Nanoscale Structure and Transport – From Atoms to Devices

by

Matthew Hiram Evans

B.A., Cornell University, 2000

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

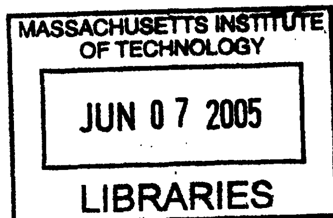
June 2005

© Massachusetts Institute of Technology 2005. All rights reserved.

Author ....
Department of Physics
April 27, 2005

Certified by ....
John D. Joannopoulos
Francis Wright Davis Professor of Physics
Thesis Supervisor

Accepted by ....
Thomas J. Greytak
Professor of Physics
Associate Department Head for Education



ARCHIVES

Nanoscale Structure and Transport –
From Atoms to Devices

by

Matthew Hiram Evans

Submitted to the Department of Physics
on April 27, 2005, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Nanoscale structures present both unique physics and unique theoretical challenges. Atomic-scale simulations can find novel nanostructures with desirable properties, but the search can be difficult if the wide range of possible structures is not well-understood. Electrical response and other non-equilibrium transport phenomena are measured experimentally, but not always simulated accurately. This thesis presents four diverse applications that demonstrate how first-principles calculations can address these challenges.

Novel boron nanotube structures with unusual elastic properties are presented. Internal degrees of freedom are identified that allow longitudinal stress to be dissipated without changing the tube's diameter, leading to high lateral stiffness.

Self-trapped hole structures in amorphous silicon dioxide are investigated in order to connect the behavior of hole currents to atomic-scale structures. Calculations on a paired-oxygen analogue to the V_k center show that such a configuration does not result in a metastable trapped-hole state.

A novel method to enable first-principles mobility calculations in ultrathin silicon-on-insulator (UTSOI) structures is presented and applied to interface roughness scattering in transistor channels. Self-consistent potentials and accurate wavefunctions and band structures allow for a direct link between measured electrical response and atomic structure. Atomic-scale interface roughness is shown to be an important limit on mobility at high carrier densities. At low carrier densities, such short-wavelength roughness results in qualitatively different mobility behavior than gradual UTSOI channel thickness fluctuations.

An effective Hamiltonian technique to calculate short-time, non-equilibrium fluctuations in quantum devices is developed. Applications to quantum dots and resonant tunneling diodes show that temporal fluctuations are reproduced well.

Thesis Supervisor: John D. Joannopoulos
Title: Francis Wright Davis Professor of Physics

Acknowledgments

This thesis would not have been possible without the help and support of numerous people. I have had the unusual privilege of working closely with two advisors during my graduate career: Professor John Joannopoulos of MIT and Professor Sokrates Pantelides of Vanderbilt University. From the first semester of "Theory of Solids", John impressed me with his careful teaching and clear presentation of the diverse concepts of condensed matter physics. This clarity carried over into my work with him. It was only with John's help that I was able to identify and explain the key results of this thesis. Sok's guidance during my time at Vanderbilt University has been invaluable. His extensive experience with semiconductor physics make my work on self-trapping and mobility possible. It was Sok's insightful questions during our first meeting at Vanderbilt that led us to study boron nanotubes. Both John and Sok have mastered what is perhaps the most important skill for a physicist: choosing good problems. They have an almost innate ability to identify interesting and feasible projects that push the boundaries of their fields. For a graduate student, this makes John and Sok ideal advisors, as they have provided a boundless supply of interesting problems. Both John and Sok are excellent speakers and writers, and have taught me how to present my results effectively. Working with two brilliant advisors has given me every opportunity to learn the skills that are necessary to succeed as a physicist. I can only hope that my current and future work shows that I've taken advantage of these chances.

There are many other people at MIT and Vanderbilt that have helped me throughout my years in graduate school. Without Margaret O'Meara, Debbie Frizzell, and Valerie Mauro, I would have succumbed to red tape or been lost on a trip long ago. Along with John, Professors Mehran Kardar and Eric Hudson of MIT made up my thesis committee. Their tough questions helped to focus and clarify my work. Professors Ron Schrimpf and Dan Fleetwood of Vanderbilt University taught me a great deal about the electrical engineering side of semiconductor devices. Dr. Xiaoguang Zhang of Oak Ridge National Laboratory was an invaluable advisor on my quantum

transport work. His ideas also inspired my approach to mobility calculations. My coworkers, graduate students and post-docs, often helped to round out ideas and provided valuable comments and criticism. In particular, I would like to thank Dr. Leonidas Tsetseris, Ryan Hatcher, Peter Bermel, Dr. Kalman Varga, and Dr. Sergey Rashkeev.

Without the love and support of my family, this long journey through graduate school would never have begun, much less have been completed. My grandfather, Dr. Hiram Evans, is the reason I've loved science for as long as I can remember. From making DNA out of pipe cleaners to finding and repairing a microscope, he was always there to teach and inspire me. I would not be where I am, nor who I am, without him. The rest of my family, my mother, brother, and grandmother, have always been behind me. They've asked about my research, though there must have been more interesting things to talk about, and provided a warm home when I just wanted to get away. Last, but certainly not least, is my lovely wife, Carmen. She has always been there for me: listening to me practice talks, putting up with me when I worked weekends, and not complaining when magazines and journal articles threatened to collapse our nightstand. As a brilliant engineer who was often genuinely interested in what I was working on, I knew that if I couldn't explain something to her, then it was my understanding that was faulty.

Contents

1 Introduction **19**

1.1 Small is Different
(with apologies to P. W. Anderson) 20

1.1.1 Nanoscale Size Effects 20

1.1.2 Applications of Nanostructures 21

1.1.3 Computational Challenges 22

1.2 Outline 23

1.2.1 Boron Nanotubes 24

1.2.2 Self-Trapping of Holes in Silicon Dioxide 26

1.2.3 Mobility in UTSOI MOSFETs 27

1.2.4 Interface Roughness Scattering 28

1.2.5 Charge and Spin Transport in Quantum Devices 29

2 Boron Nanotubes **31**

2.1 Introduction 31

2.1.1 Computational Method 32

2.2 Planar Boron Structures 33

2.2.1 Ground State Degeneracy 35

2.3 Boron Nanotubes 36

2.3.1 Structural and Electronic Properties 36

2.3.2 Elastic Properties 40

3 Self-Trapping of Holes in Silicon Dioxide 45

3.1 Hole Transport in SiO_2 45

3.1.1 The Continuous-Time Random Walk Model 46

3.1.2 The Multiple-Trapping Model 48

3.2 Hole Trapping Centers in SiO_2 49

3.2.1 Defect-Associated Traps 49

3.2.2 Impurity-Associated Traps 50

3.2.3 Self-Trapping Centers 50

4 First-Principles Mobility Calculations 59

4.1 Introduction 59

4.2 Method for Mobility Calculations 60

4.2.1 Ground-State Calculations 60

4.2.2 MOSFET Electrostatics 63

4.2.3 First Approximation: Born and Boltzmann 66

4.2.4 Green's Functions: Beyond Born and Boltzmann 69

4.2.5 Phonon Scattering 77

4.3 Electronic Structure of UTSOI Channels 82

4.3.1 Wavefunction Penetration 83

4.3.2 Volume Inversion 89

4.3.3 Self-Consistent Electrostatics 94

5 Interface Roughness Scattering 99

5.1 Introduction 99

5.1.1 Existing Models 100

5.1.2 Atomic-Scale Roughness 102

5.2 First-Principles Results 103

5.2.1 Scattering Potentials and Screening 103

5.2.2 Calculated Mobilities 105

5.2.3 Comparison to Existing Models and Experiments 107

6 Quantum Transport **111**

6.1 Introduction 111

6.1.1 Non-Equilibrium Green's Functions 115

6.1.2 Quantum Fluctuations 121

6.2 Effective Hamiltonian Method 122

6.2.1 A New Basis 123

6.2.2 Effective Hamiltonian 126

6.3 Applications 130

6.3.1 Quantum Dot 130

6.3.2 Resonant Tunneling 133

7 Conclusions **137**

7.1 Computational Design 137

7.2 Transport and Non-Equilibrium Phenomena 139

7.3 Future Work 140

7.3.1 First-Principles Mobility Calculations 140

7.3.2 Effective Hamiltonian Method 142

List of Figures

1-1 Schematic representation of a double-gate SOI MOSFET. Solid black regions denote conducting gates. The substrate is generally a heavily-doped silicon wafer. The undoped silicon channel is surrounded by the top and buried oxides. The two gates can be biased separately with voltages V_B (top gate) and V_{BG} (buried gate). Source (V_S) and drain (V_D) biases are applied to the silicon channel through conducting source and drain electrodes (not shown). Thicknesses of the various regions are shown on the right. 28

2-1 Contour plot of the electron density for boron planes, in electrons/Å^3. The ionic cores are shown as black circles. Density fluctuations near the cores are characteristic of the pseudopotential. 34

2-2 Schematic band structure for the hexagonal boron and carbon planes. 34

2-3 Flat triangular lattice electronic bands for boron (solid) and free electrons (dashed), and boron density of states. The zero of energy is set at the Fermi energy. 35

2-4 p_x and p_y orbitals drawn schematically for a single atom on a triangular lattice (in the x-y plane). 36

2-5 The stable buckled triangular plane. 36

2-6 Definition of chirality for triangular lattice nanotubes. The primitive vectors are $\vec{a}$ and $\vec{b}$. The dotted lines show the circumference vectors for the $(4,4)$ and $(4,0)$ nanotubes. 37

2-7 The $(4, 4)$ boron nanotube. Left: ball and stick structure model. Right: valence electron density isosurface at 0.67 electrons/Å^3. The ionic cores are shown as dark spheres. 38

2-8 Ball and stick cross sections of boron nanotubes. 39

2-9 Cross section of a proposed $(6, 6)$ boron nanotube structure in which the sides buckle. The tube is not symmetric with respect to the mirror planes shown by dotted lines. 39

2-10 The $(8, 0)$ boron nanotube. Left: ball and stick structure model. Right: valence electron density isosurface at 0.67 electrons/Å^3. The ionic cores are shown as dark spheres. 40

2-11 Ball and stick cross section of the $(8, 8)$ boron nanotube, with vectors showing lateral relaxations under positive longitudinal strain. Longer vectors represent larger relaxations, but are not to scale. 42

3-1 Schematic of a six-member ring from the amorphous SiO_2 network. Oxygen $2p_z$ orbitals are perpendicular to the Si-O-Si plane. σ or mixed σ-π bonds can form between $2p_z$ orbitals as the oxygen atoms are pulled together. 52

3-2 Molecular level diagram for Si-O-Si bonds in amorphous SiO_2. **A** represents the isolated O and Si atoms, with $2p$ and hybridized sp^3 orbitals. **B** represents the levels after formation of an Si-O-Si bond. The bonding orbitals are hybrids of Si sp^3 orbitals and O $2p_x$ and $2p_y$ orbitals. The O $2p_z$ orbital is perpendicular to the Si-O-Si plane and cannot interact with the Si orbitals due to symmetry. **C** represents the formation of a weak bond between O atoms on opposite sides of a ring. The non-bonding $2p_z$ orbitals split into σ-bonding and -antibonding orbitals. In the presence of a hole, the σ-antibonding orbital is half-occupied, resulting in a net energy gain from the bond of Δ. 53

3-3 Valence electron density isosurface showing hole density distributed over oxygen $2p_z$ orbitals in an amorphous SiO_2 supercell. O atoms are shown in red, Si atoms in blue. 53

3-4 Valence electron density isosurface showing hole density in a σ-antibonding orbital. O atoms are shown in red, Si atoms in blue. Pulling two oxygen atoms together towards the center of the six-member ring creates an O_2^--like molecule. 54

3-5 Statistics of oxygen-oxygen pairs in amorphous SiO_2 supercells. The supercells contained 526 total pairs with O-O distances < 5.0Å. . . . 55

3-6 Energy cost versus fractional displacement for a representative O-O pair in the presence of a hole. The oxygen atoms are moved along the line between them; fractional displacement is defined as the fractional change in the equilibrium O-O distance. 56

4-1 The valence and conduction band edges across the Si-SiO_2 interface, calculated by projecting the density of states onto planes of atoms. (Plot courtesy of S. T. Pantelides) 61

4-2 Schematic of the external potential Φ in a symmetric double-gate UTSOI MOSFET. Dotted lines divide the silicon channel and oxide regions. 65

4-3 Schematic of a single-gate MOSFET simulation in an extended supercell. The symmetric external potential Φ_{ext} reflects a constant electric field confining the carriers at each silicon-oxide interface. This results in a realistic simulation so long as the wavefunctions of interest (represented by dotted lines) do not overlap. 65

4-4 Supercell containing a 5Å-thick UTSOI silicon channel with ideal, abrupt Si-SiO_2 interfaces. 67

4-5 Schematic representation of wavefunction penetration into the oxide. The wavefunction is the solid line within the potential well defining the Si channel. The oxide is represented by the shaded regions, and the height of the oxide potential barrier is Φ_{ox}. The channel thickness is T_{Si}, and the penetration depth of the carriers is L_p. 83

4-6 Calculated conduction electron densities for 10Å-, 15Å-, and 20Å-thick channels, with the channel density fixed at $5.6 \times 10^{-11} e^-/\text{Å}^2$. The z axis is perpendicular to the Si-SiO_2 interface. Vertical lines mark the positions of the Si-SiO_2 interfaces; line styles correspond to those in the legend. Electron densities are given as linear densities, integrated over planes parallel to the interface. 85

4-7 Penetration length L_p versus $1/T_{Si}^2$, as calculated from Figure 4-6. Circles show calculated points, solid line is a linear fit. 86

4-8 Plane-averaged electron densities in a 10Å-thick silicon channel. Densities are scaled to have equal weight. The z axis is perpendicular to the Si-SiO_2 interface. 88

4-9 Schematic representation of applied bias potential and electron densities in a thin SOI device. Shaded regions represent the oxide. The dotted line shows the symmetric gate bias potential. The solid line shows the carrier density for a thicker channel, with carriers divided into front and back channels. The dashed line represents the carrier density in a volume-inverted channel. 89

4-10 Schematic of the potential well volume inversion model. The applied external potential ϕ_{ext} is shown as a dotted line, while the dashed line shows the carrier density model n. The oxide is represented by dashed region, with the oxide potential barrier assumed to be infinite. 91

4-11 Plane-averaged electron densities in a 20Å-thick silicon channel. Densities are scaled to have equal weight. The z axis is perpendicular to the Si-SiO_2 interface. 93

4-12 Schematic of a T_{Si}-thick silicon channel, T_{Ox}-thick oxide, and metal gate. The axis of the cylinder is perpendicular to the silicon-oxide and metal-oxide interfaces. The cylinder has length z and base area A. . . 94

4-13 Plane-averaged external electrostatic potentials for a 2nm-thick channel at carrier electron densities of $5.6\times10^{11}e^-/cm^2$ and $5.6\times10^{13}e^-/cm^2$. Both potentials are scaled to be equal in magnitude, to facilitate shape comparisons. The z axis is perpendicular to the interface. 96

4-14 Effective field versus conduction electron density for a 20Å-thick silicon channel. The line is a guide to the eye. 97

5-1 Schematics of an Si-SiO_2 interface showing elementary interface roughness defects. 103

5-2 Plot showing plane-averaged scattering potentials for oxygen protrusions (top) and suboxide bonds (middle), along with carrier density (bottom). The z axis is perpendicular to the Si-SiO_2 interface. Black circles mark the defect centers. 104

5-3 Plane-averaged suboxide bond scattering potential at carrier electron densities of $5.6\times10^{11}e^-/cm^2$ and $5.6\times10^{13}e^-/cm^2$. The z axis is perpendicular to the interface. The inset shows the height of the depth of the potential well at $z = 3.0$Å as a function of carrier electron density. 104

5-4 Calculated electron mobilities due to scattering from suboxide bonds and oxygen protrusions, for channel thicknesses of 10, 15, and 20Å. . 105

5-5 Plane-averaged wavefunction at the bottom of the conduction band in a 10Å-thick UTSOI channel. The z axis is perpendicular to the Si-SiO_2 interface. A black circle marks the center of the oxygen protrusion. . 106

5-6 Suboxide-bond-limited electron mobilities calculated with screened and unscreened scattering potentials. The channel thickness is 10Å. . . . 107

6-1 Schematic of a quantum device connected to two leads. The left and right leads are metals in thermal equilibrium, with chemical potentials μ_L and μ_R. The energy levels for the device, labeled $0, 1, 2, \ldots$, are discrete and widely spaced in the absence of the leads. 112

6-2 Schematic representation of a resonant tunneling diode. On the right is a representative current-voltage (I-V) curve showing negative differential resistance. A, B, and C as labeled on the I-V curve correspond to the band diagrams on the left. Metallic leads are represented by the dotted regions, barriers are dashed regions. The dashed line marks the energy of the bound-state between the barriers. After W. R. Frensley [49]. 113

6-3 Schematic representation of a single-electron transistor. The gate voltage shifts the position of the bound-state energy levels in the potential well. On the right is a representative gate voltage-conductance curve, inspired by the low-temperature data of Takahashi *et al.* [131]. Points A, B, and C on the curve correspond to the band schematics on the left. The dotted regions represent the metallic source and drain, the shaded regions are barriers. The dashed line within the potential well shows the bound-state energy level. The dotted line in B and C represents the energy level for a doubly-occupied bound state, e^2/C above the singly-occupied level due to the charging energy. 114

6-4 The Keldysh countour. The contour extends along the time (t) axis from $-\infty$ to ∞ and back to $-\infty$. The $+$ branch corresponds to increasing time, while the $-$ branch corresponds to decreasing time. . . 118

List of Tables

2.1 Summary of the cohesive energy E_{coh}, the curvature energy with respect to the buckled plane E_{curv}, the equilibrium diameter d, the modified Young's modulus Y_s (see Equation 2.1), and the Poisson ratio σ for boron nanotubes. 38

4.1 Calculated penetration length L_p versus channel thickness T_{Si}, from Figure 4-6. Measurement definitions as in Figure 4-5. 84

Chapter 1

Introduction

One atom is not the same as a collection of 10^{23} atoms. This important, if obvious, observation has motivated condensed matter physicists since the 1930's. Quantum mechanics has proven to be the key to explaining the wide variety of material properties observed in nature. The early development of band theory permitted physicists to explain the difference between metals and insulators, as well as to understand why quartz is transparent and gold is not. Equally powerful theories were developed to explain superconductivity, magnetism, superfluidity, and other phenomena unique to 10^{23}-atom systems.

From the beginning of condensed matter physics, analytical results on real materials were hard to come by. Solving Schrödinger's equation for 10^{23} interacting electrons is an impossible task. As a result, approximations are necessary to strip away unnecessary complications and illuminate essential physics. Simplifying approximations are very useful in explaining unusual phenomena observed in real materials, but determining why a particular phenomenon occurs in a particular material requires additional detail. Simple theories can explain the difference between metals and insulators, but why should carbon appear naturally in both forms? Understanding how diamond differs from graphite requires an ability to calculate how different arrangements of carbon atoms result in different properties.

The development of Density Functional Theory by Hohenberg, Kohn, and Sham [70, 82] opened the door to modern calculations of material properties from first-principles.

The translational and other symmetries of crystal structures, in conjunction with Bloch's theorem [19], allowed the 10^{23} atoms in a macroscopic crystal to be reduced to the handful of atoms in the unit cell. Many experimental efforts of the last two decades have focused not on macroscopic materials, however, but rather on nanometer-scale structures that can be engineered to display unique properties. This has presented new challenges for computational theorists, because such structures can contain hundreds or thousands of atoms and do not have lattice symmetries.

The work described in this thesis demonstrates how computational techniques based on Density Functional Theory can be used to predict and understand the novel properties of nanoscale structures. In addition, new methods are described that permit the calculation of charge and spin transport properties. The ability of theorists to calculate currents and voltages measured experimentally in nanoscale devices not only allows for direct comparison between experiment and theory, but also allows theory to contribute to the design and development of novel nanoscale electronic devices.

1.1 Small is Different
(with apologies to P. W. Anderson)

1.1.1 Nanoscale Size Effects

Nanometer-scale structures can display properties very different from their macroscopic counterparts. Macroscopic gold surfaces are non-reactive, resulting in characteristic luster of gold jewelry, but nanoscale gold clusters are very effective catalysts [151]. Carbon is found in nature in two forms: diamond, a hard insulator, and graphite, a soft semimetal. At the nanoscale, however, carbon displays a rich variety of structures [38]. The C_{60} molecule, or "buckyball", is a carbon cluster in the shape of a soccer ball, whose hollow interior and stable exterior could find a range of applications. Carbon nanotubes, hollow tubes in the shape of rolled-up graphene sheets, show great tensile strength and stiffness, and are either metallic or semiconducting depending on their structure. Nanotubes have inspired numerous ap-

plications, including nanoscale transistors [132, 12] and strong, light-weight composite materials [7, 96, 136].

The origin of these unique nanoscale structures and properties are the size effects that result from forming clusters of tens or hundreds of atoms, instead of the 10^{23} atoms in a macroscopic crystal. The surface-to-volume ratio for such nanometer-scale clusters is not small, and as a result there will be many surface atoms with different coordination and bonding patterns than those in the bulk. Surface reconstructions will be very important, and the resulting stable structures may be very different than their bulk or macroscopic surface counterparts. An extreme example are carbon nanotubes and fullerenes, discussed above. Two-dimensional sections of graphene sheets will not remain flat, but prefer to roll into tubes or balls.

Another important size effect relates to the balance between kinetic and potential energy in nanoscale structures. In macroscopic metals, valence electrons behave as nearly-free particles that screen external fields with frequencies below the plasma frequency. The screening effect is largely driven by potential energy: electrons are attracted to the "valleys" of the applied potential and repelled from the "peaks". As a result, the valleys become less deep and the peaks smaller. In nanostructures, however, kinetic energy plays a more important role. The kinetic energy cost to localize electrons within the nanostructure is already high, and the further localization necessary to respond to an applied field may be prohibitive. In small nanostructures, kinetic energy may dominate the potential energy and turn a cluster of metal atoms into a dielectric.

1.1.2 Applications of Nanostructures

By understanding the novel structures formed by atoms at the nanoscale, it is possible to design materials with desirable properties from the ground up. For example, carbon nanotubes in isolation are stiff and strong, and could make excellent materials for cars, aircraft, and other systems where strength is needed but weight is at a premium. To go from the nanometer scales of individual tubes to meter-scale beams and sheets, nanotubes can be incorporated into a polymer matrix or other flexible, light-weight

material. Optimizing the performance of these nanotube-polymer composites requires a detailed understanding of the chemical and mechanical properties of the nanotubes. Computational methods permit relatively rapid "experimentation" on candidate nanotubes to find those with the best combination of lateral and longitudinal stiffness and strength.

In addition to engineering materials that exploit nanoscale structures to achieve desirable macroscopic properties, nanometer-scale devices can be designed to continue the miniaturization trend in electronics and sensors. Novel transistor structures may be constructed out of nanotubes [76] or even from single molecules [28]. Quantum fluctuations in such nanoelectronic devices are an important influence on the behavior of individual devices and circuits. In order to properly design nanoscale circuits, computational methods must be developed that can account for charge and spin fluctuations over femtosecond time scales.

Even before such conceptual devices emerge in large-scale integrated circuits, the silicon-based technology in use today will enter the nanoscale regime. The semiconductor industry's technology roadmap predicts that the length of the conducting channel in MOSFET transistors will shrink to 10nm by 2015 [1]. New technologies may reduce the thickness of that channel to as little as 1nm in that time frame, and the thickness of the insulating gate dielectric will fall well below 10nm as well. Within the next decade, the semiconductor industry will have to confront the challenges of designing and manufacturing truly nanoscale devices. Since the industry is dependent on accurate modeling of device structures to shorten the development cycle and improve performance, computational theory must be able calculate the electrical response of nanometer-scale silicon structures.

1.1.3 Computational Challenges

Computational Design

First-principles calculations on nanoscale structures present a number of challenges. The absence of long-range translational symmetries in nanometer-scale clusters means

that very large unit cells must be used, often with significant amounts of vacuum. The wide variety of possible structures that can be realized by tens or hundreds of atoms, and their dissimilarity to bulk crystal phases, means that physical and chemical intuition is often the only guide to choosing candidate structures. Careful study of higher-dimensional structures with translational symmetries can guide the search for stable nanostructures. For example, planar structures can be conceptually "rolled up" into nanotubes. Computational design requires not only an ability to predict nanoscale structures and their properties, but also a framework within which to explain them.

Transport and Non-Equilibrium Properties

Transport properties offer an additional challenge because the field of quantum transport calculations is relatively new, and direct comparison with experimental data is challenging. Much work has been done to calculate ballistic transport through molecules, where scattering events can be neglected and contact resistance dominates. There is a need for new methods that can properly handle the diffusive transport often seen in nanoscale semiconductor devices, where scattering with defects and impurities is important. Recent experiments have succeeded in measuring femtosecond fluctuations in currents through quantum devices, and revealed interesting temporal correlations. New methods are needed that can account for this short-time behavior in quantum devices. It is the goal of this thesis to demonstrate that the challenges inherent in the calculation of nanoscale structure and transport are surmountable, and to show how first-principles computational methods can inform experiments and device design.

1.2 Outline

The work presented in subsequent chapters of this thesis illustrates how theoretical and computational methods can be used to predict and explain experimental results on nanoscale structures. The prediction of novel boron nanotube structures and elas-

tic properties shows how an understanding of atomic properties can narrow the range of candidate structures that must be investigated. The orbitals of the boron atom explain the planar structure. The buckled boron plane in turn leads to tubular structures. Structural properties of boron nanotubes reveal internal degrees of freedom that lead to unusual elastic properties. Understanding the less complex planar structures enables the computational design of novel tubular structures with unique and desirable properties.

The investigation of self-trapped hole structures in silicon dioxide is an important link in the understanding of hole transport in amorphous SiO_2 films. Experiments can validate a statistical theory, but without a link to atomic-scale structure, efforts to reduce the incidence of trapped holes and improve the radiation hardness of MOSFETs can stall. In nanoscale structures, such as thin SiO_2 gate dielectrics, atomic-scale structures can influence macroscopic transport properties.

The connection between structure and transport is again highlighted by mobility calculations in ultra-thin silicon-on-insulator MOSFETs. First-principles calculations provide an accurate description of the structure and electronic properties of ultrathin silicon channels, and by developing a method to calculate mobilities, the relation between atomic-scale structure and measured device properties can be established. Atomic-scale interface roughness scattering illustrates how scattering phenomena that are uniquely important in nanoscale channels lead to qualitatively different electrical behavior.

Quantum devices present new challenges. Such small devices are amenable to first-principles calculations, but a basic understanding of quantum transport is lacking. Existing statistical theories can describe time-dependent currents and voltages in quantum devices, but cannot account for short-time fluctuations. In nanoscale structures, important phenomena can occur at femtosecond time scales.

1.2.1 Boron Nanotubes

Chapter 2 presents novel boron nanotube structures that display tunable elastic properties. Boron is carbon's neighbor in the periodic table, with three $2p$ valence elec-

trons. Unlike carbon, however, boron's three naturally-occurring crystalline phases are complex rhombohedral structures based on 12-atom icosahedra. It stands to reason, though, that like carbon, boron could form three sp^2 bonds and have a metastable hexagonal planar phase similar to graphene. Although the sp^2-bonded hexagonal boron plane is stable, it has a low cohesive energy. A six-fold coordinated triangular plane has a much higher cohesive energy with metallic bonds and a nearly uniform electron density. Frustration in the boron $2p$ orbitals leads to a buckling instability in the flat triangular plane, making the stable planar phase of boron a six-fold coordinated pleated triangular sheet.

The pleats in the buckled plane break the six-fold symmetry of the flat triangular plane, and define a preferred direction. When the plane is rolled into a boron nanotube, the orientation of this preferred direction strongly influences the structure and properties of the resulting tube. The two classes of achiral boron nanotubes were investigated, and were found to have sharply different electronic structures and elastic properties. Carbon nanotubes have electronic properties that depend strongly on chirality: $(n, 0)$ (zigzag) tubes are semiconducting, while (n, n) (armchair) carbon nanotubes are metallic. Boron nanotubes are the elastic analogue of carbon nanotubes: $(n, 0)$ boron nanotubes are stiff longitudinally but soft laterally, while (n, n) tubes are softer longitudinally but rigid laterally. Carbon nanotubes offer tunable electronic properties, while boron nanotubes offer tunable elastic properties.

As a result of their chirality-dependent elastic properties, boron nanotubes could be used to optimize the performance of composite materials. $(n, 0)$ boron nanotubes could provide high strength and stiffness, while the lateral rigidity of (n, n) tubes could allow for optimal stress transfer from the polymer matrix to the nanotubes. A composite material constructed from both types of boron nanotubes could improve upon those made from carbon nanotubes. Chapter 2 demonstrates how nanoscale structures evolve from atomic properties, how physical intuition guides the investigation of candidate structures, and how an understanding of nanostructure properties can inform the design of engineered nanomaterials.

1.2.2 Self-Trapping of Holes in Silicon Dioxide

Chapter 3 presents an investigation of self-trapped hole structures in silicon dioxide. The behavior of holes in silicon dioxide has long been of interest. SiO_2 is the most commonly used gate insulator in Metal-Oxide-Semiconductor Field Effect Transistors (MOSFETs), separating the metallic gate from the conducting silicon channel. High-energy photons can create electron-hole pairs in SiO_2, and if the MOSFET is "turned on", electric fields generated by a biased gate can quickly separate the electron from the hole. While the mobility of electrons in the conduction band of SiO_2 is quite high, the holes move much more slowly and are easily trapped in the amorphous oxide lattice. Trapped holes result in positive charges in the oxide that can both scatter conduction electron in the silicon channel and change the on/off profile of the device. Understanding how holes generated in the oxide diffuse to the gate or channel is important to maintaining device performance in the presence of radiation.

Experimental measurements of currents generated by holes in the oxide show unusually long tails with a power-law dependence. Statistical theories of diffusive transport that could explain these tails were developed in the 1970's. These theories assumed that as the holes diffused through the oxide, they encountered trapping centers, metastable localized states that held the holes for varying amounts of time. Although the statistical theories included no atomic-scale detail, the success of these simple models in describing experimental results inspired a search for the atomic origins of trapping centers. Defects, such as oxygen vacancies or impurities such as germanium can trap holes; the density of such defects can be reduced through improvements in device processing. Another important category of trap is the self-trapping center. A localized lattice distortion traps the hole in a metastable state, often called a "small polaron". Such traps are not associated with defects or impurities and are unlikely to be removed through device processing.

Electron spin resonance measurements have identified two distinct self-trapped hole signatures. One of these structures has already been identified; the other has remained elusive for over a decade. Chapter 3 presents a candidate structure that

could represent the unidentified self-trapped hole. This structure is the analogue of the V_k hole-trapping center found in the ionic alkali halides. Calculations performed on a range of configurations in the amorphous SiO_2 lattice show that holes do not self-trap in this configuration, demonstrating that this analogue of the V_k center does not occur in the partially-covalent SiO_2 system. This result lends credence to the original conjecture that a type of Anderson localization is responsible for the as-yet unidentified self-trapping center.

1.2.3 Mobility in UTSOI MOSFETs

Chapter 4 presents a new method to calculate mobility in ultra-thin silicon-on-insulator (UTSOI) MOSFETs. A schematic representation of an SOI MOSFET is shown in Figure 1-1. Silicon-on-insulator technology is predicted to play an important role in maintaining transistor scaling along the path of Moore's Law. As channel lengths grow shorter, the thickness of the channel in a UTSOI transistor can be reduced to suppress deleterious short channel effects. Next-generation UTSOI devices may have channels as thin as one nanometer, or four atomic layers of silicon.

In order to move UTSOI MOSFETs from the laboratory to the manufacturing plant, accurate simulations of such devices must be available. Modeling and simulation of novel device structures shortens the development cycle and can aid in diagnosing manufacturing problems. Current state-of-the-art MOSFET simulations work well for bulk silicon devices and have had some success describing silicon-on-insulator devices with thick channels. Since current simulations of devices employ approximations that neglect atomic-scale detail in order to gain efficiency, problems have arisen in modeling UTSOI devices.

DFT calculations can provide an accurate quantum mechanical description of the silicon channel and oxide interfaces in a UTSOI device, but for device simulations the quantities of interest are electronic transport properties. The mobility encapsulates the net effect of all scattering processes in the channel, and is readily measured by experiments. Because the mobility is a linear transport property, ground-state energies and wavefunctions from a DFT calculation provide all quantities necessary

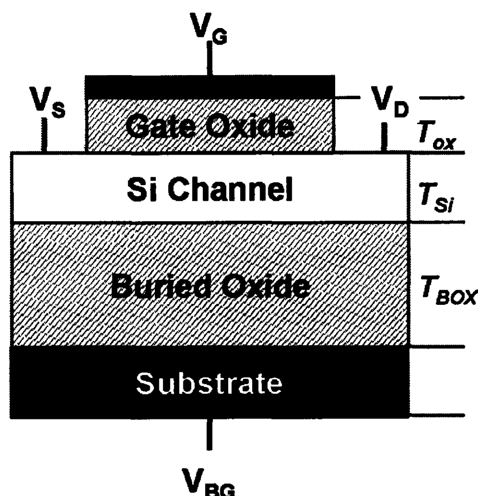


Figure 1-1: Schematic representation of a double-gate SOI MOSFET. Solid black regions denote conducting gates. The substrate is generally a heavily-doped silicon wafer. The undoped silicon channel is surrounded by the top and buried oxides. The two gates can be biased separately with voltages V_B (top gate) and V_{BG} (buried gate). Source (V_S) and drain (V_D) biases are applied to the silicon channel through conducting source and drain electrodes (not shown). Thicknesses of the various regions are shown on the right.

to calculate the mobility. Chapter 4 presents a method to calculate the mobility in a UTSOI channel, under an arbitrary gate bias and in the presence of defects, impurities, and phonons. Little computational effort is required beyond ground-state DFT calculations.

1.2.4 Interface Roughness Scattering

For UTSOI devices operating at moderate to high carrier densities, interface roughness scattering is expected to dominate the mobility. Recent models of interface roughness scattering in UTSOI channels rely on a continuum description of the interface, in which the position of the interface can vary by an arbitrary amount. It is clear that the interface position cannot vary arbitrarily, however, since it can only change by the addition or subtraction of silicon atoms at the interface. The continuum model works well in bulk devices where a single-atom fluctuation only changes the channel thickness by a small percentage, but in a one-nanometer UTSOI channel, a single displaced silicon atom changes the channel thickness by 25%!. In Chapter 5,

an accurate model of interface roughness scattering in UTSOI devices is developed, based on scattering from the elemental defects that constitute interface fluctuations. This defect scattering model marks a departure from continuum models, and represents a direct connection between the atomic-scale structure of the channel and device performance.

At low carrier densities, atomic-scale interface roughness leads to qualitative different mobility behavior than gradual changes in the channel thickness (long-wavelength interface roughness). Experiments on UTSOI MOSFETs have shown that the mobility at low carrier densities is significantly reduced relative to bulk MOSFETs. Existing models of long-wavelength roughness scattering can qualitatively reproduce this mobility reduction. Atomic-scale roughness, though it can limit the mobility at high carrier densities, is not likely to play a major role at low densities. The model of atomic-scale roughness presented in Chapter 5 thus confirms earlier speculation that the reduced mobility in UTSOI MOSFETs is due to fluctuations in the channel thickness, and not due to intrinsic roughness defects. These fluctuations can be reduced or eliminated through improved device processing techniques, leading potentially to a 100% increase in UTSOI mobility at low carrier densities.

1.2.5 Charge and Spin Transport in Quantum Devices

Chapter 6 presents a new method to calculated short-time fluctuations in the charge and spin currents through quantum devices. UTSOI technology is an important advance in silicon technology, and will improve the performance of next-generation transistors. Looking several generations into the future, transistors and other electronic device elements will shrink into a regime where quantum effects dominate. In channels less than 10 nanometers thick, for example, scattering events in the channel are unlikely and transport properties are determined largely by the energy spectrum of states in the channel. Transport through quantum dots and resonant tunneling diodes is an example of such ballistic transport that is accessible to experiments today.

Describing ballistic transport theoretically is challenging due to the need to consider both localized and extended states in the device, and to treat time-dependence

explicitly (AC currents). The Landauer-Büttiker and Non-Equilibrium (Keldysh) Green's Function (NEGF) approaches are both widely used to study time-dependent transport through quantum devices. It is difficult to treat an arbitrary bias and time dependence in the Landauer-Büttiker approach, and quantum correlation effects are often neglected. The NEGF approach can in principle provide an exact solution for quantum transport at an arbitrary bias, but including time dependence is problematic because non-equilibrium Green's functions represent a statistical average. The ergodic theorem asserts that time averaging is equivalent to a statistical phase-space average, but it is not clear *a priori* how much time is needed to properly represent a phase-space average. This need for time averaging is inconsistent with arbitrarily-fast temporal fluctuations, and highlights the need for a new method to treat the short-time behavior of quantum devices.

