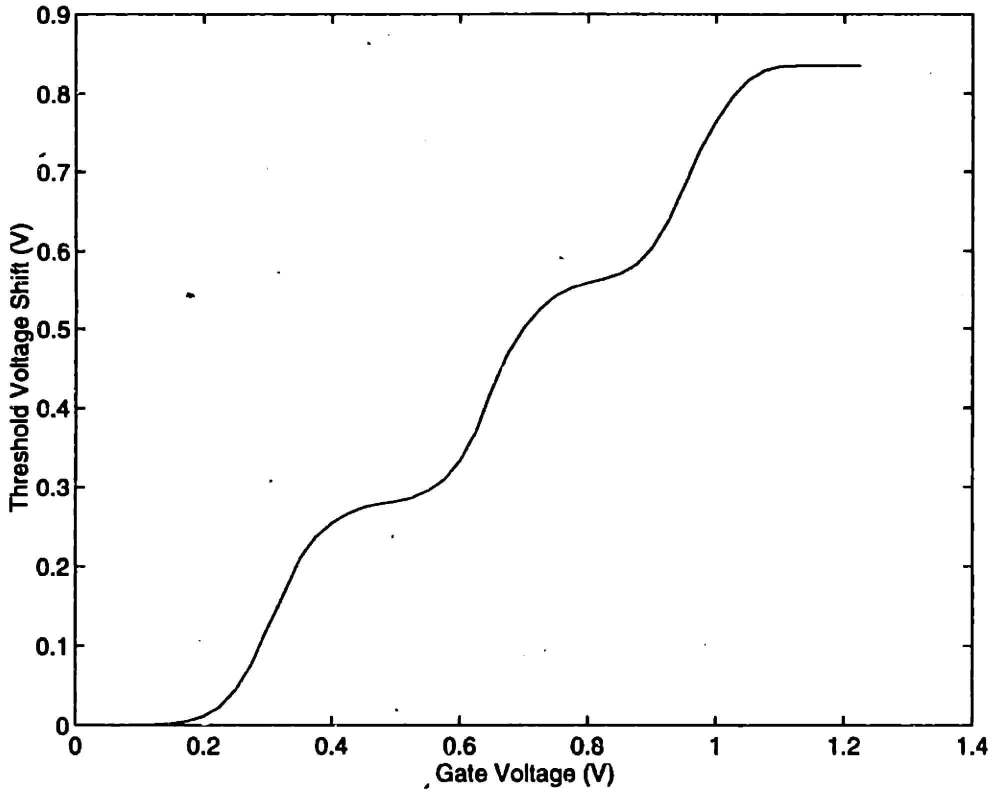

Figure 6-5: Threshold voltage shift ΔV_T as a function of gate voltage

Figure (6-5) shows the threshold voltage shift $\Delta V_T(N)$ as a function of gate voltage. Equation (6.51) was used to calculate $\Delta V_T(N)$. Figure (6-6) shows the variance (or the root mean square value or RMS value) of electron number as function of gate voltage. We see that at gate voltages where the mean electron number is an integer plus a half, the variance in electron number is also a half. This implies that the two level model presented in the last section captures the essential physics at gate voltages at which the mean number of electrons in the dot is an integer plus a half. Thus we expect that at these gate voltages the spectrum of number fluctuations will be a lorentzian given by (6.50). In our simulations we have fixed the upper limit on the number of electrons inside the dot to three (i.e. $N_o = 3$) for computational ease. Therefore, figure (6-4) shows the mean number of electrons leveling off around a gate voltage of 1.2 Volts, and the fluctuations in electron number in figure (6-6) dying off at the same voltage. This is, of course, unphysical.

