MACROSCOPIC MODELING OF QUANTUM EFFECTS IN SEMICONDUCTOR DEVICES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Venkatasubramanian Narayanan

May 2006

MACROSCOPIC MODELING OF QUANTUM EFFECTS IN SEMICONDUCTOR DEVICES

Venkatasubramanian Narayanan, Ph.D.

Cornell University 2006

This dissertation explores the use of macroscopic quantum hydrodynamic (QHD) models as tools for investigating the transport of charge carriers in semiconductor devices in the regime where quantum effects are important. The primary contributions are an elucidation of the nature of density-gradient theory in the treatment of barrier repulsion effects and confinement and a clear derivation and application of a completely macroscopic model for tunneling calculations.

Chapter 1 provides a panoramic view of the field of carrier transport modeling in semiconductors. The essential differences between classical and quantum transport are discussed. Successively less detailed models from the fundamental starting points of the Boltzmann transport equation (BTE) for classical transport and the quantum distribution functions (Wigner function and the density matrix) based methods for quantum transport. A mention is made of the various quantum hydrodynamic models without going into the details of their derivation and applicability.

Chapter 2 brings into focus the area of quantum hydrodynamic modeling of carrier transport. A detailed derivation using the method of moments is presented for each of two popular quantum hydrodynamic models currently being explored in the literature, namely the density-gradient method and the smooth quantum potential model. A summary is made of their limitations and these limitations are then shown as arising out of particular assumptions made in their derivations that could hamper their applicable regimes.

Chapter 3 presents an analysis of the boundary layers near interfaces obtained in the density-gradient theory. An integral equation for the density near such interfaces is obtained and this is used to analytically compare the DG solution with the solutions from one-electron quantum mechanics in non-degenerate conditions. Modeling of confinement in simple potential wells is then discussed using the macroscopic equations.

Chapter 4 discusses the derivation of macroscopic equations to describe quantum mechanical tunneling through large barrier potentials. Using the approximate solutions of the Schrödinger equation it is analytically shown that the density profile inside the barrier satisfies a second order differential equation, very similar to the Schrödinger equation for a carrier at a suitably chosen average energy. Use of this is made to derive a consistent macroscopic treatment of tunneling transport in the insulating barrier.

Chapter 5, the final chapter, summarizes the major contributions of this dissertation and concludes it with several suggestions for future research directions that can stem from this work.

BIOGRAPHICAL SKETCH

Venkat Narayanan entered the world a sickly baby on March 30, 1978 in Palayamkottai, a small town adjoining the city of Tirunelveli in the state of Tamil Nadu, India. His first memories are from early kindergarten - of his mother feeding him and his elder brother Vijay lunch under the tall eucalyptus trees in Arul Malar convent, a missionary school in the historical temple city of Madurai. He also vaguely remembers his brother essaying forcefully the role of *Karna*, from the *Mahabharata* in a school play and swears by that to this day as the reason for the choice of his fondest character from the great Indian epic.

The family moved to Chennai (then Madras) in 1984 (non-Orwellian), a mere few months before the tumultuous events that followed the unfortunate assassination of the then Prime Minister, the redoubtable Mrs. Indira Gandhi. The next fifteen years were spent in the hot and sometimes hotter climes of Chennai, but being of a naturally sunny disposition he paid no particular heed to the conspiracy of geography and proceeded to absorb to his capacity from the carefully-packaged-in-school-textbooks erudition of yore. Somewhere along the way, he discovered the joy of reading and has been a voracious, nay, compulsive reader ever since. He still remembers the tingling sensation that accompanied hearing Sidney Carton say to him, "…It is a far, far better rest I go to than I have ever had." for the first time and has relived it anew as forcefully in the various travails of the Scarlett O'Haras and Robin Hoods as in the cold clarity of the Jiddu Krishnamurtis and Emmanuel Kants (the latter was of course rather confused tingling, he would hasten to add, probably an aftermath of scratching one's head for an extended period of time).

At seventeen, a goodly amount of hard work, a reasonable aptitude in Mathematics and Physics and a kindly and most opportune smile from Lady Luck

interfered constructively and he rode the crest into IIT Madras. That hallowed institution taught him more than just how to evaluate Fourier transforms and solve Maxwell's equations. Amidst some outrageously talented peers he learnt modesty; from some truly dedicated and larger than life teachers, inspiration; a little bit from every experience he tried to learn about himself and learnt of that quest it's never ending nature. And finally, being naturally more inclined to Newton's laws than to Riemann's sums, he chose semiconductors as an apposite vocation to indulge his interest and went to Cornell University where he defended his thesis in August 2005.

To Amma and Appa *– For teaching me the meaning of sacrifice*

To Vijay *– For being everything a brother can and should be… and more*

To Sharmi *– For bringing joy and laughter anew into our lives*

and

To Paatti *– For teaching me the power of unconditional love*

ACKNOWLEDGMENTS

No one can claim to have produced anything of value in a vacuum and it is very seldom indeed that ideas, or scientific contributions or works of art are ever thus produced. A thesis, in being a product of sustained effort over a few years is never so and in this section I offer my deepest gratitude to any number of individuals who made this possible.

I have always believed that a great teacher is one who inspires, rather than instructs and Prof. Edwin Kan, my advisor has been one such in my long stay at Cornell. He has been an ever present guiding force, a constant friend and a pillar of hope equally through the summers of my academic joy and the winters of scientific despair. To him I owe a debt that words here cannot amply describe, for how does one ever express in five lines what would take a book of anecdotes to pay justice to ?

I sincerely thank Profs. Sandip Tiwari and Prof. Bradley Minch, both for serving on my committee and being inspirations on and off it. Prof. Tiwari's great courses on Semiconductor Physics and Device Physics, and Prof. Minch's Analog Design course, were both very significant in providing excellent foundations on those subjects and in shaping my own academic and research interests overall.

I take this opportunity to thank all my group mates at Cornell – Zengtao Liu, Yu-Min (Nick) Shen, Myongseob Kim, Gen Pei, Chungho Lee, Jami Meteer, Pingshan Wang , Weiping Ni, Jinsook Kim, Blake Jacquot, Alex Tuo-Hung Hou and Jaegoo Lee and Udayan Ganguly. Zengtao was my first mentor at Cornell and the two years that we overlapped and shared an office saw us in many interesting and enjoyable conversations on work and otherwise. Gen Pei, I greatly admire for her tremendous work ethic, irrepressibly free spirit and boundless enthusiasm for learning. Chungho and Jinsook have been the concerned older friends to whom I have turned

(and will turn) for many a sage advice when my spirits sagged. Udayan's arrival filled totally the vacuum of sorts that was left on Zengtao's departure - I am grateful for the numerous stimulating discussions and his steadfast friendship I greatly cherish.

My long sojourn at Cornell has left me with lasting friendships struck over the course of the years. Jai's friendship is something that I am proud of in that we are very similar in outlook in many things and the few differences we have had have always been the subject of numerous interesting discussions. Ganesh will always be a good friend and well-wisher and never have I known anyone like him to wish and see good in everyone and everything. Giridhar, Swamy and Kiddo the trio that entered Cornell with me, have provided great camaraderie over the years and we have together laughed at our trials and tribulations in research many a time to keep each other going – this apart from sharing each other's culinary masterpieces.

Sowmya, Madan and Venu have offered me the ultimate support that one could have hoped for, in being always around when it mattered most. In my last three years at Cornell when I really got to know them very well, they metamorphosed from friends into family. My delight in interacting with them must have been apparent, but it is only proper that I take an extra moment or two to offer a toast to them here.