Chapter 6 presents an effective Hamiltonian approach to the calculation of short-time fluctuations in quantum devices. The full Hamiltonian describing a device coupled to an arbitrary number of leads is reduced to an effective Hamiltonian treating only device degrees of freedom through a block-diagonalization of the Hamiltonian matrix in Fock space. Preliminary results on charge and spin transport in a quantum dot and resonant tunneling diode are presented and compared to previously-published computational results. The current limitations of the effective Hamiltonian method and possible ways to improve it are discussed.

Chapter 2

Boron Nanotubes

2.1 Introduction

Boron, carbon's first-row neighbor, has only three valence electrons. Its natural crystalline structure is a rhombohedral lattice with 12-atom icosahedral clusters at each lattice site [103]. Nevertheless, there are some intriguing similarities with carbon. Boron's three electrons could in principle form sp^2 hybrid orbitals that might lead to planar and tubular structures similar to those formed by carbon. Since carbon nanotubes and fullerenes [38] are metastable structures, formed only under kinetically constrained conditions [39, 135], one might envision analogous boron structures. Indeed, initial results by Boustani et al. [23, 22] have demonstrated the possibility of such metastable structures with relatively low energy cost. Crystalline [106] and amorphous [25, 145, 101] boron nanowires with diameters as small as 20 nm have recently been fabricated, suggesting that boron nanotubes may already be within the range of experimental possibility.

There is an intriguing and potentially significant difference between carbon and boron, however. Boron has only three valence electrons, so that in sp^2-bonded planar or tubular boron structures the relative occupations of the sp^2- and the π-bonded bands depends on the energetic positions and dispersions of the two bands, perhaps opening up a broader range of possibilities.

In this chapter, the electronic structure and relative stabilities of planar and tubu-

lar boron structures are examined in detail. It is found that boron does form a stable sp^2-bonded hexagonal graphene-like sheet, but a planar triangular lattice has an even larger cohesive energy, though still smaller than that of the bulk α-rhombohedral structure. The triangular planar structure has an unusual property. It is essentially a homogeneous electron gas system with a threefold-degenerate ground state. This degeneracy makes the flat triangular plane unstable with respect to buckling, which breaks the symmetry and introduces a preferred direction defined by strong σ bonds. When rolled into a tube, this preferred direction, which is not present in carbon nanotubes, defines the chirality and controls the electron density, cohesive energy, and elastic response of boron nanotubes. The properties of the $(n, 0)$ tubes (proposed in References [23] and [22]) arise from the flat plane and are very similar to carbon nanotubes. The properties of the (n, n) boron nanotubes are derived from the buckled plane and contrast sharply with carbon nanotube structures. As a result of the buckling, the curvature energies of (n, n) tubes are lower than those of $(n, 0)$ tubes and show a non-monotonic plateau structure as a function of n. The buckled sides of the larger $(4n, 4n)$ boron nanotubes allow for internal relaxations that can dissipate longitudinal stress. As a result, larger $(4n, 4n)$ tubes have a very low Poisson ratio. The resulting lateral rigidity is important for mechanical stress transfer in nanotube composite materials [126].

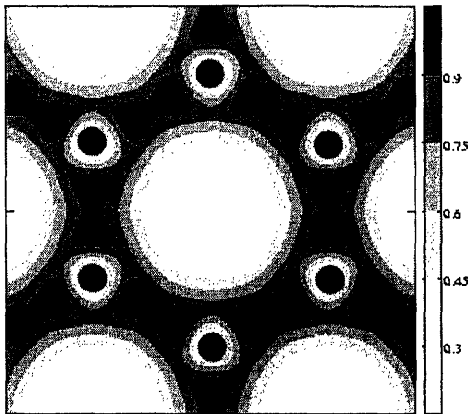
2.1.1 Computational Method

The present calculations were based on the local density approximation (LDA) to density functional theory [109], a plane wave basis set [59], and pseudopotentials [51] to represent the ionic cores. K-point meshes for Brillouin zone integration were sufficient to converge total energies to 1 mHa per atom. The boron pseudopotential accurately reproduces experimental [37] and theoretical [143] parameters for the α-rhombohedral structure.

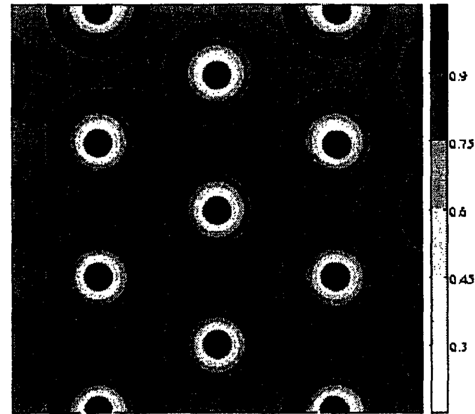
2.2 Planar Boron Structures

Figure 2-1(a) shows the in-plane electron density for the hexagonal plane. The structure, as would be expected by analogy to carbon, is sp^2-bonded, with high charge concentrations within the bond regions. However, the cohesive energy is only 5.96 eV, considerably lower than the 7.37 eV for the α-rhombohedral phase. In both boron and carbon, the three sp^2 orbitals bind in the plane, forming bonding and antibonding states that are separated by a large energy gap. The p orbitals perpendicular to the plane also bind weakly to each other (note that an isolated plane is considered here), forming bonding and antibonding bands that are degenerate at the K point in the Brillouin zone [30]. Carbon's four valence electrons fill all of the bonding bands from both the in-plane sp^2 orbitals and the π-bonded p orbitals, resulting in a zero-gap semiconducting plane. In the hexagonal boron plane, the π-bonded band crosses the sp^2 bands, and there are not enough electrons to fill them all. The result is a metallic plane without strong sp^2 binding, leading to a low cohesive energy. Figure 2-2 shows a schematic band structure for the hexagonal boron and carbon planes. The band structure of the hexagonal boron plane is similar to that of the graphene sheet, though the reduced Fermi energy makes the boron plane metallic.

Although the covalent hexagonal plane is not energetically favorable, the flat triangular plane has a cohesive energy of 6.53 eV, only 0.84 eV less than the α-rhombohedral phase. The contrast between this boron plane and graphite is striking. Figure 2-1(b) shows the in-plane electron density for the lowest energy flat triangular plane. The density is nearly uniform between the atoms, with little covalent character. Despite the nearly homogeneous electron density, the electronic bands are not free-electron-like. Figure 2-3 shows the boron bands and free electron bands from the Γ point to the X point in the Brillouin zone. The free electron mass is scaled such that the free electron Fermi energy (with three electrons per primitive cell) is equal to the boron Fermi energy.



(a) The hexagonal boron plane. Note the sp^2 bonding, also found in graphite planes.



(b) The triangular boron plane. Outside of the ionic cores, the density is nearly uniform.

Figure 2-1: Contour plot of the electron density for boron planes, in electrons/Å^3. The ionic cores are shown as black circles. Density fluctuations near the cores are characteristic of the pseudopotential.

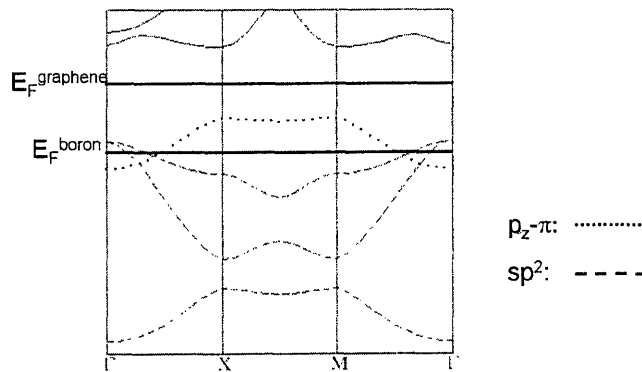


Figure 2-2: Schematic band structure for the hexagonal boron and carbon planes.

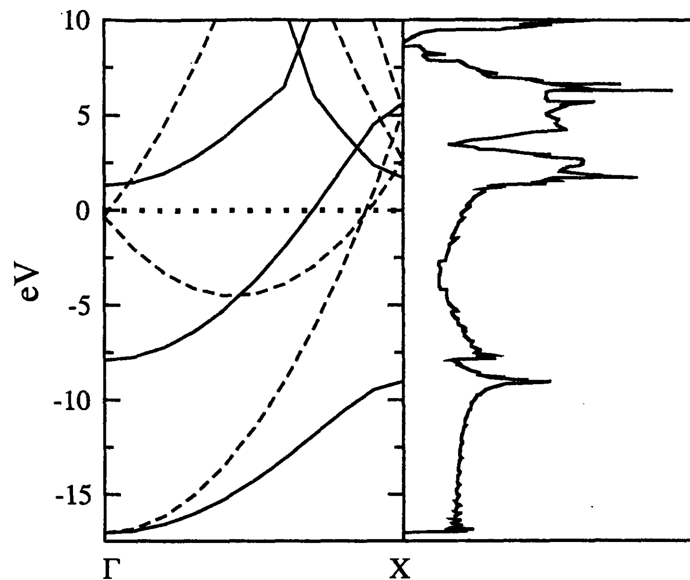


Figure 2-3: Flat triangular lattice electronic bands for boron (solid) and free electrons (dashed), and boron density of states. The zero of energy is set at the Fermi energy.

2.2.1 Ground State Degeneracy

It is important to note that the sixfold coordination of the boron atoms in the triangular phase is not compatible with the symmetries of the p orbitals (see Figure 2-4). The best that can be hoped for is to let the p_x orbital lie along a line of atoms, overlapping with the p_x orbitals of two neighbors and forming strong σ bonds. The p_y orbital will have a much smaller overlap with the four remaining neighbors, forming mixed σ-π bonds. This frustrated alignment of p orbitals leaves the ground state indeterminate: three possible σ bond directions can be chosen. In the flat triangular plane, symmetry makes these three directions equivalent, and the nearly homogeneous electron density makes a metallic bonding picture more appropriate. However, the distinction between σ and mixed σ-π bonds has important consequences when the symmetry of the flat plane is broken.

A degenerate ground state suggests that the flat plane would be unstable with respect to buckling that breaks the triangular symmetry. This instability is confirmed by the phonon dispersion [57], which has an acoustic branch with an imaginary frequency in the vicinity of the X point in the Brillouin zone. The imaginary frequency mode at the X point corresponds to vibrations that lead to the stable planar phase

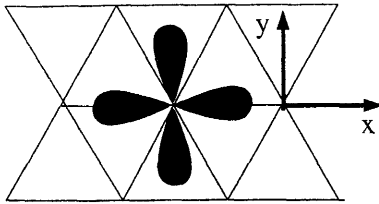
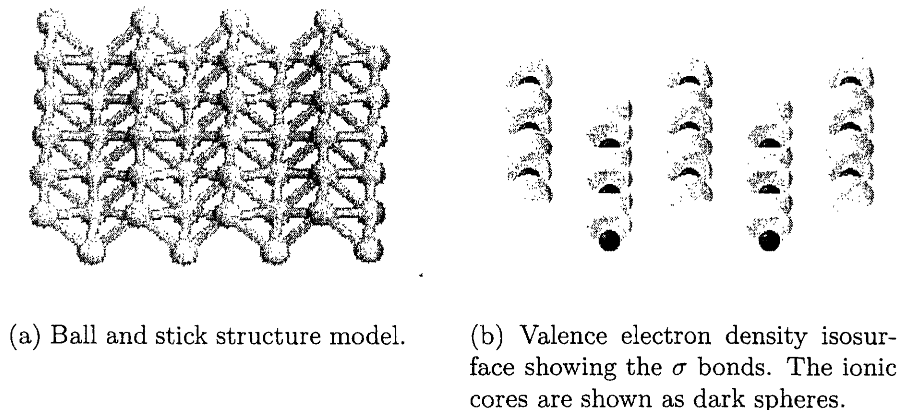


Figure 2-4: p_x and p_y orbitals drawn schematically for a single atom on a triangular lattice (in the x-y plane).



(a) Ball and stick structure model.

(b) Valence electron density isosurface showing the σ bonds. The ionic cores are shown as dark spheres.

Figure 2-5: The stable buckled triangular plane.

of boron, shown in Figure 2-5(a). Instead of the nearly homogeneous electron density of the flat triangular plane, the buckled plane shows strong directional σ bonds (see Figure 2-5(b)), as would be expected from the p-orbital model presented above. In Figure 2-5(b), the weaker mixed σ-π bonds do not contain enough electrons to appear on the isosurface. This buckled plane, which is distinct from the buckled boron sheet proposed in Reference [22], selects a preferred direction and breaks the ground-state degeneracy, raising the cohesive energy to 6.79 eV.

2.3 Boron Nanotubes

2.3.1 Structural and Electronic Properties

The preferred bonding direction in the planar phase foreshadows the importance of chirality in nanotube phases. In rolling up a triangular sheet to create a tube, the lines

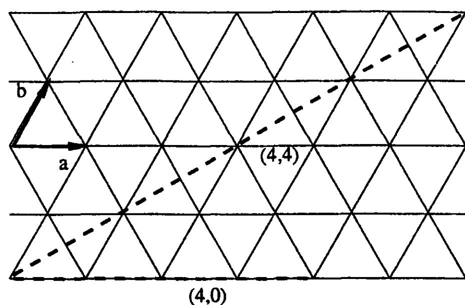


Figure 2-6: Definition of chirality for triangular lattice nanotubes. The primitive vectors are $\vec{a}$ and $\vec{b}$. The dotted lines show the circumference vectors for the $(4,4)$ and $(4,0)$ nanotubes.

of σ bonds can be chosen to run along the length of the tube, along its circumference, or to wind along it. The orientation of these bonds determines the electron density and energetics of the nanotube to an extent not seen in carbon nanotubes [69]. The chirality of a boron nanotube is defined based on the triangular lattice. An (m,n) tube is constructed by rolling a triangular plane such that the head of the lattice vector $m\vec{a} + n\vec{b}$ meets its tail; $\vec{a}$ and $\vec{b}$ are the primitive vectors of the triangular lattice. Figure 2-6 shows the vectors for a $(4,4)$ and $(4,0)$ nanotube. The dotted lines in the figure form the circumference of the tube.

Consider first the case of an (n,n) nanotube, shown in Figure 2-7. Here, the σ bond direction can be chosen to lie along the length of the tube. Given this possibility, the boron atoms will form σ bonds running along the nanotube, as the electron density isosurface in Figure 2-7 shows. The strong longitudinal bonds allow the tube to buckle laterally, emulating the buckled plane structure. Instead of the circular cross section seen in carbon nanotubes, the $(8,8)$ tube has a square cross section (shown in Figure 2-8(b)). The sides of the square are sections of the buckled plane, and the corners show only a slight distortion. In contrast, the cross section of the $(6,6)$ tube (Figure 2-8(a)) shows no buckling. With only four atoms on each side, it is not possible to buckle the sides without distorting the topology of the corners. The "buckled" structure of the $(6,6)$ tube (Figure 2-9) breaks the mirror symmetries of the actual $(6,6)$ and $(8,8)$ structures (Figures 2-8(a) and 2-8(b)), introducing chirality to an achiral tube.

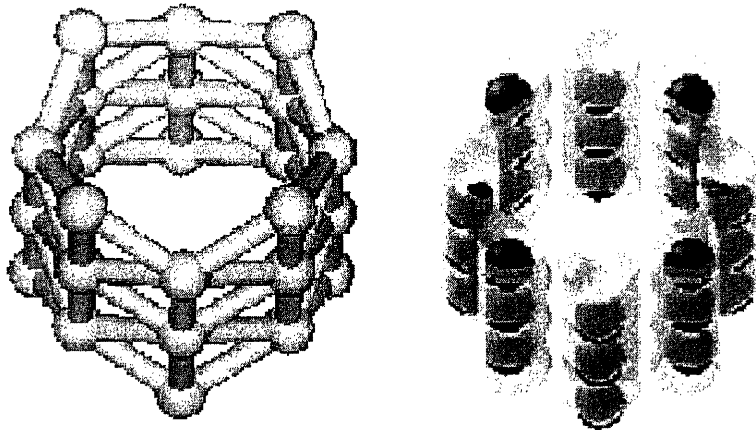


Figure 2-7: The $(4,4)$ boron nanotube. Left: ball and stick structure model. Right: valence electron density isosurface at 0.67 electrons/Å^3. The ionic cores are shown as dark spheres.

Table 2.1: Summary of the cohesive energy E_{coh}, the curvature energy with respect to the buckled plane E_{curv}, the equilibrium diameter d, the modified Young's modulus Y_s (see Equation 2.1), and the Poisson ratio σ for boron nanotubes.

Chirality	E_{coh} (eV)	E_{curv} (eV)	d (Å)	Y_s (TPa nm)	σ
(4,4)	6.71	0.08	4.34	0.29	0.5
(6,6)	6.65	0.14	5.65	0.15	0.4
(8,8)	6.76	0.03	8.48	0.22	< 0.1
(7,0)	6.36	0.43	3.99	0.49	0.2
(8,0)	6.39	0.40	4.62	0.49	0.1

Larger diameter tubes sharing the favorable buckled structure of the $(8,8)$ tube can be constructed by adding atoms to the sides of the square in pairs. Among the (n,n) boron nanotubes, therefore, $(4n,4n)$ tubes should have lower curvature energies and be more stable. This trend is suggested by the cohesive and curvature energies summarized in Table 2.1. In the table, the curvature energy is defined as the difference in cohesive energy between the tube and the plane: $E_{curv} \equiv E_{coh}^{tube} - E_{coh}^{plane}$.

An $(n,0)$ nanotube cannot align σ bonds longitudinally (see Figure 2-10). Although buckling and selecting a σ bond direction proves energetically favorable in the (n,n) tubes, it is not required by symmetry. Buckling is necessary to break the symmetry of the flat plane, but rolling up the plane into an achiral tube breaks the

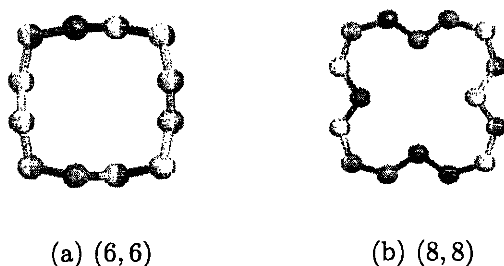


Figure 2-8: Ball and stick cross sections of boron nanotubes.

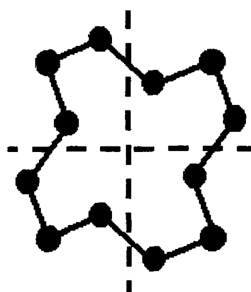


Figure 2-9: Cross section of a proposed $(6, 6)$ boron nanotube structure in which the sides buckle. The tube is not symmetric with respect to the mirror planes shown by dotted lines.

degeneracy automatically. The threefold planar degeneracy reduces to a twofold degeneracy (spiraling σ bonds related by chiral symmetry) and a non-degenerate state (σ bonds running longitudinally or laterally). It is possible for σ bonds to run along the circumference of an $(n, 0)$ tube, but the electron density of an $(8, 0)$ nanotube, shown in Figure 2-10, shows no such bonds. Instead, the density of $(n, 0)$ boron nanotubes is nearly uniform, exhibiting the free electron character of the flat triangular plane. The energetic consequences are apparent. The curvature energies of $(n, 0)$ boron nanotubes lie $0.25 - 0.4$ eV above those of the (n, n) tubes, roughly the same as the 0.26 eV cohesive energy difference between the flat and buckled triangular planes. Between the achiral limits, there may be a critical chiral angle at which boron nanotubes switch from the σ-bond-dominated electronic structure of the buckled plane to the free-electron-like structure of the flat plane. This transition could have important consequences for the behavior of boron nanotubes under torsion.

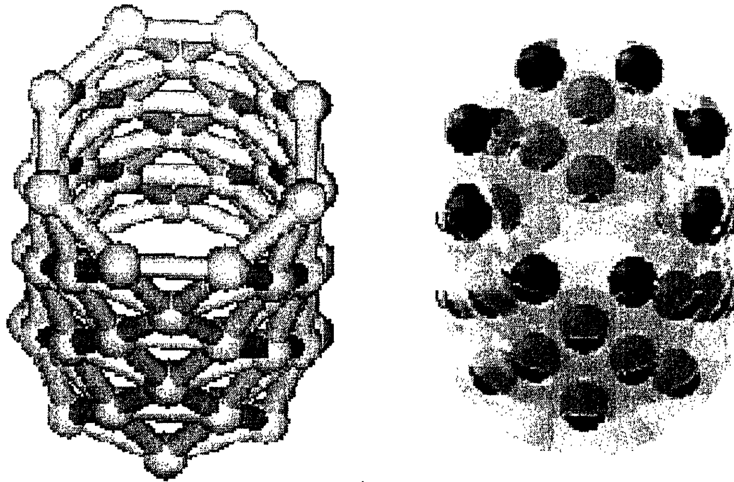


Figure 2-10: The (8,0) boron nanotube. Left: ball and stick structure model. Right: valence electron density isosurface at 0.67 electrons/Å^3. The ionic cores are shown as dark spheres.

2.3.2 Elastic Properties

Elastic properties of boron nanotubes exhibit a strong chirality dependence as well. Among the most important characteristics of a cylindrical object is the Young's modulus Y_s, defined for single walled tubes as [69]:

$$Y_s = \frac{1}{S_0}\left(\frac{\partial^2 E}{\partial \epsilon^2}\right)_{\epsilon=0}, \tag{2.1}$$

where S_0 is the equilibrium surface area, E is the total energy, and ϵ is the longitudinal strain. Since the walls of boron nanotubes are only a single atom thick, it is not possible to define the tube volume, and the modified Young's modulus must be used. For the (n,n) tubes with square cross sections, the diameter is defined as the diagonal of the square, and buckling is ignored in calculating the surface area.

Another important elastic property of a tube is the Poisson ratio σ:

$$\frac{d-d_{eq}}{d_{eq}} = -\sigma\epsilon, \tag{2.2}$$

where d is the tube diameter at strain ϵ, and d_{eq} is the equilibrium tube diameter. The

Poisson ratio measures the change in the tube's radius as it is strained longitudinally. The combination of the Young's modulus and the Poisson ratio provides information about the strength of both the longitudinal and lateral bonds of the nanotube.

The Young's modulus Y_s and Poisson ration σ were calculated for the (n, n) and $(n, 0)$ boron nanotubes discussed above. As with the binding energy, chirality plays an important role in determining these properties. It is clear from the electron density isosurface in Figure 2-7 that the strength of the (n, n) boron nanotubes arises from the bonds running along the length of the tube. Since the lateral bonds are comparatively weak, an (n, n) nanotube would be expected to expand and contract circumferentially to relieve stress, leading to a high Poisson ratio and a low Young's modulus. This is indeed the case for the small-diameter $(4, 4)$ and $(6, 6)$ tubes. For the $(4, 4)$ tube, $Y_s = 0.29$ TPa nm and $\sigma = 0.5$. The Young's modulus is comparable to boron nitride nanotubes and roughly half that of carbon nanotubes, while the Poisson ratio is nearly twice as large as the value for either carbon or boron nitride nanotubes [69].

For the $(8, 8)$ boron nanotube, however, structural dynamics play a more important role. Figure 2-8(b) shows that the cross-section of the $(8, 8)$ tube is a square with buckled sides. Using the previous definition of the Poisson ratio, we find that $\sigma = 0.1$, considerably smaller than the Poisson ratio of the $(4, 4)$ and $(6, 6)$ tubes. The diameter of the $(8, 8)$ tube does not change significantly with strain, but there is considerable lateral relaxation that results in a low Young's modulus. The simple square cross sections of the $(4, 4)$ and $(6, 6)$ tubes, as well as the circular cross sections of the $(n, 0)$ tubes, permit only uniform lateral dilation and contraction if the symmetry of the structure is to be maintained. The buckled sides of the $(8, 8)$ (and larger $(4n, 4n)$ tubes, as discussed previously) permit tube walls to relax without changing the overall square structure. Figure 2-11 shows the lateral relaxations that the $(8, 8)$ tube undergoes under positive longitudinal strain.

Internal Degrees of Freedom

Although the $(8, 8)$ boron nanotube does not change its diameter under longitudinal strain, the comparatively weak lateral bonds do allow for relaxation of the buckled

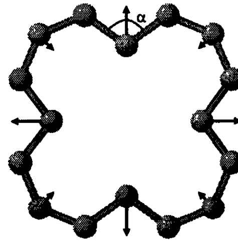


Figure 2-11: Ball and stick cross section of the $(8, 8)$ boron nanotube, with vectors showing lateral relaxations under positive longitudinal strain. Longer vectors represent larger relaxations, but are not to scale.

tube sides. As the tube is stretched, the angle α in Figure 2-11 increases. The increase in this buckling angle is also seen when the buckled plane is stretched along the lines of strong σ bonds, shown in Figure 2-5. The buckled triangular plane has a degree of freedom that is not present in either the flat triangular or the flat hexagonal plane. Buckling introduces a third dimension into the planar structure, enabling strain to be relaxed in an internal degree of freedom that does not break the planar symmetry. As the buckled plane is stretched, atoms can move perpendicular to the plane to relieve stress, an option that is forbidden by symmetry in flat planar structures. This is analogous to the fundamental difference between the phonon modes of a monatomic Bravais lattice and those of a lattice with a basis: in the latter, optical modes are present that allow relative motion without a net translation of the crystal. The larger $(4n, 4n)$ boron nanotubes with buckled sides, such as the $(8, 8)$ tube, have unique mechanical properties that arise from the presence of these internal degrees of freedom. These nanotubes are not directly analogous to macroscopic tubes: longitudinal strain is not accompanied solely by an expansion or contraction of the tube radius, but rather by an additional relaxation of the tube structure. Controlling the chirality of boron nanotubes permits stress dissipation through novel internal mechanisms unique to nanoscale structures.

The enhanced lateral rigidity seen in $(8, 8)$ boron nanotubes could prove advantageous in the design of nanotube composite materials. Composites consisting of carbon nanotubes embedded in a polymer matrix have already been fabricated and result in a high-strength, light-weight material [96, 7, 136]. As Srivastava *et al.* [126] have

pointed out, optimization of such a composite would require nanotubes with a high Young's modulus and a low Poisson ratio. The cylindrical symmetry of carbon nanotubes, only broken at high strains [120], prevents them from reducing their Poisson ratio through internal relaxations, as in $(8, 8)$ boron nanotubes. Composites using boron nanotubes could take advantage of the high Young's modulus of $(n, 0)$ tubes to provide stiffness, while the low Poisson ratio of $(4n, 4n)$ tubes could provide optimal stress transfer between the polymer matrix and the nanotubes. Boron nanotubes are the elastic analogue of carbon nanotubes: chirality dependence leads to tunable electrical properties in carbon tubes, and to tunable elastic properties in boron tubes. Just as the variation in electrical properties permits the design of carbon nanotube transistors [132, 12, 76], boron nanotubes may find use in high-strength composites and nanoelectromechanical systems [34].

In contrast to the (n, n) nanotubes, the $(n, 0)$ tubes have no dominant bonding direction. Straining the tube stresses bonds both laterally and longitudinally, making it difficult for the tube to expand or contract circumferentially. An increased Young's modulus and a decreased Poisson ratio reflects this cost. For the $(8, 0)$ boron nanotube, $Y_s = 0.49$ TPa nm and $\sigma = 0.1$. Although the radius of the $(8, 0)$ tube is only 6% larger than that of the $(4, 4)$ tube, the Young's modulus is 68% larger. This is in sharp contrast to carbon nanotubes, where tubes of similar radius have Young's moduli that differ by only a few percent [69].

In summary, the unusual properties of planar and tubular boron structures have been investigated. Although boron might be expected to form planar structures similar to carbon graphene sheets, a metallic triangular plane with a nearly homogeneous electron density has a larger cohesive energy. A threefold degenerate ground state in the flat plane makes it unstable with respect to buckling, which breaks the triangular symmetry and introduces a preferred direction. When the plane is rolled into a tube, this direction defines the chirality and controls the electronic and mechanical properties of the tube. Novel (n, n) boron nanotube structures are proposed that arise from the buckled plane and have lower curvature energies than the $(n, 0)$ tubes arising from the flat triangular plane. As a result of buckling, (n, n) boron nanotubes have a

novel internal relaxation mechanism that results in a very low Poisson ratio. The electron density differences between the flat and buckled planes explains the differences in the elastic properties of the $(n, 0)$ and (n, n) tubes. Understanding planar boron structures is crucial to understanding the unusual properties of boron nanotubes.

Chapter 3

Self-Trapping of Holes in Silicon Dioxide

3.1 Hole Transport in SiO_2

The transport properties of holes in silicon dioxide has long been of interest both experimentally and theoretically. SiO_2 is the most common gate insulator used in MOSFETs, and the electrical characteristics of the gate oxide are an important factor in device performance. Trapped charge in the gate oxide results in a shift in the threshold voltage, the gate voltage at which the transistor "turns on". Given sufficient charge build-up, the threshold voltage can fall to zero, preventing the MOSFET from turning off, and possibly disrupting the entire circuit of which it is a part.

Ionizing radiation incident on the oxide can create electron-hole pairs. The electron-hole creation energy in SiO_2 was determined experimentally to be 17 ± 1 eV [17], so that photons with energies in the keV or MeV range can be expected to produce large numbers of pairs. Such ionizing radiation is an important influence on electronics designed for space and defense applications, and a proper understanding of the dynamics of generated electrons and holes is crucial to ensuring that MOSFETs can survive incident radiation. The mobility of electrons in the conduction band of SiO_2 is quite high, so that in the presence of an applied field (such as when the MOSFET is turned on), electrons generated by ionizing radiation are swept out of the

oxide in about a picosecond [71, 105]. The holes, on the other hand, move through the oxide quite slowly and can be trapped easily, resulting in long-term shifts to the threshold voltage. The theoretical description of hole transport in amorphous SiO_2 is complicated by the range of time scales that must be considered. At 77K, measurable currents due to holes leaving the oxide persist from approximately 10^{-7} to 10 seconds after an applied radiation pulse [73]. A combination of trapping and hopping processes must be invoked to explain transit times spread over eight orders of magnitude.

3.1.1 The Continuous-Time Random Walk Model

In 1975, Scher and Montroll [123] proposed the Continuous-Time Random Walk (CTRW) model to explain observations of transient photocurrents in amorphous insulators such as As_2Se_3 [112] and the organic complex trinitrofluorenone-polyvinylcarbazole (TNF-PVK) [54]. Such photocurrent experiments are a controlled version of ionizing radiation processes in MOSFET gate insulators. A light flash creates electron-hole pairs at the surface of the insulator. An applied electric field quickly collects the electrons at the surface of the insulator, while the holes migrate to the other end, where the hole current is measured. Two key features emerged from these experiments: the transient photocurrents decayed very slowly, and when normalized by a characteristic transit time, all experimental curves could be superimposed onto a universal curve. To explain the polynomial tails and scale-free behavior of the photocurrent curves required a statistical process beyond simple diffusion.

Standard diffusion is based on a Markoffian process, in which the probability of a hop at time t depends only on the condition of the system at t. In other words, the system has no "memory". This results in the familiar relation between the mean position of the carrier distribution $\bar{l}$ and its spread σ: $\bar{l}/\sigma \propto \sqrt{t}$. Kenkre, Montroll, and Shlesinger [81] developed a generalized theory of non-Markoffian transport, based

on the following master equation:

$$d\tilde{P}/dt = \int_0^t dx\ \phi(t-x) \sum_{l'} \left[p(l-l')\tilde{P}(l',x) - p(l'-l)\tilde{P}(l,x) \right]. \quad (3.1)$$

$\tilde{P}(l,t)$ is the probability that site l will be occupied by a carrier at time t, and $p(l-l')$ gives the rate of hopping from site l' to site l. The first term on the right gives the net inflow of particles to site l, while the second term accounts for outflow. For the case of Markoffian transport, $\phi(t) \propto \delta(t)$, and the system has no memory. In general, however, the "relaxation function" ϕ describes how the previous states of the system affect the carrier distribution function at a time t.

The relaxation function is related to a distribution of hopping times $\psi(t)$ through their Laplace transforms. An exponential distribution, $\psi(t) \propto \exp(-\lambda t)$, leads to the Markoffian relaxation function described above. The exponential form corresponds to a system with a single characteristic hop that takes time $1/\lambda$ to complete. Physically, such a system would have a single trapping level, with binding energy $\hbar\lambda$. By solving Equation 3.1, Scher and Montroll demonstrated that transient photocurrent experiments characterized by scale-free behavior and long tails could be fit by a hopping time distribution of the form: $\psi(t) \propto t^{-\alpha}$, where α is a positive constant. Such a hopping time distribution, in contrast to the exponential distribution described above, has long tails and no well-defined characteristic hopping time.

Physically, a hopping time distribution with polynomial tails has a wide range of trapping levels, of varying depth. The exponential model may be appropriate for a holes hopping in a crystalline insulator such as α-SiO_2, where $\hbar\lambda$ is a measure of the overlap between localized oxygen $2p$ orbitals, but Hughes found that such a model did not explain hole transport in amorphous SiO_2 [72]. However, the CTRW model did fit the available data quite well [73], indicating that the hole transport process in amorphous SiO_2 relies on traps of various depth. Such a prediction inspired a search for potential trapping sites that continues to the present day, and that will be discussed below in further detail.

3.1.2 The Multiple-Trapping Model

Motivated by perceived shortcomings in the CTRW model, including apparent non-universal behavior in SiO_2 transport data, Curtis and Srour [36] proposed another model to explain dispersive hole transport in amorphous insulators. Their model, the multiple-trapping model, is considerably less mathematically complex than the CTRW model and predicts similar behavior over experimental time scales. The multiple-trapping model assumes that a range of traps exist at different energy levels, and begins by considering the rate of change of the concentration of trapped holes p_i at an energy level E_i:

$$dp_i/dt = c_p p N_i - e_p N_i^+. \tag{3.2}$$

p is the concentration of holes in the valence band, c_p is the capture probability, N_i is the number of trapping centers at energy E_i, $e_p \propto \exp(-E_i/k_B T)$ is the emission probability from the trap, and N_i^+ is the number of traps containing a hole. The number of holes is assumed to be small so that the number of traps not containing a hole N_i^0 is approximately equal to the total number of traps N_i.

Equation 3.2 must be solved for each level i assumed to exist in the system. The other constraint comes from conservation of charge, modified by the number of holes carried out of the insulator by the applied field:

$$dp/dt + \sum_i dp_i/dt = -k_1 p F, \tag{3.3}$$

where k_1 relates to the free carrier mobility, and F is the applied electric field.

Solving Equations 3.2 and 3.3 together with an exponential distribution of trap densities, $N_i = N_0 \exp(-E_i/\beta)$, with β a parameter, Curtis and Srour were able to fit the available experimental data on amorphous SiO_2. The success of this model relies on the same principle as that of the CTRW model: a wide range of trap depths. A complete picture of hole transport in amorphous SiO_2 requires identifying the atomic-scale origin of these traps. First-principles calculations, in conjunction with experiments, can identify trapping centers and provide an energy and hopping time

distribution to be used in conjunction with either the CTRW model or the multiple-trapping model for transport calculations.

3.2 Hole Trapping Centers in SiO_2

Trapping centers for holes in SiO_2 can be separated into three general categories: defect-associated traps, impurity-associated traps, and self-trapping sites. This section will describe defect and impurity traps briefly, and then focus on self-trapping centers (also referred to as "small polarons"), in which a localized lattice distortion in the presence of a hole creates a metastable trapped-hole state.

3.2.1 Defect-Associated Traps

Structural defects in SiO_2 are related to either oxygen vacancies, additional oxygen atoms, or under-coordinated oxygens. All three elemental defect types are associated with hole trapping centers. Oxygen vacancies have been well-studied, since they are associated with the prominent E' peaks in electron spin resonance (ESR) spectra (E'_1 in crystalline SiO_2 [45] and E'_γ [90, 146] and E'_δ [64] in amorphous SiO_2). Recent theoretical studies have established the atomic-scale structure of both the E'_1 centers in crystalline α-SiO_2 [20] and the E'_γ and E'_δ centers in amorphous SiO_2 [94, 104]. The oxygen vacancy results in a strained Si-Si bond in the neutral configuration that can break in the presence of a hole, resulting in local structures that contain either three- or fivefold coordinated silicons.