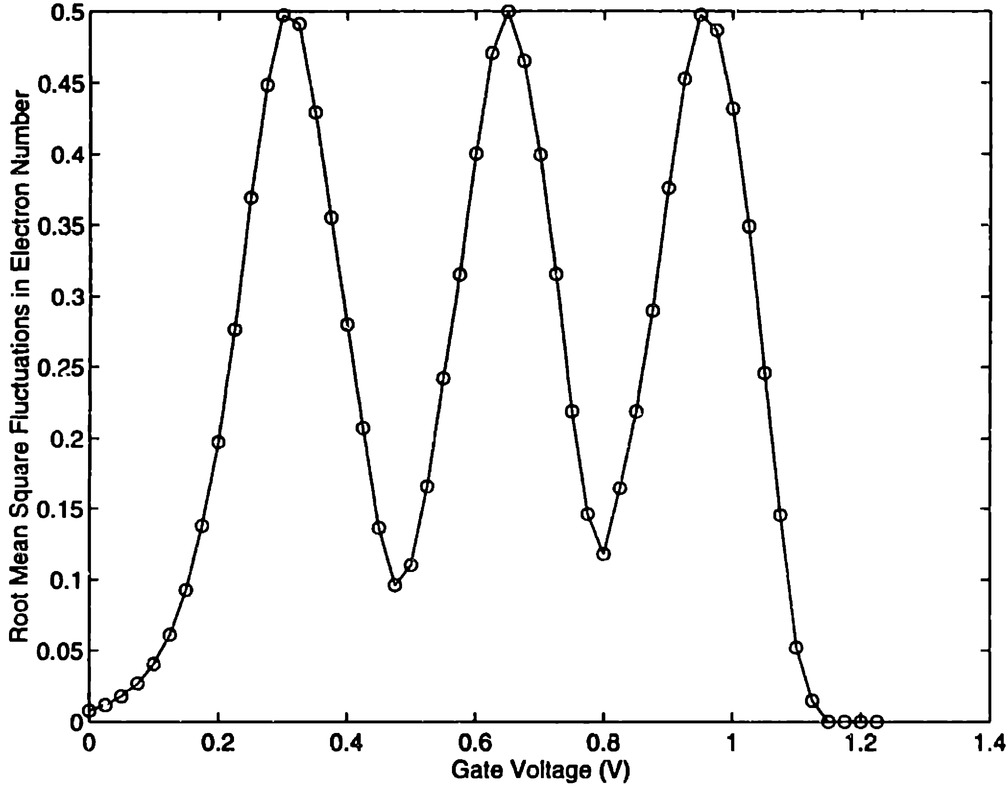

Figure 6-6: Variance of electron number in the dot as a function of gate voltage

6.8.2 Numerical Results for the Time Dependent Case

In practice a gate voltage pulse is applied to charge the dot. Our analysis above has assumed that the pulse is sufficiently long so that the time dependent solution of equation (6.30) reaches its stationary value, and we have used the stationary probabilities p_N^s given by equations (6.34) and (6.35) to calculate the mean number of electrons and the variance in electron number as a function of gate voltage. However, our formalism is sufficiently powerful to incorporate pulses of arbitrary short duration. Suppose a square voltage pulse is applied to the gate at time $t = 0$ of duration T. The coupled equations in (6.30) can be solved with appropriate boundary conditions to yield the time-dependent probabilities $p_N(t)$. If the dot at time $t = 0$ was empty then boundary conditions would be $p_0(t = 0) = 1$ and $p_N(t = 0) = 0$ for $N \neq 0$. The time dependent mean number of electrons $m_N(t)$ inside the dot is

$$m_N(t) = \sum_{N=0}^{N_o} N p_N(t) \tag{6.82}$$

The mean number of electrons at time $t = T$ is $m_N(t = T)$. Thus, the threshold voltage shift at the end of the pulse is $\Delta V_T(N = m_N(t = T))$. In general, therefore, the threshold voltage shifts would depend upon the magnitude and also the duration of the pulse.

Here we first present time dependent results for the case when a 3.0 Volt square pulse at time $t = 0$ is applied to the gate of the single dot device which has been described above. Figure (6-7) shows the mean number of electrons in the dot as a function of time following the application of the gate pulse. It can be seen from the figure that it takes longer and longer time to put more and more electrons in the dot. This can easily be understood from the fact that initially when the pulse is applied there are no electrons inside the dot. Consequently, the potential drop across the injecting oxide is large and the electric field in the injecting oxide is also large resulting in large coupling constants. Electrons are injected into high energy states of the dot, and the density of states at high energies is also large and also these states are all empty. Thus the rate of injecting electrons into the dot is also large. As the dot gets filled up with electrons, the potential drop across the injecting oxide becomes smaller and the coupling constants become smaller. Injection also now takes place in relatively lower energy states of the dot, where the density of states is lower and some of them are, of course, occupied by the electrons already present in the dot. In addition, the presence of electrons inside the dot also changes the threshold voltage of the device, which means that there are also less electrons available in the channel that can tunnel into the dot. All these factors add to make the injection of additional electrons more and more slower. Figure (6-8) shows the corresponding threshold voltage shift as a function of time.