Asha for Education and Spicmacay are great organizations and my involvement with these provided the extracurricular spice to Cornell life. I thank the numerous friends I made during the course of those activities for their commitment to a good cause and offer my sincerest wishes that they continue to hold the flame of Gandhian service aloft.

And finally, I thank my most wonderful family – my parents, my brother, sister-in-law and my grandmother, who really need no mention. They have always been at the background of my thoughts and actions and to them rightly this thesis is dedicated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

IC	Integrated Circuit
BJT	Bipolar Junction Transistor
MOSFET	Metal-oxide-semiconductor field effect transistor
CMOS	Complementary metal-oxide-semiconductor
BTE	Boltzmann Transport Equation
PDE	Partial Differential Equation
NEGF	Non-equilibriuim Green's Function Formalism
PME	Pauli Master Equation
QTBM	Quantum Transmitting Boundary Method
DFT	Density-Functional Theory
LDA	Local Density Approximation
GGA	Generalized Gradient Approximation
SOI	Silicon-on-Insulator
WKB	Wentzel-Kramers-Brillouin Approximation
QHD	Quantum Hydrodynamic Equations
DG	Density Gradient Theory
QM	Quantum-Mechanics
DOS	Density of States
MIM	Metal-Insulator-Metal
UTBSOI	Ultra-thin body Silicon-on-Insulator

LIST OF SYMBOLS

$f(r,p)$	Classical phase-space distribution function
m^*	Effective mass of electrons
k,k'	Wave vectors
h	Planck's constant (1.054×10^{-34} Js)
n	Electron density
p	Hole density, Classical momentum
$\boldsymbol{\Pi}$	Momentum density in an electron gas
W	Energy density in an electron gas
$\boldsymbol{P}$	Stress tensor in an electron gas
τ_m	Momentum relaxation time for electrons
τ_w	Energy relaxation time for electrons
k_B	Boltzmann's constant (1.38×10^{-23} J.K^{-1})
T	Absolute Temperature
$\Psi_{n,k}$	Bloch states (in band n, and with wavevector k)
$E_{n,k}$	Energy of an electron in a lattice in band k and wavevector k
$W_{n,R}$	Wannier function centered on unit cell at R, for band n
V_t	Thermal voltage (26 mV at room temperature of 300 K)
λ_{th}	Thermal wavelength (electron wavelength at energy k_BT)
H	The Hamiltonian (energy) operator
ψ_i, ϕ_i	The i'th eigenvector of the Schrodinger equation
$U(t,t_0)$	The evolution operator for the wavefunction from t_0 to t
$G^n(k,k_1,t,t_1)$	Retarded time Green's function of the Schrodinger equation
$G^p(k,k_1,t,t_1)$	Advanced time Green's function of the Schrodinger equation
a_k^+	The many-body particle annihilation operator

a_k	The many-boxy particle creation operator
Σ_R, Σ_A	The retarded and advanced self-energy functions
$\Sigma_{in}, \Sigma_{out}$	The particle in-scattering and out-scattering functions
$\rho(x,x')$	The density-matrix of an ensemble in real-space representation
$f_w(x,p)$	The Wigner function in quantum mechanical phase space
β	The inverse temperature ($1/k_BT$)
E_F, k_F, λ_F	Fermi energy, Fermi wavevector and Fermi wavelength
$U(r)$	The potential energy
$V(r)$	The electrostatic potential
Z	The partition function
ε	Small parameter for perturbation analysis of the Bloch equation
μ_n, μ_p	Electron and hole mobilities
D_n, D_p	Electron and hole diffusivities
J_n, J_p	Electron and hole current densities
b_n, b_p	Gradient effect parameters for electrons and holes
r	Fitting parameter for the b_n and b_p expressions
s	Square root of electron density
γ	The tunneling recombination velocity at contacts
F	The electric field
E_b	Energy barrier height at semiconductor-insulator interface
k_b	Electron wavevector corresponding to the barrier height E_b
T_{WKB}	WKB transmission probability
P	The probability density of a single pure state wavefunction
S	Phase of a pure state wavefunction in Planck's constant units
$F_{1/2}$	The Fermi integral of order half

CHAPTER 1

CARRIER TRANSPORT MODELS

1.1 CMOS SCALING AND MOORE'S LAW

The invention of the planar processed integrated circuit (IC) [1] following that of the bipolar junction transistor (BJT) [2] is arguably the greatest technological breakthrough to have occurred in the last five decades, in terms of impact on everyday life. One would be hard pressed today to find a single area of human endeavor, in which electronics, as implemented in the integrated circuit does not play a role. The growth of integrated circuits in turn has been aided by numerous other inventions along the way, primary among which are those of the metal-oxide-semiconductor-field-effect-transistor (MOSFET) [3] and complementary-metal-oxide-semiconductor (CMOS) technology [4]. Tremendous improvements in IC performance have been achieved in the last three decades by steadily scaling down the size of the MOS transistor that has become the *de-facto* fundamental building block.

In 1965, Gordon Moore predicted [5] that the density of transistors in an IC chip was likely to double every 18 to 24 months. This prediction based on empirical data which has since been dubbed Moore's law has held out remarkably well for the better part of the last three decades.

The problem of designing devices for use in integrated circuits is an optimization problem with numerous constraints on performance, even for the case when they are meant to act only as switches in digital circuits. For most of the last three decades, the trend in device scaling has almost faithfully followed a universal scaling rule [6] that is derived from the assumption of simple electrostatics and the classical physics of ideal gases for transport. However, with device sizes entering the deep sub-micron (< 0.1 μm gate length) regime, several additional elements of the

detailed transport physics have required attention to understand device behavior adequately. These range from purely classical effects caused by the onset of ballisticity in the transport, such as hot-carrier effects [7], velocity overshoot [8], and impact ionization [9], to purely quantum mechanical effects due to the wave nature of the charge carriers in semiconductors [10,11].

This dissertation will deal with macroscopic models for the physics of semiconductor devices in the regime where quantum mechanical effects, such as quantization and tunneling are important. The rest of this chapter is devoted to introducing the reader to notions of classical and quantum transport and to give a brief overview of this very diverse field and current research.

1.2 TRANSPORT MODELS

There are two fundamental levels at which one can understand and describe the behavior of any many-body physical system. The first, is at a microscopic level – i.e. one can try to understand the dynamics of a single constituent particle in the system in the presence of external forces and then to extend this to when there are many such particles which interact with each other. This is physics, as described by statistical mechanics. The second description that one can attempt is at a macroscopic level, where one is not interested so much in the details of the driving forces and interactions, but their aggregate effects, on suitably defined averages (densities) in the system. This description is epitomized, for instance, in equilibrium by the laws of thermodynamics. Quite obviously, in order to have overall consistency, the macroscopic description should be derivable from the statistical mechanics of the system under consideration. A macroscopic description of the physics can however be obtained independently from experimental results as well. Indeed, the ideal gas laws

were known long before the fundamental laws of classical statistical mechanics were codified by Boltzmann [12] and subsequent workers.

The statistical mechanical description of transport in many-body systems is conveniently expressed in terms of kinetic equations [13,14]. Kinetic equations, describe the evolution in time and approach to equilibrium, of distribution functions (reduced to a single particle description or otherwise) defined on the system. The macroscopic descriptions of transport processes on the other hand are interpreted as conservation equations for physical quantities – charge, momentum, energy, energy flux etc. They can therefore be regarded as fundamental laws, provided the right relations for the various generalized driving terms that appear in these conservation equations are found experimentally or from the more detailed statistical descriptions [15]. These ideas remain independent of the differences between classical and quantum mechanics.