Additional oxygen atoms form what is known as a "peroxy radical", a Si-O-O-Si bond in place of the standard Si-O-Si bond. A hole can trap in the O-O bond, making the trapping center effectively an O_2^- molecule [50, 63]. Under-coordinated oxygen atoms result in "non-bridging oxygen hole centers" (NBOHCs). An NBOHC can trap a hole at the singly-coordinated oxygen atom without significant structural realignment [63, 127].

3.2.2 Impurity-Associated Traps

Silicon substitutional dopants in SiO_2 can result in hole traps on neighboring oxygen atoms. Germanium in α-SiO_2 breaks the symmetry of the Si-O-Si bond, and permits a hole to trap on an oxygen atom in an Si-O-Ge bond [67, 107]. Boron dopants in SiO_2 also result in hole trapping centers, although in this case the trapping is also associated with a strong lattice relaxation. Pacchioni, Vezzoli, and Fanciulli [108] found that the Si-O-B bond breaks, resulting in a hole trapped on a singly-coordinated oxygen. The boron atom relaxes back into the plane of its three other oxygen neighbors, and remains threefold-coordinated.

3.2.3 Self-Trapping Centers

The best-known example of a self-trapped hole is the V_k center in the alkali halides, in which a hole binds two adjacent negative ions into a negatively-charged molecule-like configuration (see, for example, Reference [10]). Self-trapping in SiO_2 was first predicted by Mott [102] in 1977. Experiments in the 1980's [61, 32] established the existence of self-trapped holes in amorphous SiO_2, but concluded that self-trapping did not occur in crystalline α-quartz. [67]. Griscom's evaluation of electron spin resonance (ESR) measurements [62] divided self-trapped holes in amorphous SiO_2 into two classes: STH_1, a hole trapped in the $2p$ orbital of a normal bridging oxygen, and STH_2, a hole apparently delocalized over the $2p$ orbitals of adjacent oxygens. Theoretical calculations based on cluster models of amorphous SiO_2 [40, 152, 79, 107] have confirmed the atomic structure of STH_1. The structure of STH_2, however, has remained elusive. Pacchioni and Basile [107] proposed a model for STH_2, but their relaxed structure contracts the O-Si-O bond angle to 84° from the tetrahedral value of 109°. This large relaxation in the presence of a hole disagrees with Griscom's prediction, based on ESR data, of small atomic motions in the vicinity of STH_2. Cluster calculations provide a relatively flexible environment that may permit large distortions that would be unlikely to occur under the constraints of the amorphous network.

With the exception of the 630-atom cluster of Zhang *et al.*, who considered only the STH_1 structure, the calculations described above did not treat clusters large enough to contain the five-, six-, and seven-member rings that constitute the amorphous network of SiO_2. The absence of rings means that previous calculations could not consider an important potential candidate for the STH_2 structure: oxygen atoms on opposite sides of a ring bound together into an O_2^- molecule-like state in the presence of a hole. This O_2^--like state has the hole distributed over two adjacent oxygen atoms, satisfying an important criterion derived by Griscom from ESR data. Since oxygen atoms in SiO_2 are negative ions, the O_2^--like state is the direct analogue of the V_k center. While V_k centers are well known in alkali halides with ionic bonds, it remains to be seen whether such a self-trapped hole state can occur in partially-covalent systems such as SiO_2.

In this chapter, first-principles calculations based on amorphous SiO_2 supercells containing five-, six-, and seven-member rings are used to investigate the structure and energetics of the O_2^--like self-trapped hole state. It is found that the O_2^--like state is not even metastable, meaning that this V_k center analogue does not occur in partially-covalent SiO_2. This result supports Griscom's original conjecture [62] that STH_2 is a type of Anderson localized state [4], localized by disorder in the amorphous network and not by atomic relaxations.

Computational Method

The present calculations used density functional theory [70] within both the local-density (LDA) [82] and generalized-gradient (GGA) [110] approximations, a plane wave basis [109], and ultrasoft potentials [142] to represent the ions. These methods are implemented in the VASP computer code [83]. Supercells, used in previous work [13, 94], were used to represent amorphous SiO_2. The cutoff energy for the plane wave basis set was 29 Ry, and the Brillouin zone was sampled at the Gamma point. A uniform negative background charge was used to maintain neutrality in the supercell.

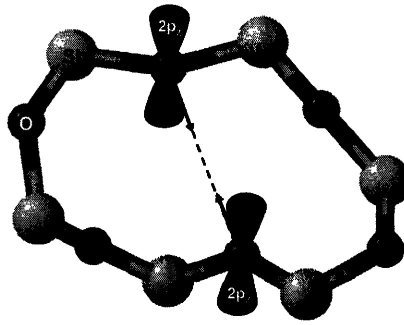


Figure 3-1: Schematic of a six-member ring from the amorphous SiO_2 network. Oxygen $2p_z$ orbitals are perpendicular to the Si-O-Si plane. σ or mixed σ-π bonds can form between $2p_z$ orbitals as the oxygen atoms are pulled together.

O_2^--like Bonding

Figure 3-1 shows schematically a six-member ring like those found in amorphous SiO_2. In the presence of a hole, oxygen atoms on opposite sides of the ring can come together, as shown by the arrows in the figure, to form an O_2^- molecule-like state. This configuration is a small polaron state that could explain Griscom's observation of STH_2. That this O_2^--like state is energetically favorable can be understood from the simplified molecular orbital diagram in Figure 3-2. In the undistorted ring, the ground state of the neutral system, the doubly-occupied oxygen $2p_z$ orbitals on opposite sides of the ring do not interact. As the oxygen atoms move toward the center of the ring, the $2p_z$ orbitals can form a weak σ bond, splitting the formerly degenerate $2p_z$ orbitals into σ_b and σ_a, bonding and anti-bonding levels. In the neutral system, both orbitals are doubly-occupied and no energy gain is expected. In the presence of a hole, however, the σ_a orbital will be only singly-occupied and we expect a cohesive energy gain of Δ. Figure 3-3 shows that the hole is distributed among oxygen $2p_z$ orbitals in the undistorted ring configuration. When the oxygen atoms are pulled together, the hole resides in a σ-antibonding orbital as shown in Figure 3-4.

As the oxygen atoms move closer together, Δ will increase. However, the overall energy of the system also includes the effect of distorting the Si-O-Si bonds, which will lead to a decrease in cohesive energy. The cohesive energy of the distorted ring in the presence of a hole is determined by the interplay between the positive $2p_z$

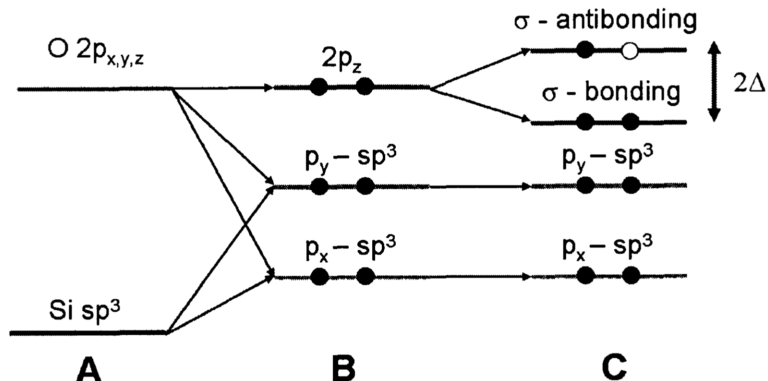


Figure 3-2: Molecular level diagram for Si-O-Si bonds in amorphous SiO_2. **A** represents the isolated O and Si atoms, with $2p$ and hybridized sp^3 orbitals. **B** represents the levels after formation of an Si-O-Si bond. The bonding orbitals are hybrids of Si sp^3 orbitals and O $2p_x$ and $2p_y$ orbitals. The O $2p_z$ orbital is perpendicular to the Si-O-Si plane and cannot interact with the Si orbitals due to symmetry. **C** represents the formation of a weak bond between O atoms on opposite sides of a ring. The non-bonding $2p_z$ orbitals split into σ-bonding and -antibonding orbitals. In the presence of a hole, the σ-antibonding orbital is half-occupied, resulting in a net energy gain from the bond of Δ.

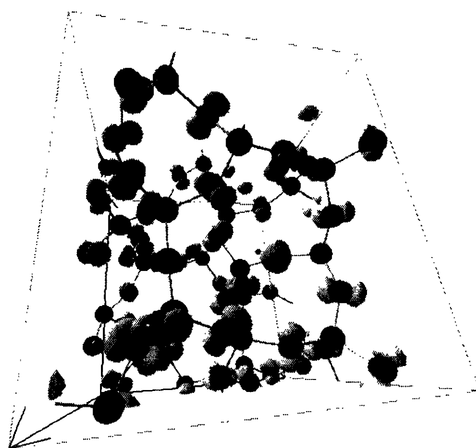


Figure 3-3: Valence electron density isosurface showing hole density distributed over oxygen $2p_z$ orbitals in an amorphous SiO_2 supercell. O atoms are shown in red, Si atoms in blue.

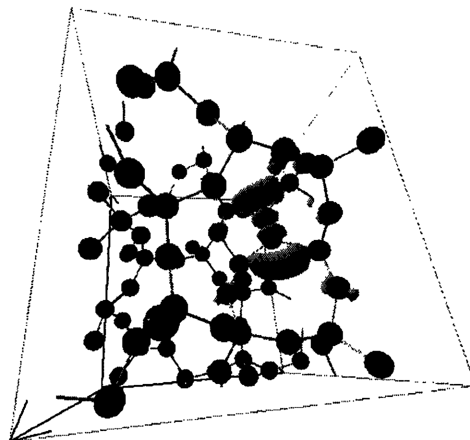
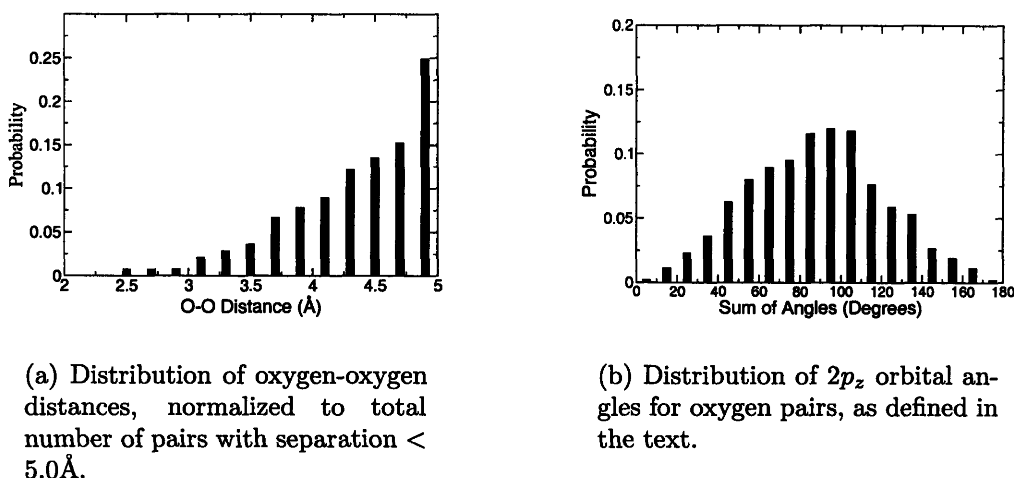


Figure 3-4: Valence electron density isosurface showing hole density in a σ-antibonding orbital. O atoms are shown in red, Si atoms in blue. Pulling two oxygen atoms together towards the center of the six-member ring creates an O_2^--like molecule.

binding energy and the negative contribution from stretching the Si-O-Si bonds. As a result, there is no guarantee that the distorted ring will represent a stable small polaron. The most likely configuration for a self-trapped hole is one in which the oxygen atoms are close together in the undistorted ring, in order to minimize the energy cost of stretching the Si-O-Si bonds, and in which the opposite $2p_z$ orbitals are aligned, in order to maximize the σ bond energy Δ. With these criteria in mind, we can sample ring configurations in various amorphous SiO_2 supercells to find those most likely to result in trapping configurations with positive cohesive energies.

Oxygen Pair Statistics

Figure 3-5 shows histograms of both the O-O distances across the ring and the alignment angle. If the $2p_z$ orbitals of oxygen atoms 1 and 2 make angles α_1 and α_2 with respect to the line between atoms 1 and 2, then we define the alignment angle A to be $A = \alpha_1 + \alpha_2$. The alignment angle is a convenient figure of merit that represents the degree of σ bonding between oxygen atoms. A pure σ bond is found when $A = 0°$, while a pure π bond is found when $A = 180°$. Four 72-atom supercells were sampled, with densities ranging from 1.99 to 2.14 g/cm^3, reflecting the variation in local density expected within amorphous material. It is important to note that the smallest



(a) Distribution of oxygen-oxygen distances, normalized to total number of pairs with separation < 5.0Å.

(b) Distribution of $2p_z$ orbital angles for oxygen pairs, as defined in the text.

Figure 3-5: Statistics of oxygen-oxygen pairs in amorphous SiO_2 supercells. The supercells contained 526 total pairs with O-O distances < 5.0Å.

O-O distance across a ring found within these cells is 2.56 Å, almost twice the O_2^- bond length of 1.34 Å as measured in solution [124]. Thus, the binding between $2p_z$ orbitals can be expected to be weak, and must be balanced by a correspondingly small Si-O-Si bond distortion if the O_2^--like hole trapping state is to be stable.

Self-Trapping Energetics

Figure 3-6 shows the energy in the presence of a hole as a function of O-O distance for a likely candidate structure with a neutral-cell equilibrium O-O distance of 2.73 Å and a $2p_z$ alignment angle of 95.61°. The energetic trends discussed below are seen in all other configurations that were investigated. Energies are calculated with respect to the zero-displacement configuration in the presence of a hole. There is no local minimum at non-zero displacement, meaning that there is not even a metastable hole-trapping configuration.

The absence of O_2^--like self-trapped hole configurations in the sampled supercells provides strong evidence for the absence of such self-trapping sites. The distribution of O-O distances shown in Figure 3-5 demonstrates that oxygen separations obey a roughly exponential distribution. Fitting the measured distribution (from 526 oxygen pairs) permits us to predict the fraction of oxygen pairs with separations of 1.34Å.

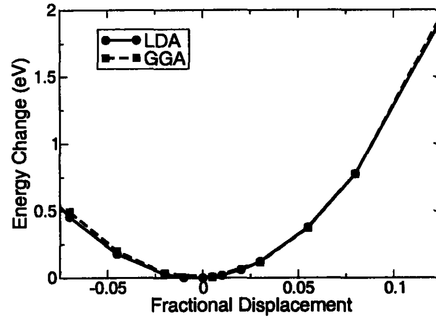


Figure 3-6: Energy cost versus fractional displacement for a representative O-O pair in the presence of a hole. The oxygen atoms are moved along the line between them; fractional displacement is defined as the fractional change in the equilibrium O-O distance.

From this fit, 0.13% of oxygen pairs will have a bond length of 1.34Å. However, the exponential fit overestimates the fraction of pairs with bond lengths less than 3.0Å. For example, we find that 0.76% of pairs have a bond length between 2.8Å and 3.0Å; the fitted distribution predicts 2.15%. In amorphous SiO_2, ring configurations with O-O distances $<$ 3.0Å are rare, and it thus becomes very difficult to predict the incidence of pairs with very short O-O distances.

The bond length of the O_2^- ion in solution is not a perfect proxy for the expected bond length of the O_2^--like V_k center analog in SiO_2, but it does provided an approximate distance over which the $2p_z$ bonding is strong. It is likely that the O-O distance required for bonding in amorphous SiO_2 is even shorter, since energy gain from $2p_z$ bonding is balanced by the energy cost to stretch the Si-O-Si bonds, and the alignment of the $2p_z$ orbitals may only permit a mixed $\sigma - \pi$ bond. As a result, given that O_2^--like self-trapped hole states do not occur in common ring configurations in amorphous SiO_2, potential anomalously short O-O distances are unlikely to account for the observed concentration of STH_2 trapped holes.

In conclusion, a potential analog to the V_k center in amorphous SiO_2 was investigated that could explain the atomic structure of the measured STH_2 self-trapped hole. Through first-principles calculations and a sampling of amorphous SiO_2 supercells, it has been established that the O_2^--like self-trapped hole state is not even metastable at normal oxygen pair separations. Though it is not possible to rule out the existence

of anomalous ring configurations in the amorphous network that could permit self-trapped hole states, it is demonstrated statistically that such configurations are too rare to explain observed STH_2 concentrations. This result lends credence to Griscom's original conjecture that the STH_2 self-trapped hole is an Anderson-localized state.

Chapter 4

First-Principles Mobility Calculations

4.1 Introduction

For four decades, industrial modeling of metal-oxide-semiconductor field-effect transistors (MOSFETs) has been based on a series of approximations that suppress atomic-scale detail. Quantum effects in thin channels are treated in the effective-mass approximation, with the oxide interfaces represented by infinite potential barriers (see Reference [6] for a review). Roughness at the interface is described by continuum models [114, 5] that account only for long-wavelength fluctuations. Monte Carlo simulations based on these approximations [46, 47, 150] have had success in describing bulk-silicon MOSFETs, but the continued evolution of silicon microelectronics deep into the nanoscale regime presents new challenges. Novel device structures such as ultrathin-body silicon-on-insulator (UTSOI) MOSFETs [35, 33, 27] will likely be needed to continue performance gains in next-generation devices [1]. UTSOI MOSFETs have been fabricated with channels as thin as one nanometer, four atomic layers of silicon [41, 140]. At these nanoscale dimensions, traditional device modeling that suppresses atomic-scale detail is no longer adequate. In this chapter, a method is presented that permits the measured electrical response of UTSOI devices to be calculated directly from first-principles, including full atomic and electronic structure

detail.

Of central importance to MOSFET modeling is the channel mobility as function of gate bias, the source-drain conductance normalized by the channel carrier density. The mobility encapsulates all the scattering processes that influence the carriers, and is a measure of how "fast" an average carrier can traverse the channel. The present method permits this scattering problem to be solved using accurate wavefunctions and self-consistent scattering potentials calculated within Density Functional Theory (DFT) [70]. Calculating the mobility using the Boltzmann equation and the Born approximation permits a direct comparison between first-principles results and existing theories. In addition, a Green's function-based method is described that allows for systematic improvements to the Born-Boltzmann method.

4.2 Method for Mobility Calculations

4.2.1 Ground-State Calculations

The present ground-state calculations were based on Density Functional Theory (DFT) [70] in the Local Density Approximation (LDA) [82]. A plane-wave basis [109] was used, and the ionic cores were represented with ultra-soft pseudopotentials [142]. These approaches are implemented in the VASP code [83]. Modifications to this code were made to enable mobility calculations, as described in the following sections. The supercell approach to describe the silicon channel and oxide interfaces is not directly applicable in the presence of a gate bias (external electric field). Solutions to this difficulty are discussed in Section 4.2.2.

A well-known limitation of the standard implementation of DFT is that the empty one-electron states in the Kohn-Sham formalism [82] do not constitute a good representation of the conduction bands of a semiconductor or insulator. The largest effect is a significant underestimation of the value of the energy gap. Figure 4-1 shows the variation of the valence and conduction band edges across the Si-SiO_2 interface. The calculated valence-band discontinuity at the Si-SiO2 interface is only 2.8 eV compared

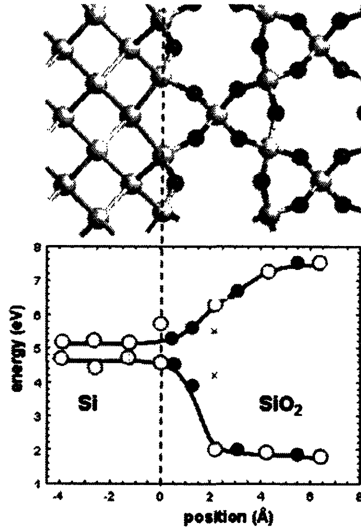


Figure 4-1: The valence and conduction band edges across the Si-SiO_2 interface, calculated by projecting the density of states onto planes of atoms. (Plot courtesy of S. T. Pantelides)

with the experimental value of 4.5 eV. The calculated conduction-band discontinuity is 2.3 eV compared with the experimental value of 3.3 eV [2]. Since the band offsets are underestimated by DFT in both the LDA and the Generalized-Gradient Approximation (GGA) [110], both conduction band electrons and valence band holes in silicon will see a reduced barrier to tunnel into the oxide. Although treating the oxide interface as a potential barrier is not an approximation employed in first-principles DFT calculations, the reduced offsets would nonetheless be expected to correlate into increased wavefunction penetration into the oxide. This phenomenon would be manifest in calculations of scattering from defects and impurities in the oxide, such as suboxide bonds (discussed in Chapter 5). A carrier wavefunction that penetrates further into the oxide results in a larger overlap with oxide defect/impurity scattering potentials, leading to a reduced mobility.

There are several available paths to overcome this potential difficulty and improve the accuracy of calculated mobilities with respect to experimental measurements. First, the band gap can be corrected with a "scissor operator" [55, 65] that shifts the conduction bands rigidly with respect to the valence bands. The utility of this method is based on experience with bulk semiconductors, in which the LDA and GGA

reproduce correctly the shape of the conduction bands, but with a reduced band gap. Figure 4-1 demonstrates that by projecting the density of states onto planes of atoms parallel to the interface, the valence and conduction band edges can be calculated as a function of position. Using this technique, it is possible to construct a position-dependent scissor operator that shifts the conduction bands on the silicon and oxide sides of the interface independently. This additional degree of freedom can result in a calculation that correctly reproduces band gaps and the conduction band offset, with little computational complexity beyond the LDA or GGA.

Another possibility is to abandon the relative simplicity of the LDA and GGA and pursue calculations in a higher-order electronic structure scheme, such as the GW approximation [68, 74, 84] or the Self-Interaction-Corrected (SIC) LDA [111, 66]. These methods yield energy gaps that agree well with experiment, but at the cost of considerable computational complexity. No GW or SIC-LDA calculations have yet been reported on the 100-200 atom supercells used in this work to represent UTSOI channels.

It is important to note that the mobilities calculated using the method presented below are calculated in channels of infinite length. That is, there is no contribution to the mobility from source or drain contact resistance, and the external gate potential is not influenced by source and/or drain fields. This is desirable from a physical point of view, since scattering in the channel is a separate process from contact resistance, and should be understood independently. However, it is not possible to completely eliminate contact resistance in real devices, and so experimental measurements will never correspond exactly to these theoretical calculations, even in the limit of long channels. The most important direct insight from these mobility calculations is the dependence of mobility on gate field, channel thickness, and carrier concentration. The band gap and band alignment difficulties described above should not affect relative mobilities calculated at different channel thicknesses or gate fields. Although it will be desirable in the future to combine this method for mobility calculations with a higher-order electronic structure method, the quantities relevant to comparison with experiment should be accurate within the LDA.

4.2.2 MOSFET Electrostatics

The influence of gate bias on the electronic structure of a silicon channel modeled with a supercell is a challenging issue due to the constraints of periodicity. The electrostatics of a symmetric double-gate MOSFET will be discussed in detail below. In this configuration, the chemical potential of the source (or drain) serves as the voltage reference. Both the front and back gates are biased symmetrically such that the potential on the front and back interfaces of the silicon channel are equal, but raised or lowered with respect to the source potential. Maintaining this potential difference requires the presence of free charge within the channel, which accumulates when the gate bias lowers the conduction bands of the channel relative to the source. [134]

DFT calculations in a periodic supercell do not directly model the presence of source and drain electrodes that can contribute carriers as the gate bias changes. The number of electrons in the supercell remains constant, making it necessary to use an inverse approach to double-gate MOSFET modeling. In the following discussion, the case of an n-type channel will be considered, in which the carriers are electrons in the conduction band. To model a p-type channel, electrons are removed from the valence band and the compensating background charge is negative. Excess electrons are introduced into the conduction band at the beginning of the calculation, and the resulting conduction electron density and electrostatic potential are calculated self-consistently.

Periodicity introduces an additional complication, however. Consistent electrostatics requires that the supercell be neutral; a charged cell would result in a divergent electrostatic contribution to the energy. Thus, when excess electrons are added to the conduction band, a uniform positive background charge must also be included to neutralize the supercell. Adding an external potential that satisfies the proper boundary conditions allows the contribution of this positive background to be subtracted out and the proper electrostatics of the charged supercell recovered. The Hartree potential in the calculation satisfies the following equation, where the z direction is

perpendicular to the Si-SiO2 interface, and variation in the x-y plane is neglected:

$$\frac{d^2\phi}{dz^2} = \frac{-e}{\epsilon_0}\left(n_{\text{ion}}(z) - n_{\text{val}}(z) - n_{\text{cond}}(z) + n_{\text{bg}}\right), \tag{4.1}$$

where n_{ion} is the positive charge of the ion cores, n_{val} is the negative charge of the valence electrons, n_{cond} is the negative charge of the conduction electrons, and n_{bg} is the uniform positive background charge. By construction, $\int dz\,(n_{\text{ion}}(z) - n_{\text{val}}(z)) = 0$ and $\int dz\,(n_{\text{bg}} - n_{\text{cond}}(z)) = 0$. If a constant external potential is added that satisfies the Hartree equation for a uniform negative background charge,

$$\frac{d^2\phi_{\text{ext}}}{dz^2} = \frac{-e}{\epsilon_0}(-n_{\text{bg}}), \tag{4.2}$$

then the total potential $\phi + \phi_{\text{ext}}$ satisfies the proper electrostatic equation for the charged channel:

$$\frac{d^2(\phi + \phi_{\text{ext}})}{dz^2} = \frac{-e}{\epsilon_0}\left(n_{\text{ion}}(z) - n_{\text{val}}(z) - n_{\text{cond}}(z)\right). \tag{4.3}$$

Since n_{bg} is constant, it is trivial to solve Equation 4.2 for the external potential that obeys the proper boundary conditions for the charged cell:

$$\phi_{\text{ext}} = \frac{e n_{\text{bg}}}{2\epsilon_0} z^2. \tag{4.4}$$

With the conduction electrons confined to the silicon channel, and the oxide represented as a dielectric medium, the resulting external potential is given in Figure 4-2.

Single-gate or non-symmetrically-biased MOSFETs present different challenges for periodic-supercell electronic structure calculations. A biased gate results in an electric field on the channel that confines electrons to the vicinity of the Si-SiO_2 interface. A constant electric field results in a linear scalar potential that is not periodic and thus cannot be accommodated without extending the supercell. Figure 4-3 shows how a constant electric field can be accommodate in an extended supercell. The external

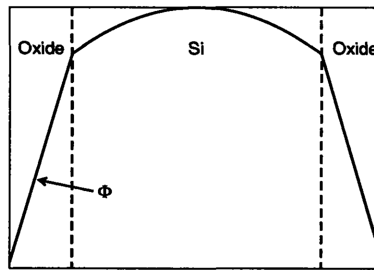


Figure 4-2: Schematic of the external potential Φ in a symmetric double-gate UTSOI MOSFET. Dotted lines divide the silicon channel and oxide regions.

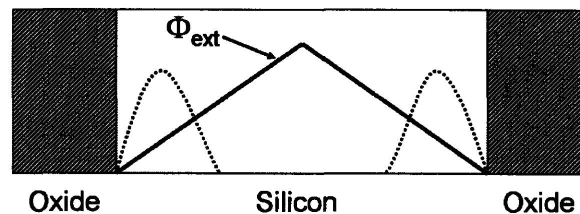


Figure 4-3: Schematic of a single-gate MOSFET simulation in an extended supercell. The symmetric external potential Φ_{ext} reflects a constant electric field confining the carriers at each silicon-oxide interface. This results in a realistic simulation so long as the wavefunctions of interest (represented by dotted lines) do not overlap.

potential Φ_{ext} is periodic, and reflects electric fields of opposite direction confining electrons at each silicon-oxide interface. The middle region of the supercell is not realistic, but so long as the wavefunctions of interest (the bottom of the conduction band or the top of the valence band) are confined to the interfaces, this is not likely to be a problem. The difficulty with simulating a single-gate MOSFET in this configuration is the size of the supercell. In bulk MOSFETs, the wavefunctions extend approximately 100Å from the interface, requiring a supercell at least 20nm in the direction perpendicular to the interfaces. Such a large cell is beyond the reach of present-day first-principles calculations.

Another method to simulate a single-gate MOSFET is to use a gauge transformation that transforms the linear scalar potential into a periodic, but time-dependent, vector potential. In particular, a constant electric field can be implemented by using

either a scalar potential ϕ,

$$\vec{E} = -\vec{\nabla}\phi \qquad \phi = -E_x x - E_y y - E_z z, \tag{4.5}$$

or by using a time-dependent vector potential $\vec{A}$,

$$\vec{E} = -\frac{\partial \vec{A}(t)}{\partial t} \qquad \vec{A} = -\vec{E}t. \tag{4.6}$$

Self-consistent electronic structure calculations in the presence of a time-dependent field, such as $\vec{A}$ in Equation 4.6, are possible within the Time-Dependent Density Functional Theory (TDDFT) formalism [118]. $\vec{A} = -\vec{E}t$ is turned on starting from $t = 0$, and the system is simulated until a quasi-stationary state is reached. This quasi-stationary state reflects the *time-independent* state that would be found if the constant electric field were included via a scalar potential. The end state is only quasi-stationary, since at long enough times, the electrons confined to the interface would overcome tunnel through the potential barrier into the oxide.

4.2.3 First Approximation: Born and Boltzmann

Since carriers traversing the channel in a MOSFET can scatter off of defects and impurities, calculating the mobility entails solving a scattering problem. As such, the first step in a mobility calculation is to calculate the relevant wavefunctions $\{\Psi_{nk}\}$, energy bands $\{\epsilon_{nk}\}$, and Kohn-Sham potential V_{ideal} for a reference, "unperturbed" system. For UTSOI channels, the unperturbed system is a supercell with an ideal, abrupt Si-SiO_2 interface, as shown in Figure 4-4. The relevant bands and wavefunctions are those at the bottom of the conduction band for electron carriers, and those at the top of the valence band for holes. Since the length and width of the channel are large compared to its width, there is effectively two-dimensional periodicity in the plane of the interface, making $\vec{k}$ a two-dimensional wavevector. The band index n is analogous to the "sub-bands" introduced by Stern and Howard [128], except that an SOI channel is confined to two-dimensions by the oxide interface even in the absence

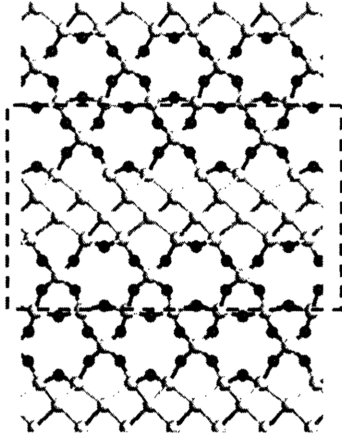


Figure 4-4: Supercell containing a 5Å-thick UTSOI silicon channel with ideal, abrupt $Si\text{-}SiO_2$ interfaces.

of an electric field.

Next, a scattering center (defect or impurity) is introduced into the supercell, and the interface+defect/impurity structure is relaxed into its lowest energy state, using either a conjugate-gradient-type algorithm [109] or Car-Parrinello molecular dynamics [26, 85]. The Kohn-Sham potential V_{defect} of this supercell is calculated. The self-consistent scattering potential ΔV can then be constructed: $\Delta V = V_{\text{defect}} - V_{\text{ideal}}$.

Current state-of-the-art MOSFET simulations [46, 47, 150] solve the semiclassical Boltzmann equation using Monte Carlo techniques. The scattering models used in these simulations (see Reference [6] for canonical examples) are based on the Born approximation for the scattering matrix and Fermi's Golden Rule for the scattering rate. Using the accurate wavefunctions and bands calculated for the ideal UTSOI channel, and the self-consistent scattering potential ΔV, we can calculate the scattering matrix within the Born approximation:

$$T_{mn}(\vec{k}, \vec{k'}) = \left\langle n\vec{k'} \middle| \Delta V \middle| m\vec{k} \right\rangle. \tag{4.7}$$

$T_{mn}(\vec{k}, \vec{k'})$ gives the probability amplitude for scattering from state $(m\vec{k})$ to state $(n\vec{k'})$.

In order to calculate the mobility, we can solve the Boltzmann equation to linear order in the applied source-drain electric field [153, 95]. To first order in the electric

field $\vec{E}$, the Boltzmann equation gives the occupancy of state $(n\vec{k})$:

$$f_n(\vec{k}) = f_0(\epsilon_n(\vec{k})) + \frac{e\tau_n(\vec{k})}{\hbar}\left(\vec{E}\cdot\vec{\nabla}_k\epsilon_n(\vec{k})\right)\frac{\partial f_0(\epsilon)}{\partial\epsilon}\bigg|_{\epsilon=\epsilon_n(k)}, \tag{4.8}$$

where $\tau_n(\vec{k})$ is the lifetime (relaxation time) for state $(n\vec{k})$, and f_0 is the equilibrium Fermi distribution. The mobility is defined by $(\langle\vec{v}\rangle)_\alpha = \mu_{\alpha\beta}(\vec{E})_\beta$, where $\langle\vec{v}\rangle$ is the average velocity calculated according to the above distribution $f_n(\vec{k})$.

The mobility tensor component $\mu_{\alpha\beta}$ for an applied field in the β direction resulting in a current in the α direction is given by:

$$\mu_{\alpha\beta} = -e\sum_n\int\frac{d^2k}{(2\pi)^2}\,\tau_n(\vec{k})\left(\vec{v}_n(\vec{k})\right)_\alpha\left(\vec{v}_n(\vec{k})\right)_\beta\frac{\partial f_0(\epsilon)}{\partial\epsilon}\bigg|_{\epsilon=\epsilon_n(k)}, \tag{4.9}$$

where $\vec{v}_n(\vec{k}) = \vec{\nabla}_k\epsilon_n(\vec{k})/\hbar$ is the group velocity of state $(n\vec{k})$.

Now the relaxation times $\{\tau_n(\vec{k})\}$ must be calculated. Fermi's Golden Rule gives the rate for scattering from state $(m\vec{k})$ to state $(n\vec{k}')$ under the influence of a single defect or impurity:

$$R_{mn}(\vec{k},\vec{k}') = \frac{2\pi}{\hbar}\left|T_{mn}(\vec{k},\vec{k}')\right|^2\delta(\epsilon_n(\vec{k}')-\epsilon_m(\vec{k})). \tag{4.10}$$

The inverse lifetime of an electron in state $(n\vec{k})$ is given by summing over the rates for transitions to all possible final states $(m\vec{k}')$. If correlated scattering from multiple defects/impurities does not occur, an assumption that is valid to first order in the defect/impurity density n_d, the scattering rates for multiple defects/impurities simply add [95]. It multiple-scattering events do occur, it is possible to include several defects/impurities in the supercell and calculate a scattering potential ΔV that reflects the combination of scatterers. Adding the scattering rates, the inverse lifetime of state $(n\vec{k})$ in the presence of a defect/impurity density n_d is:

$$\frac{1}{\tau_n(\vec{k})} = n_d\sum_m\int\frac{d^2k'}{(2\pi)^2}R_{mn}(\vec{k}',\vec{k}) = \frac{2\pi n_d}{\hbar}\sum_m\int\frac{d^2k'}{(2\pi)^2}\left|T_{mn}(\vec{k}',\vec{k})\right|^2\delta(\epsilon_n(\vec{k})-\epsilon_m(\vec{k}')). \tag{4.11}$$

A more rigorous calculation of the inverse lifetime (see, for example, References [95] and [153]) gives a nearly identical expression, but includes an additional weight factor $(1-\cos\theta)$, where where θ is the angle between $\vec{v}_n(\vec{k})$ and $\vec{v}_m(\vec{k}')$. Physically, this factor reflects the fact that backscattering has a greater effect on mobility than small-angle scattering. Thus, the final expression for the mobility is:

$$\mu_{\alpha\beta} = \frac{-e\hbar}{2\pi n_d}\sum_n \int \frac{d^2k}{(2\pi)^2}\left[\sum_m \int \frac{d^2k'}{(2\pi)^2}\delta(\epsilon_n(\vec{k})-\epsilon_m(\vec{k}'))\left|T_{mn}(\vec{k}',\vec{k})\right|^2(1-\cos\theta)\right]^{-1} \times \left(\vec{v}_n(\vec{k})\right)_\alpha \left(\vec{v}_n(\vec{k})\right)_\beta \left.\frac{\partial f_0(\epsilon)}{\partial\epsilon}\right|_{\epsilon=\epsilon_n(k)} . \quad (4.12)$$

4.2.4 Green's Functions: Beyond Born and Boltzmann

Calculating mobility using the Born approximation and the Boltzmann equation has the benefit of being simple conceptually, and permits direct comparison between first-principles results and those based on existing models and Monte Carlo device simulations. The atomic-scale structure of the Si-SiO_2 interface and quantum mechanical effects are included through the wavefunctions and self-consistent scattering potential. Both elements are important to accurate simulations of UTSOI devices, and both are absent in current state-of-the-art models. However, several restrictive assumptions have been made. For instance, Equation 4.7 assumes that the scattering potential ΔV is small. For traditional devices with thick channels, a single impurity or interface fluctuation is likely to be a small perturbation on the conduction electrons, and the assumption of weak ΔV can be expected to yield accurate results. For a UTSOI MOSFET, where the channel can be as thin as four atomic layers, a single oxygen protrusion or suboxide bond will change the channel thickness by 25% in the space of a few Angstroms. In the nanoscale regime, first-order perturbation theory may not be adequate to accurately describe device physics, making it necessary to identify ways to obtain more accurate solutions.