Figure (6-9) shows the mean number of electrons inside the dot when a gate voltage pulse of -3.0 Volt is applied to remove the charge stored inside the dot. The various curves are for different initial number of electrons inside the dot. From the figure it is obvious that the rate of discharge is higher if the dot initially contained a larger number of electrons. This is because more charge inside the dot results in higher electric field inside the injecting oxide when a negative gate voltage is applied to

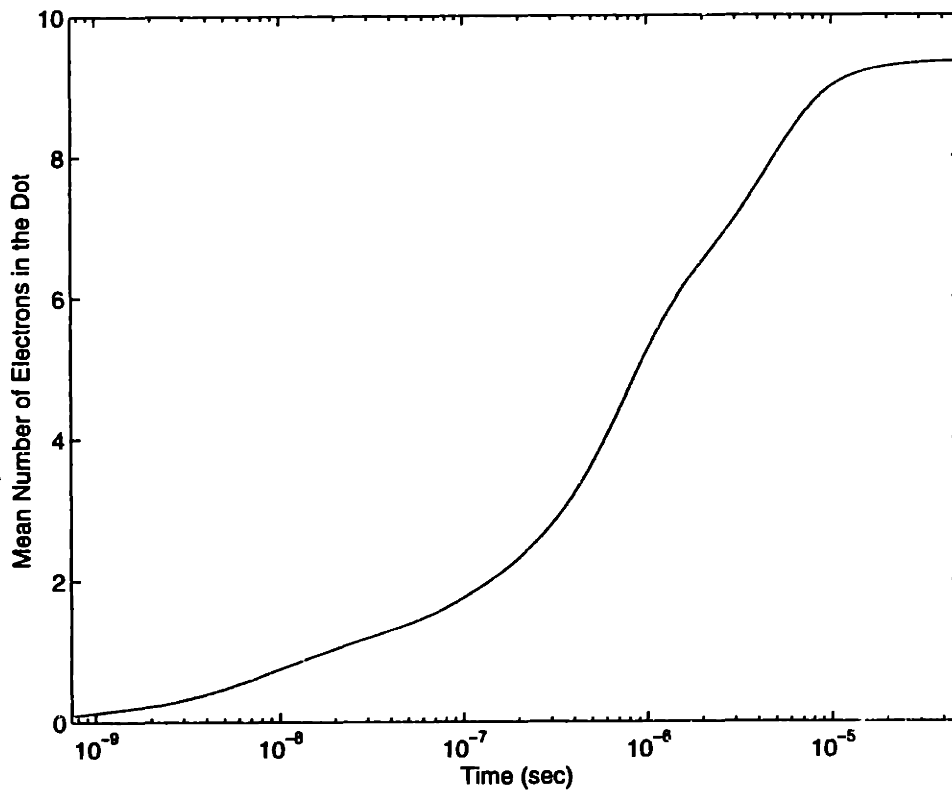

Figure 6-7: Mean number of electrons in the dot as a function of time on application of a 3.0 Volt pulse at the gate.