In the context of semiconductors, the many-particle system that one requires a description of, are the charge carriers – the electrons. The presence of the crystal lattice is usually acknowledged for a description of near equilibrium transport, by [16] i) the effective mass approximation, which assumes a parabolic band structure close to the edge of the energy bands obtained by solving the Schrödinger equation with the crystal periodic potential, and ii) the introduction of positively charged holes to account for the negative curvature of the valence bands. It is worthwhile to remember that in this description, the carrier wavefunctions described by solving the effective mass Schrödinger equation are the modulating functions for the set of Wannier functions [17-19] on the lattice. This is revisited in more detail in the Section 1.4, when we discuss quantum transport.

The transport properties of electrons and holes are respectively, in the effective mass approximation, those of interacting negatively and positively charged gases in an

external potential. The external potential, *V*, can be obtained in the Hartree approximation, by solving the Poisson equation with charge contributions from the mobile charges (electrons and holes) and the fixed charges (ionized dopants). Depending on the particular conditions that we are interested in, a classical or a quantum mechanical description of these (already semiclassical because of the effective mass approximation) gases will then be appropriate. These descriptions can in turn be statistical or macroscopic as has already been mentioned above.

1.3 CLASSICAL TRANSPORT

1.3.1 The Boltzmann Transport Equation

The kinetic equation for a classical gas is the Boltzmann Transport Equation (BTE), which describes the time evolution of the classical phase space distribution function *f(r,p)*, due to the action of the external forces (drift) and because of inhomogeneities in the distribution function in real space (diffusion). The equation is expressed as [16,20],

$$\frac{\partial f}{\partial t} + \vec{v}.\nabla f + \frac{F}{m^*}.\nabla_p f = \left.\frac{\partial f}{\partial t}\right|_{coll} \qquad (1.3.1)$$

The solution of this equation is the fundamental problem in classical transport theory. The right hand side of this equation yields the change in the distribution function due to the random external forces (scattering) that are encountered by the carriers. When this is zero, the BTE represents nothing but the conservation of phase space volumes (Liouville's theorem) under the evolution determined by Hamilton's equations of classical dynamics [21]. The scattering term is frequently written as an integral operator on the distribution function as under,

$$\left.\frac{\partial f}{\partial t}\right|_{coll} = \int_{-\infty}^{\infty} dk' \left[S(k,k') f(k)\left(1 - f(k')\right) - S(k',k) f(k')\left(1 - f(k)\right) \right] \qquad (1.3.2)$$

Detailed models for the different scattering mechanisms, due to impurities, acoustic and optical phonons, surface roughness etc. are available in the literature [16]. The scattering terms above represent the dominant term in the BTE and virtually all the tremendous variety in classical transport phenomena are associated with the very different solutions to the BTE that are obtained in regimes where scattering is important or relatively unimportant.

The BTE as written above is a semiclassical equation, because the effective mass m^* and since the scattering probabilities that appear in (1.3.2) are calculated from detailed quantum mechanics from the band structure and time dependent perturbation theory (Fermi's golden rule) [22], while the equation itself represents a purely classical evolution.

The BTE, with the scattering term (1.3.2) is a very difficult equation to solve numerically even in the effective mass approximation, since it is an integro-differential equation in the six coordinate phase space and time. Direct solutions using discretized versions of (1.3.1) are computationally unfeasible, except in simple one dimensional problems. Solution using the expansion in terms of basis functions (e.g. spherical harmonics) [23] has also been performed in some cases. The Monte-Carlo method is most often adopted for its solution [24] in multi-dimensional problems, self-consistently with the Poisson equation. The field of Monte-Carlo modeling of semiconductor devices using the BTE is very rich in itself and has seen extensive work [25-33] in the last two decades, on both the numerical aspects and towards the inclusion of more accurate physical models. Tremendous advances have been made in the ability to include the accurate band structure [25-27], different mechanisms of

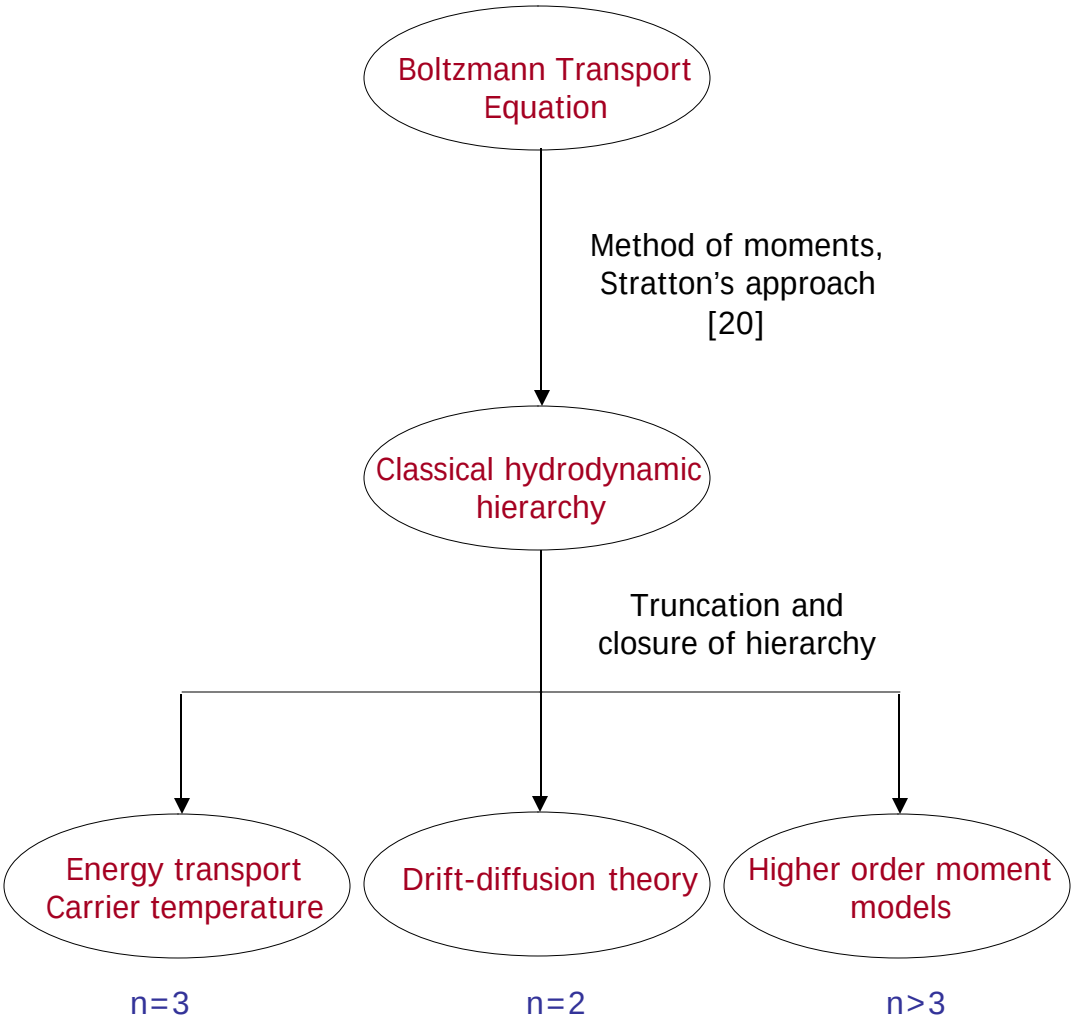

Figure 1.1 A representation of the classical transport theories at different levels of detail, from the kinetic equation (BTE) to the macroscopic equations from the truncation of the hydrodynamic set of equations

scattering (e.g. carrier-carrier scattering) [28,29], different materials [30] et cetera in this time. The Monte-Carlo method has served to clarify the understanding of several problems related to issues as diverse as mobility parameterization [28-31], high energy tails in the carrier energy distributions [33], detailed effects of band structure [32] etc that would have been otherwise difficult to analyze. In recent years, there have also been several attempts to include the effects of quantum mechanics namely quantization and tunneling, in classical Monte-Carlo transport simulations – these however are obviously not based on fundamental quantum mechanical transport equations (since the BTE is classical), but added on as corrections to the classical equations for particle dynamics and statistics, using numerical or analytical solutions of the one-electron Schrödinger equation [34].