A systematic way to improve Equation 4.12 for the mobility is to solve the problem through a finite-temperature Green's function approach [95]. This method will

be introduced below, and it will be demonstrated that it reduces to the equations of Section 4.2.3 in the appropriate limits. Green's functions provide a powerful computational framework that permits an exact quantum solution of transport in the presence of defects and impurities. The cost, however, is increased conceptual complexity and potential complications due to numerical convergence. At present, first-principles mobility calculations have been performed using only the approach of Section 4.2.3, to enable direct comparison with existing models and to highlight the importance of atomic-scale channel structures and accurate quantum-mechanical wavefunctions and scattering potentials. In the future, however, calculations within the Green's function formalism may be able to probe the limits of semiclassical transport and treat transport in the smallest devices.

The Green's function approach begins by defining a Hamiltonian in second-quantized notation that contains all of the physics described in Section 4.2.3:

$$H = \sum_{nk} \epsilon_n(\vec{k}) c^\dagger_{nk} c_{nk} + \sum_{mk} \sum_{nk'} S(\vec{k}' - \vec{k}) \left\langle n\vec{k}' \middle| \Delta V \middle| m\vec{k} \right\rangle c^\dagger_{nk'} c_{mk}. \tag{4.13}$$

The structure factor $S(\vec{k}) = \sum_{R_d} \exp(-i\vec{k} \cdot \vec{R_d})$ contains all information about the random defect/impurity positions. Given this Hamiltonian, we can solve for the finite-temperature (Matsubara) Green's function $G_{mn}(\vec{k}, \vec{k}', i\omega)$ [95], which encapsulates all transport and scattering properties of the carriers. Physically-relevant results are obtained by averaging over the random defect/impurity positions, which makes the Green's function translationally invariant. Expanding the Green's function in powers of ΔV, this average becomes an average over products of $S(\vec{k})$: $\left\langle S(\vec{k_1}) \ldots S(\vec{k}_N) \right\rangle$. Although these quantities cannot be evaluated in general, they are tractable to first order in the defect/impurity density. This limit corresponds to neglecting correlated scattering events between different defects/impurities, and is the equivalent assumption to that made in Equation 4.11. Although the averaged retarded Green's function is diagonal in the wavevector $\vec{k}$ as a consequence of translational invariance, it still

has off-diagonal elements in the band index:

$$\Big[\mathbf{G}(\vec{k},\omega)\Big]_{mn} = \Big[\mathbf{G}_0^{-1}(\vec{k},\omega) - \mathbf{\Sigma}(\vec{k},\omega)\Big]_{mn}^{-1}. \tag{4.14}$$

The diagonal unperturbed Green's function is given by:

$$\Big[\mathbf{G}_0^{-1}(\vec{k},\omega)\Big]_{mn} = \Big((\omega + i\eta) - (\epsilon_n(\vec{k}) - \mu)\Big)\,\delta_{mn}, \tag{4.15}$$

where η is a small positive constant and μ is the chemical potential. The self-energy is linear in the defect/impurity density n_d, and is obtained by solving the following integral equation:

$$\Gamma_{mn}(\vec{k},\vec{k}',\omega) = \Big\langle n\vec{k}'\Big|\Delta V\Big|m\vec{k}\Big\rangle + \sum_{pk''} \frac{\Big\langle n\vec{k}'\Big|\Delta V\Big|p\vec{k}''\Big\rangle\,\Gamma_{mp}(\vec{k},\vec{k}'',\omega)}{(\omega + i\eta) - (\epsilon_p(\vec{k}'') - \mu)}, \tag{4.16}$$

where the singularity at $\omega = \epsilon_p(\vec{k}'') - \mu$ is handled by taking the limit $\eta \to 0^+$, as would be the case for a retarded Green's function. Once Γ is calculated, the self-energy is given by: $\Sigma_{mn}(\vec{k},\omega) = n_d\Gamma_{mn}(\vec{k},\vec{k},\omega)$.

There is a direct connection between the self-energy and the scattering matrix. The structure of Equation 4.16 is the same as the equation for the scattering (or transfer) matrix from elementary scattering theory via the Lippmann-Schwinger equation [119]. In fact, we have: $T_{mn}(\vec{k},\vec{k}') = \Gamma_{mn}(\vec{k},\vec{k}',\omega = \epsilon_n(\vec{k}))$. As a direct result of taking the low-defect-density limit when averaging over defect positions, which results in each defect being an independent scatterer, the first-order solution to Equation 4.16 is just $\Big\langle n\vec{k}'\Big|\Delta V\Big|m\vec{k}\Big\rangle$, the result we obtained in the Born approximation in Equation 4.7. Solving Equation 4.16 exactly provides the exact scattering matrix. In ultrathin-body devices where atomic-scale defects are large perturbations, this exact solution may result in different physics than the Born approximation.

Within the Green's function formulation, it is also possible to compute the lifetime of a state $(n\vec{k})$. The Green's function in the time domain is proportional to the probability amplitude for a state which was created at time $t = 0$ to be destroyed after

a time τ. After Fourier transforming the retarded Green's function in Equation 4.14, there will be a decaying exponential factor $\exp(-|\operatorname{Im}[\Sigma]|\tau/\hbar)$. The imaginary part of the self-energy corresponds to the inverse lifetime of the state. If Equation 4.16 is solved to second order in ΔV, and off-diagonal elements in the band index are neglected, then:

$$\frac{1}{\hbar}\operatorname{Im}\left[\Sigma_{nn}(\vec{k},\omega=\epsilon_n(\vec{k}))\right] = \frac{2\pi n_d}{\hbar}\sum_m \int \frac{d^2k'}{(2\pi)^2}\left|\left\langle n\vec{k}\middle|\Delta V\middle|m\vec{k}'\right\rangle\right|^2 \delta(\epsilon_n(\vec{k})-\epsilon_m(\vec{k}')), \tag{4.17}$$

which is identical to the inverse lifetime given in Equation 4.11, with $T_{mn}(\vec{k},\vec{k}')$ calculated within the Born approximation. Higher-order or exact solutions to the self-energy can yield lifetimes that correspond to the exact scattering matrix.

Mobility via the Kubo Formula

To go beyond the semiclassical Boltzmann equation for transport, the conductivity can be calculated from the Kubo formula [95]. The Kubo formula is a linear response formula, meaning that the calculated mobility is only valid for small source-drain voltages. The conductivity for an applied electric field in the α direction and a current response in the β direction is given by:

$$\sigma_{\alpha\beta} = -\lim_{\omega\to 0}\frac{1}{\omega}\operatorname{Im}\left[\Pi^{\text{ret}}_{\alpha\beta}(\omega)\right], \tag{4.18}$$

where Π^{ret} is the retarded current-current correlation function. Within the Matsubara finite-temperature formalism, Π^{ret} is calculated by first calculating the time-ordered correlation function Π as a function of the Matsubara frequency. All internal frequency integrations are sums over Matsubara frequencies, and it is not until the end of the calculation that the analytic continuation $i\omega \to \omega + i\eta$ is made and the retarded correlation function recovered.

The current-current correlation function can be evaluated exactly in the low-defect-density limit that was previously invoked. A significant simplification is achieved by using the current operator appropriate for free particles: $\vec{j} = \sum_{nk}(-\hbar e\vec{k}/m)c^\dagger_{nk}c_{nk}$.

This expression avoids off-diagonal terms in the band index and is appropriate when the gradients of the conduction band Bloch functions in the direction perpendicular to the interface are negligible. With this approximation,

$$\Pi_{\alpha\beta}(i\omega) = \frac{-1}{2\Omega} \sum_{nk} \frac{1}{\beta} \sum_{i\omega'} \left(\frac{\hbar e}{m}\right)^2 \left(\vec{k}\right)_\alpha \left(\vec{X}_n(\vec{k}, i\omega, i\omega')\right)_\beta, \tag{4.19}$$

where Ω is the volume of the system and $\{i\omega\}$ are the Matsubara frequencies. As $\Omega \to \infty$, the sum over $\vec{k}$ becomes an integral over the first Brillouin zone. $\beta = 1/kT$ is the standard thermal factor.

In order to calculate the auxiliary vector function $\vec{X}$, a new scalar function Λ is defined as follows:

$$\vec{k}\Lambda_n(\vec{k}, i\omega, i\omega') \equiv \sum_{mk'} \vec{X}_m(\vec{k}', i\omega, i\omega') n_d \Gamma_{mn}(\vec{k}', \vec{k}, i\omega) \Gamma_{nm}(\vec{k}, \vec{k}', i\omega + i\omega'). \tag{4.20}$$

With this definition, $\vec{X}$ becomes:

$$\vec{X}_n(\vec{k}, i\omega, i\omega') = \vec{k} \sum_m \left[1 + \Lambda_m(\vec{k}, i\omega, i\omega')\right] G_{mn}(\vec{k}, i\omega + i\omega') G_{nm}(\vec{k}, i\omega'). \tag{4.21}$$

Finally, Equations 4.20 and 4.21 combine to give a scalar integral equation for Λ:

$$\Lambda_n(\vec{k}, i\omega, i\omega') = \sum_{mpk'} \frac{\vec{k} \cdot \vec{k}'}{k^2} n_d \Gamma_{mn}(\vec{k}', \vec{k}, i\omega) \Gamma_{nm}(\vec{k}, \vec{k}', i\omega + i\omega') \left[1 + \Lambda_p(\vec{k}, i\omega, i\omega')\right] \times G_{pm}(\vec{k}, i\omega + i\omega') G_{mp}(\vec{k}, i\omega'). \tag{4.22}$$

By solving Equation 4.22, we can evaluate $\vec{X}$. The advantage of this approach is that it avoids solving a vector integral equation for $\vec{X}$ in favor of solving a simpler scalar integral equation for Λ.

Evaluating the internal sum over Matsubara frequencies in Equation 4.19 is difficult since it requires knowledge of Γ as a function of Matsubara frequencies. In some cases, Γ, calculated numerically from Equation 4.16, can be fit to an analytic form, allowing the Matsubara frequency sums to be evaluated directly. Another approach

can be used effectively in large supercells. A large supercell has a correspondingly small Brillouin zone, and as a result, the matrix elements of the scattering potential $\left\langle n\vec{k}'\middle|\Delta V\middle|m\vec{k}\right\rangle$ will be nearly constant as a function of $\vec{k}$ and $\vec{k}'$ for a given pair of band indices (m, n). Neglecting the variation of these matrix elements with the wavevector implies that Γ is independent of $\vec{k}$ and $\vec{k}'$ as well. For a k-independent Γ, calculating Π is trivial, and the isotropic conductivity (via Equation 4.18) takes a simple form:

$$\sigma = -\left(\frac{\hbar e}{2mS_{cell}}\right)^2 \sum_{mn} \int_{-\infty}^{\infty} d\epsilon A_{mn}(\epsilon) A_{nm}(\epsilon) \frac{\partial f_0(\epsilon - \mu)}{\partial \epsilon}, \tag{4.23}$$

where S_{cell} is the Si-SiO_2 interface area in the supercell. $A_{mn}(\epsilon)$ is the spectral function of the retarded Green's function from Equation 4.14: $A_{mn}(\epsilon) = -2\,\mathrm{Im}[G_{mn}(\epsilon)]$. Since the potential matrix elements are k-independent, wavevector dependence in G_{mn} can only enter through the energies $\{\epsilon_n(\vec{k})\}$. In formulating Equation 4.23, band energies were assumed to be constant across the Brillouin zone as well.

In the general k-dependent case, Equation 4.22 can be solved to yield Π. From Π and the electron density (input to the initial electronic structure calculation), the conductivity and the mobility can be calculated. This procedure includes the factor $(1 - \cos\theta)$ that is missing in the inverse lifetime calculated from Fermi's Golden Rule or the Green's function above. In addition, the off-diagonal elements of the self-energy are properly taken into account when there is more than one conduction band. Equation 4.17 corresponds directly to the Boltzmann equation solution, but at the cost of assuming the self-energy to be diagonal in the band index. For a spatially-localized potential, the off-diagonal elements can be substantial and the proper carrier lifetime must be obtained from the Green's function. When carriers can scatter between conduction bands, solving the Kubo formula includes important physics that are neglected in simple solutions of the Boltzmann equation.

The Kubo formula allows the direct calculation of mobility without assumptions about the scattering rate. It is complicated to solve either analytically or numerically, but allows for the freedom to include correlated scattering between different impurities

and other quantum effects that are beyond the reach of the Boltzmann equation. With this systematic Green's function approach based on electronic structure calculations, fully-quantum transport calculations in ultrathin-body devices are possible. Future work will determine the range of validity of perturbation theory and semiclassical transport in nanoscale channels.

Basis State Convergence

A critical issue that affects the numerical accuracy of the mobility calculations is the number of relevant conduction or valence band states included in the Hamiltonian (Equation 4.13). The potential ΔV couples states of different band index, as well as those with different wavevector. For a diffuse potential that is slowly varying and extends over many supercells (such as the $1/r$ impurity potential used in textbook problems), matrix elements of ΔV are nearly diagonal in the band index. Thus, the Hamiltonian need only contain enough conduction or valence band states to accommodate the carriers. In the present calculations, however, ΔV is generally localized within a supercell and can be rapidly varying in space. As a result, ΔV can couple the states included in the Hamiltonian to those outside of the chosen subset of conduction or valence bands. Convergence of the mobility with respect to the number of conduction- or valence-band states included in the Hamiltonian must be carefully studied to ensure that important ΔV-induced transitions are not being neglected.

An alternative approach is to include Bloch-like states in the Hamiltonian that couple to the relevant conduction or valence states through ΔV, and behave as an additional type of "particle" in the Hamiltonian. Consider the function space $\mathbb{H}$ spanned by the eigenstates of the Hamiltonian. The relevant conduction or valence bands that contain the carriers form a subspace of $\mathbb{H}$: $\mathbb{H}_C \subset \mathbb{H}$. The other bands that are not directly relevant to the transport process are contained in the subspace $\mathbb{H}_O$, so that $\mathbb{H} = \mathbb{H}_C \oplus \mathbb{H}_O$. ΔV in general couples states in $\mathbb{H}_C$ to those in $\mathbb{H}_O$ as described above, meaning that a Hamiltonian such as that in Equation 4.13 neglects potentially important scattering processes because it artificially restricts the mobility calculation to states in $\mathbb{H}_C$. As mentioned above, one way to include these scattering processes

is to increase the number of bands included in the Hamiltonian, i.e. increase the size of the subspace $\mathbb{H}_C$. The difficulty with this solution is that extended Bloch states, while an accurate depiction of carrier states in a transport problem, do not efficiently represent $\mathbb{H}_O$. Matrix elements of the potential $\left\langle n\vec{k}'\middle|\Delta V\middle|m\vec{k}\right\rangle$ behave as Fourier transforms of the potential at wavevector $\vec{k}-\vec{k}'$ due to the extended nature of the Bloch wavefunctions. Localized potentials couple many bands because wavevectors within the first Brillouin zone are insufficient to represent the Fourier transform of a potential localized within the supercell. This is analogous to familiar problem of needing many plane waves to represent a very localized function: the smaller a function's spread in real space, the larger its spread in reciprocal space.

The way around this problem is to recognize that $\mathbb{H}_O$ does not need to be represented using Bloch states. It is possible to choose a set of basis states $\{\Phi_{\alpha k}\}$ that span $\mathbb{H}_O$, but that also provide a compact representation of ΔV. For any state $(n\vec{k})$ in $\mathbb{H}_C$, matrix elements of the potential $\left\langle n\vec{k}\middle|\Delta V\middle|\alpha\vec{k}'\right\rangle$ will be non-negligible only for $\alpha \in 1\ldots N$. Thus, only those conduction or valence bands that are occupied by carriers need to be included in $\mathbb{H}_C$, and scattering outside this subset is accounted for by including a small number N of states from $\mathbb{H}_O$.

Basis states that span $\mathbb{H}_O$ can be chosen to be Bloch-like functions $\Phi_{\alpha k}$, written as linear combinations of localized functions chosen to accurately represent the potential:

$$\Phi_{\alpha k}(\vec{r}) = \sum_{R} e^{i\vec{k}\cdot\vec{R}} w_\alpha(\vec{r}-\vec{R}), \tag{4.24}$$

where $\vec{R}$ is a lattice vector and $w_\alpha(\vec{r})$ is a function localized within the supercell. The most important restriction on $\{\Phi_{\alpha k}\}$ is that these states be orthogonal to the states $\{\Psi_{nk}\}$ included in $\mathbb{H}_C$. $\Psi_{nk} = \exp(i\vec{k}\cdot\vec{r})u_{nk}(\vec{r})$ is a Bloch function, so that the orthogonality condition reduces to:

$$\int_{cell} d^3r e^{-i\vec{k}\cdot\vec{r}} u^*_{nk}(\vec{r}) w_\alpha(\vec{r}). \tag{4.25}$$

The most straightforward way to satisfy Equation 4.25 is to choose a set of orthonor-

mal functions $\{w_\alpha\}$ (for example, harmonic oscillator functions) and then use Gram-Schmidt [8] or another technique to orthogonalize $\{w_\alpha\}$ to the Bloch functions in $\mathbb{H}_C$.

Once a satisfactory set of orthonormal functions $\{\Phi_{\alpha k}\}$, $\alpha \in 1 \ldots N$, is chosen, the unperturbed Hamiltonian can be diagonalized within this subspace. This diagonalization provides a set of "energies" $\{E_{\alpha k}\}$ for the Bloch-like "particles", and simplifies the resulting Hamiltonian. The modified Hamiltonian including the Bloch-like pseudo-particles can now be written:

$$
\begin{aligned}
H = \sum_{nk} \epsilon_n(\vec{k}) c^\dagger_{nk} c_{nk} + \sum_{mk} \sum_{nk'} S(\vec{k}' - \vec{k}) \left\langle n\vec{k}' \middle| \Delta V \middle| m\vec{k} \right\rangle c^\dagger_{nk'} c_{mk} + \\
\sum_{\alpha k} E_\alpha(\vec{k}) d^\dagger_{\alpha k} d_{\alpha k} + \sum_{\alpha k} \sum_{\beta k'} S(\vec{k}' - \vec{k}) \left\langle \beta\vec{k}' \middle| \Delta V \middle| \alpha\vec{k} \right\rangle d^\dagger_{\beta k'} d_{\alpha k} \\
\sum_{\alpha k} \sum_{nk'} S(\vec{k}' - \vec{k}) \left\langle n\vec{k}' \middle| \Delta V \middle| \alpha\vec{k} \right\rangle c^\dagger_{nk'} d_{\alpha k} + \text{h.c.}, \quad (4.26)
\end{aligned}
$$

where $d_{\alpha k}/d^\dagger_{\alpha k}$ creates/destroys a pseudo-particle in state $\Phi_{\alpha k}$, and h.c. represents the Hermitian conjugate of the previous term. The Bloch-like functions $\{\Phi_{\alpha k}\}$, as described above, are chosen to be orthogonal to the relevant conduction- or valence-band states.

The Hamiltonian in Equation 4.26 can be used in place of the previous Hamiltonian (Equation 4.13). In calculating the Green's function, the new terms in the Hamiltonian containing $d_{\alpha k}$ and $d^\dagger_{\alpha k}$ result in an additional term in the self-energy that is straightforward to evaluate. The additional complication of these Bloch-like states is only necessary if convergence of the mobility requires an excessive number of conduction or valence band states to be included in the Hamiltonian.

4.2.5 Phonon Scattering

The calculation of electron-phonon matrix elements from first principles is a relatively recent endeavor [91, 121, 122, 141], based on Density-Functional Perturbation Theory (DFPT) [15, 60, 58]. The phonon spectrum is calculated via DFPT, and used in

conjunction with perturbed wavefunctions to yield matrix elements. This method is rigorous and straightforward, and yields important information about the coupling of electron states to particular phonon modes (important, for example, in the study of phase transitions [92]).

For simulations of transport in devices, carrier lifetimes and mobilities are required. The lifetime of a particular state reflects the probability for that state to make a transition into any other state; in the context of phonon scattering lifetimes, it reflects a sum over phonon modes. Phonon modes are the frequency-space analog of the time-dependent vibrations of a solid at finite temperature. The DFPT methods described above calculate phonon spectra by diagonalizing the dynamical matrix: a frequency-space approach. A real-time method to calculate the phonon spectrum also exists [77]: run a molecular dynamics simulation, and Fourier transform the dynamical structure factor. Since transport properties reflect a sum over phonon modes, this sum can be calculated in either real or frequency space. This is the essence of the dynamical phonon scattering method: avoid direct calculation and diagonalization of the dynamical matrix by integrating real-time dynamics.

The general approach to calculate mobilities from a molecular dynamics simulation is based on time-dependent perturbation theory. If the ions are moving, then the electrons see a time-dependent potential:

$$V_{\text{ion}}(t) = \sum_i V_i(\vec{r} - \vec{R}_i(t)), \tag{4.27}$$

where $\{\vec{R}_i(t)\}$ are the time-dependent ionic positions and $V_i(\vec{r})$ is the potential of the ith ion. The equilibrium ionic positions are denoted by $\{\vec{R}_i^0\}$. If the ions remain close to their equilibrium positions (as is the case for a solid far from its melting point), then the time-dependent perturbation ΔV will be small:

$$\Delta V(t) = \sum_i \left\{ V_i(\vec{r} - \vec{R}_i(t)) - V_i(\vec{r} - \vec{R}_i^0) \right\}. \tag{4.28}$$

The unperturbed Hamiltonian is constructed using the equilibrium ionic positions:

$$H_0 = T + V_{\text{e-e}} + V_{\text{ext}} + \sum_i V_i(\vec{r} - \vec{R}_i^0), \tag{4.29}$$

where T is the kinetic energy operator and $V_{\text{e-e}}$ is the effective electron-electron interaction potential (e.g. $V_{\text{Hartree}} + V_{\text{xc}}$ for DFT). The external potential V_{ext} was included in the Hamiltonian to emphasize that this phonon scattering approach is compatible with applied gate fields such as those described in Section 4.2.2.

Using first-order time-dependent perturbation theory, the rate of scattering from state $(m\vec{k}')$ to state $(n\vec{k})$ can be written down:

$$R_{m\vec{k}' \to n\vec{k}}(T) = \frac{1}{\hbar^2} \frac{\partial}{\partial T} \left| \int_0^T dt \; e^{i(\epsilon_{nk} - \epsilon_{mk'})t/\hbar} \langle n\vec{k} | \Delta V(t) | m\vec{k}' \rangle \right|^2 . \tag{4.30}$$

The ions begin moving at $t = 0$, and the dynamical simulation (using first-principles molecular dynamics (MD) [26, 85]) runs for a time T. Since the electron density at each time step is calculated self-consistently, the effects of coupling between the polar interface optical phonon modes and the channel carriers are included automatically. This phonon-plasmon coupling contributes to "remote phonon scattering", which Fischetti, Neumeyer and Cartier identified as an important limitation on the mobility of MOSFETs with high-κ dielectrics [48]. The origin of the high dielectric constant κ is the high bond polarizability. Optical phonon modes are associated with bond stretching, and as a result thermal vibrations lead to oscillating dipoles. These dipole fluctuations couple strongly to the charge carriers in the channel and can be a primary source of scattering and dissipation. A major limitation of this dynamical phonon scattering method is that a simulation run for a finite time T will not capture phonon modes with frequencies smaller than $2\pi/T$. However, even relatively short simulations will capture important interface optical modes.

Connection to Existing Phonon Models

The theory of phonons begins with a straightforward approximation to the time-dependent scattering potential $\Delta V(t)$. If the ions remain close to their equilibrium positions, then the time-dependent ion coordinates can be expressed as a sum of the equilibrium position and a small deviation: $\vec{R}_i(t) = \vec{R}_i^0 + \delta\vec{R}_i(t)$. To first order in $\delta\vec{R}_i$, the perturbation potential is:

$$\Delta V(t) = -\sum_i \delta\vec{R}_i(t) \cdot \vec{\nabla} V_i(\vec{r} - \vec{R}_i^0). \tag{4.31}$$

The deviation $\delta\vec{R}$ can be expressed as a sum of phonon modes [113]:

$$\delta\vec{R}_i(t) = \sum_{\vec{q}\lambda} \left(1/\sqrt{NM_i}\right) \left(\sqrt{\hbar/2\omega_{q\lambda}}\right) e^{i\vec{q}\cdot\vec{R}_i^0} \vec{\epsilon}_{q\lambda} \times \\ \left\{ e^{-i\omega_{q\lambda}t} \langle\chi|a_{q\lambda}|\chi\rangle + e^{i\omega_{-q\lambda}t} \langle\chi|a^\dagger_{-q\lambda}|\chi\rangle \right\}, \tag{4.32}$$

where N is the number of unit cells and M_i is the mass of the ith ion. $\{\omega_{q\lambda}\}$ are the phonon frequencies, labeled by the wavevector $\vec{q}$ and polarization λ. $\vec{\epsilon}_{q\lambda}$ is the polarization vector. The exponential time dependence comes from the time dependence of the operators $a_{q\lambda}$ and $a^\dagger_{-q\lambda}$ in the Heisenberg picture, with the phonon Hamiltonian containing only harmonic terms. $|\chi\rangle$ is the ground-state phonon wavefunction.

The primary constraint on $|\chi\rangle$ is that the expectation value of the number operator should give the Bose-Einstein occupancy:

$$\langle\chi|a^\dagger_{q\lambda} a_{q\lambda}|\chi\rangle = \frac{1}{\exp(\beta\omega_{q\lambda}) - 1}. \tag{4.33}$$

In order to evaluate Equation 4.32 it is only necessary to evaluate certain expectations values of $|\chi\rangle$, and not the wavefunction itself. The simplest choice for the matrix elements $\langle\chi|a_{q\lambda}|\chi\rangle$ and $\langle\chi|a^\dagger_{-q\lambda}|\chi\rangle$ is to take the square root of the Bose-Einstein occupancy:

$$\langle\chi|a_{q\lambda}|\chi\rangle = \langle\chi|a^\dagger_{q\lambda}|\chi\rangle = \frac{1}{\sqrt{\exp(\beta\omega_{q\lambda}) - 1}}. \tag{4.34}$$

If $\langle\chi|a^\dagger_{q\lambda}a_{q\lambda}|\chi\rangle \simeq \langle\chi|a^\dagger_{q\lambda}|\chi\rangle\langle\chi|a_{q\lambda}|\chi\rangle$, then Equation 4.33 is satisfied. Evaluating $\langle\chi|a^\dagger_{q\lambda}a_{q\lambda}|\chi\rangle$ in this manner assumes that the phonon wavefunction is a *coherent state* (see Reference [87], Chapter 21, for a brief introduction). A coherent state $|\chi\rangle$ is an eigenstate of the annihilation operator: $a_{q\lambda}|\chi\rangle = c_{q\lambda}|\chi\rangle$, where $c_{q\lambda}$ is a constant eigenvalue. Coherent states are commonly used in the study of radiation to represent quantum analogues of classical radiation fields, when large numbers of photons are present. Since $|\chi\rangle$ represents a large number of phonons in a room-temperature solid, it is appropriate to treat $|\chi\rangle$ as a coherent state.

Using Equation 4.34 for the expectation values of $a_{q\lambda}$ and $a^\dagger_{-q\lambda}$, it is possible to evaluate Equation 4.30 for the phonon scattering rate. The time dependence in $\delta\vec{R}_i(t)$ enters only through the exponentials $\exp(-i\omega_{q\lambda}t)$ and $\exp(i\omega_{-q\lambda}t)$, so that the time integral in Equation 4.30 can be evaluated mode-by-mode:

$$I^{\pm}_{q\lambda}(T) = \int_0^T dt\, e^{i(\epsilon_{nk}-\epsilon_{mk'})t/\hbar}e^{\pm i\omega_{q\lambda}t} = \frac{1}{i((\epsilon_{nk}-\epsilon_{mk'})/\hbar \pm \omega_{q\lambda})}\left[e^{i((\epsilon_{nk}-\epsilon_{mk'})/\hbar\pm\omega_{q\lambda})T} - 1\right]. \tag{4.35}$$

The potential matrix elements are time-independent:

$$f_{q\lambda} = \sum_i (1/M_i)e^{i\vec{q}\cdot\vec{R}^0_i}\langle n\vec{k}|\vec{\epsilon}_{q\lambda}\cdot\vec{\nabla}V_i(\vec{r}-\vec{R}^0_i)|m\vec{k}'\rangle. \tag{4.36}$$

In terms of the $\{I_{q\lambda}(t)\}$ and $\{f_{q\lambda}\}$, the scattering rate is given by:

$$R_{m\vec{k}'\to n\vec{k}}(T) = \frac{1}{\hbar^2}\frac{\partial}{\partial T}\Bigg|\sum_{q\lambda}\left(1/\sqrt{N}\right)\left(\sqrt{\hbar/2\omega_{q\lambda}}\right) f_{q\lambda}\times \left\{\frac{1}{\sqrt{\exp(\beta\omega_{q\lambda})-1}}I^-_{q\lambda}(T) + \frac{1}{\sqrt{\exp(\beta\omega_{-q\lambda})-1}}I^+_{-q\lambda}(T)\right\}\Bigg|^2. \tag{4.37}$$

Taking the square magnitude, $|\ldots|^2$, introduces cross terms in $\vec{q}$ and λ. In the $T \to \infty$ limit, however, the time derivative of those terms vanish. The diagonal

terms are straightforward to evaluate in the $T \to \infty$ limit:

$$\lim_{T\to\infty} |I_{q\lambda}^{\pm}(T)|^2 = \lim_{T\to\infty} \frac{4}{\left((\epsilon_{nk} - \epsilon_{mk'})/\hbar \pm \omega_{q\lambda}\right)^2} \sin^2\left[\frac{\left((\epsilon_{nk} - \epsilon_{mk'})/\hbar \pm \omega_{q\lambda}\right)T}{2}\right]$$
$$= 2\pi\hbar T \delta\left(\epsilon_{nk} - \epsilon_{mk'} \pm \hbar\omega_{q\lambda}\right), \quad (4.38)$$

where $\delta(\ldots)$ is the Dirac delta function. In the $T \to \infty$ limit, the expression for the phonon scattering rate in terms of phonon modes is:

$$R_{m\vec{k}'\to n\vec{k}}(T\to\infty) = \frac{2\pi}{\hbar}\sum_{q\lambda}(1/N)\left(\hbar/2\omega_{q\lambda}\right) f_{q\lambda}\times$$
$$\left\{\frac{1}{\exp(\beta\omega_{q\lambda})-1}\delta\left(\epsilon_{nk} - \epsilon_{mk'} - \omega_{q\lambda}\right) + \frac{1}{\exp(\beta\omega_{-q\lambda})-1}\delta\left(\epsilon_{nk} - \epsilon_{mk'} + \omega_{-q\lambda}\right)\right\}. \quad (4.39)$$

Equation 4.39 is a familiar expression for phonon scattering, involving both absorption (the first term in braces) and emission (the second term) of phonons. Momentum conservation is enforced through the matrix elements $f_{q\lambda}$. The disadvantage to using this expression for the scattering rate is that it does not include phonon-plasmon coupling, since the "scattering potential" is $\vec{\nabla}V_i$, which is not calculated self-consistently.

4.3 Electronic Structure of UTSOI Channels

In this section, results from the ground-state DFT-LDA calculations that provide the unperturbed wavefunctions and energy bands are discussed and compared to the predictions of a free electron model of the channel currently used in state-of-the-art simulations. Understanding the electronic structure of UTSOI channels as a function of carrier density (or gate bias) is an important first step to explaining calculated mobilities. Traditional UTSOI MOSFET simulations treat the channel as a potential well perpendicular to the Si-SiO_2 interface. The oxide interfaces are generally represented by infinite potential barriers. Based on solutions of the one-dimensional Schrödinger

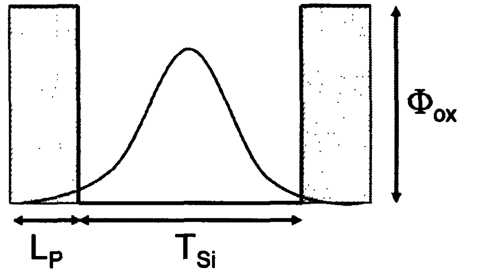


Figure 4-5: Schematic representation of wavefunction penetration into the oxide. The wavefunction is the solid line within the potential well defining the Si channel. The oxide is represented by the shaded regions, and the height of the oxide potential barrier is Φ_{ox}. The channel thickness is T_{Si}, and the penetration depth of the carriers is L_p.

equation in this potential well, two phenomena unique to UTSOI channels have been described: wavefunction penetration and volume inversion. Wavefunction penetration into the oxide and volume inversion will be treated in depth later in this section, and compared to free-electron results. In addition, the self-consistent screening of the external potential will be discussed.

The subsequent results on wavefunction penetration and volume inversion specialize to the case of electron carriers, i.e. the relevant wavefunctions are those at the bottom of the conduction band. Similar trends will be found in systems with hole carriers. Electron densities and potentials are not symmetric about the center of the channel because the two Si-SiO_2 interfaces are not equivalent.

4.3.1 Wavefunction Penetration

Figure 4-5 demonstrates schematically the phenomena of wavefunction penetration into the oxide. As discussed in Section 4.2.1, the offset between the conduction band of Si and that of SiO_2 represents a potential barrier that confines the electrons at the bottom of the conduction band to the Si channel. In most traditional simulations that solve the Schrödinger equation for effective-mass electrons in the channel (see, Reference [149] for an example), the oxide interface potential barrier is infinite. As a result, the carriers are completely confined within the channel.

As the calculated conduction electron densities in Figure 4-6 show, the oxide

Table 4.1: Calculated penetration length L_p versus channel thickness T_{Si}, from Figure 4-6. Measurement definitions as in Figure 4-5.

T_{Si} (Å)	L_p (Å)
10	1.8468
15	1.3405
20	1.1886

interface barrier should not be thought of as infinite. There is significant penetration of the carrier wavefunctions into the oxide region. In order to compare the calculated density with the predictions of the effective-mass model, the penetration length was determined from the plot in Figure 4-6. For concreteness, the width of the channel was chosen to be exactly T_{Si} (10Å, 15Å, or 20Å), although the exact position of the Si-SiO_2 interface is arbitrary at the atomic scale. The penetration length was determined to be the distance from the interface position at which the linear electron density n_e was less than $10^{-5}e^-/\text{Å}$. A channel density of $5.6 \times 10^{-11}e^-/\text{Å}^2$ was chosen to minimize the influence of the external bias, making the potential well nearly "square", as in Figure 4-5. Table 4.1 gives the penetration length as a function of channel thickness.

For comparison to the calculated values in Table 4.1, consider an electron in a square well such as that in Figure 4-5. Since the conduction electron density is low, only the lowest bound state in the well must be considered. Matching boundary conditions at the edges of the well gives a transcendental equation for the wavevector k:

$$\tan\left(\frac{T_{Si}k}{2}\right) = \frac{\kappa}{k}, \tag{4.40}$$

where $\kappa = \sqrt{(2m/\hbar^2)(\Phi_{ox} - E)}$, and the energy E is a function of k: $E = \hbar^2k^2/2m$. Solving Equation 4.40 for k gives a set of allowed wavevectors that satisfy the boundary conditions, and a corresponding set of allowed energies. Unlike the infinite square well, where $\Phi_{ox} \rightarrow \infty \Rightarrow \kappa \rightarrow \infty$, Equation 4.40 admits only a finite number of bound-state solutions (with $E < \Phi_{ox}$). In the limit $\Phi_{ox} \gg E$, which is likely to be appropriate for the ground state of the well, the allowed wavevectors will be given by

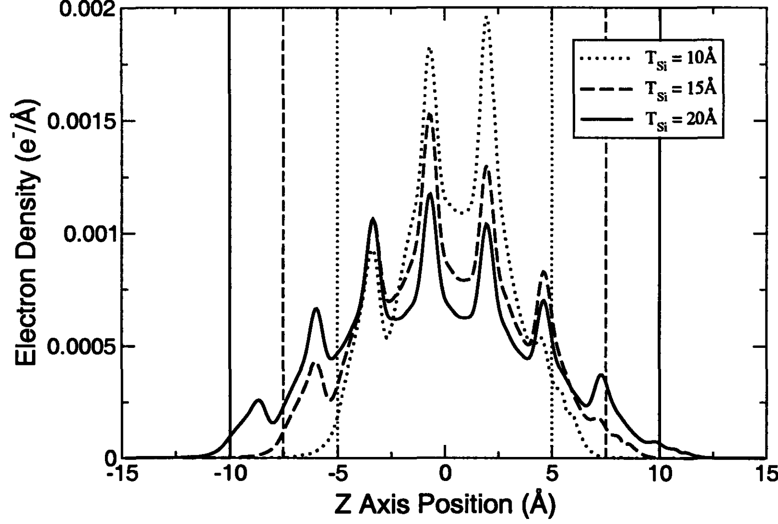


Figure 4-6: Calculated conduction electron densities for 10Å-, 15Å-, and 20Å-thick channels, with the channel density fixed at $5.6 \times 10^{-11} e^-/\text{Å}^2$. The z axis is perpendicular to the Si-SiO_2 interface. Vertical lines mark the positions of the Si-SiO_2 interfaces; line styles correspond to those in the legend. Electron densities are given as linear densities, integrated over planes parallel to the interface.

the simpler equation $\tan(T_{Si}k/2) = 0$. Using this approximation, the ground-state energy is $E_0 = \hbar^2\pi^2/2mT_{Si}^2$.