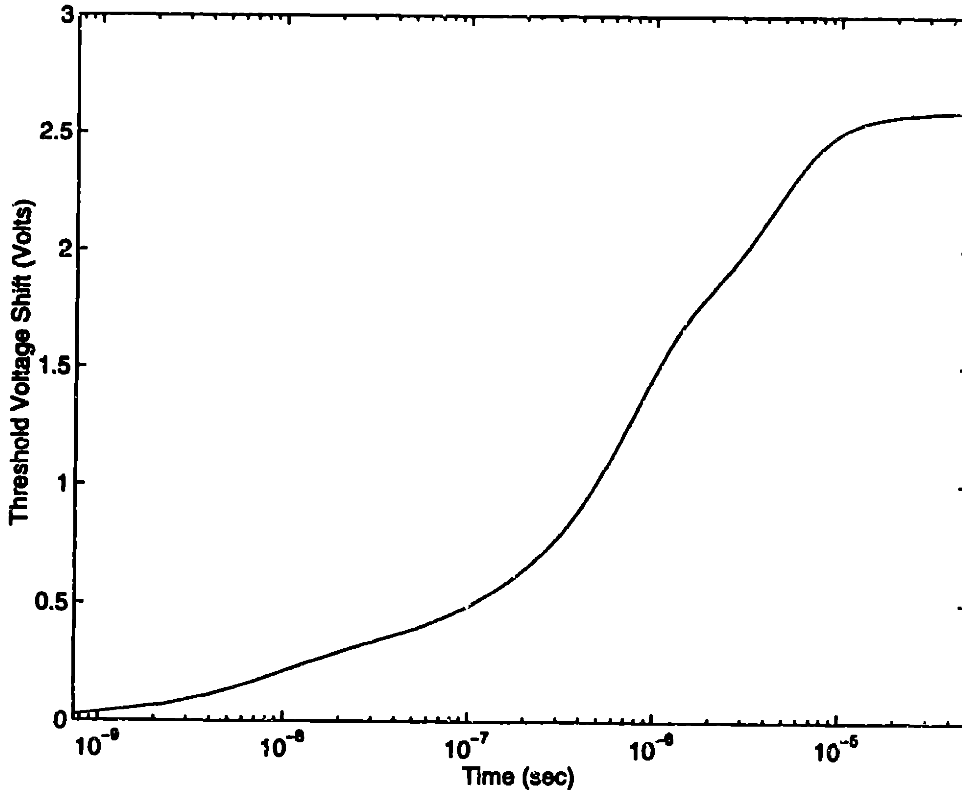

Figure 6-8: Shift in the threshold voltage of the device as a function of time.

discharge the device. Consequently, the ejection rate is higher. However, irrespective of the initial number of electrons in the dot, it takes about almost a micro second to completely discharge the device.

From the time dependent results presented here it is seen that if the dot is to be charged with one electron then this can be achieved by applying a pulse of 3.0 Volts at the gate for just 20ns. But discharging the dot will require a pulse of -3.0 Volts at the gate for at least $2\mu s$.

6.9 Conclusion

In this chapter we have presented a comprehensive analytical treatment of tunneling processes that take place when quantum dots is coupled to a two dimensional electron gas. We derived a set of coupled differential equations (6.30) for the probabilities $p_N(t)$ for the dot to contain N electrons. These probabilities depend upon the transition rates in (6.31). The transition rates in turn depend upon device dimensions like the