1.3.2 Classical Hydrodynamic Modeling and Closure Relations

In spite of the tremendous advances made in the Monte-Carlo modeling of devices using the BTE, the method remains more an exploratory tool than an engineering tool suitable for use in device design. This is owing to the tremendous computational complexity involved in the calculation of the full-band Monte-Carlo solution. Instead the pride of place as far as use in engineering design with the classical transport equations is concerned, has been accorded to the macroscopic transport equations obtained from the BTE, viz. the hydrodynamic equations.

The term “hydrodynamic” is a general qualification used to refer to the various equations that are obtained by averaging out the detailed momentum space information [20] from the BTE. Any distribution function (or a probability distribution) can be expressed in terms of its moments [35] and therefore its evolution can modeled by the evolution of its moments. The easiest way to get a PDE description in real space is to extract the moments of the BTE, by integrating out the

momentum space information systematically, after multiplying the equation with different orders of the momentum, $\{p^i, I = 0,1,2\ldots\}$. The first three moments from this operation, yield the equations of conservation of charge density (n), momentum density (Π), and energy density (W) respectively and can be written as under,

$$\begin{aligned}
&\frac{\partial n}{\partial t} + \frac{1}{m}\nabla.\vec{\Pi} = 0 \\
&\frac{\partial \vec{\Pi}}{\partial t} + \nabla.(u\vec{\Pi} - \vec{\vec{P}}) = -n\nabla U - \frac{\Pi}{\tau_m} \\
&\frac{\partial W}{\partial t} + \nabla.(\vec{u}W - \vec{u}.\vec{\vec{P}} - \vec{Q}) = \vec{u}.\nabla U - \frac{(W - W_0)}{\tau_w}
\end{aligned} \tag{1.3.3}$$

A simple relaxation time approximation [16] for the BTE has been assumed in deriving the above, with momentum and energy relaxation times given by τ_m and τ_w. The quantity u represents the centroid of the non-equilibrium distribution function f and Q is a heat flux term. P represents the stress tensor of the gas given by

$$\vec{\vec{P}} = \left\langle \frac{(\vec{p} - m^*\vec{u})(\vec{p} - m^*\vec{u})}{m^*} \right\rangle \tag{1.3.4}$$

Frequently, the distribution function in most regions of a device does not differ appreciably in functional form from the equilibrium Fermi-Dirac distribution, except for a small shift in the centroid in momentum space. In this case, the higher order moments remain unchanged from their equilibrium values and the system can be modeled to an excellent approximation by the first few moment equations only.

As is seen from the Eqns (1.3.3) each equation in the hydrodynamic hierarchy has a higher order moment as one of its driving terms, due to the diffusion term (second term) in the BTE. For instance, the first moment equation, the momentum

balance equation retains the divergence of the stress tensor *P*, which is a second moment of the full *non-equilibrium* distribution function. When the system is modeled using the hydrodynamic equations to a certain order *(n-1)*, a closure relation, has to be found for the *n*'th order driving term appearing in the highest order moment equation, in terms of the lower order moments, to truncate the hierarchy. Since the non-equilibrium distribution function is not known (it is what we are solving for using the BTE), we need to make an intelligent assumption as to the form of the closure relation. The method that is most often adopted is to assume that the functional form for the highest order moment from equilibrium holds *locally* under non-equilibrium conditions as well. The closure relations can therefore be interpreted as *equations of state* for the electron and hole gases – i.e. a constitutive relation, which holds for the particular kind of system, in a particular type of ensemble, independent of the transport equations.

The truncation of the hierarchy at *n=2* yields the drift-diffusion equations [16,20]. The closure relation for the stress tensor *P* is obtained from the equilibrium Boltzmann distribution function and is isotropic,

$$\left\langle \frac{(\vec{p}-m^{*}u)(\vec{p}-m^{*}u)}{m^{*}} \right\rangle = -nk_{B}T.\delta_{ij} \qquad (1.3.5)$$

The stress tensor relation is therefore, the same as the pressure/density relationship (equation of state) of an ideal gas [12]. The drift-diffusion equations therefore, correspond to assuming that the electrons and holes are ideal interacting gases in the semiconductor. The convective term inside the divergence term for the stress tensor is usually ignored in classical transport, since it is much smaller than the stress tensor term, on account of the average carrier momentum $m^{*}u$ being much smaller than the thermal momentum.

The drift-diffusion equations have been the bulwark of semiconductor device modeling for engineering design in the last two decades, due to their simplicity and their ease of numerical implementation. They have also formed the backbone of all - physical and semi-physical compact device models for circuit simulation. Consequently they have attracted great attention [36,37] over the last few decades on virtually every aspect of their derivation, numerical implementation and parameterization for different materials, device structures and geometries.

The applicability of the drift-diffusion equations depends on the careful parameterization of the average quantities appearing in them, the mobilities and the generation-recombination terms in terms of the local dependent variables, i.e. the electron and hole densities and the electrostatic potential. A closure relation such as (1.3.5) assumes that carrier acceleration by the electric field is not very important and that there is adequate scattering to keep the distribution function locally close to an equilibrium Maxwellian like form.

Higher order truncations of the hierarchy can be performed with different sets of closure relations. These equations, partially model the carrier acceleration in the presence of an electric field that is not captured in the drift-diffusion model since it assumes that the local form of the distribution function is very close to an equilibrium distribution function shifted in momentum coordinates [16]. Energy transport equations can be obtained by closing the system of equations at $n=3$. These can further be simplified to the electron temperature model, through an assumption of a Maxwell-like distribution function with a different temperature for the carriers [20]. Modeling of semiconductor transport using as many as six moment equations have been reported in the literature [38-40]. However beyond a point these equations lose the physical simplicity and elegance of the moment equation approach, since it is hard to get an intuitive grasp of very high order moments and their behavior, especially

given the number of parameters that are introduced in these equations to model the average behavior. These parameters themselves are often calibrated against full band Monte-Carlo simulations [40], when they can no longer be described using simple extrapolations from equilibrium.

1.4 QUANTUM TRANSPORT – STATISTICAL APPROACHES

1.4.1 Validity of Effective Mass Theorem and Classical Transport Theory

The carriers in semiconductors, in equilibrium occupy the energy levels that are obtained from a solution of the Schrödinger equation in the periodic lattice potential V_L [17], i.e. the Bloch states $\Psi_{n,k}$. The states are labeled by the crystal momentum k in the first Brillouin zone and the band index, n.

$$\left(-\frac{\hbar^2}{2m}+V_L(r)\right)\Psi_{n,k}(r)=E_{n,k}\Psi_{n,k}(r) \tag{1.4.1}$$

The potential V_L has the periodicity of the lattice, i.e. if Q_i is the translation vector that takes from the unit cell at the origin to the i'th cell then,

$$V_L(r+R_i)=V(r) \tag{1.4.2}$$

The probability of occupation of these states is given by the Fermi function of their energy eigenvalues. These states are delocalized over the entire lattice. As mentioned earlier, these Bloch states can be written in terms of Wannier functions, W_k [17-19] defined on the lattice for a particular band index n as,

$$\Psi_{n,k}(r)=\frac{1}{\sqrt{N}}\sum_j \exp\left(\frac{i\vec{k}.\vec{R}_i}{\hbar}\right)W_k(r-R_j) \tag{1.4.3}$$