The decaying part of a bound-state solution in the oxide region is given by $\psi \propto \exp(-\kappa x)$. Thus, $1/\kappa$ is proportional to the decay length. To distinguish this characteristic length from the previously defined penetration length L_p, $1/\kappa \equiv \lambda$. L_p and λ are proportional to each other, but not necessarily equal. L_p measures the decay length for the electron density, while λ measures wavefunction decay. Because only the ground-state of the potential well is considered, however, the density decay length is just $\lambda/2$. The factor of two can be absorbed into the proportionality constant connecting λ to L_p. Using the expression for κ given above, and substituting the ground-state energy E_0 for E, we find:

$$\kappa_0 = \frac{1}{\lambda_0} = \sqrt{\frac{2m\Phi_{ox}}{\hbar^2}\left(1 - \frac{\hbar^2\pi^2}{2m\Phi_{ox}}\frac{1}{T_{Si}^2}\right)}. \tag{4.41}$$

$\hbar^2/2m\Phi_{ox} \equiv \Lambda^2$ defines a characteristic length that is very small when Φ_{ox} is large. $\Lambda/T_{Si} \ll 1$ based on the previous assumption that $E_0 \ll \Phi_{ox}$. Thus, we can

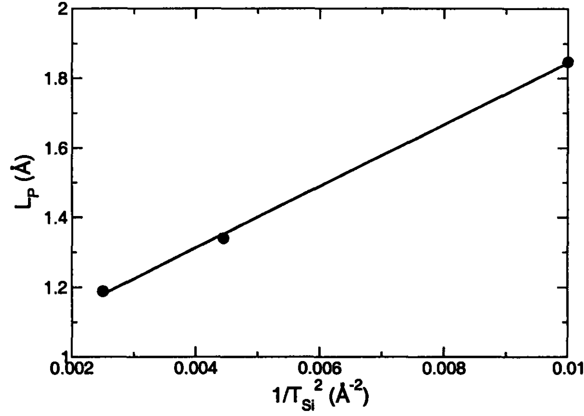


Figure 4-7: Penetration length L_p versus $1/T_{Si}^2$, as calculated from Figure 4-6. Circles show calculated points, solid line is a linear fit.

expand the square root in Equation 4.41 to find an expression for the characteristic penetration length as a function of T_{Si}.

$$\lambda_0 = \frac{1}{\kappa_0} = \Lambda \left(1 + \frac{\pi^2}{2}\left(\frac{\Lambda}{T_{Si}}\right)^2\right). \tag{4.42}$$

If the previous assumption $E_0 \ll \Phi_{ox}$ is valid, and if the calculated conduction band densities behave similarly to effective-mass electrons with a constant interface potential barrier, then the penetration length should be linear in $1/T_{Si}^2$. Figure 4-7 shows that this linear relationship holds. A further prediction of the effective-mass model is a relationship between the slope m and intercept b of $\lambda_0(1/T_{Si}^2)$. Because the only undetermined constant in Equation 4.42 is Λ, $m/b^3 = \pi^2/2$. As was previously discussed, L_p is not equal to λ_0, but is proportional to it: $L_p = \alpha\lambda_0$. Based on this constant, the measured ratio of the slope to the cube of the intercept is: $\tilde{m}/\tilde{b}^3 = \pi^2/2\alpha^2$. From the fit in Figure 4-7, $\tilde{m}/\tilde{b}^3 = 100.427 = 20.351 * (\pi^2/2)$, and thus $\alpha = 0.222$.

The effective-mass model that treats the oxide interface as a potential barrier does predict the variation of the penetration length with channel thickness. The accuracy of the $L_p(1/T_{Si}^2)$ fit indicates that the Si-SiO_2 interface can be modeled as a potential barrier of constant height across at least a limited range of T_{Si} values. The intercept $\tilde{b}$ of the $L_p(1/T_{Si}^2)$ curve is given by: $\tilde{b} = \alpha\Lambda = \alpha\sqrt{\hbar^2/2m\Phi_{ox}}$. In the

effective-mass model, the electron mass m is an effective mass. To make this clear, subsequent discussion will refer to the effective mass as m^*, and the unrenormalized electron rest mass as $m_0 = 9.11 \times 10^{-31}$kg. Using the fitted intercept from Figure 4-7, $\tilde{b} = 0.95875$Å, Φ_{ox} is given by:

$$\Phi_{ox} = \left(\frac{m_0}{m^*}\right) \frac{\alpha^2 \hbar^2}{2 m_0 \tilde{b}^2} = \left(\frac{m_0}{m^*}\right) 0.20368\text{eV}. \tag{4.43}$$

The relevant effective masses in silicon are the longitudinal and transverse effective masses at the bottom of the conduction band (the valleys along the $\Gamma - X$ direction in the Brillouin zone): $m_{long} = 0.98 m_0$ and $m_{trans} = 0.19 m_0$ [129]. Using these values for the effective mass, we find $\Phi_{ox}^{long} = 0.208$eV and $\Phi_{ox}^{trans} = 1.072$eV. These two values highlight difficulties with the effective-mass model. First, the justification for treating the interface as a potential barrier was that the conduction band edge of the oxide was offset from the edge of silicon conduction band. Since the bands define the energy landscape for effective-mass electrons, both the longitudinal and transverse $\Gamma - X$ valleys should have equal barrier heights, being of equal energy. In this model, Φ_{ox} depends explicitly on the effective mass, which is not equal for the longitudinal and transverse valleys. Second, neither calculated value of Φ_{ox} is comparable to the calculated conduction band edge offset of 2.3eV (discussed in Section 4.2.1). This discrepancy may be due to the lack of a sharp transition between the Si and SiO_2 conduction bands. As shown in Figure 4-1, it takes roughly 5Å for the conduction band edge to rise from its Si value to the SiO_2 value. The density penetration length is 1-2Å, so the effective band offset may be less than the difference in bulk band edges.

Both of these difficulties likely arise from the basic conceptual difficulty of the effective-mass model: the discontinuity of the effective mass between Si and SiO_2. Both the conduction- and valence-band effective masses differ in Si and SiO_2, and this makes constructing a rational kinetic energy operator a challenge. The easiest way around this difficulty is to make the Si-SiO_2 interface an infinite potential barrier, confining the electrons to the silicon channel, but this approach neglects important wavefunction penetration effects. In MOSFET simulations that solve the Schrödinger

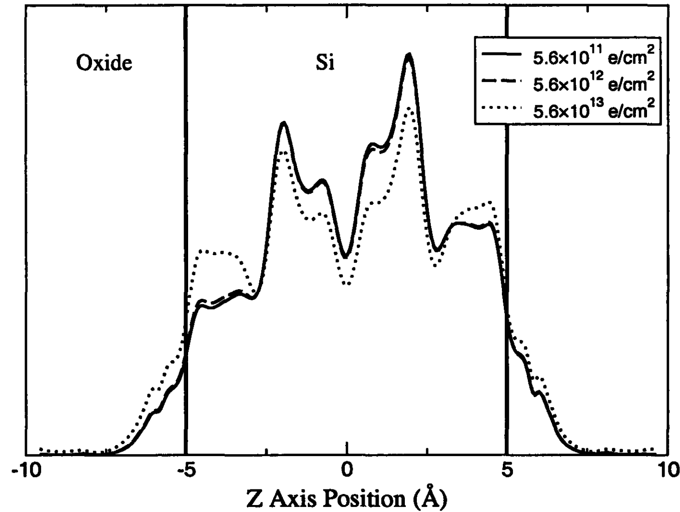


Figure 4-8: Plane-averaged electron densities in a 10Å-thick silicon channel. Densities are scaled to have equal weight. The z axis is perpendicular to the Si-SiO_2 interface.

equation with a finite interface barrier, the effective mass of Si is used in both the channel and oxide regions. This is not likely to accurately represent wavefunction penetration since the kinetic energy in the oxide region will be either over- or underestimated (the effective mass for conduction electrons in SiO_2 is $0.85m_0$ [99, 97], between the longitudinal and transverse effective masses in Si). In order to fit the calculated penetration length, the value of the oxide potential barrier must be adjusted. As a result, there is no accurate way to derive the parameters for an effective-mass model of the channel from first-principles.

Another important consideration is the penetration length as a function of carrier electron density, or equivalently, applied gate bias. The effective-mass model discussed above predicts no dependence on gate bias, but that is solely a result of assuming that the potential well defining the channel is square. This assumption completely neglect any external fields, and is only valid in the low-carrier-density (low-gate-field) limit. Figure 4-8 shows the conduction electron density for a 10Å-thick channel. In order to show how the shapes of the density profiles vary, all three electron densities are scaled to have equal weight. As a result, the ordinate of Figure 4-8 is arbitrary.

Even on increasing the conduction electron density by an order of magnitude, from 5.6×10^{11} to $5.6 \times 10^{12} \mathrm{e}^-/\mathrm{cm}^2$, the density profile hardly changes. At lower carrier

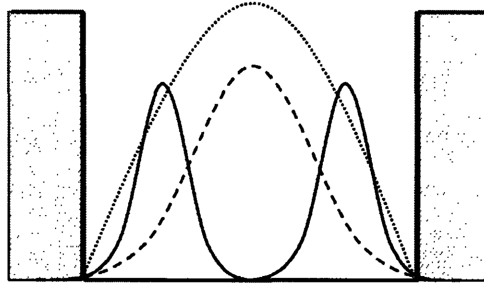


Figure 4-9: Schematic representation of applied bias potential and electron densities in a thin SOI device. Shaded regions represent the oxide. The dotted line shows the symmetric gate bias potential. The solid line shows the carrier density for a thicker channel, with carriers divided into front and back channels. The dashed line represents the carrier density in a volume-inverted channel.

densities, the applied gate field does not appreciably affect the electronic structure of the channel. Although at 5.6×10^{13}e$^-$/cm^2 additional charge begins to accumulate at the oxide interfaces, the penetration length remains largely unchanged. For thin channels such as that in Figure 4-8, kinetic energy is clearly the dominant factor in determining the charge distribution in the channel. Although the potential energy provided by the external gate field favors extra charge at the interface, the kinetic energy cost to further localize the carriers cannot be overcome in a thin channel. In thicker channels the interplay between potential and kinetic energy results in appreciable changes to the density profile as the carrier density is increased. The electronic structure of thicker channels will be discussed below, and the concept of *volume inversion* will be introduced.

4.3.2 Volume Inversion

In a standard, single-gate MOSFET the carriers are confined to a narrow region near the silicon-oxide interface. In a symmetric double-gate device, the applied bias (the dotted line in Figure 4-9) creates potential wells at both oxide interfaces. In a thick SOI device, this creates two carrier channels, one at the front and one at the back interface. The solid line in Figure 4-9 shows the electron density that would be expected for such a two-channel configuration. For a given bias, the thick double-gate device carries twice as much current as a single-gate device.

In thin SOI devices, however, the two-channel model of a double-gate SOI MOSFET breaks down. The dashed line in Figure 4-9 shows the carrier density in a thin SOI device. The electron density has a peak in the center of the silicon region instead of near the interfaces, and as a result, there is only a single channel. This configuration is referred to as *volume inversion*, and was introduced by Balestra *et al.* [14] as a means to improve performance in double-gate SOI MOSFETs. Electrons concentrated near the oxide interfaces scatter more frequently from interface defects and charges trapped in the oxide. Electrons in a volume-inverted channel are concentrated away from the interfaces, and thus can avoid frequent interaction with interface and oxide defects. As a result, the mobility [44] of volume-inverted double-gate UTSOI MOSFETs is enhanced relative to the same devices operated in single-gate mode. As shown by the two-channel model for thick SOI devices, a double-gate MOSFET will have twice as many carriers as a single-gate device for the same bias voltage. Since mobility is normalized to the carrier density, this doubling does not by itself lead to an increased mobility.

The phenomenon of volume inversion is easily understood through the potential well model introduced in Section 4.3.1. The ground-state wavefunction of the potential well is $\psi_0 = \sqrt{2/T_{Si}}\cos(\pi z/T_{Si})$ (neglecting wavefunction penetration), which has a peak at the center of the silicon region and nodes at the interfaces. At low carrier densities, only ψ_0 is occupied, and thus the density has the volume-inverted form shown in Figure 4-9. The situation is less clear, however, at higher carrier densities and in thicker channels. Higher carrier densities mean that potential well states above the ground state are occupied, and the accompanying bias potential makes it energetically-favorable for carriers to concentrate near the interface. Thicker channels reduce the kinetic energy cost to localize carriers near the interfaces. The range of carrier densities and channel thicknesses over which volume inversion will be observed is a result of a balance between the potential energy gain from staying near the interface and the kinetic energy cost of localization.

In order to elucidate how the boundaries of the volume inversion regime depend on T_{Si} and the carrier density n_e, a simplified potential well model of the channel will be

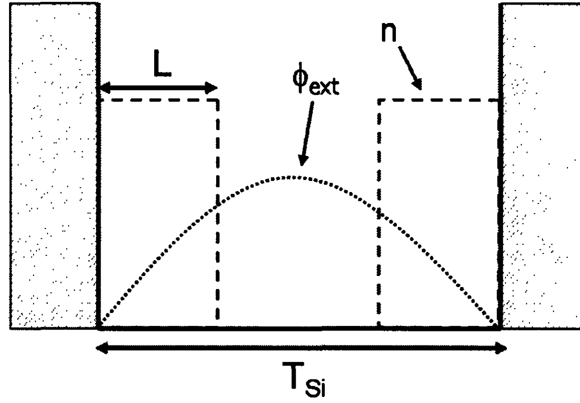


Figure 4-10: Schematic of the potential well volume inversion model. The applied external potential ϕ_{ext} is shown as a dotted line, while the dashed line shows the carrier density model n. The oxide is represented by dashed region, with the oxide potential barrier assumed to be infinite.

introduced. This model is shown schematically in Figure 4-10. As in Section 4.3.1, the well width is T_{Si}, but the oxide interface barrier is taken to be infinite. n_e, as measured for SOI MOSFETs, is a two-dimensional carrier density. Channels of different widths will be compared at constant n_e, which is constant over the width of the channel. The density of electrons per unit width of the channel is given by n_e/T_{Si}. The applied bias potential is taken to be $\phi_{\text{ext}}(z) = (-qn_e/2\epsilon_{Si}T_{Si})z^2$, where $q = +1(-1)$ for holes (electrons), and ϵ_{Si} is the dielectric constant for bulk silicon. The expression for ϕ_{ext} above is calculated assuming a constant carrier density in the channel; non-uniform carrier densities lead to only minor changes in the potential.

The potential energy due to the applied bias potential for a carrier density distribution $n(z)$ is given by: $PE = \int_{-T_{Si}/2}^{T_{Si}/2} dz\phi_{ext}(z)\,(qn(z))$. $n(z)$ is a three-dimensional density with the constraint: $\int_{-T_{Si}/2}^{T_{Si}/2} dzn(z) = n_e$. For computational ease and clarity, a model form for the carrier density is used, as shown in Figure 4-10. The density is symmetric and characterized by constant-density regions of with L, $0 < L \leq T_{Si}/2$, adjacent to the oxide interfaces. L characterizes the degree of localization of the carriers. Volume inversion is characterized by $L = T_{Si}/2$, with two independent channels

for $L < T_{Si}/2$. Using this form for the density, we can calculate the potential energy:

$$PE = -\left(\frac{q^2 n_e^2 T_{Si}}{24\epsilon_{Si}}\right)\left(\frac{T_{Si}}{2L}\right)\left[1-\left(1-\frac{2L}{T_{Si}}\right)^3\right]. \tag{4.44}$$

The kinetic energy can be estimated from the degree of localization of the carrier density, via the Heisenberg uncertainty relation: $\Delta x \Delta p \sim \hbar$. The characteristic width L of the density in Figure 4-10 gives a localization length $\Delta x = 2L$, and thus $\Delta p \sim \hbar/2L$. Taking the momentum p to be Δp, the kinetic energy $p^2/2m$ can be calculated:

$$KE = \frac{n_e \hbar^2}{2m(2L)^2} = \frac{n_e \hbar^2}{2m T_{Si}^2}\left(\frac{T_{Si}}{2L}\right)^2. \tag{4.45}$$

The total energy $E = KE + PE$ depends on the width of the carrier density through the dimensionless variable $x \equiv 2L/T_{Si}$:

$$E(x) = \left(\frac{n_e \hbar^2}{2m T_{Si}^2}\right)\frac{1}{x^2} - \left(\frac{q^2 n_e^2 T_{Si}}{24\epsilon_{Si}}\right)(x^2 - 3x + 3). \tag{4.46}$$

Since $2L \le T_{Si}$, $x \in (0,1]$. Depending on the values of $\alpha \equiv (n_e \hbar^2/2m T_{Si}^2)$ and $\beta \equiv (q^2 n_e^2 T_{Si}/24\epsilon_{Si})$, $E(x)$ has two types of behavior over the range $(0,1]$. The first occurs when $E(x)$ has no minimum for $x \in (0,1]$. In this case, since $E(x \to 0) \to \infty$, the minimum energy is found at $x = 1 \Rightarrow 2L = T_{Si}$. When $E(x)$ has no minimum in $(0,1]$, the channel is volume-inverted, since the localization length of the carrier density is equal to the channel width. The second regime occurs when $E(x)$ has a minimum for $x \in (0,1]$. In this case, the localization length is less than T_{Si}, and the carrier density splits into two uncoupled channels. This situation is the one depicted schematically in Figure 4-10.

From $E(x)$ as given in Equation 4.46, it is possible to determine the boundaries of the volume inversion regime. As $x \to 0$, $\partial E/\partial x \to -\infty$. At $x = 1$, $\partial E/\partial x = -2\alpha + \beta$. Since $E(x)$ is continuous in the region $(0,1]$, if $\partial E/\partial x(x = 1) > 0$, then there must be some $x \in (0,1]$ at which $\partial E/\partial x = 0$. Thus, the volume inversion regime is characterized by $2\alpha > \beta$, which implies that $E(x)$ has no minimum for $x \in (0,1]$. Note that, by the definitions above, α and β are always positive. The volume inversion

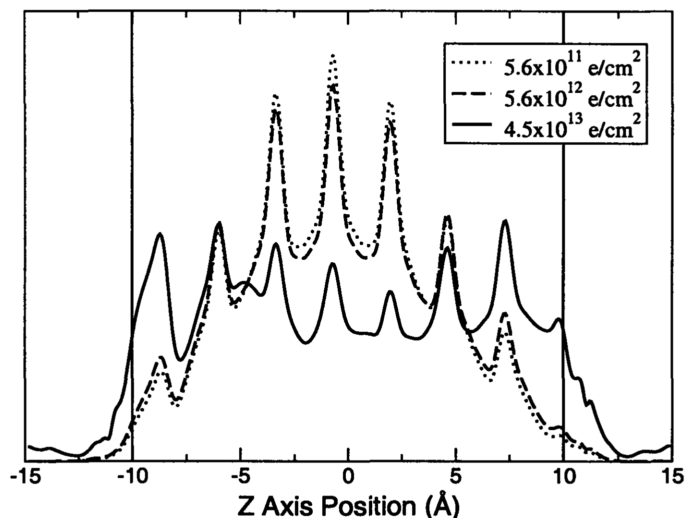


Figure 4-11: Plane-averaged electron densities in a 20Å-thick silicon channel. Densities are scaled to have equal weight. The z axis is perpendicular to the Si-SiO_2 interface.

regime is defined by:

$$n_e T_{Si}^3 \leq \frac{12\hbar^2 \epsilon_{Si}}{q^2 m}. \tag{4.47}$$

Equation 4.47 states that thinner channels operating at lower carrier densities are more likely to be volume-inverted. As Figure 4-8 shows, a thin 10Å channel is always within the volume inversion regime, even as the conduction electron density is increased by two orders of magnitude. The situation for thicker channels is different, however. Figure 4-11 shows the scaled conduction electron densities for a 20Å-thick channel, over the same order-of-magnitude range as in Figure 4-8. At $4.5 \times 10^{13} e^-/cm^2$, electron density is qualitatively different. A considerable amount of charge has built up near the oxide interfaces, and the density is nearly uniform across the channel instead of being peaked in the center. For a 20Å channel, this carrier density marks the boundary of the volume inversion regime. Increasing the density of conduction electrons drives more charge toward the interface, and eventually results in two separate channels, with a node instead of a peak at the center of the silicon region.

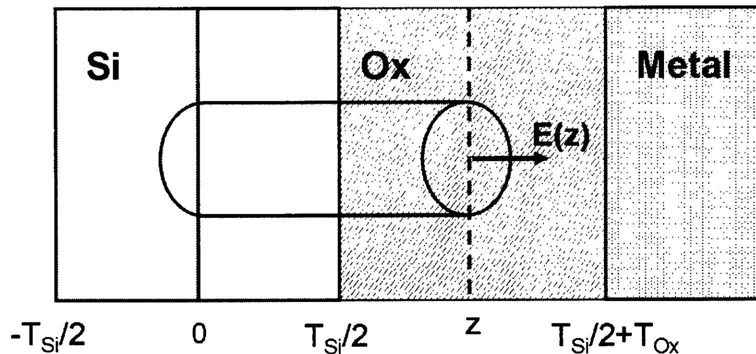


Figure 4-12: Schematic of a T_{Si}-thick silicon channel, T_{Ox}-thick oxide, and metal gate. The axis of the cylinder is perpendicular to the silicon-oxide and metal-oxide interfaces. The cylinder has length z and base area A.

4.3.3 Self-Consistent Electrostatics

In a MOSFET, the carriers in the channel are coupled to the free charge in the metal gate. A biased gate results in an electric field across the channel that is screened by the carriers in the channel. In this section, the macroscopic electrostatic equations that govern the channel and gate charges are discussed, and compared to calculated results. Figure 4-12 shows a schematic representation of a silicon channel. Based on this model, the electric field as a function of the channel charge $qn(z)$ and the metal gate charge σ will be calculated. The width of the channel is T_{Si}, and the width of the gate oxide is T_{Ox}. A symmetric double-gate MOSFET has zero electric field at $z = 0$, the center of the channel. Since the metallic gate is assumed to be a good conductor, an applied gate bias will accumulate charge only near the gate-oxide interface. The accumulated gate charge will be represented by a surface charge σ.

The cylinder in Figure 4-12 defines a surface $\mathcal{S}$ containing a volume $\mathcal{V}$ over which the Poisson equation will be integrated. Since the system is symmetric in the plane of the interfaces, the only variation of the electric field E and the channel charge n will be in the z-direction. In the regime of macroscopic electrostatics, the oxide and silicon channel will have static dielectric constants ϵ_{Ox} and ϵ_{Si}, respectively. The volume integral of the Poisson equation gives:

$$\int_{\mathcal{V}} d^3r \vec{\nabla} \cdot \vec{D} = \int_{\mathcal{V}} d^3r \; n_{tot}(\vec{r}) \Rightarrow \int_{\mathcal{S}} d\vec{S} \cdot \vec{D} = Q_{tot}. \qquad (4.48)$$

The left end of the cylinder is at $z = 0$, where $\vec{D} = \epsilon_{Si}\vec{E} = 0$. Thus, since the system varies only in the z direction, Equation 4.48 gives:

$$\begin{aligned} E(z) &= \frac{q}{\epsilon_{Si}} \int_0^z dz'\ n(z') && 0 \leq z \leq T_{Si}/2 \\ &= \frac{q}{\epsilon_{Ox}} \int_0^{T_{Si}/2} dz'\ n(z') && T_{Si}/2 \leq z \leq T_{Si}/2 + T_{Ox}. \end{aligned} \tag{4.49}$$

Integrating Equation 4.48 out to $z > T_{Si}/2 + T_{Ox}$ gives an important condition connecting the total charge $N \equiv 2\int_0^{T_{Si}/2} dzn(z)$ in the silicon channel to the surface charge in the gate. The field at $z = 0$ is zero, as is the charge for $z > T_{Si}/2 + T_{Ox}$, since the gate is assumed to be a good conductor. This implies that the total charge in the cylinder is zero and thus: $\sigma = -qN/2$.

Equation 4.49 evaluated at $z = T_{Si}/2$ implies that a surface charge $\sigma' = (qN/2)(1/\epsilon_{Ox} - 1/\epsilon_{Si})$ must exist at the silicon-oxide interface in order to ensure continuity of the electric field (the total charge in the channel must still equal the gate charge: $N/2 + \sigma' = \sigma$. The field $E(z)$ in general must be calculated self-consistently with the channel charge $qn(z)$. The self-consistency condition is satisfied in the DFT calculations described above. Here, the qualitative features of the self-consistent field are discussed and compared to calculated results.

If the silicon channel were a perfect conductor, the field in the gate oxide would be screened completely by the surface charge: $\sigma' = -\sigma$. As a result, there will be a discontinuity in the electric field (and hence a kink in the scalar potential) at the silicon-oxide interface. $E(z) = \sigma'/\epsilon_{Ox}$ for $T_{Si}/2 < z \leq T_{Si}/2 + T_{Ox}$ and $E(z) = 0$ for $z < T_{Si}/2$. Silicon, of course, is not a perfect conductor, but in thicker channels with higher carrier densities, more charge will build up at the interface to screen the oxide field. This phenomenon is most evident outside of the volume inversion regime, as illustrated in Figure 4-11. Wavefunction penetration into the oxide blurs the line between the silicon and oxide regions and acts to smooth the kink in the scalar potential. Qualitatively, as channel thicknesses and carrier densities increase, the silicon will more closely resemble a perfect conductor and the kink in the scalar potential at the silicon-oxide interface will become more pronounced. Outside of the

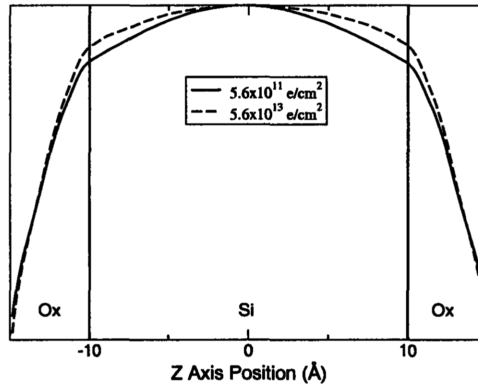


Figure 4-13: Plane-averaged external electrostatic potentials for a 2nm-thick channel at carrier electron densities of $5.6 \times 10^{11}e^-/cm^2$ and $5.6 \times 10^{13}e^-/cm^2$. Both potentials are scaled to be equal in magnitude, to facilitate shape comparisons. The z axis is perpendicular to the interface.

volume inversion regime, the potential inside the channel will have a smaller curvature, making the electric field close to zero.

Figure 4-13 shows the self-consistent external potential ϕ_{ext}, averaged over the plane parallel to the interface, for carrier densities of $5.6 \times 10^{11}e^-/cm^2$ and $5.6 \times 10^{13}e^-/cm^2$. As shown in Figure 4-11, the former density is well within the volume inversion regime, while the latter is at the boundary of this regime. The electric field $E(z)$ discussed above is given by this scalar potential: $E(z) = -d\phi_{ext}/dz$. The self-consistent external potential as a function of channel charge qN is calculated as follows: $\phi_{ext}(qN) = \Phi(qN) - \Phi(0)$, where the total electrostatic potential $\Phi = \phi_{ion} + \phi_H + \phi^0_{ext}$ is the sum of the ionic potential, the Hartree potential due to the electrons, and the unscreened external potential. $\Phi(0)$, the total electrostatic potential with no channel charge, has no unscreened external potential contribution.

The intuition provided by the macroscopic electrostatic model discussed above is confirmed by Figure 4-13. At a carrier density of $5.6 \times 10^{13}e^-/cm^2$, the conduction band electrons are free to move close to the interface and screen the field in the oxide. As a result, the curvature of the potential in the silicon channel is reduced relative to the potential at $5.6 \times 10^{11}e^-/cm^2$, implying a smaller electric field in the channel. As the channel moves out of the volume inversion regime, more charge builds up a the silicon-oxide interface, and the applied gate field in the oxide is screened more

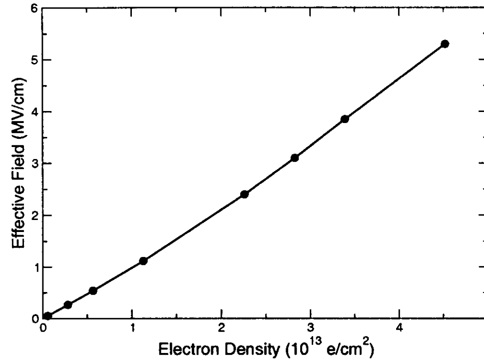


Figure 4-14: Effective field versus conduction electron density for a 20Å-thick silicon channel. The line is a guide to the eye.

efficiently.

An important quantity that is often measured experimentally and calculated theoretically is the "effective field", the applied gate field averaged over the silicon channel. For a symmetric double-gate MOSFET, the effective field $\bar{E}$ is defined by:

$$\bar{E} \equiv \frac{\int_{-\infty}^{0} dz\ E(z)n(z)}{\int_{-\infty}^{0} dz\ n(z)}. \tag{4.50}$$

$E(z)$ and $n(z)$ are the plane-averaged electric field and carrier density, respectively, discussed above. The coordinate system used in Equation 4.50 is the same as that shown in Figure 4-12.

Figure 4-14 shows the effective field as a function of conduction electron density for a 20Å-thick channel. It is linear in the conduction electron density, as would be expected for an undoped channel. An alternative expression for the effective field, which is more easily measured experimentally, is: $\bar{E} = (Q_b + \eta Q_c)/\epsilon_{Si}$, where Q_b is the "bulk charge", and Q_c is the charge in the channel [9]. For an SOI MOSFET, the bulk charge is zero, implying that the effective field is linear in the channel charge. For evaluating the results of self-consistent DFT calculations, using the channel carrier density directly is more convenient. However, for comparison to experimental measurements, it is important to be able to calculate the effective field as described above.

Chapter 5

Interface Roughness Scattering

5.1 Introduction

Interface roughness scattering becomes increasingly important in thinner UTSOI channels, since the carriers are confined in closer and closer proximity to the oxide interface. Current roughness models treat interface fluctuations as a long-wavelength phenomenon, with the thickness of the channel varying gradually over tens of nanometers. These thickness fluctuations can be reduced by improved device processing and the use of shorter channels. However, atomic-scale interface roughness will always be present, as a consequence of strain at the interface and entropy [24, 21]. Current models do not address atomic-scale roughness scattering, but the consequence of having even a single atom out of place in a 1nm-thick channel can be severe. This chapter describes a first-principles approach to atomic-scale interface roughness scattering, based on the methods of Chapter 4. This model has important consequences for future UTSOI device development.

The elemental defects that contribute to interface roughness are suboxide bonds, an Si-Si bond on the oxide side of the interface, and oxygen protrusions, a Si-O-Si bond on the silicon side of the interface. Calculated mobilities based on scattering from these defects allow us to address an important experimental result on UTSOI devices. At low carrier densities, the mobility measured for single- and double-gate UTSOI MOSFETs is significantly reduced relative to the universal mobility curve [52,

31, 130, 144] that describes bulk devices [140, 138, 44, 115]. Our results demonstrate that atomic-scale interface roughness is not responsible for this reduction in mobility, and thus imply that long-wavelength interface-thickness fluctuations are a primary contributor.

5.1.1 Existing Models

Interface roughness scattering is expected to be the dominant form of scattering in UTSOI devices with channels thinner than 4Å [56, 42]. Current models of interface roughness scattering [149, 53, 42] trace back to the ideas of Ando [5] and Prange and Nee [114], and provide a continuum picture of interface roughness. The discussion below follows Esseni [42], and demonstrates how scattering matrix elements can be calculated from these continuum models. At each point $\vec{r}$ in the plane of the interface, the location of the interface is assumed to be displaced by a random amount $\Delta(\vec{r})$. Δ is taken to be a Gaussian-distributed random variable, with average $\bar{\Delta}$ and correlation length Λ:

$$\langle \Delta(\vec{r})\Delta(\vec{r}')\rangle = \bar{\Delta}^2 \exp\left(|\vec{r}-\vec{r}'|^2/\Lambda^2\right). \tag{5.1}$$

$\bar{\Delta}$ and Λ are parameters that can be used to fit the model to experimental data.

If the channel is represented by a square well of width L, and the interfaces represented by potential barriers of height Φ_{ox}, then the unperturbed (zero roughness) potential is:

$$V_0(z) = \Phi_{\text{ox}}\left[\Theta(-z) + \Theta(z-L)\right], \tag{5.2}$$

where Θ is the Heaviside step function, and the z axis is perpendicular to the interface. If fluctuations occur only at the front interface ($z = 0$), then the full potential (including roughness) at a point $\vec{r}$ in the plane of the interface is:

$$V(z,\vec{r}) = \Phi_{\text{ox}}\left[\Theta(-z+\Delta(\vec{r})) + \Theta(z-L)\right]. \tag{5.3}$$

The interface roughness perturbation potential is given by the difference between the

full and unperturbed potentials:

$$V_{\text{IR}}(z,\vec{r}) = \Phi_{\text{ox}}\left[\Theta(-z+\Delta(\vec{r})) - \Theta(-z)\right]. \tag{5.4}$$

The scattering matrix elements $\langle n\vec{k}|V_{\text{IR}}|m\vec{k}'\rangle$ can now be calculated. For the square-well potential, the eigenfunction $|n\vec{k}\rangle$ in real space is given by: $\psi_{nk}(z,\vec{r}) = \exp(i\vec{k}\cdot\vec{r})u_n(z)$, where $\vec{k}$ is a two-dimensional wavevector in the plane of the interface. Using these wavefunctions, the scattering matrix element is:

$$\langle n\vec{k}|V_{\text{IR}}|m\vec{k}'\rangle = \frac{\Phi_{\text{ox}}}{A}\int_A d^2r\; e^{i(\vec{k}'-\vec{k})\cdot\vec{r}} I_{mn}[\Delta(\vec{r})] \tag{5.5}$$

$$I_{mn}[\Delta(\vec{r})] = \int_0^{\Delta(r)} dz\; u_n^*(z)u_m(z), \tag{5.6}$$

where A is the area in the interface plane. If $\Delta(\vec{r})$ is small, the wavefunction can be approximated by its behavior in the vicinity of the interface. From Equation 4.58 of Reference [6], the wavefunction near the interface can be written as:

$$u_n \simeq L_0 \frac{du_n(0)}{dz}\exp(z/L_0), \tag{5.7}$$

where $L_0^2 = \hbar^2/2m^*\Phi_{\text{ox}}$, and $du_n(0)/dz$ is the wavefunction derivative at the interface as $\Phi_{\text{ox}} \to \infty$. m^* is the effective mass. Using Equation 5.7 for $u_n(z)$, I_{mn} can be evaluated explicitly:

$$I_{mn}[\Delta(\vec{r})] = L_0^2 \frac{du_n^*(0)}{dz}\frac{du_m(0)}{dz}\left(\frac{L_0}{2}\right)\left[e^{2\Delta(\vec{r})/L_0} - 1\right]. \tag{5.8}$$

Under the assumption that Δ is small, the exponential in Equation 5.8 can be expanded to first order in Δ, giving $I_{mn}[\Delta(\vec{r})] \propto \Delta(\vec{r})$. Substituting the first-order I_{mn} into Equation 5.6, an simple expression for the interface roughness scattering matrix element can be written:

$$\langle n\vec{k}|V_{\text{IR}}|m\vec{k}'\rangle = \frac{1}{A}\left[\frac{\hbar^2}{2m^*}\frac{du_n^*(0)}{dz}\frac{du_m(0)}{dz}\right]\int_A d^2r\; e^{i(\vec{k}'-\vec{k})\cdot\vec{r}}\Delta(\vec{r}). \tag{5.9}$$

Fermi's Golden Rule is used to calculate the scattering rate from the square of the scattering matrix element, averaged over all possible random interface fluctuations. Based on Equation 5.1 for the real-space distribution function, the average can be evaluated explicitly:

$$\frac{1}{A^2}\int_A d^2r\, e^{i\vec{q}\cdot\vec{r}} \int_A d^2r'\, e^{i\vec{q}\cdot\vec{r}'} \langle\Delta(\vec{r})\Delta(\vec{r}')\rangle = \frac{\pi\bar{\Delta}^2\Lambda^2}{A}\exp\left(|\vec{q}|^2\Lambda^2/4\right) \equiv \frac{1}{A}S_{\mathrm{IR}}(\vec{q}). \quad (5.10)$$

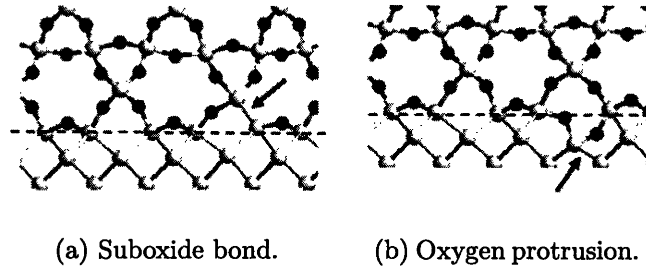
Using this integral, the final expression for the averaged square of the scattering matrix element can be written:

$$|\langle n\vec{k}|V_{\mathrm{IR}}|m\vec{k}'\rangle|^2 = \frac{1}{A}\left[\frac{\hbar^2}{2m^*}\frac{du_n^*(0)}{dz}\frac{du_m(0)}{dz}\right]^2 S_{\mathrm{IR}}(\vec{q}). \quad (5.11)$$

If the power spectrum S_{IR} has significant weight at small $\vec{q}$, as is true for most MOSFET simulations, then the interface roughness consists largely of gradual changes in the channel thickness (long-wavelength interface roughness).