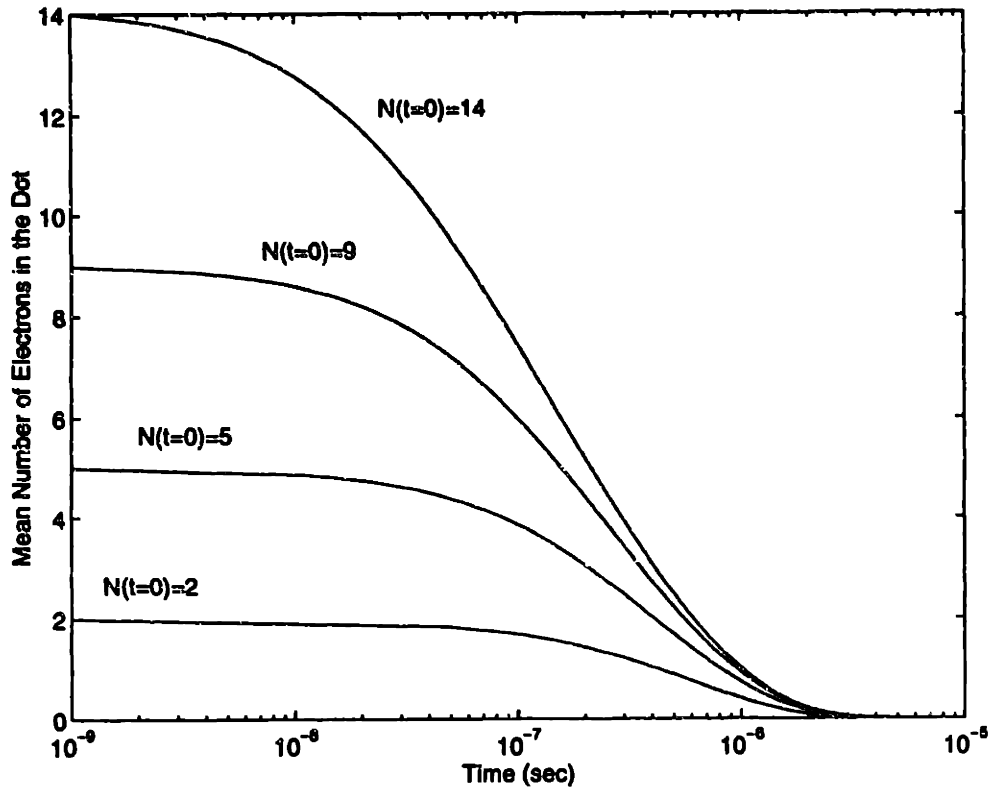

Figure 6-9: Mean number of electrons in the dot as a function of time on application of a -3.0 Volt pulse at the gate. The four curves are for different initial number of electrons in the dot.

dot size, oxide thickness e.t.c. and also on the applied gate bias. We showed that the coupled differential equations for $p_N(t)$ could be solved for any given initial conditions and applied gate voltage, and their solution completely described the time dependent behavior of the device. Thus these differential equations provide an important tool, not only to model the charging and discharging of the dots when write and erase pulses are applied to the gate, but also to study the charge retention abilities of these devices.

Chapter 7

Conclusion

In this thesis we have presented a comprehensive theoretical treatment of tunneling processes in Silicon/Silicon-dioxide systems. In chapter one we described the semi-classical model for calculating tunneling currents through oxides. We mentioned the shortcomings of that model and in chapter three we presented a fully self-consistent quantum mechanical model for describing tunneling in MOS devices. We introduced the concept of electron life times and showed how they may be used in computing tunneling currents from quasi bound states. In chapter four we presented the results of our numerical calculations and compared them with experimental measurements. We showed that our theoretical calculations agreed remarkably well with experimental data. Our calculations showed that very large tunneling currents ($> 5\text{A/cm}^2$) can be obtained in very thin ($\sim 15\mathring{A}$) oxides.

In chapter five we discussed two important issues related to modeling of electron tunnel transport in oxides : the mid-gap dispersion relation in SiO_2, and the effect of image forces. We presented a crystalline WKB approximation which generalized the usual WKB approximations to account for the complete mid-gap dispersion relation. We presented a novel method to describe the time-dependent response of a two dimensional electron gas. We also showed in the same chapter that barrier reduction effects due to image forces are certainly not negligible in MOS devices but the uncertainty in the experimental determination of oxide thicknesses by state of the art techniques make the experimental verification of image force effects almost

impossible.

Finally, in chapter six we presented a theoretical analysis of quantum dots coupled with the channel of MOS devices. We developed theoretical models to predict the time-independent and time-dependent behavior of these devices. These models have not yet been tested against experimental data.

The work done in this thesis certainly does not exhaust all possible research avenues. Further research in theoretical modeling of tunneling processes in Si/SiO_2 systems may be directed along the following lines :

- The correct boundary conditions on the electron wavefunction at the Si/SiO_2 interface are presently not known. It might be possible to obtain an accurate theoretical formulation of these boundary conditions from first principles.

- It remains unclear whether bulk effective mass theory may be used for very thin (a few atomic layers wide) oxides. The mid-gap dispersion relation in SiO_2 may change as the oxide thickness becomes too small. Clearly more work remains to be done to study such effects.

- The effective mass theory for SiO_2 has been formulated on the basis of the band structure of crystalline α-quartz. SiO_2 in MOS devices is amorphous. It remains unclear how well does the crystalline formulation work for amorphous thin films.