In the presence of an external potential V_{ext} the solution for the total wavefunction can be written in terms of the Wannier function basis as,

$$\Psi_q(r) = \sum_j \varphi_q(R_j) W_k(r - R_j) \tag{1.4.4}$$

The *discrete* coefficients φ_q with the new quantum number q in this defined only in each unit cell, can then be extended to the continuous space r and can be shown to satisfy the following equation [19], when V_{ext} is slowly varying over one unit cell,

$$\left[E_{n,k}(-i\hbar \frac{\partial}{\partial r}) + V_{ext}(r) \right] \varphi_q(r) = E_q \varphi_q(r) \tag{1.4.5}$$

This is just a re-statement of the effective mass approximation when the band $E_{n,k}$ is parabolic in k. This suggests that the description of electron wavefunctions has been shifted from the complete Bloch form to the coefficients φ_q which modulate the Wannier functions. The Wannier functions form a highly localized basis and their spatial extent is limited to the unit cells of the lattice – they can be visualized as playing the same role in a lattice that delta functions do in the absence of a periodic potential. Therefore the effective mass approximation above is valid for potential variations which are small relative to the magnitude of the crystal potential over a unit cell size (i.e. over a distance of the order of the lattice constant).

If the effective mass approximation is assumed to be valid we then have, in the absence of an external potential, a description for electrons and holes in terms of plane waves, as seen from a comparison of (1.4.3) and (1.4.4). These plane waves are nearly at a continuum, since the quantized wavevectors k_i are determined by the total size of the crystal, V according to the requirement that the wavefunctions go to zero at the edges of the crystal. Since the macroscopic size of the crystal V is usually very large

compared to the unit cell volume (which is decided by the lattice constant *a*) the wavevectors can be assumed to be a continuum. There are exactly *N* states for each band in the Brillouin zone, where *N* is the total number of unit cells in the crystal.

When we have a near continuum of free-carrier like eigenstates (plane waves), one can define localized wavepackets as a superposition (with some weight function *w*) of closely separated eigenfunctions with different wavevectors in a small neighborhood of some *k*.

$$g_k(r) = \int_{k-\Delta k}^{k+\Delta k} w(\vec{k}_1).\exp\left(i\vec{k}_1\vec{r}\right)dk_1 \qquad (1.4.6)$$

Now, according to Ehrenfest's theorem [41] in quantum mechanics, the *average* position of any such wavepacket and its *average* momentum obey the laws of classical mechanics. Hence the above wavepacket, which has a finite extent in real and momentum space, can be interpreted as a classical particle, provided we examine it over length and momentum scales which are large compared to the respective spreads. This is the only sense in which the concept of a particle has to be understood in quantum mechanics and therefore in semiconductors.

We can now state under what conditions classical transport theory as described in the previous section can be a useful approximation to the physics. One can stipulate the following general guidelines as necessary for the transport to be describable in semiclassical terms.

a) The length scales over which quantities (may be potential, scattering probabilities etc.) vary significantly must be much larger than the delocalization length of the wavepackets that one can construct with the occupied (delocalized) eigenstates.

b) Collisions between wavepackets at k and k_1 must be local in space and instantaneous in time, compared to time and length scales considered [16]
c) Scattering must be completely randomizing and the carriers must not retain any phase information for sufficiently long time evolutions.

In particular, for instance, close to equilibrium and for Boltzmann statistics, (i.e. non-degenerate conditions) only wavevectors smaller than corresponding to the thermal energy k_BT, i.e.

$$k_{th} = \sqrt{\frac{2m^* k_B T}{\hbar^2}} \tag{1.4.7}$$

are available to construct wavepackets. This implies that one cannot use the classical theory to describe phenomena that require a spatial resolution of better than about

$$\lambda_{th} = \sqrt{\frac{\hbar^2}{2m^* k_B T}} \tag{1.4.8}$$

since this is a practical limit on the smallest wavepacket that can be constructed at room temperature. For silicon and its large longitudinal effective mass [16] (0.916 m_0) the above length scale works out to around 1.3 nm at room temperature. Typically the length is around a few nanometers, since the average effective mass is smaller than this due to the effect of the small transverse mass (0.19 m_0). This puts an obvious restriction on the variation scale of the external potential – i.e. the potential variation in this length scale must be much smaller than the thermal energy k_BT in order for a classical description based on the BTE to be adequate.

We can therefore understand from the above discussions that the validity of the effective mass approximation and the semiclassical theory of transport hinge on

similar restrictions. For the effective mass theory to be valid, one requires the variation of the potential to be small compared to the crystal potential over a lattice constant, while for the semiclassical theory to be applicable one requires that variation to be small compared to the thermal voltage over a thermal wavelength. Since the crystal potential [17] is usually large compared to the thermal voltage V_t (0.0259 V at 300 K) in the unit cell (being due to the screened ion cores) and the lattice constant is around 0.5 nm (i.e. smaller than the thermal wavelength as determined by the effective masses), there is likely to be a significant regime where the effective mass theory holds for the description of carrier wavefunctions, but the classical transport description for the constructed wavepackets does not. For all the discussions that follow, we will assume that we are considering the semiconductor transport properties in this regime, so that we do not have to take into account the lattice explicitly.

It is worthwhile to reiterate at this point that the above discussions pertain only to the classical or quantum mechanical description of transport and have nothing to do with the microscopic or macroscopic models that we may employ for such a description.

1.4.2 Coherent Evolution

A complete understanding of the dynamics of a single particle system is achieved from the knowledge of its wavefunction at any initial time t_0 and the Hamiltonian of the system. The Schrödinger equation can then be solved for the time evolution and all properties at time t can be extracted from the resulting wavefunction. The same information can of course, be obtained by solving the time-independent Schrödinger equation for the spectrum of the Hamiltonian, writing the initial state wavefunction as a superposition of these eigenstates and then using the individual evolutions of the eigenstates with energy E according to the rule [41],

$$\Psi(x,t) = \psi(x,t_0).\exp\left(-\frac{iEt}{\hbar}\right) \tag{1.4.9}$$

The above discussion assumes that the Hamiltonian itself is not directly dependent on time. If it is time dependent then the evolution cannot be solved for trivially using the spectrum at any one time alone as above. We can instead write the total wavefunction at time *t*, as the action of an evolution operator, *U(t,t$_0$)* on the initial state, i.e.

$$\Psi(x,t) = U(t,t_0)\psi(x,t_0) \tag{1.4.10}$$

The evolution operator *U* satisfies the initial value equation,

$$i\hbar\frac{d}{dt}U(t,t_0) = H(t)U(t,t_0) \ , \ U(t_0,t_0) = 1 \tag{1.4.11}$$

This operator equation has to be integrated explicitly in order to determine the final state [42]. The spectrum of the Hamiltonian changes with time, and the Hamiltonian operators at different times t_1 and t_2 do not in general commute with one another. Quite obviously then a particle in an eigenstate of the initial Hamiltonian does not stay in an eigenstate of the Hamiltonian at some other time during the evolution.

For a many particle system, the dynamics in the presence of a time independent, deterministic Hamiltonian can be understood in much the same way as above for the single particle case. The system is completely described by the Hamiltonian and the initial state – which in the absence of exact information about the full many body wavefunctions is approximated by the linear combination of the product of single particle wavefunctions in a mean field, as in the Hartree or Hartree-Fock approximation [17]. Symmetry (or antisymmetry) relevant for Fermions or Bosons can be introduced by the linear combinations that are admitted as

wavefunctions, for instance through the use of Slater determinants. Coherent transport phenomena, involving two-level systems, quantum Hall effect [43] etc. can be readily handled using these methods.