5.1.2 Atomic-Scale Roughness

The picture of continuous, random displacements of the interface may accurately reflect bulk silicon MOSFETs with thick channels, but it is problematic for thin SOI channels. At the atomic-scale, the interface position can fluctuate only by adding or subtracting silicon or oxygen atoms. In the extreme case of a one-nanometer thick UTSOI channel, displacing a single atom changes the thickness of the channel by 25%! The reality of nanoscale channel thicknesses precludes assuming that interface displacements are small and continuous. Previous work [24] has identified the elemental defects that contribute to interface roughness: a Si-Si bond on the oxide side of the interface, called a *suboxide bond* (Figure 5-1(a)), and a Si-O-Si bond on the silicon side of the interface, called an *oxygen protrusion* (Figure 5-1(b)). The suboxide bond consists of a missing oxygen atom on the oxide side of the interface, while the oxygen protrusion consists of an extra oxygen atom on the silicon side. A combination of these two defects leads to a fluctuating interface.



(a) Suboxide bond. (b) Oxygen protrusion.

Figure 5-1: Schematics of an Si-SiO_2 interface showing elementary interface roughness defects.

5.2 First-Principles Results

5.2.1 Scattering Potentials and Screening

In order to investigate the effects of suboxide bonds and oxygen protrusions on mobility in UTSOI channels, relaxed suboxide bond and oxygen protrusion structures have been calculated for channels with thicknesses ranging from 10-25Å. The Born approximation (Equation 4.7) was used for the scattering matrix and the mobility was calculated by solving the Boltzmann equation (Equation 4.9). Figure 5-2 shows the self-consistent scattering potentials for the oxygen protrusion and suboxide bond in a 10Å channel, together with the conduction electron density. The potentials were calculated at a conduction electron density of $5.6 \times 10^{12} \mathrm{e}^-/\mathrm{cm}^2$. The oxygen protrusion consists of an additional oxygen ion, and so provides an attractive potential on the silicon side of the interface. The suboxide bond, consisting of a missing oxygen ion, provides a repulsive potential on the oxide side. The electron density accurately reflects the penetration of the conduction electron wavefunctions into the oxide region, often neglected in MOSFET simulations but important for the accurate description of scattering in the channel.

Calculating the scattering potentials self-consistently at each value of the carrier density includes screening effects that are generally treated only through a semi-empirical dielectric constant (see, for example, Section IIB in Reference [56]). At higher conduction electron densities, electrons will accumulate in the attractive regions of the potential and be repelled from the repulsive regions. As a result, the

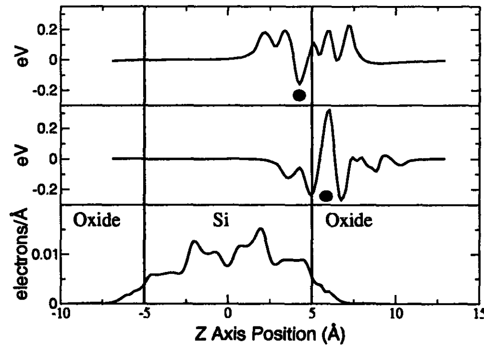


Figure 5-2: Plot showing plane-averaged scattering potentials for oxygen protrusions (top) and suboxide bonds (middle), along with carrier density (bottom). The z axis is perpendicular to the Si-SiO_2 interface. Black circles mark the defect centers.

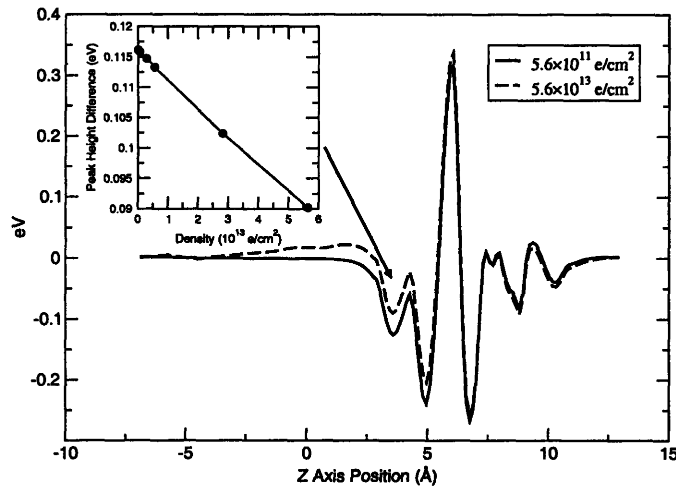


Figure 5-3: Plane-averaged suboxide bond scattering potential at carrier electron densities of $5.6 \times 10^{11}\mathrm{e}^-/\mathrm{cm}^2$ and $5.6 \times 10^{13}\mathrm{e}^-/\mathrm{cm}^2$. The z axis is perpendicular to the interface. The inset shows the height of the depth of the potential well at $z = 3.0$Å as a function of carrier electron density.

"peaks" and "valleys" of the scattering potential are softened, and the overall magnitude of the potential reduced. Figure 5-3 shows the suboxide bond scattering potential as the conduction electron density varies over three orders of magnitude. The inset of the figure plots the depth of the first attractive potential region, showing a linear decrease with increasing electron density.

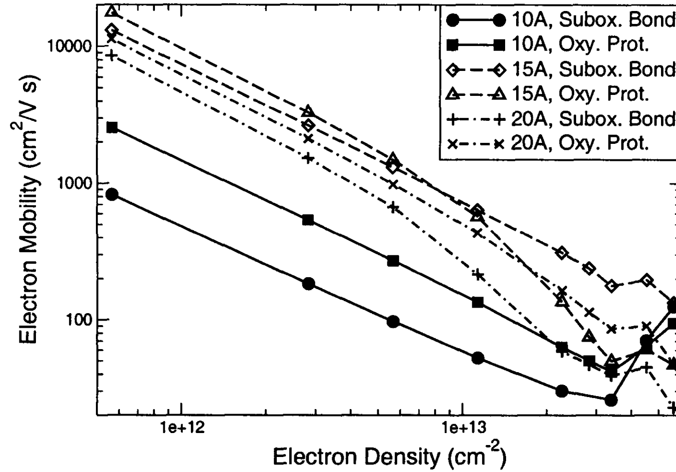


Figure 5-4: Calculated electron mobilities due to scattering from suboxide bonds and oxygen protrusions, for channel thicknesses of 10, 15, and 20Å.

5.2.2 Calculated Mobilities

Figure 5-4 shows the calculated mobilities due to suboxide bonds and oxygen protrusions, for 10Å-, 15Å-, and 20Å-thick channels. The density of both suboxide bonds and oxygen protrusions is taken to be 5.6×10^{11}defects/cm^2, the same order of magnitude as observed interface trap densities [138]. The defect density will in general depend on film growth and device processing conditions, but as shown by Equation 4.11, the defect density only provides an overall scaling factor for the mobilities.

The most surprising feature of Figure 5-4 is the fact that at lower electron densities, for each channel thickness, the oxygen protrusion-limited mobility is greater than the suboxide bond-limited mobility. This is counterintuitive because, as shown in Figure 5-2, the oxygen protrusion and suboxide bond scattering potentials are of roughly equal strength, while the oxygen protrusion lies within the silicon channel and thus overlaps more strongly with the electron density. As a result, oxygen protrusions would be expected to be stronger scatterers than suboxide bonds. The explanation for the unexpectedly high oxygen protrusion-limited mobility lies in the sp^3-antibonding character of the states at the bottom of the conduction band in silicon. The extra oxygen atom that constitutes the protrusion lies near the middle of the Si-Si bond, where an sp^3-antibonding wavefunction has a node. Figure 5-5 shows that this intu-

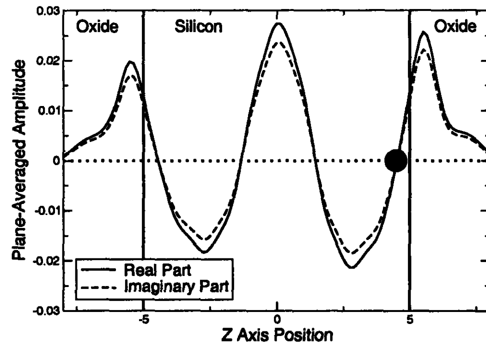


Figure 5-5: Plane-averaged wavefunction at the bottom of the conduction band in a 10Å-thick UTSOI channel. The z axis is perpendicular to the Si-SiO_2 interface. A black circle marks the center of the oxygen protrusion.

ition is valid for a UTSOI channel. Since the center of the oxygen protrusion coincides with a node in the wavefunctions at the bottom of the conduction band, the overlap with the scattering potential is reduced and an increased mobility results.

Two competing effects influence the shape of the mobility curves in Figure 5-4. The first, discussed in Section 4.3.2, is volume inversion. At higher electron densities, the additional potential energy gained by drawing the electrons toward the silicon-oxide interfaces more than compensates for the kinetic energy cost of further localizing the electrons. As a result, the electron density is peaked at the interfaces instead of in the center of the channel. This qualitative change in the shape of the electron distribution is shown in Figure 4-11. Increasing the electron density at the interfaces increases the overlap with the suboxide bond and oxygen protrusion scattering potentials, and decreases the mobility. Since the volume inversion effect is more pronounced in thicker channels, it is the 15Å- and 20Å-thick channels that show a marked decrease in mobility at higher electron densities.

The second effect, which dominates the behavior of thin channels, is potential screening. As discussed in Section 5.2.1, the scattering potential is "softer" at higher electron densities. This reduction in the strength of the scattering potentials increases the mobility at higher electron densities. In the 10Å-thick channel, in which volume inversion is not a pronounced effect, the mobilities begin to increase around $4 \times 10^{13} e^-/cm^2$ due in part to potential screening. In ultrathin channels, the balance between screening and volume inversion determines the mobility at high carrier

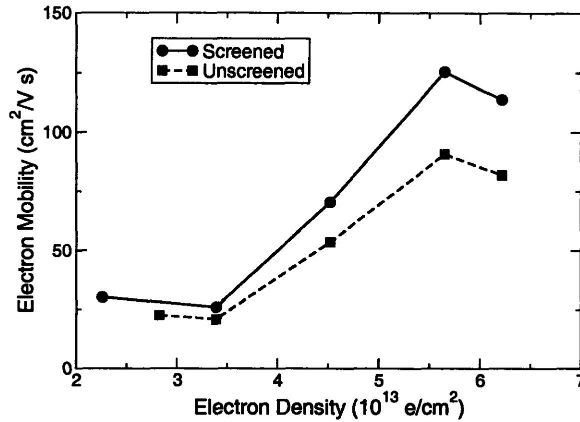


Figure 5-6: Suboxide-bond-limited electron mobilities calculated with screened and unscreened scattering potentials. The channel thickness is 10Å.

densities, leading to qualitatively different behavior in channels of different thickness.

Unscreened Mobilities

Without self-consistent screening, the softening effect would be absent and the resulting mobilities would be unphysically low at higher carrier densities. To illustrate the importance of this effect, Figure 5-6 shows calculated suboxide-bond-limited mobilities with screened scattering potentials and without. The unscreened scattering potential is taken to be the potential calculated self-consistently at a conduction electron density of $5.6 \times 10^{11} e^-/cm^2$. The scattering potential calculated at this electron density is used with all subsequent calculations.

At higher conduction electron densities, the screening of the potential becomes more effective and the difference between the screened and unscreened mobilities grows larger. The unscreened mobilities are consistently lower than the screened mobilities due to the deeper "valleys" and higher "peaks" of the unscreened potential, as discussed in Section 5.2.1.

5.2.3 Comparison to Existing Models and Experiments

The electron mobility measured in UTSOI MOSFETs remains constant or even decreases with decreasing electron density in the $\lesssim 10^{13} e^-/cm^2$ range [140, 138, 44, 115], falling well below the universal mobility curve (UMC) that describes bulk MOS-

FETs [52, 31, 130, 144]. Theoretical work [53, 43, 42], based on the interface roughness models described in Section 5.1.1, has reproduced qualitatively this UMC deviation. In the literature, scattering off of these interface fluctuations is considered to be distinct from "T_{Si} fluctuation scattering" [117, 140, 138, 139], in which variations in the channel thickness affect the eigenspectrum of the channel states. However, Esseni [42] has shown that both types of scattering show the same dependence on channel thickness, with the same numerical prefactor. In fact, since the origins of both models lie in fluctuations of the interface position, there is little reason to consider them independent scattering mechanisms. In the following discussion, scattering caused by interface fluctuations will be referred to as *long-wavelength roughness scattering.*

The present model, based on atomic-scale defects, reflects short-wavelength roughness. As pointed out by Uchida *et al.* [140] and Ernst *et al.* [41], it is difficult to fabricate long-channel UTSOI MOSFETs with uniform channel thickness. Improved device processing and shorter channel lengths can be expected to lead to greater uniformity and significantly decrease long-wavelength roughness. Suboxide bonds and oxygen protrusions, on the other hand, are elemental interface roughness defects. Short-wavelength roughness will likely always be present due to strain at the Si-SiO_2 interface [21, 24]. As shown in Fig. 5-4, the mobility for suboxide bonds and oxygen protrusions at all channel thicknesses is proportional to $1/n_e$ for low electron densities ($\lesssim 10^{13}\text{e}^-/\text{cm}^2$), where n_e is the conduction electron density. At low electron densities, short- and long-wavelength roughness lead to mobility contributions that can differ by orders of magnitude.

Simulations of a 25Å-thick UTSOI channel based on Esseni's interface roughness scattering model [42] show that at room temperature and electron densities below $10^{13}\text{e}^-/\text{cm}^2$, the phonon and long-wavelength roughness scattering contributions to the mobility are roughly equal. As a result, the total mobility μ_{tot}, given by Mathiessen's rule, is: $1/\mu_{\text{tot}} = 1/\mu_{\text{phonon}} + 1/\mu_{\text{LWR}} \approx 2/\mu_{\text{phonon}}$, where μ_{LWR} is the mobility due to long-wavelength roughness scattering. If improved device fabrication processes can reduce or eliminate long-wavelength interface thickness fluctuations, only short-wavelength roughness due to oxygen protrusions and suboxide bonds will remain.

Thus, the total mobility becomes: $1/\mu_{\text{tot}} = 1/\mu_{\text{phonon}} + 1/\mu_{\text{SWR}} \approx 1/\mu_{\text{phonon}}$, since $1/\mu_{\text{SWR}} \approx 0$ at low electron densities. Due to the rapid increase in short-wavelength roughness mobility with decreasing electron densities, reducing UTSOI channel thickness fluctuations could result in up to a 100% gain in mobility at low carrier densities.

Chapter 6

Quantum Transport

6.1 Introduction

Current state-of-the-art MOSFETs have dimensions that are well within the nanoscale regime. The length and width of the channel are roughly 40nm, and as discussed in Chapter 4, the channel can be as thin as 1-2nm. Despite the continued shrinking of silicon technology, the active channel region of MOSFETs still contains roughly $100,000$ atoms. If channel is represented as a potential well, then the level spacing in a MOSFET-sized box is roughly 10^{-5}eV. The connection to the source and drain regions further broadens the levels into a continuum. Despite the small size of modern MOSFETs, semiclassical transport calculations that require a continuous band structure are still accurate.

What changes as devices shrink further? A very small device (with characteristic dimensions of 1-10nm), weakly coupled to leads, will have a discrete spectrum of states. If the temperature is less than the spacing between levels in the device, the transport process depends crucially on the properties of each individual level. Due to the small size of the device, Coulomb interactions between electrons on the device are strong and can shift significantly the single particle spectrum. As a result, correlations can be very important to transport in small devices. A device that displays some or all of the properties described above will be referred to as a *quantum device.* Figure 6-1 shows a schematic of a quantum device connected to two leads,

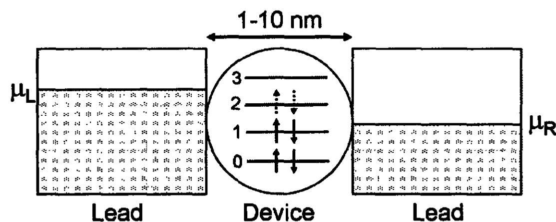


Figure 6-1: Schematic of a quantum device connected to two leads. The left and right leads are metals in thermal equilibrium, with chemical potentials μ_L and μ_R. The energy levels for the device, labeled $0, 1, 2, \ldots$, are discrete and widely spaced in the absence of the leads.

and introduces concepts that will be important in subsequent discussions. Both the left and the right lead are assumed to be metals in thermal equilibrium, described by chemical potentials μ_L and μ_R, respectively. In the absence of the leads, the device has a discrete spectrum of levels, labeled $0, 1, 2, \ldots$ in the figure. The left and right chemical potential determine the occupancies of the device levels. Levels 0 and 1 are doubly-occupied since they lie below the chemical potentials of both the left and right leads. The occupancy of level 2 is more ambiguous. Since it lies between μ_L and μ_R, the number of electrons in level 2 will be a dynamic quantity that depends on the transport process.

Examples of such quantum devices include resonant tunneling diodes [29, 125, 49], quantum dots [11, 133], and single-electron transistors [137, 3, 131]. All three types of devices share the characteristics described above, but each device demonstrates unique physics that result in unusual electrical responses. The resonant tunneling diode consists of a confined structure with a discrete bound state, separated from two leads by barriers (see Figure 6-2, part A). At a low applied bias (part B), the energy level of the bound state is aligned with the occupied states in both the left and right lead, allowing a current to flow as electrons tunnel from the left lead to the right via the bound state. At higher voltages (part C), the bound-state energy falls below the occupied states in the left lead, preventing resonant tunneling through the bound state. As a result, the current decreases as voltage increases, leading to negative differential resistance.

Quantum dots are small semiconductor nanostructures ($\lesssim$ 100nm in diameter)

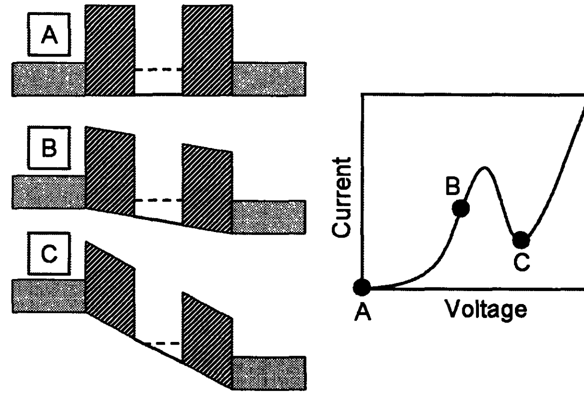


Figure 6-2: Schematic representation of a resonant tunneling diode. On the right is a representative current-voltage (I-V) curve showing negative differential resistance. A, B, and C as labeled on the I-V curve correspond to the band diagrams on the left. Metallic leads are represented by the dotted regions, barriers are dashed regions. The dashed line marks the energy of the bound-state between the barriers. After W. R. Frensley [49].

that contain only a few free electrons. The dot contains thousands or millions of atoms and as a result, the effective mass theory applies and permits electrons in the conduction band to be thought of as free carriers with an effective mass. For GaAs, the conduction band effective mass m^* is only $0.07m_0$, resulting in an effective Bohr radius of $a_0^* = \hbar^2/2m^*e^2 = 0.76nm$. The small effective mass of the free carriers in a quantum dot results in a new, larger length scale for the dot. With a proper confining potential, the quantum dot can appear to be a single artificial atom [11], with associated shell-filling effects [133] as electrons are added. Coulomb interactions between electrons as they are added and removed from the dot result in temporal correlations in the current that have recently been measured [93].

The single-electron transistor (SET) is similar to the resonant tunneling diode described above, but with the addition of a gate to control the position of the energy levels in the dot. When an electron hops on to a bound-state energy level in the potential well, it costs an amount of energy, called the "charging energy", equal to e^2/C, where C is the capacitance of the well. Figure 6-3 shows schematically the energy levels and gate voltage-conductance curve for a SET. If the bound-state level is above the Fermi level in the leads (Figure 6-3, part A), then no electrons can tunnel through from the source to the drain, and the conductance is zero. As the gate voltage

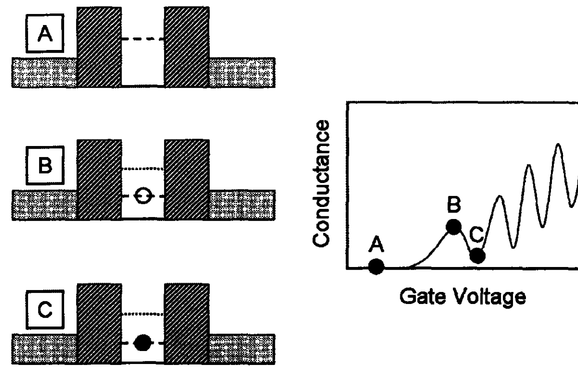


Figure 6-3: Schematic representation of a single-electron transistor. The gate voltage shifts the position of the bound-state energy levels in the potential well. On the right is a representative gate voltage-conductance curve, inspired by the low-temperature data of Takahashi *et al.* [131]. Points A, B, and C on the curve correspond to the band schematics on the left. The dotted regions represent the metallic source and drain, the shaded regions are barriers. The dashed line within the potential well shows the bound-state energy level. The dotted line in B and C represents the energy level for a doubly-occupied bound state, e^2/C above the singly-occupied level due to the charging energy.

changes, the bound-state energy level aligns with the source/drain Fermi level, and the conductance peaks (part B). As the gate voltage increases further, the lowest bound-state level is singly-occupied, and it costs an additional energy e^2/C to add another electron. As a result, the levels in the well are not aligned with the leads and the conductance decreases. Increasing gate voltage leads to a series of conductance peaks due to the finite gap between potential well levels induced by the charging energy.

Calculating the transport properties of such quantum devices is difficult since semi-classical approximations are clearly inapplicable and many of the unique features of such devices occur only a finite bias, making quantum linear response calculations (using the Kubo formula) insufficient. The effectively infinite size of the leads and the lack of periodicity in the device-plus-leads system makes first-principles calculations difficult. The most promising approach to such a problem is to construct a Hamiltonian from first-principles, including the level spectrum of the isolated device, the densities of states in the leads, and the coupling between them. For time-independent calculations or those that seek only the long-time behavior of the system,

Non-Equilibrium (Keldysh) Green's Function techniques, described in Section 6.1.1 below, may be used. If the short-time behavior, including quantum fluctuations as in Reference [93], of the system is desired, new approaches must be developed. This chapter discusses progress toward such a theory of short-time fluctuations in quantum devices, based on an effective Hamiltonian for the device degrees of freedom.

6.1.1 Non-Equilibrium Green's Functions

Quantum Field Theory methods were first used to solve non-equilibrium many-body problems by Martin and Schwinger in 1959 [98]. Field-theoretic techniques emerged in their modern form, based on Green's functions and diagrammatic expansions, through the work of Kadanoff and Baym [78], Keldysh [80], and Langreth [88]. It is not the intent of this chapter to present a complete review of the Non-Equilibrium Green's Function theory or its numerous applications, but rather to briefly review its major points and examine its limitations. For a complete review with an emphasis on electron transport, the reader is referred to Reference [116].

A system in equilibrium is a time-independent system, from the standpoint of measurable quantities. Once a system reaches equilibrium, currents, voltages, temperatures, and other parameters remain constant. In order to drive the system out of equilibrium, some time-dependent potential must be applied. The basic description of a non-equilibrium system rests on a three-part Hamiltonian:

$$\mathcal{H} = H_0 + H_{\text{int}} + H'(t). \tag{6.1}$$

H_0 is the non-interacting part of the Hamiltonian; as in the traditional Green's function context, H_0 is assumed to exactly solvable. H_{int} represents the interparticle interactions, e.g. the Coulomb interaction. $H'(t)$ is an arbitrary time-dependent potential, turned on at $t = 0$, that serves to drive the system out of equilibrium. The time-independent components of the Hamiltonian are given by H: $H = H_0 + H_{\text{int}}$, $\mathcal{H} = H + H'(t)$.

For a system in equilibrium, e.g. if $H'(t) = 0$, the expectation value of any observ-

able O is given by: $\langle O\rangle = \mathrm{Tr}[\rho(H)O]$, where $\rho(H) = (\mathrm{Tr}[\exp(-\beta H)])^{-1}\exp(-\beta H)$ is the density matrix. Non-equilibrium statistical physics is concerned with calculating time-dependent expectation values given by:

$$\langle O(t)\rangle = \mathrm{Tr}[\rho(H)O_{\mathcal{H}}(t)], \tag{6.2}$$

for $t > 0$. $O_{\mathcal{H}}$ is the operator O in the Heisenberg representation with respect to the full, time-dependent Hamiltonian $\mathcal{H}$: $O_{\mathcal{H}} = \exp(i\mathcal{H}t)O\exp(-i\mathcal{H}t)$.

The equilibrium density operator in Equation 6.2 seems to be an unusual choice to represent a non-equilibrium statistical average, but is simply a consequence of using the Heisenberg representation to time-evolve a system that is initially in equilibrium. At times $t < 0$, $H'(t) = 0$ and the system is in equilibrium. As such, the system is in a mixed state characterized by the density matrix $\rho(H) : |\Psi_0\rangle = \sum_\alpha \exp(-\beta E_\alpha)|\alpha\rangle$, where $\{|\alpha\rangle, E_\alpha\}$ are the eigenstates and eigenvalues of H. For $t > 0$, after H' is "turned on", the system's state Ψ can evolve with time, as in the Schrödinger picture, or the operators can evolve with $|\Psi_0\rangle$ fixed. It is important to realize that in either context, the density operator $\rho(H)$ does not behave as an operator in the Heisenberg sense, i.e. it does not evolve with time. Instead, it is part of the wavefunction and remains fixed in the Heisenberg picture.

The first step in calculating non-equilibrium expectation values using Equation 6.2 is to transform to the interaction representation with respect to the time-independent component of the Hamiltonian H. In the theory of equilibrium finite-temperature (Matsubara) Green's functions, the expectation value of O in this representation is:

$$\langle O\rangle = \frac{\langle\Psi_0|\mathrm{T}\left[S(-\infty,\infty)O_H(t)S(\infty,-\infty)\right]|\Psi_0\rangle}{\langle\Psi_0|\mathrm{T}\left[S(-\infty,\infty)S(\infty,-\infty)\right]|\Psi_0\rangle}, \tag{6.3}$$

with the time evolution operator $S(t_0, t)$ given by:

$$S(t_0,t) = \mathrm{T}\left[\exp\left(-i\int_{t_0}^{t} dt'\; H'_H(t')\right)\right]. \tag{6.4}$$

T is the time-ordering operator, and O_H is the operator O in the interaction represen-

tation with respect to H: $O_H = \exp(iHt)O\exp(-iHt)$. $H'_H(t)$ is the time-dependent component of the full Hamiltonian $\mathcal{H}$ in the interaction representation; this operator now has both explicit time-dependence and additional time-dependence from the exponential factors.

Equilibrium systems do not allow for irreversible effects as the system evolves from $t = -\infty$ to $+\infty$. Thus, $\Psi(t \to \infty) = \exp(i\alpha)\Psi(t \to -\infty)$, where α is a constant phase. This phase relationship of course excludes systems such as superconductors in which the interacting final state is of a different symmetry than the non-interacting initial state, but all other systems can take advantage of this convenient relation. As a result, the expectation value simplifies further:

$$\langle O \rangle = \frac{\langle \Psi_0 | \mathrm{T}\left[O_H(t) S(\infty, -\infty)\right] | \Psi_0 \rangle}{\langle \Psi_0 | S(\infty, -\infty) | \Psi_0 \rangle}. \tag{6.5}$$

In the non-equilibrium case, however, irreversible effects are unavoidable, and no simple relation connects the final and initial states. Keldysh [80] provided the insight necessary to successfully generalize the equilibrium Green's function formalism to non-equilibrium situations. Instead of evolving the system from $-\infty$ to ∞, with time moving in a single direction, it is possible to follow a path from $-\infty$ to ∞ and back to $-\infty$. As a result of this choice, the only state that must be well-defined *a priori* is the state at $t \to -\infty$. This is, of course, the initial equilibrium state of the system, before the time-dependent Hamiltonian $H'(t)$ is turned on.

Non-equilibrium expectation values, like their equilibrium counterparts, can be calculated using Green's function techniques (see Reference [95] for an extensive textbook treatment). A one- or two-body operator O will require the calculation of familiar one- and two-body Green's functions. Green's functions possess simple perturbation expansions (in $H'(t)$ and H_{int}), and as such are easier to calculate than direct expectation values such as Equation 6.2. The Green's function with the most convenient perturbation expansion is the contour-ordered Green's function $G_{\mathrm{T_c}}(1, 1')$:

$$G_{\mathrm{T_c}}(1, 1') = \frac{\langle \Psi_0 | \mathrm{T_c}\left[\psi_H(1)\psi_H^\dagger(1') S(-\infty, \infty) S(\infty, -\infty)\right] | \Psi_0 \rangle}{\langle \Psi_0 | \mathrm{T_c}\left[S(-\infty, \infty) S(\infty, -\infty)\right] | \Psi_0 \rangle}. \tag{6.6}$$

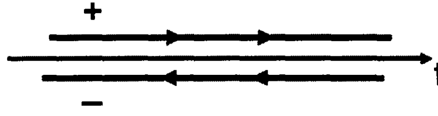


Figure 6-4: The Keldysh countour. The contour extends along the time (t) axis from $-\infty$ to ∞ and back to $-\infty$. The + branch corresponds to increasing time, while the − branch corresponds to decreasing time.

$\psi_H(1)$ is a field operator in the interaction picture at the space and time position $1 = (\vec{x}_1, t_1)$. The time-ordering operator $\mathrm{T_c}$ now orders times along the contour in Figure 6-4, introduced by Keldysh [80]. Times that are further along the contour are placed to the left. The times in $S(\infty, -\infty)$ lie on the + branch, and those in $S(-\infty, \infty)$ lie on the − branch. Using the Keldysh contour, a new time-evolution operator can be introduced:

$$S_c = \mathrm{T_c}\left[\exp\left(-i\int_c dt'\ H'_H(t')\right)\right], \tag{6.7}$$

where $\int_c$ represents integration along the Keldysh contour shown in Figure 6-4.

To account for all of the possible time-orderings, it is necessary to introduce two new Green's functions, in addition to the time-ordered and anti-time-ordered Green's functions familiar from equilibrium theory. Since this chapter is concerned with electron transport, the Green's functions for fermions will be presented. The time- and anti-time-ordered Green's functions retains their familiar forms:

$$\begin{aligned} G_{\mathrm{T}}(1,1') &= -i\langle \mathrm{T}\left[\psi_{\mathcal{H}}(1)\psi^{\dagger}_{\mathcal{H}}(1')\right]\rangle \\ G_{\widetilde{\mathrm{T}}}(1,1') &= -i\langle \widetilde{\mathrm{T}}\left[\psi_{\mathcal{H}}(1)\psi^{\dagger}_{\mathcal{H}}(1')\right]\rangle. \end{aligned} \tag{6.8}$$

$\widetilde{\mathrm{T}}$ is the anti-time-ordering operator that puts earlier times to the left. Field operators are represented in the Heisenberg picture with respect to the complete (time-dependent) Hamiltonian $\mathcal{H}$. Angle brackets correspond to expectation values with respect to the state $|\Psi_0\rangle$ introduced above. The two new Green's functions corre-

spond to a fixed ordering of $\psi(1)$ and $\psi^\dagger(1')$:

$$
\begin{aligned}
G^<(1,1') &= i\langle\psi^\dagger_{\mathcal{H}}(1')\psi_{\mathcal{H}}(1)\rangle \\
G^>(1,1') &= -i\langle\psi_{\mathcal{H}}(1)\psi^\dagger_{\mathcal{H}}(1')\rangle.
\end{aligned}
\tag{6.9}
$$

The need for the new Green's functions in Equation 6.9 comes from using contour-ordering instead of time-ordering. Consider two times, t and t'. If both t and t' are located on the $+$ branch of the Keldysh contour, then the contour-ordering operator $\mathrm{T_c}$ will put t to the left of t' if $t > t'$, and vice-versa, corresponding to the time-ordering operator. If both t and t' are both on the $-$ branch, then $\mathrm{T_c}$ will place t' to the left of t if $t > t'$, corresponding to anti-time-ordering. These two situations, where both times are on the same branch of the Keldysh contour, lead to the time- and anti-time-ordered Green's functions in Equation 6.8. If t is on the $+$ branch and t' on the $-$ branch, however, then t' will always be to the left. In the opposite case, where t' is on the $+$ branch and t on the $-$, t will always be on the left. These situations correspond to the new, fixed-order Green's functions in Equation 6.9. Because the time-evolution operator S_c involves times on both the $+$ and $-$ branch, all four Green's functions will enter into the expansion.

The contour-ordered Green's function $G_{\mathrm{T_c}}$ can be represented by a two-by-two matrix that accounts for the four possible combinations of t and t' on the two branches of the Keldysh contour. Each component of $G_{\mathrm{T_c}}$ corresponds to one combination, shown here schematically:

$$
G_{\mathrm{T_c}}(1,1') = \begin{pmatrix} G_{++} & G_{+-} \\ G_{-+} & G_{--} \end{pmatrix}.
\tag{6.10}
$$

G_{+-} is the component with t_1 on the $+$ branch and t'_1 on the $-$ branch, and so forth. As discussed above, these $G_{\pm\pm}$ components correspond to the four Green's functions

introduced in Equations 6.8 and 6.9:

$$G_{\mathrm{T_c}}(1,1') = \begin{pmatrix} G_{\mathrm{T}}(1,1') & G^{<}(1,1') \\ G^{>}(1,1') & G_{\bar{\mathrm{T}}}(1,1') \end{pmatrix}. \tag{6.11}$$

$G_{\mathrm{T_c}}$ can be readily expanded in this two-by-two matrix representation, called Keldysh space. Interactions in Keldysh space are likewise represented by matrices. Using the ideas of Keldysh, the full machinery of equilibrium Green's function perturbation theory can be applied without change to non-equilibrium systems. The only change is that each Green's function line, interaction line and vertex represents a two-by-two matrix instead of a scalar function. Linear transformations in Keldysh space can simplify the matrix representation of interactions. The widely-used rotation introduced by Larkin and Ovchinnikov [89], for example, makes a one-body interaction a diagonal matrix.