- We showed in chapter five that in MOS devices image force correction to barrier height is not small, specially as the oxide thicknesses become small. Our formulation for describing the time-dependent response of 2-DEG was not self-consistent (i.e. we did not take into account the effect of the induced charge density in the 2-DEG on the motion of the point charge). Given the rather large correction to barrier height that comes from the image potential in MOS devices with thin oxides, it is important that a self-consistent theory for dynamic image potential be developed for MOS devices.

- In modeling quantum dots we have used a two capacitor model for calculating

coulomb energy. This expression may not yield accurate results for the quantum dot memory devices. More accurate numerical calculations based upon equations described in chapter six are desired.

- In calculating the energies and wavefunctions of electrons in quantum dots we did not use a self-consistent formulation. A fully self-consistent solution which solves for the energies and wavefunctions in the 2-DEG as well in the dot may be interesting. Moreover, such a methodology will also be useful in calculating coulomb charging energies more accurately, as described in chapter six.

We are certain that with the recent advances in nano technology and with the innovations of quantum effect structures in Silicon, the research done in this thesis would prove to be valuable, and will motivate additional research in the yet unexplored area.

Bibliography

[1] A. H. Wilson, Proc. Roy. Soc. London, 136A, 487 (1932).

[2] J. Maserjian , *Historical Perspective on Tunneling in SiO_2*, Private Manuscript obtained from Max Fischetti, IBM.

[3] R. H. Fowler and L. Nordheim, Proc. Roy. Soc. London, 119A, 173 (1928).

[4] R. Holm, J. Appl. Phys., 22, 569 (1951).

[5] E. L. Murphy and R. H. Good, Phys. Rev., 102, 1464 (1956).

[6] M. Lenzlinger and E. H. Snow, *Fowler-Nordheim Tunneling into Thermally Grown SiO_2* , J. Appl. Phys., 40, 278 (1968).

[7] G. Krieger and R. M. Swanson, *Fowler-Nordheim Electron Tunneling in Thin Si-SiO_2-Al Structures* , J. Appl. Phys., 52, 5710 (1981).

[8] Z. A. Weinberg, *Tunneling of Electrons from Si into Thermally Grown SiO_2* , Solid State Electronics, 20, 11 (1977).

[9] Z. A. Weinberg, *On Tunneling in Metal-Oxide-Silicon Structures*, J. Appl. Phys., 53, 5052 (1982).

[10] P. Olivo, J. Sune, and B. Ricco, *Determination of the Si-SiO_2 Barrier Height from the Fowler-Nordheim Plot* , IEEE Elec. Dev. Lett., 12, 620 (1991).

[11] Yannis P. Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill, NY (1987).

[12] S. M. Sze, *Physics of Semiconductor Devices*, Wiley, NY (1981).

[13] J. Stoer, R. Bulirsch, *Introduction to Numerical Analysis*, Springer-Verlag, NY (1980).

[14] T. Ando, A. Fowler, F. Stern, Rev. Mod. Phys., 54, 437 (1982).

[15] F. Stern, Phys. Rev., 5, 4891 (1972). F. Stern, Phys. Rev., 163, 816 (1967). F. Stern, J. Comp. Phys., 6, 56 (1970).

[16] S. E. Laux, F. Stern, Appl. Phys. Lett., 91, 49 (1986).

[17] F. Stern, Phys. Rev., 33, 960 (1974).

[18] J. Sune, P. Olivo, B. Ricco, J. Appl. Phys., 1, 337 (1991).

[19] J. Sune, P. Olivo, B. Ricco, IEEE Trans. Elec. Dev., 39, 1732 (1991).

[20] W. Harrison, *Electronic Structure and the Properties of Solids*, Freeman, San Francisco (1980).

[21] John W. Negele and Henri Orland, *Quantum Many Particle Systems*, Addison-Wesley, NY (1988).

[22] L. S. Schulman, *Techniques and Applications of Path Integration*, Wiley, NY (1981).

[23] D. Langreth, *Linear and Non-Linear Response Theory with Applications* in *Linear and Non-Linear Electron Transport in Solids*, ed.'s J. T. Devreese and V. E. Doren, Plenum, NY (1976).

[24] J. Bardeen, Phys. Rev. Lett., 6, 57 (1961).