1.4.3 Non-Equilibrium Green's Functions

Finite temperature, interactions and scattering processes complicate the above simple picture considerably. At finite temperature, in equilibrium, the laws of statistical mechanics require a Fermionic system to be distributed in energy according to the Fermi-Dirac distribution, irrespective of the exact nature of the scattering processes. This amounts to a certain loss of coherence in the system, since the system is no longer in one single pure state given by a linear combination of eigenfunctions of the Hamiltonian – only a probability distribution for the occupation of states different in energy can be given. The initial state of a system must thus be described as an incoherent superposition of pure states. The problem of quantum transport is the description of the time evolution of this initial state (possibly an equilibrium state) in the presence of interactions, dissipation and external driving forces.

A simple picture of the various processes that must be studied in order to understand the transport properties is as follows. Consider one particle, occupying at some time t_0 the state k in some suitable basis set. There is a Hamiltonian H_{ext}, due to the externally applied, driving perturbation to the system. The state begins to evolve according to the coherent evolution given by this Hamiltonian, from some initial state which is statistically defined. As a result of this coherent evolution, there is an amplitude for the particle to propagate to state k_1 at some time t_1. There may be a scattering event at some random time t_2 in between, to another intermediate state k_2, which might in turn evolve to reach k_1 at t_1. The randomness in the time t_1 will mean a loss of coherence in the evolution. In addition, if the system is open (i.e. current can

flow through the system), the particle may escape to one of the contacts, in which case, for current conservation, another coherent evolution of a particle injected from one of the other contacts will begin and this might lead to an amplitude for the state k_1 at t_1 as well. This latter evolution due to the contacts will also be incoherent, since they are usually in equilibrium and the distribution of carriers injected into an active device will only depend on the Fermi energy of the contact.

The evolution of a quantum mechanical system, at finite temperature in the presence of a time varying Hamiltonian is then most generally written in terms of correlation functions that take the role of the distribution function in a classical description [43-45]. For instance in the free carrier basis set (plane waves), labeled by wavevectors $\{k\}$ we can define the following correlation functions in terms of the creation and annihilation operators [44] a_k and a_k^+

$$G^n(k,k_1,t,t_1) = \left\langle a_{k_1}^+(t_1)a_k(t)\right\rangle$$
$$G^p(k,k_1,t,t_1) = \left\langle a_k(t)a_{k_1}^+(t_1)\right\rangle \qquad (1.4.12)$$

The language of second quantization [44] has been used here, but basically what the above represent are the amplitudes of the state k at time t given unit amplitude in state k_1 at t_1, for the cases when the time t_1 is earlier or later than t (for a Hermitian system, these contain identical information).

In steady state, the above functions depend only on the difference between the times ($\tau = t\text{-}t_1$). Besides, in steady state the spectrum is static in time and so the time evolution from t_1 to t, depends on the state at t_1 similar to Eqn (1.4.9). This allows us to identify the conjugate variable of ($t\text{-}t_1$) (in the sense of a Fourier transform) as the energy. The correlation function can also be written in terms of the energy variable by performing the Fourier transform of the above with respect to the time difference τ.

$$G^n(k,k_1,t,E)=\frac{1}{\hbar}\int_0^{\infty}G^n(k,k_1,t,\tau)\exp\left(-\frac{iE\tau}{\hbar}\right)d\tau \qquad (1.4.13)$$

The quantities G^n and G^p contain all the information that is required of a system in non-equilibrium for calculating average values. For instance, the density $n(r,t)$ is given by the diagonal elements in space integrated over energy,

$$n(r,t)=\frac{1}{2\pi}\int_0^{\infty}G^n(r,r,t,E)dE \qquad (1.4.14)$$

Other quantities, such as currents can be calculated in a similar manner through integrations in energy.

The kinetic equation for the above two correlation functions is the fundamental result of the Non-Equilibrium Green's Function formalism (NEGF). This equation plays the same role in quantum transport theory that the BTE plays in classical transport theory. For the sake of brevity we simply present the equations here in matrix form in steady state, without going into the details of their derivation [43,44] (each of the quantities have the same matrix structure as the correlation functions above).

$$G^n=G^R\Sigma^{in}G^A \quad , \quad G^p=G^R\Sigma^{out}G^A \qquad (1.4.15)$$

The quantities appearing in the above equations have rather simple meanings – the functions G_R and G_A are the retarded and advanced Green's functions respectively of the modified Schrödinger equation. These represent the propagation of the single particle excitations in the Hamiltonian field, in the presence of dissipative interactions or open boundary conditions.

$$G^R = \left[EI - H - \Sigma^R\right]^{-1} \quad , \quad G^A = \left[EI - H - \Sigma^A\right]^{-1}$$
$$H_{eff} = H + \Sigma^R \quad , \quad \Sigma^A = (\Sigma^R)^+ \tag{1.4.16}$$

The self-energy functions Σ^R and Σ^A model the effects of random scattering due to impurities, phonons and surface roughness, as well as that of the open boundaries. All these mechanisms have the effect of introducing a non-Hermitian part to the Hamiltonian, which cause the single particle eigenstates to decay in time. The lifetime of a state *k* is determined by the imaginary part of the eigenvalues introduced by the self-energy functions.

The functions Σ^{in} and Σ^{out} represent the in-scattering and out-scattering amplitudes respectively – these are simply generalizations of the semiclassical scattering function $S(k,k')$ and $S(k',k)$ that appear in the Boltzmann Transport equation. In the NEGF, all these quantities must be *inputs* estimated independently using first order perturbation theory.

As can be realized from (1.4.15-16), a self-consistent numerical solution of the Poisson equation and the NEGF equations, is bound to be even more computationally intensive than the BTE due to the additional energy coordinate, and the integrations with respect to energy (1.4.14) which are required to calculate the carrier densities at each iteration. The matrices that are to be inverted for each energy in (1.4.16) are in general not sparse and have high bandwidth (since we do not know the basis set in which they are nearly diagonal) when scattering terms are included [45]. Besides, since the energy eigenvalues are not known apriori, very fine energy grids are often required to resolve quantization energies around which the density of states vary very rapidly.

Because of the above reasons of computational intractability, there are relatively few instances in which the above equations have been applied generally to semiconductor devices. These are mostly one-dimensional or quasi one-dimensional investigations of lateral transport in MOSFETs or double-gate devices for quasi-ballistic transport [46], or with very simple scattering models [47,48,50]. For quasi-ballistic transport, the NEGF equations are really unnecessary and they reduce to those of the Landauer formalism [43]. A notable exception is [51] in which the 2-D equations including the full phonon-scattering self-energies have been solved for an idealized double-gate geometry. The inclusion of realistic device geometries, band-structure and the effect of all the various scattering mechanisms is still a long way from realization.

1.4.4 Density Matrices and Wigner Functions

As noted in the previous section, in order to completely describe quantum transport in the presence of a time varying Hamiltonian, one requires a description in terms of the two time correlation functions, $G^n(k,k_1,t,t_1)$. The correlation between two different times, in turn represents the effect of the detailed *time varying* spectrum through (1.4.13). Inelastic processes which transfer carriers from one energy to another, cause a part of this correlation, while the coherent Hamiltonian evolution causes the other part.

The time correlation caused by the inelastic processes, decay extremely fast in the time difference ($\tau = t\text{-}t_1$). The typical time for which this correlation exists, is of the order of the scattering time, which in turn can be approximated by the uncertainty principle to be around,

$$\tau_{sc} = \frac{\hbar}{\Delta E} \tag{1.4.17}$$

where ΔE represents the energy difference between the two "levels" taking part in the process. For energies relevant in semiconductors, (of the order of eV), these times are of the order of femtoseconds. Tunneling times, scattering times etc. are all of this typical duration.