Limitations of the Green's Function Method

The non-equilibrium Green's function (NEGF) formalism introduced above is a powerful theoretical tool for treating quantum devices. Despite the difficulty in calculating high-order terms in the perturbation expansion, numerous groups have used NEGF techniques to study resonant-tunneling diodes [86], single-molecule transistors [148], and other quantum systems. However, the statistical nature of the NEGF method limits its applicability to high-frequency responses. The density matrix $\rho(H)$ implies an uncorrelated ensemble average over accessible states of the system. The ergodic theorem asserts that this ensemble average over many equivalent systems is the same as a time average over a single system. For systems in equilibrium, this statement is generally uncontroversial because all quantities of interest are time-independent.

Non-equilibrium systems, however, are explicitly time-dependent. The quantities of interest display measurable steady-state and transient behavior before the system settles into equilibrium. The time-dependence in the NEGF method enters only through the Hamiltonian $H'(t)$, enabling the study of systems under AC bias [75] but not reproducing the correlated fluctuations observed in experiments even under DC

bias [93]. The NEGF method is based on the ensemble average, while experiments are generally performed on a single systems over long (laboratory) time scales. The ergodic theorem must be valid in order to ensure that quantities calculated using theoretical statistical approaches are the same as those measured experimentally. For non-equilibrium phenomena, this equivalence requires that measurable quantities (i.e. currents, voltages) vary over time scales that are long compared to the time required for a statistical average. Denoting the time needed for the system to sample all of the states in its accessible phase space by τ, and the characteristic time scale over which the system's quantities of interest vary by T, the ergodic hypothesis for non-equilibrium systems holds if $T \gg \tau$.

If $T \lesssim \tau$, then the ergodic hypothesis will not hold, and statistical averaging will smooth out the short-time behavior of the system. Such a situation could result from probing the response of systems to very high-frequency radiation, or if quantum fluctuations are measured in real time. In these cases, an approach that can follow this short-time behavior without averaging is needed.

6.1.2 Quantum Fluctuations

Recent experiments have measured the tunneling of electrons in quantum dots in real time [93]. These results have sufficient time resolution to record temporal correlations in the measured current and voltage noise. Such correlations were predicted by Ben-Jacob and Gefen [16] on the basis of a model tunnel junction Hamiltonian. In order to model more realistic systems that correspond to experimental configurations, the system Hamiltonian must include not only the quantum device (tunnel junction or dot), but also the leads that connect the device to the measurement and control apparatus. While a quantum device in isolation is a closed system, the device and leads together constitute an open system without simple translational symmetry.

The approach described below maps this complex open problem onto a simpler problem that, while still open, deals only with the device degrees of freedom. The degrees of freedom in the leads are traced out by assuming that the leads remain in equilibrium. Results calculated within this approximation are discussed, and a

method to systematically extend the equilibrium-lead approximation is presented.

6.2 Effective Hamiltonian Method

The Hamiltonian for the device-plus-leads system is that used originally by Meir and Wingreen [100], and subsequently by many others. The leads are taken to be non-interacting, but the device Hamiltonian can contain arbitrary many-body interactions. With the Hamiltonian for the isolated device given by H_{int}, the device-plus-leads Hamiltonian is:

$$H = H_{\text{int}}(\{c_n^\dagger\};\{c_n\}) + \sum_{k_i} E_{k_i} d_{k_i}^\dagger d_{k_i} + \sum_{k_i n} V_{k_i n} d_{k_i}^\dagger c_n + \text{h.c.}, \tag{6.12}$$

where $\{c_n^\dagger\}$ and $\{c_n\}$ form a complete and orthogonal set of creation and annihilation operators for the device, and $d_{k_i}^\dagger$ creates an electron in state $|k_i\rangle$ in the lead labeled by the subscript i (e.g. for the system shown in Figure 6-1, $i = \text{L}$ or $i = \text{R}$). $\{V_{k_i n}\}$ determine the coupling between the leads and the device. "h.c." refers to the Hermitian conjugate of the preceding term. These couplings can be calculated from first-principles, and will depend on the details of the lead-device interface: bonding patterns, the presence of insulating tunnel barriers, etc. First-principles calculations can also provide the energy spectra of the device and leads in isolation, allowing a direct connection between the atomic-scale structure of the system and the Hamiltonian in Equation 6.12.

The currents flowing into and out of the device can be calculated from expectation values of the appropriate current operator. Each state $|k_i\rangle$ in lead i provides a channel through which current can flow into the device. The current from channel k_i into the device can be related to the time rate of change of the occupancy of that state:

$$I_{k_i} = -\frac{d\langle n_{k_i}\rangle}{dt} = -\frac{d\langle d_{k_i}^\dagger d_{k_i}\rangle}{dt} = \frac{ie}{\hbar}\sum_n \left(V_{k_i n}\langle d_{k_i}^\dagger c_n\rangle - V_{k_i n}^*\langle c_n^\dagger d_{k_i}\rangle\right), \tag{6.13}$$

where $\langle \dots \rangle$ indicates the expectation value with the (unknown) wavefunction of the

device-plus-leads system. Equation 6.13 is derived directly from the Heisenberg equation of motion for the number operator: $d(d^\dagger_{k_i} d_{k_i})/dt = (-i/\hbar)\left[H, d^\dagger_{k_i} d_{k_i}\right]$. The expression for the current in terms of expectation values leads directly to an expression for the current operator:

$$\hat{I}_{k_i} = \frac{ie}{\hbar}\sum_n \left(V_{k_i n} d^\dagger_{k_i} c_n - V^*_{k_i n} c^\dagger_n d_{k_i} \right). \tag{6.14}$$

The total current into the device from lead i is given by the sum over all channels k_i: $I_i = \sum_{k_i} \langle \hat{I}_{k_i} \rangle$.

6.2.1 A New Basis

A sufficiently large lead can be viewed as a reservoir in thermal equilibrium. As such, each lead is associated with a chemical potential μ_i and temperature T. The entire device-plus-leads system is assumed to be in thermal equilibrium, so that the temperature is constant across all leads. Chemical potential differences are maintained by external biases. The electrons in the leads are non-interacting, allowing the wavefunction within the lead to be written as:

$$|\mu_i\rangle = \prod_{k_i} |\psi_{k_i}\rangle, \tag{6.15}$$

where

$$|\psi_{k_i}\rangle = |k_i, 0\rangle \sqrt{1 - f_{k_i}} + |k_i, 1\rangle e^{i\phi_{k_i}} \sqrt{f_{k_i}}. \tag{6.16}$$

$|k_i, 0\rangle$ means that state $|k_i\rangle$ is unoccupied, and $|k_i, 1\rangle$ that state $|k_i\rangle$ is occupied: i.e., $|k_i, 1\rangle = d^\dagger_{k_i}|k_i, 0\rangle$. $f(k_i) = f(E_{k_i} - \mu_i)$ is the Fermi distribution function. The random phase ϕ_{k_i} reflects fluctuations of the states in the leads, and the effect of the device on the lead states is neglected. It is easily verified that $|\psi_{k_i}\rangle$ corresponds to the state $|k_i\rangle$ being occupied according to the Fermi distribution: $\langle \psi_{k_i} | n_{k_i} | \psi_{k_i} \rangle = f(k_i)$.

The orthonormal basis in the leads is completed by the following wave functions:

$$|\{l_i\}\rangle = \prod_{k_i \notin \{l_i\}} |\psi_{k_i}\rangle \prod_{\{l_i\}} |\chi_{l_i}\rangle, \tag{6.17}$$

where

$$|\chi_{l_i}\rangle = |l_i, 0\rangle \sqrt{f_{l_i}} - |l_i, 1\rangle e^{i\phi_{l_i}} \sqrt{1 - f_{l_i}}. \tag{6.18}$$

$\{|\{l_i\}\rangle\}$ constitute a set of states that reflect excitations above the thermal "ground state" $|\mu_i\rangle$. This is most easily seen by taking the limit $T \to 0$, in which the Fermi distribution function becomes a step function. In this case, for an energy E_{k_i} below the chemical potential (Fermi energy) μ_i, $|\psi_{k_i}\rangle$ is the electron state and $|\chi_{k_i}\rangle$ the hole state. The opposite is true for E_{k_i} above the Fermi energy. The $\{|\psi_{k_i}\rangle, |\chi_{k_i}\rangle\}$ basis set defined in Equations 6.16 and 6.18 is a unitary transformation from the $\{|k_i, 0\rangle, |k_i, 1\rangle\}$ basis set defined previously. The creation and annihilation operators for $\{|k_i, 0\rangle, |k_i, 1\rangle\}$, $\{d^\dagger_{k_i}, d_{k_i}\}$, are transformed into $\{b^\dagger_{k_i}, b_{k_i}\}$, for which the new "vacuum" state is $|\mu_i\rangle$. The transformation is:

$$d^\dagger_{k_i} = \sqrt{f_{k_i}(1 - f_{k_i})}(1 - 2b^\dagger_{k_i} b_{k_i}) - (1 - f_{k_i}) b^\dagger_{k_i} + f_{k_i} b_{k_i}, \tag{6.19}$$

and

$$d^\dagger_{k_i} d_{k_i} = f_{k_i} + (1 - 2f_{k_i}) b^\dagger_{k_i} b_{k_i} - \sqrt{f_{k_i}(1 - f_{k_i})}(b^\dagger_{k_i} + b_{k_i}), \tag{6.20}$$

where the phase factor $e^{i\phi_{k_i}}$ has been absorbed into $b^\dagger_{k_i}$. After the transformation, the Hamiltonian in Equation 6.12 becomes:

$$\begin{aligned} H = H_{\text{int}}(\{c^\dagger_n\}; \{c_n\}) &+ \sum_{k_i n} \sqrt{f_{k_i}(1 - f_{k_i})} V_{k_i n} c_n + \text{h.c.} \\ &+ \sum_{k_i n} \left\{ -2\sqrt{f_{k_i}(1 - f_{k_i})} b^\dagger_{k_i} b_{k_i} - (1 - f_{k_i}) b^\dagger_{k_i} + f_{k_i} b_{k_i} \right\} V_{k_i n} c_n + \text{h.c.} \\ &+ \sum_{k_i} E_{k_i} \left\{ f_{k_i} + (1 - 2f_{k_i}) b^\dagger_{k_i} b_{k_i} - \sqrt{f_{k_i}(1 - f_{k_i})}(b^\dagger_{k_i} + b_{k_i}) \right\}, \end{aligned} \tag{6.21}$$

The thermal ground state $|\mu_i\rangle$ and set of excitations $\{|\{l_i\}\rangle\}$ for each lead pro-

vide an intuitive partition of the Hilbert space for the device-plus-leads system. For simplicity, consider a quantum device connected to a single lead. States in the device-plus-leads Hilbert space are products of a lead state, for example $|\mu\rangle$, and a device state: $|\text{system}\rangle = |\text{device}\rangle|\mu\rangle$. The Hamiltonian can be written as a block matrix:

$$H = \begin{pmatrix} \langle\mu|H|\mu\rangle & \langle\mu|H|l_1\rangle & \dots \\ \langle l_1|H|\mu\rangle & \langle l_1|H|l_1\rangle & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}. \tag{6.22}$$

Each block, such as $\langle l_1|H|\mu\rangle$, is itself a Hamiltonian containing only device degrees of freedom. The element of H in the first row and first column is $\langle\mu|H|\mu\rangle$, an effective device Hamiltonian in which the lead degrees of freedom have been eliminated by taking an expectation value with the thermal ground state of the lead. This effective Hamiltonian does not necessarily accurately represent the device-plus-leads system because the device couples excited states of the lead to the ground state (via the off-diagonal blocks of H). In principle, the matrix in Equation 6.22 can be block-diagonalized. The element in the first row and first column of the block-diagonal $\bar{H}$ is the accurate effective Hamiltonian for the device degrees of freedom, representing an expectation value taken with the true thermal ground state of the lead.

The block-diagonalization problem presented above can be simplified by introducing projection operators (defined below for an N-lead system):

$$P = \prod_i \left(|\mu_1\rangle|\mu_2\rangle \dots |\mu_N\rangle\right)\left(\langle\mu_1|\langle\mu_2| \dots \langle\mu_N|\right), \tag{6.23}$$

and $Q = 1 - P$. Using these projection operators, a new block form for the Hamiltonian can be written:

$$H = \begin{pmatrix} PHP & PHQ \\ QHP & QHQ \end{pmatrix}. \tag{6.24}$$

PHP corresponds to the first row-first column matrix element of H in Equation 6.22. The form of H in Equation 6.24 makes it clear that a full block-diagonalized form of the Hamiltonian is not needed if the leads are taken to be in thermal equilibrium. It is

only necessary to transform to a new Hamiltonian $\bar{H}$ in which the off-diagonal blocks $P\bar{H}Q$ and $Q\bar{H}P$ are zero. The QHQ block that describes matrix elements of the Hamiltonian with excited states of the leads does not have to be block-diagonalized.

The projected components of the Hamiltonian are easily evaluated in the $\{b^\dagger_{k_i}, b_{k_i}\}$ basis of Equation 6.21:

$$\begin{aligned} H_0 = PHP = H_{\text{int}}(\{c^\dagger_n\};\{c_n\}) + \sum_{k_i} f_{k_i} E_{k_i} + \\ \sum_{k_i n} \sqrt{f_{k_i}(1-f_{k_i})} V_{k_i n} c_n + \text{h.c.}, \end{aligned} \tag{6.25}$$

$$QHQ = H_0 + \sum_{k_i} \left\{ (1-2f_{k_i})E_{k_i} - 2\sqrt{f_{k_i}(1-f_{k_i})} \sum_n (V_{k_i n} c_n + V^*_{k_i n} c^\dagger_n) \right\} b^\dagger_{k_i} b_{k_i}, \tag{6.26}$$

$$QHP = \sum_{k_i} \left\{ f_{k_i} \sum_n V^*_{k_i n} c^\dagger_n - (1-f_{k_i}) \sum_n V_{k_i n} c_n - \sqrt{f_{k_i}(1-f_{k_i})} E_{k_i} \right\} b^\dagger_{k_i}, \tag{6.27}$$

and $PHQ = (QHP)^\dagger$. The PHP component of the Hamiltonian, H_0, is simply the device-plus-leads Hamiltonian (Equation 6.12) with the operators $\{d^\dagger_{k_i}, d_{k_i}\}$ replaced by their expectation values evaluated with the thermal equilibrium states $\{|\psi_{k_i}\rangle\}$ (Equation 6.16). $\sum_{k_i} f_{k_i} E_{k_i}$ is the energy of the isolated lead i in thermal equilibrium.

6.2.2 Effective Hamiltonian

In order to find the effective Hamiltonian that corresponds to equilibrium leads, it is necessary to find the matrix U that block-diagonalizes the Hamiltonian H. In the matrix notation introduced above:

$$\bar{H} = \begin{pmatrix} P\bar{H}P & 0 \\ 0 & Q\bar{H}Q \end{pmatrix} = U \begin{pmatrix} PHP & PHQ \\ QHP & QHQ \end{pmatrix} U^\dagger, \tag{6.28}$$

where $UU^\dagger = U^\dagger U = 1$. If the off-diagonal matrix elements of H, PHQ and QHP, are small in magnitude or negligible for another reason, then perturbative approaches to this diagonalization may be effective. For example, since $PHQ \propto b^\dagger$, a first-order

solution to Equation 6.28 (with terms proportional to $(PHQ)(QHP)$) will correspond to including only lead states with a single excitation above the thermal ground state. The first-order solution may be valid if multiple-excitation processes in the leads are unimportant.

As a first approach to the problem presented in Equation 6.28, consider the zeroth-order solution: $PHQ = 0$ and $U = 1$. In this case, the effective Hamiltonian is simply H_0, given in Equation 6.27. The $\sum_{k_i} f_{k_i} E_{k_i}$ term that gives the thermal ground-state energy of the leads can be ignored, since it provides only a constant energy shift. Without this constant term, the effective Hamiltonian is:

$$H_{\text{eff}} = H_{\text{int}}(\{c_n^\dagger\}; \{c_n\}) + \sum_{k_i n} \sqrt{f_{k_i}(1 - f_{k_i})} V_{k_i n} c_n + \text{h.c.}. \tag{6.29}$$

Given this effective Hamiltonian, the current operator, presented in its general form in Equation 6.14, can be evaluated. The effective current operator will be derived below in two manners. The first employs the projection operator P discussed above to isolate the device degrees of freedom. The second starts from the charge continuity equation, considering only changes in the charge on the device. For higher-order effective Hamiltonians, the projection method may prove cumbersome, although it provides a more direct connection to the full device-plus-leads system.

Current Operator – Projection Method

As with the Hamiltonian, the current operator for the device-plus-leads system can be written as a block matrix, employing the projection operators P and Q:

$$I = \begin{pmatrix} PIP & PIQ \\ QIP & QIQ \end{pmatrix} \tag{6.30}$$

Employing the same zeroth-order approximation used above, in which PIQ and QIP are taken to be zero, the current operator for the device degrees of freedom is given by $I_{\text{eff}} = PIP$. In the $\{b_{k_i}^\dagger, b_{k_i}\}$ basis, the current operator for the device-plus-leads

system (from Equation 6.14) is:

$$\hat{I}_{k_i} = \frac{ie}{\hbar}\sum_n V_{k_i n}\left\{\sqrt{f_{k_i}(1-f_{k_i})}(1-2b^\dagger_{k_i}b_{k_i}) - (1-f_{k_i})b^\dagger_{k_i} + f_{k_i}b_{k_i}\right\}c_n + \text{h.c.} \quad (6.31)$$

Given $\hat{I}_{k_i}$ in this basis, it is straightforward to evaluate the effective current operator $I_{\text{eff}} = PIP$ in the zeroth-order approximation:

$$I_{\text{eff}} = \frac{ie}{\hbar}\sum_n V_{k_i n}(\sqrt{f_{k_i}(1-f_{k_i})})c_n + \text{h.c.} \quad (6.32)$$

Current Operator – Charge Continuity

Consider the charge continuity equation, integrated over a volume Ω surrounding the quantum device, with bounding surface $\partial\Omega$:

$$\frac{\partial Q}{\partial t} + \int_{\partial\Omega} \vec{I}\cdot\hat{n} = 0, \quad (6.33)$$

where Q is the total charge contained inside Ω, and $\hat{n}$ is normal to $\partial\Omega$. If Ω is chosen to contain the device but not the leads, then Q is the charge in the device and $\int_{\partial\Omega}\vec{I}\cdot\hat{n}$ is the sum of currents going out of the device into the attached leads. In operator form, $Q = (-\text{e})\langle\frac{\partial\hat{N}}{\partial t}\rangle$, where $\hat{N} = \sum_n c^\dagger_n c_n$ is the number operator and $\langle\ldots\rangle$ denotes an expectation value with the ground state of the device. Eliminating the expectation values gives the continuity equation in operator form :

$$(-\text{e})\frac{\partial\hat{N}}{\partial t} = \sum_i \hat{I}_i, \quad (6.34)$$

where $\hat{I}_i$ gives the current going into the device from lead i (consistent with the sign convention of Equation 6.13).

The time evolution of $\hat{N}$ is given by the Heisenberg equation of motion with respect to the effective Hamiltonian for the quantum device (Equation 6.29). For ease of notation, the effective Hamiltonian can be written without reference to the

lead states $\{|k_i\rangle\}$:

$$H_{eff} = H_{\text{int}}(\{c_n\}, \{c_n^\dagger\}) + \sum_{i,n} V_{in} c_n + \text{h.c.}, \tag{6.35}$$

where

$$V_{in} = \sum_{k_i} \sqrt{f_{k_i}(1 - f_{k_i})} V_{k_i n}. \tag{6.36}$$

describes the coupling of all the states in lead i to state n in the device. Evaluating the Heisenberg equation of motion:

$$i\hbar \frac{\partial \hat{N}}{\partial t} = \left[\hat{H}_{\text{int}}, \hat{N}\right] + \sum_{i,n} \left(V_{in} c_n - V_{in}^* c_n^\dagger\right). \tag{6.37}$$

Since H_{int} is the Hamiltonian for the isolated quantum device, particle number must be conserved: $\left[\hat{H_{int}}, \hat{N}\right] = 0$. Number conservation simplifies Equation 6.37 , allowing the current operator to be identified by comparing the Heisenberg equation of motion to the continuity equation:

$$(-e)\frac{\partial \hat{N}}{\partial t} = \frac{ie}{\hbar} \sum_{i,n} \left(V_{in} c_n - V_{in}^* c_n^\dagger\right) = \sum_i \hat{I}_i. \tag{6.38}$$

The operator which gives the current flowing out of the device and into lead i is thus:

$$\hat{I}_i = \frac{ie}{\hbar} \sum_n \left(V_{in} c_n - V_{in}^* c_n^\dagger\right). \tag{6.39}$$

Using the definition of V_{in} (Equation 6.36), the components $\hat{I}_{k_i}$ can be readily identified. The resulting current operators are equivalent those derived from the projection operator method above. Since the current operator is derived explicitly from the charge continuity equation, current is conserved.

6.3 Applications

The two sections below illustrate potential applications of the effective Hamiltonian method, using the simple zeroth-order solution of Equation 6.29. Coherent charge and spin dynamics in a quantum dot connected to a single lead are described well, but the zeroth-order solution is shown to be inadequate to describe steady-state current in a resonant-tunneling device connected to two leads.

6.3.1 Quantum Dot

Consider a quantum dot with a single bound state connected to a single lead with chemical potential μ. The basis for the quantum dot is $\{|0\rangle, |\uparrow\rangle, |\downarrow\rangle, |2\rangle\}$, where 0 and 2 indicate the number of electrons on the site, and $\uparrow$ and $\downarrow$ indicate the spin orientation of the electron when there is a single electron on the site. The effective Hamiltonian for the system is

$$H_{\text{eff}} = \begin{pmatrix} -\mu & V_1 & V_2 & 0 \\ V_1^* & \epsilon-\mu-U/2 & 0 & V_1 \\ V_2^* & 0 & \epsilon-\mu-U/2 & V_2 \\ 0 & V_1^* & V_2^* & 2(\epsilon-\mu) \end{pmatrix}, \tag{6.40}$$

where

$$V_j = \sum_k e^{-i\phi_k^{(j)}} \sqrt{f_k(1-f_k)} V_k, \quad j = 1, 2. \tag{6.41}$$

The phase ϕ_k may be different for states that couple to $|\uparrow\rangle$ than those that couple to $|\downarrow\rangle$. In order to distinguish them, these couplings are labeled with the index $j = 1(\uparrow), 2(\downarrow)$. The on-site interaction term is calculated using:

$$U n_\uparrow n_\downarrow - \frac{U}{2}(n_\uparrow + n_\downarrow). \tag{6.42}$$

The $U/2$ term is taken from the symmetric Anderson Hamiltonian to eliminate processes that empty or doubly-fill the localized state. There are four eigenstates. If the

energy levels are such that $\epsilon \leq \mu$ and $\epsilon + U/2 > \mu$, then at equilibrium the ground state should have one electron on the quantum dot. If $U \gg |\epsilon - \mu|$ and $U \gg V$ (the strong-correlation limit), then to second order in V_j the four eigenenergies are

$$E_{\pm} = \epsilon - \mu - \frac{U}{2} + \frac{2}{U}\left(|V_1|^2 + |V_2|^2 \pm |V_1^2 + V_2^2|\right), \tag{6.43}$$

$$E_0 = -\frac{2}{U}\left(|V_1|^2 + |V_2|^2 \pm |V_1^2 + V_2^2|\right), \tag{6.44}$$

and

$$E_2 = 2(\epsilon - \mu) - \frac{2}{U}\left(|V_1|^2 + |V_2|^2 \mp |V_1^2 + V_2^2|\right), \tag{6.45}$$

where the first set of the $\pm$ and $\mp$ signs for E_0 and E_2 is taken if $\epsilon - \mu > 2|V_1^2 + V_2^2|/U$, and the second set is taken if $\epsilon - \mu < 2|V_1^2 + V_2^2|/U$.

As expected, the two singly-occupied states are lower in energy. Their separation, $4|V_1^2 + V_2^2|/U$, is the exchange splitting due to coupling to the electrodes, and is equal to the exchange term in the Kondo Hamiltonian. An examination of the ground state wave function confirms that the Kondo effect is indeed contained in this effective Hamiltonian. To illustrate the Kondo effect, the bound state in the dot is assumed to couple to only a single conduction channel in the lead. As a result of this simplified coupling Equation 6.41 becomes:

$$V_j = V e^{i\phi^{(j)}}. \tag{6.46}$$

The Fermi distribution function factor $\sqrt{f(1-f)}$ has been absorbed into V. Consequently, the two coupling constants, V_1 and V_2, have the same magnitude and differ by a phase factor $\exp(i\phi)$, with $\phi = \phi^{(1)} - \phi^{(2)}$. It is then straightforward to find the ground state wavefunction. To first order in the coupling, it is

$$\psi_{\pm} = \frac{1}{\sqrt{2}}(|\uparrow\rangle \pm |\downarrow\rangle) - \frac{\sqrt{2}(V_1 \pm V_2)}{U}|0\rangle - \frac{\sqrt{2}(V_1^* \pm V_2^*)}{U}|2\rangle, \tag{6.47}$$

where the $+$ is taken when $\cos\phi < 0$, and the $-$ is taken otherwise. The ground state thus has zero net spin. This implies that the complete wave function in both the

quantum dot region and the electrode region should be (neglecting the order V_j/U terms),

$$\frac{1}{\sqrt{2}}\left(|\uparrow\rangle_{\rm QD}|\text{lead}\downarrow\rangle - \frac{\cos\phi}{|\cos\phi|}|\downarrow\rangle_{\rm QD}|\text{lead}\uparrow\rangle\right), \tag{6.48}$$

where "QD" indicates the wave function on the quantum dot and $|\text{lead}\rangle$ is the single-electron wavefunction in the lead that couples to the bound state in the dot.

The dynamics of a wavefunction on the quantum dot can be illustrated by the following examples. First consider a wavefunction for a single electron on the dot whose initial state has a spin polarization of $|\uparrow\rangle$. The time evolution of the wavefunction ψ ($\psi(t=0) = |\uparrow\rangle$) can be expressed as a linear combination of the two singly occupied states:

$$\begin{aligned}\psi(t) &= \frac{1}{\sqrt{2}}\left(\psi_{\pm}e^{iE_{+}t/\hbar} + \psi_{\mp}e^{iE_{-}t/\hbar}\right) \\ &\approx \left(|\uparrow\rangle\cos\omega t \pm i|\downarrow\rangle\sin\omega t\right)e^{i(E_{+}+E_{-})t/2\hbar},\end{aligned} \tag{6.49}$$

where $\omega = (E_{+} - E_{-})/2\hbar = 2|V_1^2 + V_2^2|/U\hbar$, and the choices of the $\pm$ signs depend on the sign of $\cos\phi$ as discussed earlier. Equation 6.49 predicts a time-dependent oscillation of the spin on the dot, $S(t) = \langle\psi(t)|n_{\uparrow} - n_{\downarrow}|\psi(t)\rangle = \cos 2\omega t$. This oscillation was obtained numerically in Reference [147] by integrating the time-dependent Schrödinger equation.

The time-dependent dynamics of the charge can also be found if the initial state mixes the $E_{\pm}$ with the E_0 and E_2 states. The oscillation frequency of the charge is much higher, determined by the separation between $E_{\pm}$ and E_0, E_2, which is approximately $U/2$. In order to find the magnitude of the oscillation, it is necessary to find the wavefunctions for the E_0 and E_2 states. To first order in the coupling, these states are:

$$\psi_0 = |0\rangle + \frac{2V_1^*}{U}|\uparrow\rangle + \frac{2V_2^*}{U}|\downarrow\rangle \tag{6.50}$$

$$\psi_2 = |2\rangle + \frac{2V_1}{U}|\uparrow\rangle + \frac{2V_2}{U}|\downarrow\rangle. \tag{6.51}$$

Consider now a wavefunction that is initially a mixed-occupancy state: $\psi(0) = a\psi_- + b\psi_2$. The time dependence of this wavefunction is

$$\psi(t) = a\psi_- e^{iE_- t/\hbar} + b\psi_2 e^{iE_2 t/\hbar}. \tag{6.52}$$

The current flowing into the quantum dot can be computed from Equation 6.39:

$$I(t) = \frac{ie}{\hbar}\left(V_1\langle c_\uparrow\rangle + V_2\langle c_\downarrow\rangle - V_1^*\langle c_\uparrow^\dagger\rangle - V_2^*\langle c_\downarrow^\dagger\rangle\right) = \frac{2\sqrt{2}e}{\hbar}|ab(V_1 - V_2)|\sin(\Omega t + \phi'), \tag{6.53}$$

where $\Omega = U/2\hbar$, ϕ' is the phase of $a^*b(V_1 - V_2)$, and higher order terms in V_j/U have been omitted. The time variation of the charge is,

$$Q(t) = e\left(\langle c_\uparrow^\dagger c_\uparrow\rangle + \langle c_\downarrow^\dagger c_\downarrow\rangle\right) = e\left[1 + |b|^2 - \frac{4\sqrt{2}|ab(V_1 - V_2)|}{U}\cos(\Omega t + \phi')\right], \tag{6.54}$$

Compared to the spin oscillation, the charge oscillation is more rapid in frequency, and with a much smaller amplitude. These results are in agreement with the numerical results of Reference [147].

6.3.2 Resonant Tunneling

A simple example of steady-state current is resonant tunneling through a single bound state, as in a resonant tunneling diode or single-electron transistor. The single bound state is connected to two leads, with chemical potentials μ_1 and μ_2. Within the zeroth-order effective Hamiltonian model presented in Section 6.2.2, it is not possible to calculate the effective chemical potential of the bound state. The effective chemical potential is defined here to be μ_{eff}, with $\mu_1 \leq \mu_{\text{eff}} \leq \mu_2$ assumed. In order to highlight resonance effects without addition complications, electrons will be assumed to be spinless. Thus, there are only two states of the resonant level: $|0\rangle$ (unoccupied) and $|1\rangle$ (occupied). In this basis, taking the bound state energy to be ϵ, the Hamiltonian

is:

$$H_{eff} = \begin{pmatrix} -\mu_{\text{eff}} & V_1 + V_2 \\ V_1^* + V_2^* & \epsilon - \mu_{\text{eff}} \end{pmatrix}. \tag{6.55}$$

The couplings $V_{1,2}$ are defined as in Section 6.3.1:

$$V_j = \sum_{k_j} e^{-i\phi_{k_j}} \sqrt{f_{k_j}(1 - f_{k_j})} V_{k_j}, \;\; j = 1, 2. \tag{6.56}$$

For ease of notation, define $V = V_1 + V_2$.

The eigenvalues and eigenvectors of H_{eff} are:

$$E_\pm = (1/2)\left[(\epsilon - \mu_{\text{eff}}) \pm \sqrt{(\epsilon - \mu_{\text{eff}})^2 + 4|V|^2}\right] \tag{6.57}$$

$$|\pm\rangle = \frac{1}{E_\pm^2 + |V|^2} \begin{pmatrix} V \\ E_\pm \end{pmatrix}. \tag{6.58}$$

The $-$ state is the ground state for all values of ϵ and μ_{eff}. The current operators from leads 1 and 2 into the dot are given by:

$$\hat{I}_{1,2} = \frac{ie}{\hbar} \begin{pmatrix} 0 & V_{1,2} \\ -V_{1,2}^* & 0 \end{pmatrix}, \tag{6.59}$$

with matrix elements:

$$\langle \pm | \hat{I}_1 | \pm \rangle = \left(\frac{2e}{\hbar}\right) \frac{E_\pm}{E_\pm^2 + |V|^2} \text{Im}\{V_1^* V_2\} \tag{6.60}$$

$$\langle \pm | \hat{I}_2 | \pm \rangle = \left(\frac{2e}{\hbar}\right) \frac{E_\pm}{E_\pm^2 + |V|^2} \text{Im}\{V_2^* V_1\} \tag{6.61}$$

Since $\text{Im}\{V_1^* V_2\} = -\text{Im}\{V_2^* V_1\}$, the divergence of the current $\vec{\nabla} \cdot \vec{I} = \langle \pm | I_1 | \pm \rangle + \langle \pm | I_2 | \pm \rangle$ vanishes for both the $|+\rangle$ and $|-\rangle$ states. From the continuity equation, a current with vanishing divergence implies that $\partial Q / \partial t = 0$, where Q is the total charge in the bound state. This condition corresponds to a steady state, as would be expected if the wavefunction is an eigenstate of the effective Hamiltonian.

To investigate the behavior of the steady-state current, the eigenenergies and

currents can be evaluated in three relevant limits: $\epsilon < \mu_{\text{eff}}$, $\epsilon > \mu_{\text{eff}}$, and $\epsilon = \mu_{\text{eff}}$. The couplings V_1 and V_2 are assumed to be equal in magnitude, and differ only by a phase: $V_{1,2} = V_0 e^{i\phi_{1,2}}$, where V_0 is real. As a result, $\text{Im}\{V_1^* V_2\} = V_0^2 \sin(\phi_2 - \phi_1)$, and $|V|^2 = |V_1 + V_2|^2 = 4V_0^2 \cos^2((\phi_1 + \phi_2)/2)$.

When $\epsilon < \mu_{\text{eff}}$, $|+\rangle \approx |0\rangle$, and $|-\rangle \approx |1\rangle$. If the bound state energy lies well below the chemical potentials of both the left and right electrodes, the ground state of the device has an electron occupying the bound state. To second order in V_0:

$$E_+ = \frac{|V|^2}{|\epsilon - \mu_{\text{eff}}|}, \qquad I_1^+ = \left(\frac{2e}{\hbar}\right)\left(\frac{1}{|\epsilon - \mu_{\text{eff}}|}\right) V_0^2 \sin(\phi_2 - \phi_1) \tag{6.62}$$

$$E_- = -|\epsilon - \mu_{\text{eff}}| - \frac{|V|^2}{|\epsilon - \mu_{\text{eff}}|}, \quad I_1^- = -\left(\frac{2e}{\hbar}\right) V_0^2 \sin(\phi_2 - \phi_1). \tag{6.63}$$

For $\epsilon > \mu_{\text{eff}}$, $|+\rangle \approx |1\rangle$, and $|-\rangle \approx |0\rangle$. When the bound state energy is well above the chemical potentials, the bound state is empty in the ground state. To second order in V_0:

$$E_+ = |\epsilon - \mu_{\text{eff}}| + \frac{|V|^2}{|\epsilon - \mu_{\text{eff}}|}, \quad I_1^+ = \left(\frac{2e}{\hbar}\right)\left(\frac{1}{|\epsilon - \mu_{\text{eff}}|}\right) V_0^2 \sin(\phi_2 - \phi_1) \tag{6.64}$$

$$E_- = -\frac{|V|^2}{|\epsilon - \mu_{\text{eff}}|}, \qquad I_1^- = \left(\frac{2e}{\hbar}\right) V_0^2 \sin(\phi_2 - \phi_1). \tag{6.65}$$

When $\epsilon = \mu_{\text{eff}}$, the $|+\rangle$ and $|-\rangle$ states are mixtures of $|0\rangle$ and $|1\rangle$. To second order in V_0:

$$E_\pm = \pm|V|^2, \quad I_1^\pm = \pm\left(\frac{2e}{\hbar}\right) V_0^2 \sin(\phi_2 - \phi_1). \tag{6.66}$$

In all three cases considered above, the currents do not depend on the difference in chemical potential between the two leads. As such, even in the limit of small bias, no conductance can be defined. Furthermore, all currents are proportional to $\sin(\phi_2 - \phi_1)$. This phase difference between the coupling constants reflects quantum fluctuations, and should vary with time. A proper steady-state current should take into account this variation through an average over the phase difference. Unfortunately, the equal-weight phase average is zero: $\int_0^{2\pi} d\phi \sin\phi = 0$. In the absence of long-time phase correlations between the two electrodes, $I_1^\pm = -I_2^\pm = 0$ in the steady state.

The zeroth-order effective Hamiltonian, despite its success in predicting temporal correlations (Section 6.3.1), is unsuitable to steady-state current calculations. This is a direct result of the zeroth-order approximation. The off-diagonal blocks of the full Hamiltonian, PHQ and QHP were assumed to be negligible, resulting in a ground-state lead wavefunction that was simply a product of isolated-lead thermal ground states:

$$|\text{lead}\rangle_{0^{\text{th}}-\text{order}} = \prod_{i=1,N} |\mu_i\rangle, \tag{6.67}$$

for an N-lead system. A full block-diagonalization of the Hamiltonian results in a ground state that mixes the thermal ground state of the leads with excited states. This mixing reflects the influence of the other leads on each individual lead. The zeroth-order solution does not include this coupling, and as such cannot account for any dependence on differences in the lead chemical potentials. Higher-order effective Hamiltonians should be able to reproduce steady-state currents as well as additional details of the temporal current correlations.