[25] C. B. Duke, *Tunneling in Solids*, Academic, NY (1969).

[26] L. V. Keldysh, Soviet Phys. JETP, 20, 1018 (1965).

[27] Ned S. Wingreen, A. Jauho, Y. Meir, *Time Dependent Transport through Mesoscopic Structures*, Phys. Rev., B48, 8487 (1993).

[28] Gerald Mahan, *Many Particle Physics*, Plenum, NY (1990).

[29] D. J. DiMaria, E. Carter, *Imapct Ionization, Trap Creation, Degradation, and Breakdown in Silicon Dixoide Films on Silicon*, Reasearch Report no. RC 18336(80343), I.B.M. T. J. Watson Research Center, 1992.

[30] Private communication with Sandip Tiwari, IBM T. J. Watson Research Center.

[31] Hussein I. Hanafi, Sandip Tiwari, Stuart Burns, *A Scalable Low Power Vertical Memory*, IEEE Proceedings of IEDM, pg. 657, 1995.

[32] See for example references [2, 6, 8, 9]

[33] Seiichi Iwata and Akitoshi Ishizaka, *Spectroscopic Ananlysis of the Si/SiO_2 System and Correlation with MOS Device Characteristics*, J. Appl. Phys., 79, 6653 (1996).

[34] Private communication with Doug Buchanan, IBM T. J. Watson Research Center.

[35] Ed. Pieter Balk, *The Si-SiO_2 System* , Elsevier Press, 1988.

[36] Neil W. Ashcroft and David N. Mermin, *Solid State Physics*, Saunders College, Philadelphia (1976).

[37] H. Jones, *The Theory of Brillouin Zones and Electronic States in Crystals*, North-Holland, Netherlands (1975).

[38] J. R. Chelikowsky and M. A. Schluter, Phys. Rev., B13, 826 (1976).

[39] M. Jonson, *The Dynamical Image Potential for Tunneling Electrons*, Solid St. Comm., 33, 743 (1980).

[40] D. B. Tran Thoai and M. Sunjic, *Dynamical Effects in Electron Tunneling*, Solid St. Comm., 77, 955 (1991).

[41] M. Sunjic and L. Marusic, *Dynamical Effects in Electron Tunneling*, Phys. Rev. B44, 9092 (1991).

[42] P. A. Serena, M. Soler and N. Garcia, *Self-Consistent Image Potential in a Metal Surface*, Phys. Rev. B34, 6767 (1986).

[43] G. Binnig, N. Garcia and H. Rohrer, *Electron-Metal Surface Interaction Potential with Vaccum Tunneling*, Phys. Rev. B30, 4816 (1984).

[44] G. D. Mahan, *Electron Interaction with Surface Modes*, private manuscript.

[45] M. Sunjic and L. Marusic, *Dynamical Effects in Electron Tunneling*, Solid St. Comm., 84, 123 (1992).

[46] M. Sunjic, C. Toulouse and A. A. Lucas, *Dynamical Corrections to Image Potentials*, Solid St. Comm., 11, 1629 (1629).

[47] Peter J. Feibelman, *Inclusion of Dynamics in the Ion-Metal Surface Interaction* , Surface Sc., 27, 438 (1971).

[48] Peter J. Feibelman, C. B. Duke and A. Bagchi, *Microscopic Description of Electron-Solid Interaction at a Surface*, Phys. Rev. B5, 2436 (1973).

[49] Haug Hartmut and S. W. Koch, *Quantum Theory of the Optical and Electronic Properties of Semiconductors*, World Scientific, Singapore (1994).

[50] D. Vollhardt and P. Wolfle, *Self-Consistent Theory of Anderson Localization* in *Electronic Phase Transitions*, ed.'s W. Hanke and Y. V. Kcpaev, Elsevier, NY (1992).

[51] D. Belitz and T. R. Kirkpatrick, *The Anderson-Mott Transition*, Rev. M. Phys., 66, 261 (1994).

[52] Stephen Adler, *Quantum Theory of the Dielectric Constant of Real Solids*, Phys. Rev., B126, 413 (1962).

[53] C. W. Gardiner, *Quantum Noise*, Springer-Verlag, Berlin (1991).