If the rate of variation of the Hamiltonian is small, i.e. for some norm if

$$\left|\frac{\partial H}{\partial t}\right| << \frac{\Delta E}{\tau_{sc}} \tag{1.4.18}$$

then the only reason transitions occur between states ΔE apart in this time scale is due to the inelastic processes. Then for time scales much longer than these, we can think of the transitions as occurring between energy states, defined by the instantaneous Hamiltonian (i.e. quasi-statically) and ignore the time correlation. In this approximation (the Markov approximation, since the system now has no memory), the properties of the system can be described by the single-time correlation function, $G^n(k,k_1,t,t)$. This quantity is also called the density-matrix of the system.

$$\rho(k,k_1,t) = G^n(k,k_1,t,t) = \left[G^p(k,k_1,t,t)\right]^+ \tag{1.4.19}$$

The density matrix can thus be used to describe a quantum mechanical system with instantaneous scattering processes. All the detailed information on the scattering processes available from the self-energy functions in the above equations has to be averaged in some form to introduce the irreversibility in the transport equation. One such approximation for the transport equation is obtained by assuming that the density matrix evolves according to the Heisenberg equation of motion in the Hamiltonian, with an additional term due to scattering, similar to the spirit of the BTE i.e.

$$i\hbar\frac{\partial\rho}{\partial t}=\left[H(t),\rho(t)\right]+\left.\frac{\partial\rho}{\partial t}\right|_{coll} \tag{1.4.20}$$

Different models for the collision term [51-54] can then be used in order to solve this equation numerically. The problem of determining quantum mechanically consistent scattering models to use in (1.4.20) above is very much an open problem, as is that of a good choice of basis states for representing the density matrix in the above equation for numerical simplicity. Using a Bloch representation [52,53] or a Wannier representation [54], is particularly useful, if interband transition effects (such as Zener tunneling) are important, although for single-band transport, one might use basis states from the solution of the one-electron, effective mass Schrödinger equation. Unfortunately, since the above equation is not generally in the form of a master equation (i.e. a differential, or integro-differential equation, with diagonal dominance) one cannot use Monte-Carlo like methods for a general solution as with the BTE.

When the density matrix is represented in the real space basis, a special operation, the Weyl transform [55], can be performed to yield a phase space distribution similar to the classical phase space distribution. In this case, the density matrix is written as a correlation function of two space variables x and x_1, i.e. $\rho(x,x_1)$. Then the quantity, $f_w(r,p,t)$ defined by

$$r\rightarrow\frac{x+x'}{2}\qquad \eta\rightarrow x-x'$$
$$f(r,p,t)=\int_{-\infty}^{\infty}\rho(r,\eta,t).\exp\left(-\frac{ip\eta}{\hbar}\right)d\eta \tag{1.4.21}$$

is called the Wigner distribution function and has properties very similar to a classical phase space distribution function, for evaluating average values of observables [55,56]. The function however is not a true distribution function, since it is not

positive definite. The equation (1.4.20) for the density matrix transforms in this case to the Wigner-Boltzmann equation [57,58],

$$\frac{\partial f_w}{\partial t} + \vec{v}.\nabla_r f_w + \frac{2}{\hbar}\sin\left[\frac{\hbar}{2}\nabla_r\nabla_p\right]U(r) f_w = \left.\frac{\partial f_w}{\partial t}\right|_{coll} \qquad (1.4.22)$$

The similarity of this equation to the BTE (1.3.1) is striking – it is straightforward to note that the limit of small Planck's constant yields the BTE. The effects of quantum mechanics enter through the non-local nature of the interaction with the potential due to the presence of an infinity of derivatives.

The Wigner function and its evolution equation have been the object of much interest owing to the attractive interpretation as a distribution function. Several one-dimensional problems have been simulated using (1.4.22), notably, those involving the resonant tunneling diode and its negative differential conductivity [57-61]. These workers have investigated aspects of numerical implementation, scattering models, modifications required for inclusion of band structure, transient effects etc., using direct discretized solvers for (1.4.22). Since (1.4.20) as pointed out for the density matrix, is not generally a master equation, it does not readily admit of Monte-Carlo like methods for solution. This is the prime reason for the very few instances that exist for applications in multiple dimensions. An interesting exception is the series of investigations by Ferry and co-workers [62-64] where the Monte-Carlo method is generalized to solve (1.4.22) by ascribing to the particles an additional sign property to account for the negative values of the Wigner function – this method appears to provide meaningful results, when physical quantities do not vary too rapidly in space, although this analysis has also been limited to one dimension.

The Wigner function and the density matrix are the basis for the derivation of the various macroscopic quantum transport approaches and will be re-visited in more detail in the next chapter.

1.4.5 The Pauli Master Equation

One possible, natural way of assigning basis states and treating the scattering term in the density matrix evolution (1.4.20), is to use the set of instantaneous Hamiltonian eigenstates and then to use the Fermi Golden rule [22] for the scattering rates from one eigenstate to another. This approach leads naturally to the Pauli Master Equation (PME) for the time variation of the *diagonal* elements of the density matrix in the Hamiltonian representation.

$$\frac{\partial \rho_{ii}}{\partial t} = \sum_{j \neq i} \left[W_{ji} \rho_{jj} - W_{ij} \rho_{ii} \right] + \left(\frac{\partial \rho_{ii}}{\partial t} \right)\Bigg|_{con} \tag{1.4.23}$$

W_{ij} is the probability per unit time of a transition occurring from the state labeled i to the state j as given by the Fermi Golden rule of first order time dependent perturbation theory (E_{int} is the energy of the scattering mechanism, e.g. a phonon energy)

$$W_{ij} = \frac{2\pi}{\hbar} \left| \left\langle \psi_i \left| H_{\text{int}} \right| \psi_j \right\rangle \right|^2 \delta\left(E_i - E_j + E_{\text{int}} \right) \tag{1.4.24}$$

A general equation for transport using the density matrix, as pointed out previously requires the solution of Eqn (1.4.20) or (1.4.22), thus including the off diagonal elements as well. However, in the weak scattering limit, for a closed system (i.e. no open boundaries), it has been shown by Van Hove [66] that the off-diagonal elements in the density matrix remain negligible under weak scattering interactions and Hamiltonian evolution, *provided* that the initial state is quasi-diagonal in the Hamiltonian eigenstates. The PME ignores the non-diagonal elements of the density

matrix, by assuming that the scattering terms only re-distribute the eigenstate populations. Recently, this result has been extended to open systems by Fischetti [67], under the assumptions that the dephasing length due to the external particle reservoirs is large compared to the dimensions over which transport is considered (i.e. the physical dimensions of the device). For degenerately doped silicon, this length is atleast around 50 nm [68] and so the PME can be used for analyzing devices that are of the order of this length or smaller.

The biggest advantage that the PME offers over the full density matrix or Wigner function transport equation is that it is amenable to a Monte-Carlo solution (being a master equation with diagonal dominance) and hence the enormous work that has been performed on those methods can be directly applied to it. Several very encouraging recent results have appeared in the literature including the effects of band structure, with realistic geometries [67-69] to investigate the onset of ballisticity and interference effects etc. at small geometries in two dimensions and steady state. They suffer from a similar drawback as the Monte-Carlo method in terms of computational cost – the PME is even more intensive to solve numerically because of the requirement of solving the full multidimensional eigenvalue problem at each time step instead of simple semiclassical dynamical equations as with the BTE. This is a formidable problem [70] by itself.