Chapter 7

Conclusions

The preceding chapters have provided a glimpse of the rich variety of phenomena that occur in structures with nanometer-scale dimensions. In this regime, atomic-scale structure cannot be ignored and significant changes can happen in the span of femtoseconds. Computational methods are essential to the study of nanoscale systems, but the well-tested techniques for isolated molecules and bulk solids are no longer sufficient. Two of the emerging challenges in the simulation of nanostructures are the computational design of low-dimensional structures and the calculation of non-equilibrium transport phenomena. Calculations and novel methodology presented in this thesis show how both of these challenges can be addressed. The small size and complex behavior of nanoscale systems provide fertile ground for computational techniques to directly address experimental results.

7.1 Computational Design

Chapter 2 demonstrates that the unique properties of boron nanotubes can be understood through consideration of planar boron structures and the boron atom. The prediction of novel boron nanotube structures came directly from the study of boron planes. The two-dimensional periodicity of the planar structures greatly reduced the computational cost of these studies. A robust understanding of how the boron p orbitals lead to the planar buckling instability was achieved with relatively little

computational effort. This picture of the behavior of the buckled plane led directly to the prediction of the differences between the $(n, 0)$ and (n, n) boron nanotubes. Work on boron planes significantly narrowed the possible range of boron nanotube structures. Studying higher-dimensional analogues of low-dimensional nanostructures can minimize computational costs and speed the identification and design of desirable properties.

The internal degrees of freedom identified in (n, n) boron nanotubes are representative of a more general phenomenon. Since nanostructures are not homogeneous (individual atoms cannot be "smeared out"), the macroscopic elastic constants that are measured experimentally do not always have a clear connection to microscopic structure. Although a macroscopic tube under lateral strain must expand or contract its diameter to relieve stress, a nanotube has other mechanisms for relaxation. The atoms that constitute the walls of the nanotube can rearrange themselves without changing the overall diameter of the tube, as with optical phonons that involve no net translation of the crystal. This permits stress to be dissipated in a manner that is not reflected in the Poisson ratio. A straightforward connection between the Young's modulus and the Poisson ration is expected for a macroscopic tube, but nanotubes designed with internal degrees of freedom can break that connection, and achieve low Young's moduli in conjunction with low Poisson ratios. Low Poisson ratios imply high longitudinal stiffness that can be desirable in composite materials.

Computational design of nanostructures is challenging due to the wide variety of possible structures a small collection of atoms can adopt. In order to efficiently study that phase space, the planar and solid analogues of low-dimensional nanostructures should be studied and understood first. Accurate first-principles calculations can yield detailed information about novel structures and properties, and the atomic-scale mechanisms that bring them about. Such mechanisms can be useful guides in designing further nanostructures that display desirable properties.

7.2 Transport and Non-Equilibrium Phenomena

Chapters 4 and 6 presented novel methods to calculate transport properties of ultrathin transistors and quantum devices. Currents will flow in systems driven out of equilibrium, breaking symmetry and presenting computational challenges beyond those of equilibrium, ground-state calculations. The method for mobility calculations presented in Chapter 4 focuses on the linear transport regime and steady-state transport. Experiments on MOSFETs rarely probe the strongly non-linear regime of the source-drain current, so that mobility calculations can make a direct connection to experiments without undue complication. Fully non-linear transport with an arbitrary time dependence can be treated through the effective Hamiltonian method presented in Chapter 6, but the system must be represented by a simplified Hamiltonian that can obscure the atomic-scale structure.

First-principles mobility calculations represent a significant advance over existing state-of-the-art MOSFET mobility models. In ultrathin SOI devices, the atomic-scale structure plays an important role since even single-atom defects and impurities represent large deviations from the ideal channel. Using ground-state DFT calculations to obtain the scattering potential and wavefunctions permits a first-principles scattering matrix to be calculated. This matrix is used in conjunction with the dispersion relation to solve the Boltzmann equation to linear order or the Kubo formula to calculate the mobility. The resulting mobility is calculated entirely from first-principles, and enables a direct connection to be made between the measured electrical response and atomic-scale defects and impurities.

The effective Hamiltonian method allows for rapid temporal correlations in quantum devices to be calculated efficiently. By tracing out the lead degrees of freedom, the effective Hamiltonian focuses on the behavior of electrons occupying the device's states. Currents and bound state occupancies can be expressed in terms of eigenstates with a well-defined time dependence, allowing temporal correlations to be understood as a superposition of eigenmodes. The matrix representation of the full system (device and leads) Hamiltonian provides the framework for perturbative approaches to

constructing the effective Hamiltonian. Although the zeroth-order effective Hamiltonian only accounts for a limited range of phenomena, higher-order approaches can be readily developed for many quantum devices.

Chapter 5 demonstrated that mobility calculations can identify the atomic-scale origin of the observed electrical response in UTSOI MOSFETs. By distinguishing atomic-scale from long-wavelength interface roughness, first-principles calculations allowed the origin of anomalously low mobility in UTSOI channels to be discerned. As a result of this new understanding, a path to improving mobility at low effective fields by up to 100% was identified. Chapter 3 showed how first-principles calculations can help identify the atomic and electronic structure of trapping centers in amorphous insulators, providing greater insight into the transport behavior of these materials. Although the atomic structure of the STH_2 trapping center remains unresolved, conclusive evidence for the absence of a V_k center analogue in SiO_2 was obtained.

7.3 Future Work

The first-principles mobility calculation and effective Hamiltonian methods have numerous applications that have only begun to be explored. In this section, future methodological improvements and technologically-important applications will be discussed. The computational design of nanostructures and the calculation of their relevant transport properties is a relatively young field. With increasing computational power and the further development of novel methods, future advances that provide insight into both basic physics and technology seem certain to arrive.

7.3.1 First-Principles Mobility Calculations

In order to go beyond linear transport calculations and explore the behavior of the source-drain current at high bias, the Hamiltonian presented in Section 4.2.4 can be solved using non-equilibrium Green's Functions (discussed in Section 6.1.1). A simple, non-interacting model for the source and drain electrodes will have to be

constructed, but in sufficiently long channels this should not greatly influence the transport properties. Effects such as contact resistance that are strongly dependent on device processing can be parameterized to fit observed UTSOI MOSFET data. Constructing a UTSOI channel Hamiltonian from first-principles and then employing NEGF methods maintains the direct link between atomic-scale structure and transport properties.

The ability to connect atomic-scale defects directly to measured electrical responses through the mobility, transconductance and other properties provides a unique opportunity to improve device characterization techniques. The calculated mobilities for each defect and impurity have unique dependences on gate bias and temperature, and a particular combination of defects should have characteristic features that can be measured experimentally. The electrical "signature" of a particular combination of defects can be compared to experimentally-measured mobility curves, as a function of temperature and gate bias. If the mobilities due to a range of possible defects and impurities in a devices are calculated, then an straightforward electrical response measurement could determine the concentrations of defects and impurities, providing a simple and non-invasive way to characterize the quality of a device at the atomic scale.

Beyond interface roughness, there are many other novel scattering mechanisms in UTSOI channels. High-κ dielectrics will replace SiO_2 as the gate insulator of choice in both bulk and SOI MOSFETs. These novel materials have complicated interfaces with silicon that are not yet well-understood [18]. New structural defects in addition to suboxide bonds and oxygen protrusions may play an important role in limiting the mobility of UTSOI devices with high-κ dielectrics. Careful studies of how high-κ-Si interface defects affect mobility can influence device development by revealing which silicon-insulator structures perform best. In addition, first-principles phonon scattering calculations (described in Section 4.2.5) can be used to understand the plasmon-phonon interaction in UTSOI high-κ structures. Although scattering with polar optical modes will be more severe with high-κ dielectrics than with SiO_2, certain device designs may reduce the strength of scattering. For example, the use of

an SiO_2 buffer layer to remove the high-κ interface from the Si channel may reduce the plasmon-phonon interaction, but could introduce structural defects that limit the mobility. Making these choices in the design of future UTSOI MOSFETs requires accurate and flexible modeling tools that incorporate accurate atomic structures and quantum-mechanical effects. The first-principles mobility calculation method can find use in these, and other, applications.

7.3.2 Effective Hamiltonian Method

An important extension to the zeroth-order effective Hamiltonian method presented in Section 6.2.2 is to block-diagonalize the full Hamiltonian to higher orders. Higher-order effective Hamiltonians will include couplings, via the quantum device, between the leads and could account for steady-state currents. Although constructing an operator via a perturbation expansion is more complicated than evaluating a function, further study should uncover efficient methods for perturbative block-diagonalization. Physical intuition about the leads and/or working in the low-temperature limit may lead to simple solutions for higher-order effective Hamiltonians.

Calculating temporal fluctuations in the current and occupancies of a quantum device is an important task, since recent experiments have shown that measurements can be made at femtosecond time scales. Although fluctuations in a single device can reveal interesting physics, correlated fluctuations may play a more fundamental role in the behavior of quantum circuits. The eigenvalues of the effective Hamiltonian reveal that current and spin fluctuations occur at characteristic frequencies determined by the nature of the device and its couplings to the leads. Densely-packed devices in a nanoscale electronic circuit may couple to each other and enhance fluctuations at certain frequencies. These resonant fluctuations could drive the circuit out of equilibrium and significantly change its electrical behavior. In designing macroscopic circuits, devices that are electrically isolated are truly independent, and the response of a circuit is the sum of its parts. In quantum circuits, however, devices that are close together may couple to each other and behave as a new, composite entity. A proper theory of fluctuations, such as that provided by the effective Hamiltonian method,

can describe these couplings and inform the design of quantum electronic circuits.

Bibliography

[1] International Technology Roadmap for Semiconductors (2003). Available at http://public.itrs.net.

[2] J. L. Alay and M. Hirose. The valence band alignment at ultrathin SiO_2/Si interfaces. *J. Appl. Phys.*, 81:1606, 1997.

[3] D. Ali and H. Ahmed. Coulomb blockade in a silicon tunnel junction device. *Appl. Phys. Lett.*, 64:2119, 1994.

[4] P. W. Anderson. Absence of diffusion in certain random lattices. *Phys. Rev.*, 109:1492, 1958.

[5] T. Ando. Screening effect and quantum transport in a silicon inversion layer in strong magnetic field. *J. Phys. Soc. Japan*, 43:1616, 1977.

[6] T. Ando, A. B. Fowler, and F. Stern. Electronic properties of two-dimensional systems. *Rev. Mod. Phys.*, 54:437, 1982.

[7] R. Andrews, D. Jacques, A. M. Rao, T. Rantell, F. Derbyshire, Y. Chen, J. Chen, and R. C. Haddon. Nanotube composite carbon fibers. *Appl. Phys. Lett.*, 75:1329, 1999.

[8] G. B. Arfken and H. J. Weber. *Mathematical Methods for Physicists.* Academic, San Diego, 4th. edition, 1995.

[9] N. D. Arora and L. M. Richardson. *MOSFET Models for VLSI Circuit Simulation: Theory and Practice.* Springer-Verlag, New York, 1993.

[10] N. W. Ashcroft and N. D. Mermin. *Solid State Physics.* Saunders College Publishing, Fort Worth, 1976.

[11] R. C. Ashoori. Electrons in artificial atoms. *Nature*, 379:413, 1996.

[12] P. Avouris, J. Appenzeller, R. Martel, and S. J. Wind. Carbon nanotube electronics. *Proc. IEEE*, 91:1772, 2003.

[13] T. Bakos, S. N. Rashkeev, and S. T. Pantelides. Reactions and diffusion of water and oxygen molecules in amorphous SiO_2. *Phys. Rev. Lett.*, 88:055508, 2002.

[14] F. Balestra, S. Cristoloveanu, M. Benachir, and J. Brini. Double-gate silicon-on-insulator transistor with volume inversion: A new device with greatly enhanced performance. *IEEE Electron Dev. Lett.*, 8:410, 1987.

[15] S. Baroni, P. Giannozzi, and A. Testa. Green's-function approach to linear response in solids. *Phys. Rev. Lett.*, 58:1861, 1987.

[16] E. Ben-Jacob and Y. Gefen. New quantum oscillations in current driven small junctions. *Phys. Lett. A*, 108:289, 1985.

[17] J. M. Benedetto and H. E. Boesch, Jr. The relationship between Co^{60} and 10 keV X-ray damage in MOS devices. *IEEE Trans. Nucl. Sci*, 33:1318, 1986.

[18] G. Bersuker, P. Zeitzoff, G. Brown, and H. R. Huff. Dielectrics for future transistors. *Materials Today*, page 26, Jan. 2004.

[19] F. Bloch. Uber die quantenmechanik der electronen in kristallgittern. *Z. Physik*, 52:555, 1928.

[20] M. Boero, A. Pasquarello, J. Sarnthein, and R. Car. Structure and hyperfine parameters of E_1' centers in *alpha*-quartz and in vitreous SiO_2. *Phys. Rev. Lett.*, 78:887, 1997.

[21] A. Bongiorno, A. Pasquarello, M. S. Hybertsen, and L. C. Feldman. Transition structure at the Si(100)-SiO_2 interface. *Phys. Rev. Lett.*, 90:186101, 2003.

[22] I. Boustani, A. Quandt, E. Hernández, and A. Rubio. New boron based nanostructured materials. *J. Chem. Phys.*, 110:3176, 1999.

[23] I. Boustani, A. Quandt, and A. Rubio. Boron quasicrystals and boron nanotubes: Ab initio study of various B_{96} isomers. *J. Solid State Chem.*, 154:269, 2000.

[24] R. Buczko, S. Pennycook, and S. T. Pantelides. Bonding arrangements at the Si-SiO_2 and SiC-SiO_2 interfaces and a possible origin of their contrasting properties. *Phys. Rev. Lett.*, 84:943, 2000.

[25] L. Cao, J. Liu, C. Gao, Y. Li, X. Li, Y. Q. Wang, Z. Zhang, Q. Cui, G. Zou, L. Sun, and W. Wang. Synthesis of well-aligned boron nanowires and their structural stability under high pressure. *J. Phys.: Condens. Matter*, 14:11017, 2002.

[26] R. Car and M. Parrinello. Unified approach for molecular dynamics. *Phys. Rev. Lett.*, 55:2471, 1985.

[27] G. K. Celler and S. Cristoloveanu. Frontiers of silicon-on-insulator. *J. Appl. Phys.*, 93:4955, 2003.

[28] A. R. Champagne, A. N. Pasupathy, and D. C. Ralph. Mechanically adjustable and electrically gated single-molecule transistors. *Nano Lett.*, 5:305, 2005.

[29] L. L. Chang, L. Esaki, and R. Tsu. Resonant tunneling in semiconductor double barriers. *Appl. Phys. Lett.*, 24:593, 1974.

[30] J. C. Charlier, X. Gonze, and J.-P. Michenaud. First-principles study of the electronic properties of graphite. *Phys. Rev. B.*, 43:4579, 1991. And references therein.

[31] K. Chen, H. C. Wann, P. K. Ko, and C. Hu. The impact of device scaling and power supply change on CMOS gate performance. *IEEE Trans. Electron Dev.*, 17:202, 1996.

[32] P. V. Chernov, E. M. Dianov, V. N. Karpechev, L. S. Kornienko, I. O. Morozova, A. O. Rybaltovskii, V. O. Sokolov, and V. B. Sulimov. Spectroscopic manifestation of self-trapped holes in silica. Theory and experiment. *Phys. Status Solidi B*, 156:663, 1989.

[33] J.-P. Colinge. *Silicon-on-Insulator Technology: Materials to VLSI.* Kluwer, Boston, 2nd. edition, 1997.

[34] H. G. Craighead. Nanoelectromechanical systems. *Science*, 290:1532, 2000.

[35] S. Cristoloveanu and S.S. Li. *Electrical Characterization of Silicon-On-Insulator Materials and Devices.* Kluwer, Dordrecht, 1995.

[36] O. L. Curtis, Jr. and J. R. Srour. The multiple-trapping model and hole transport in SiO_2. *J. Appl. Phys.*, 48:3819, 1977.

[37] J. Donohue. *The Structures of the Elements.* Wiley, New York, 1974. And references therein.

[38] M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund. *Science of Fullerenes and Carbon Nanotubes.* Academic, San Diego, 1996.

[39] T. W. Ebbesen and P. M. Ajayan. Large-scale synthesis of carbon nanotubes. *Nature*, 358:220, 1992.

[40] A. H. Edwards. Theory of the self-trapped hole in a-SiO_2. *Phys. Rev. Lett.*, 71:3190, 1993.

[41] T. Ernst, S. Cristoloveanu, G. Ghibaudo, T. Ouisse, S. Horiguchi, Y. Ono, Y. Takahashi, and K. Murase. Ultimately thin double-gate SOI MOSFETs. *IEEE Trans. Electron Dev.*, 50:830, 2003.

[42] D. Esseni. On the modeling of surface roughness limited mobility in SOI MOSFETs and its correlation to the transitor effective field. *IEEE Trans. Electron Dev.*, 51:394, 2004.

[43] D. Esseni, A. Abramo, L. Selmi, and E. Sangiorgi. Study of low field electron transport in ultra-thin single and double gate SOI MOSFETs. *IEDM Tech. Dig.*, page 719, 2002.

[44] D. Esseni, M. Mastrapasqua, G. K. Celler, C. Fiegna, L. Selmi, and E. Sangiorgi. An experimental study of mobility enhancement in ultrathin SOI transistors operated in double-gate mode. *IEEE Trans. Electron Dev.*, 50:802, 2003.

[45] F. J. Feigl, W. B. Fowler, and K. L. Yip. Oxygen vacancy model for the E' center in SiO_2. *Solid State Commun.*, 14:225, 1974.

[46] M. V. Fischetti and S. Laux. Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects. *Phys Rev. B*, 38:7465, 1988.

[47] M. V. Fischetti and S. Laux. Monte Carlo study of electron transport in silicon inversion layers. *Phys Rev. B*, 48:2244, 1993.

[48] M. V. Fischetti, D. Neumayer, and E. Cartier. Effective electron mobility in Si inversion layers in metal-oxide-semiconductor systems with a high-k insulator: The role of remote phonon scattering. *J. Appl. Phys.*, 90:4587, 2001.

[49] W. R. Frensley. Wigner-function model of a resonant-tunneling semiconductor device. *Phys. Rev. B*, 36:1570, 1987.

[50] E. J. Friebele, D. L. Griscom, and M. Stapelbroek. Fundamental defect centers in glass: The peroxy radical in irradiated, high-purity, fused silica. *Phys. Rev. Lett.*, 42:1346, 1979.

[51] M. Fuchs and M. Scheffler. Ab initio pseudopotentials for electronic structure calculations of poly-atomic systems using density-functional theory. *Comput. Phys. Commun.*, 119:67, 1999.

[52] F. Gamiz, J. Lopez-Villanueva, J. Banqueri, J. Carceller, and P. Cartujo. Universality of electron mobility curves in MOSFETs: A Monte Carlo study. *IEEE Trans. Electron Dev.*, 42:258, 1995.

[53] F. Gámiz, J. B. Roldan, J. A. López-Villanueva, P. Cartujo-Cassinello, and J. E. Carceller. Surface roughness at the Si-SiO_2 interfaces in fully depleted silicon-on-insulator inversion layers. *J. Appl. Phys.*, 86:6854, 1999.

[54] W. D. Gill. Drift mobilities in amorphous charge-transfer complexes of trinitrofluorenone and poly-n-vinylcarbazole. *J. Appl. Phys.*, 43:5033, 1972.

[55] R.W. Godby, M. Schluter, and L.J. Sham. Self-energy operators and exchange-correlation potentials in semiconductors. *Phys. Rev. B*, 37:159, 1988.

[56] A. Gold. Electronic transport properties of a two-dimensional electron gas in a silicon quantum-well structure at low temperature. *Phys. Rev. B*, 35:723, 1987.

[57] X. Gonze. First-principles responses of solids to atomic displacements and homogeneous electric fields: Implementation of a conjugate-gradient algorithm. *Phys. Rev. B.*, 55:10337, 1997.

[58] X. Gonze, D. C. Allan, and M. P. Teter. Dielectric tensor, effective charges, and phonons in *alpha*-quartz by variational density-functional perturbation theory. *Phys. Rev. Lett.*, 68:3603, 1992.

[59] X. Gonze, J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J.-Y. Raty, and D. C. Allan. First-principles computation of material properties: the ABINIT software project. *Comp. Mat. Sci.*, 25:478, 2002.

[60] X. Gonze and J.-P. Vigneron. Density-functional approach to nonlinear response coefficients of solids. *Phys. Rev. B*, 39:13120, 1989.

[61] D. L. Griscom. Self-trapped holes in amorphous silicon dioxide. *Phys. Rev. B*, 40:4224, 1989.

[62] D. L. Griscom. Electron spin resonance characterization of self-trapped holes in amorphous silicon dioxide. *J. Non-Cryst. Solids*, 149:137, 1992.

[63] D. L. Griscom and E. J. Friebele. Fundamental defect centers in glass: ^{29}Si hyperfine structure of the nonbridging oxygen hole center and the peroxy radical in a-SiO_2. *Phys. Rev. B*, 24:4896, 1981.

[64] D. L. Griscom and E. J. Friebele. Fundamental radiation-induced defect centers in synthetic fused silicas: Atomic chlorine, delocalized E' centers and triplet state. *Phys. Rev. B*, 34:7524, 1986.

[65] F. Gygi and A. Baldereschi. Quasiparticle energies in semiconductors: Self-energy correction to the local-density approximation. *Phys. Rev. Lett.*, 62:2160, 1989.

[66] N. Hamada and S. Ohnishi. Self-interaction correction to the local-density approximation in the calculation of the energy band gaps of semiconductors based on the full-potential linearized augmented-plane-wave method. *Phys. Rev. B*, 34:9042, 1986.

[67] W. Hayes and T. J. L. Jenkin. Charge-trapping properties of germanium in crystalline quartz. *J. Phys. C*, 19:6211, 1986.

[68] L. Hedin. New method for calculating the one-particle Green's function with application to the electron-gas problem. *Phys. Rev.*, 139:A796, 1965.

[69] E. Hernández, C. Goze, P. Bernier, and A. Rubio. Elastic properties of C and $B_xC_yN_z$ composite nanotubes. *Phys. Rev. Lett.*, 80:4502, 1998.

[70] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:864, 1964.

[71] R. C. Hughes. Charge-carrier transport phenomena in amorphous SiO_2: Direct measurement of the drift mobility and lifetime. *Phys. Rev. Lett.*, 30:1333, 1973.

[72] R. C. Hughes. Hole mobility and transport in thin SiO_2 films. *Appl. Phys. Lett.*, 26:436, 1975.

[73] R. C. Hughes. Time-resolved hole transport in a-SiO_2. *Phys. Rev. B*, 15:2012, 1977.

[74] M. Hybertsen and S. Louie. First-principles theory of quasiparticles: Calculation of band gaps in semiconductors and insulators. *Phys. Rev. Lett.*, 55:1418, 1985.

[75] A.-P. Jauho, N. S. Wingreen, and Y. Meir. Time-dependent transport in interacting and noninteracting resonant-tunneling systems. *Phys. Rev. B*, 50:5528, 1994.

[76] A. Javey, J. Guo, Q. Wang, M. Lundstrom, and H. Dai. Ballistic carbon nanotube field-effect transistors. *Nature*, 424:654, 2003.

[77] P. T. Jochym, M. Sternik, and K. Parlínski. Evaluation of dispersion relations of crystals from molecular-dynamics. *Comp. Mat. Sci.*, 6:339, 1996.

[78] L. P. Kadanoff and G. Baym. *Quantum Statistical Mechanics*. Benjamin, New York, 1962.

[79] C. Kaneta. Trap generation induced by local distortion in amorphous silicon dioxide film. *Jpn. J. Appl. Phys.*, 35:1540, 1996.

[80] L. V. Keldysh. Diagram technique for nonequilibrium processes. *Sov. Phys. – JETP*, 12:1018, 1965.

[81] V. M. Kenkre, E. W. Montroll, and M. F. Shlesinger. Generalized master equations for continuous-time random walks. *J. Stat. Phys.*, 9:45, 1973.

[82] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:1133, 1965.

[83] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169, 1996.

[84] W. Ku and A. Eguiluz. Band-gap problem in semiconductors revisited: Effects of core states and many-body self-consistency. *Phys. Rev. Lett.*, 89:126401, 2002.

[85] K. Laasonen, A. Pasquarello, R. Car, C. Lee, and D. Vanderbilt. Car-Parrinello molecular dynamics with Vanderbilt ultrasoft pseudopotentials. *Phys. Rev. B*, 47:10142, 1993.

[86] R. Lake and S. Datta. Nonequilibrium Green's-function method applied to double-barrier resonant-tunneling diodes. *Phys. Rev. B*, 45:6670, 1992.

[87] R. H. Landau. *Quantum Mechanics II.* Wiley-Interscience, New York, 1996.

[88] D. C. Langreth. Linear and nonlinear response theory with applications. In J. T. Devreese and V. E. van Doren, editors, *Linear and Nonlinear Electron Transport in Solids.* Plenum, New York, 1976.

[89] A. I. Larkin and Yu. N. Ovchinnikov. Nonlinear conductivity of superconductors in the mixed state. *Sov. Phys. – JETP*, 41:960, 1975.

[90] P. M. Lenahan and P. V. Dressendorfer. Hole traps and trivalent silicon centers in metal/oxide/silicon devices. *J. Appl. Phys.*, 55:3495, 1984.

[91] A. Y. Liu and A. A. Quong. Linear-response calculation of electron-phonon coupling parameters. *Phys. Rev. B*, 53:R7575, 1996.

[92] A. Y. Liu, A. A. Quong, J. K. Freericks, E. J. Nicol, and E. C. Jones. Structural phase stability and electron-phonon coupling in lithium. *Phys. Rev. B*, 59:4028, 1999.

[93] W. Lu, Z. Ji, L. Pfeiffer, K. W. West, and A. J. Rimberg. Real-time detection of electron tunnelling in a quantum dot. *Nature*, 423:422, 2003.

[94] Z.-Y. Lu, C. J. Nicklaw, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides. Structure, properties, and dynamics of oxygen vacancies in amorphous SiO_2. *Phys. Rev. Lett.*, 89:285505, 2002.

[95] G. D. Mahan. *Many-Particle Physics.* Plenum Press, New York, 2nd. edition, 1990.

[96] A. A. Mamedov, N. A. Kotov, M. Prato, D. M. Guldi, J. P. Wicksted, and A. Hirsch. Molecular design of strong single-wall carbon nanotube/polyelectrolyte multilayer composites. *Nature Mater.*, 1:190, 2002.

[97] L. F. Mao, J. L. Wei, Ch. H. Tan, and M. Zh. Xu. Determination of the effective mass of ballistic electrons in thin silicon oxide films using tunneling current oscillations. *Solid State Comm.*, 114:383, 2000.

[98] P. C. Martin and J. Schwinger. Theory of many-particle systems. I. *Phys. Rev.*, 115:1342, 1959.

[99] J. Maserjian. Tunneling in thin MOS structures. *J. Vac. Sci. Tech.*, 11:996, 1974.

[100] Y. Meir and N. S. Wingreen. Landauer formula for the current through an interacting electron region. *Phys. Rev. Lett.*, 68:2512, 1992.

[101] X. M. Meng, J. Q. Hu, Y. Jiang, C. S. Lee, and S. T. Lee. Boron nanowires synthesized by laser ablation at high temperature. *Chem. Phys. Lett.*, 370:825, 2003.

[102] N. F. Mott. Silicon dioxide and the chalcogenide semiconductors: similarities and differences. *Adv. Phys.*, 26:363, 1977.

[103] A. E. Newkirk. In R. M. Adams, editor, *Boron, Metallo-Boron Compounds, and Boranes.* Wiley, New York, 1964.

[104] C. J. Nicklaw, Z.-Y. Lu, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides. The structure, properties, and dynamics of oxygen vacancies in amorphous SiO_2. *IEEE Trans. Nuc. Soc.*, 49:2667, 2002.

[105] T. R. Oldham and F. B. McLean. Total ionizing dose effects in MOS oxides and devices. *IEEE Trans. Nucl. Sci.*, 50:483, 2003.

[106] C. J. Otten, O. R. Lourie, M.-F. Yu, J. M. Cowley, M. J. Dyer, R. S. Ruoff, and W. E. Buhro. Crystalline boron nanowires. *J. Am. Chem. Soc.*, 124:4564, 2002.

[107] G. Pacchioni and A. Basile. Calculated spectral properties of self-trapped holes in pure and Ge-doped SiO_2. *Phys. Rev. B*, 60:9990, 1999.

[108] G. Pacchioni, M. Vezzoli, and M. Fanciulli. Electronic structure of the paramagnetic boron oxygen hole center in B-doped SiO_2. *Phys. Rev. B*, 64:155201, 2001.

[109] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.*, 64:1045, 1992.

[110] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 80:890, 1996.

[111] J. P. Perdew and A. Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23:5048, 1981.

[112] G. Pfister. Pressure-dependent electronic transport in amorphous As_2Se_3. *Phys. Rev. Lett.*, 33:1474, 1974.

[113] D. Pines. *Elementary Excitations in Solids.* Perseus Books, Reading, MA, 1999.

[114] R. E. Prange and T.-W. Nee. Quantum spectroscopy of the low-field oscillations in the surface impedence. *Phys. Rev.*, 168:779, 1968.

[115] M. Prunnila, J. Ahopelto, and F. Gamiz. Electron mobility in ultrathin silicon-on-insulator layers at 4.2K. *Appl. Phys. Lett.*, 84:2298, 2004.

[116] J. Rammer and H. Smith. Quantum field-theoretical methods in transport theory of metals. *Rev. Mod. Phys.*, 58:323, 1986.

[117] Z. Ren, P. M. Solomon, T. Kanarsky, B. Doris, O. Dokumaci, P. Oldiges, R. A. Roy, E. C. Jones, M. Ieong, R. J. Miller, W. Haensch, and H.-S. P. Wong. Examination of hole mobility in ultra-thin body SOI MOSFETs. *IEDM Tech. Dig.*, page 51, 2002.

[118] E. Runge and E. K. U. Gross. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.*, 52:997, 1984.

[119] J. J. Sakurai. *Modern quantum mechanics.* Addison-Wesley, Reading, MA, 1994.

[120] G. G. Samsonidze, G. G. Samsonidze, and B. I. Yakobson. Kinetic theory of symmetry-dependent strength in carbon nanotubes. *Phys. Rev. Lett.*, 88:065501, 2002.

[121] S. Y. Savrasov. Linear-response theory and lattice dynamics: A muffin-tin-orbital approach. *Phys. Rev. B*, 54:16470, 1996.

[122] S. Y. Savrasov and D. Y. Savrasov. Electron-phonon interactions and related physical properties of metals from linear-response theory. *Phys. Rev. B*, 54:16487, 1996.

[123] H. Scher and E. W. Montroll. Anomalous transit-time dispersion in amorphous solids. *Phys. Rev. B*, 12:2455, 1975.

[124] H. Seyeda and M. Jansen. A novel access to ionic superoxides and the first accurate determination of the bond distance in O_2^-. *J. Chem. Soc., Dalton Trans.*, page 875, 1998.

[125] T. C. L. G. Sollner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker, and D. D. Peck. Resonant tunneling through quantum wells at frequencies up to 2.5 THz. *Appl. Phys. Lett.*, 43:588, 1983.

[126] D. Srivastava, C. Wei, and K. Cho. Nanomechanics of carbon nanotubes and composites. *Appl. Mech. Rev.*, 56:215, 2003.

[127] M. Stapelbroek, D. L. Griscom, E. J. Friebele, and G. H. Sigel, Jr. Oxygen-associated trapped-hole centers in high-purity fused silicas. *J. Non-Cryst. Solids*, 32:313, 1979.

[128] F. Stern and W. E. Howard. Properties of semiconductor surface inversion layers in the electric quantum limit. *Phys. Rev.*, 163:816, 1967.

[129] S. M. Sze. *Physics of Semiconductor Devices.* Wiley, New York, 2nd. edition, 1981.

[130] S.-I. Takagi, A. Toriumi, M. Iwase, and H. Tango. On the universality of inversion layer mobility in Si MOSFETs: Part I - effects of substrate impurity concentration. *IEEE Trans. Electron Dev.*, 41:2357, 1994.

[131] Y. Takahashi, M. Nagase, H. Namatsu, K. Kurihara, K. Iwdate, Y. Nakajima, S. Horiguchi, K. Murase, and M. Tabe. Conductance oscillations of a Si single electron transistor at room temperature. *IEDM Tech. Dig.*, page 938, 1994.

[132] S. J. Tans, R. M. Verschueren, and C. Dekker. Room-temperature transistor based on a single carbon nanotube. *Nature*, 393:49, 1998.

[133] S. Tarucha, D. G. Austing, T. Honda, R. J. van der Hage, and L. P. Kouwenhoven. Shell filling and spin effects in a few electron quantum dot. *Phys. Rev. Lett*, 77:3613, 1996.

[134] Y. Taur. Analytic solutions of charge and capacitance in symmetric and asymmetric double-gate MOSFETs. *IEEE Trans. Electron Dev.*, 48:2861, 2001.

[135] A. Thess, R. Lee, P. Nikolaev, H. Dai, P. Petit, J. Robert, C. Xu, Y. H. Lee, S. G. Kim, A. G. Rinzler, D. T. Colbert, G. E. Scuseria, D. Tománek, J. E. Fischer, and R. E. Smalley. Crystalline ropes of metallic carbon nanotubes. *Science*, 273:483, 1996.

[136] E. T. Thostenson, Z. Ren, and T.-W. Chou. Advances in the science and technology of carbon nanotubes and their composites: a review. *Composites Sci. and Tech.*, 61:1899, 2001.

[137] J. R. Tucker. Complementary digital logic based on the "coulomb blockade". *J. Appl. Phys.*, 72:4399, 1992.

[138] K. Uchida, J. Koga, and S. Takagi. Experimental study on carrier transport mechnisms in double-and single-gate ultrathin-body MOSFETs. *IEDM Tech. Dig.*, page 805, 2003.

[139] K. Uchida and S.-I. Takagi. Carrier scattering induced by thickness fluctuation of silicon-on-insulator film in ultrathin-body metal-oxide-semiconductor field-effect transistors. *Appl. Phys. Lett.*, 82:2916, 2003.

[140] K. Uchida, H. Watanabe, A. Kinoshita, J. Koga, T. Numata, and S. Takagi. Experimental study on carrier transport mechanism in ultrathin-body SOI n- and p-MOSFETs with SOI thickness less than 5nm. *IEDM Tech. Dig*, page 47, 2002.

[141] R. van Leeuwen. First-principles approach to the electron-phonon interaction. *Phys. Rev. B*, 69:115110, 2004.

[142] D. Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B*, 41:7892, 1990.

[143] N. Vast, S. Baroni, G. Zerah, J. M. Besson, A. Polian, M. Grimsditch, and J. C. Chervin. Lattice dynamics of icosahedral *alpha*-boron under pressure. *Phys. Rev. Lett.*, 78:693, 1997.

[144] S. Villa, A. L. Lacaita, L. M. Perron, and R. Bez. A physically-based model of the effective mobility in heavily-doped n-MOSFETs. *IEEE Trans. Electron Dev.*, 45:110, 1998.

[145] Y. Q. Wang and X. F. Duan. Amorphous feather-like boron nanowires. *Chem. Phys. Lett.*, 367:495, 2003.

[146] W. L. Warren, J. Kanicki, F. C. Rong, and E. H. Poindexter. Paramagnetic point defects in amorphous silicon dioxide and amorphous silicon nitride thin films. *J. Electrochem. Soc.*, 139:880, 1992.

[147] Y.-J. Xiong and S.-J. Xiong. Spin and charge oscillations of a dot due to interaction with itinerant electrons. *Phys. Rev. B*, 64:033305, 2001.

[148] Y. Xue, S. Datta, and M. A. Ratner. First-principles based matrix Green's function approach to molecular electronic devices: general formalism. *Chem. Phys.*, 281:151, 2002.

[149] S. Yamakawa, H. Ueno, K. Taniguchi, C. Hamaguchi, K. Miyatsuji, K. Masaki, and U. Ravaioli. Study of interface roughness dependence of electron mobility in Si inversion layers using the Monte Carlo method. *J. Appl. Phys.*, 79:911, 1996.

[150] P. D. Yoder and K. Hess. First-principles Monte Carlo simulation of transport in Si. *Semi. Sci. Tech*, 9:852, 1994.

[151] B. Yoon, H. Hakkinen, and U. Landman. Interaction of O_2 with gold clusters: molecular and dissociative adsorption. *J. Phys. Chem. A*, 107:4066, 2003.

[152] X. Zhang, C. K. Ong, and A. M. Stoneham. Stability of a self-trapping hole in α-quartz. *J. Phys.: Condens. Matter*, 6:5647, 1994.

[153] J. M. Ziman. *Electrons and Phonons*. Oxford, 1960.