1.4.6 Self-Consistent Schrödinger-Poisson Solutions

The PME represents transport in the weak scattering limit, when the dephasing lengths in the contacts are larger than the device dimensions. If scattering, in the active device regions of this length is ignored completely, then the PME need not be solved and the only system that needs to be solved is that constituted by the multidimensional Schrödinger equation and the Poisson equation [70].

In this limit, all the scattering occurs at the contacts only and this can be modeled by assuming that the contacts inject a thermalized distribution into the device, consistent with their Fermi levels, as in the Landauer formalism [43]. All interference and quantization effects in the active device region are explicitly included in this treatment. Frequently, open boundaries are treated using the quantum transmitting boundary method (QTBM), due to Lent and Kirkner [73], in which the transition from device eigenstates to traveling wave like states in the leads is carried out elegantly.

The Schrödinger-Poisson system has been investigated extensively with respect to semiconductor devices. Use of the effective mass approximation in conjunction with this was pioneered by Stern [74] for an investigation of inversion layer physics. Inversion layer quantization and barrier repulsion in a MOSFET [75], thin-film SOI devices [76], double gate devices [77], resonant tunneling diodes [78], single band tunneling [79], quantum wells [80] etc. have all been explored in the last decade using these minor variations of these methods. Classical drift-diffusion simulators often use an intermediate Schrödinger-Poisson solver [81] in regions where quantum effects are important, to correct the charge densities and potential in the grid. Typically this is done only in one-dimensional slices in the device for computational ease by making use of the inherent anisotropy in the devices.

1.5 QUANTUM TRANSPORT – MACROSCOPIC APPROACHES

The previous section dealt with the various levels of detail at which statistical transport calculations can be carried out. All the methods described have the commonality of being very expensive computationally for use in practical engineering design situations.

1.5.1 Density-Functional Theory

Prior to the 1960s it was believed that a quantum mechanical description of a many-body system, necessarily involved an explicit or implicit solution of the many-body Schrödinger equation. All approximate treatments were believed to originate in the inability to render an exact solution to this equation and the consequent requirement to make assumptions on their nature.

In a series of path breaking papers [82-84], Walter Kohn and co-workers showed in the early 1960s that this viewpoint is unnecessary. They proved that the single particle density, in the quantum mechanical many-body ground state can be viewed as a fundamental quantity and that *all* observables in this ground state can be expressed as *universal* functionals of this density [82]. This shift in emphasis from the solution of the many-body Schrödinger equation, which is a function of all the N coordinates, to a single-particle density, which is a function of position only, represents a tremendous theoretical triumph and simplifies the numerical calculation of ground state properties. This new formalism, with the density as a fundamental variable and all other quantities expressed as functionals has come to be known as Density Functional theory (DFT).

In DFT, the problem is shifted from one of determining the eigenfunctions of the many-body Hamiltonian, to determining the universal functionals that represent the self-consistent energies due to Coulombic interactions as well as due to exchange and correlation. The approximations made in DFT are all due to the unknown nature of these universal functionals and in practice, empirical assumptions are made for these based on one of several available treatments for specific cases [85] – viz. for the homogeneous electron gas (the local density approximation – LDA), or the slowly varying inhomogeneous gas (generalized gradient approximation – GGA) etc. These

are in turn conveniently interpreted as equations of state for the quantum mechanical system at zero temperature.

Approximate extensions of DFT to include degeneracy, excited states [88] etc. are known and are now standard. DFT is now, the tool of choice for investigation of electronic properties of many-body systems. The correlation and exchange energy contributions in DFT are only of importance for very high densities – they will therefore be crucial for an understanding of metals but are relatively unimportant for semiconductors where the carrier concentrations do not often exceed 10^{20} cm^{-3}.

Since the many-body ground state, is one coherent pure state (or several pure states at the same energy when there is degeneracy) and is theoretically, *exactly* described using the macroscopic single particle density, it is obvious that there is no contradiction between the two independent facets of quantum effects on the one hand and a macroscopic description on the other. DFT is however, in the rigorous sense, only applicable to zero temperature systems, and has no flow or transport processes included in it explicitly.

1.5.2 Thomas-Fermi Theory

The best known semiclassical approximation to quantum mechanics is the Thomas-Fermi approximation [85,86], which is an equilibrium theory obtained under the assumption of a slowly varying potential (and thus a slowly varying density). This can be viewed as a special case of DFT where the many-body system is described by the occupation of single particle eigenstates and exchange and correlation effects are ignored. The Thomas-Fermi relation between the potential and the density hinges at the core of the theory,

$$n = \frac{8\pi}{3h^3}(2mq)^{3/2}\left[E_F - V(r)\right]^{3/2} \qquad (1.5.1)$$

The relation is obtained by considering a homogeneous electron gas (i.e. in a constant potential) and then allowing the potential to vary slowly in space. The above relation is equivalent to assuming a kinetic energy functional in terms of the density as determined from the homogeneous electron gas, in the DFT, viz.

$$T_{TF}[n] = C_1 n(r)^{5/3} \tag{1.5.2}$$

where C_1 is a constant.

The Thomas-Fermi theory has been phenomenally successful in approximating the solutions to many-body problems in quantum mechanics [87]. Originally proposed for an understanding of the electronic structure of atoms, it has since been used for an understanding of the properties of systems as diverse as molecules, solids, and even stars. Several material properties influenced by charge carriers including the dielectric constant, screening etc. [87,88] have been investigated in this approximation, in the linear response regime.

The von-Weizsäcker correction [89] to the kinetic energy functional of Thomas-Fermi theory contains gradient corrections due to the inhomogeneities in the density, and has been shown to yield excellent values for electronic states in heavy atoms.

$$T_{TF-W}[n] = C_1 n(r)^{5/3} + C_2 \frac{|\nabla n(r)|^2}{n(r)} \tag{1.5.3}$$

C_1 is the same as in Thomas-Fermi theory, and C_2 is frequently considered an adjustable parameter [90] to fit experimental results for the ground state properties.

Though strictly an equilibrium theory, the Thomas-Fermi equation of state above has been used in hydrodynamic descriptions of transport at low temperatures

[91,92], by considering the above relation for the kinetic energy density as the requisite closure relation required. This leads to a quasi-static generalization of the density-functional theory.

It can be seen that the Thomas-Fermi relations (1.5.3) above, are derivable (at zero temperature) from the following assumption for the equilibrium Wigner function,

$$f_w(r,p) = \frac{1}{1+\exp\left[\beta\left(\frac{p^2}{2m}+V(r)-E_F\right)\right]} \tag{1.5.4}$$

The above function can be obtained formally from a complete quantum description as the lowest order term in the expansion [94] for the equilibrium Wigner function in the Planck's constant. Higher order approximations to the equation of state above can be made for finite (low) temperature by using the Sommerfield expansion [17] for the Fermi function above. These establish the fundamental derivation of Thomas-Fermi theory from a statistical description. The above equation will reduce to the Boltzmann distribution function for high temperature and will therefore yield the very familiar classical equation of state (1.3.5) for an ideal gas and thus the drift-diffusion equations when used in the hydrodynamic equations.

1.5.3 Macroscopic Quantum Transport Approaches

The descriptions of the statistical approaches to quantum transport in previous sections have repeatedly stressed the tremendous computational cost involved in these calculations, particularly in multiple dimensions. These are in addition to the difficulties involved in describing the various inputs to these transport descriptions, viz. the self-energies, scattering functions and open boundaries, at the microscopic levels. This is especially owing to the fact that it is very difficult to calibrate these

microscopic quantities from experiments (as compared to macroscopic parameters such as mobility or effective mass for instance), which by their very nature, usually yield aggregate information.

The great success of DFT and Thomas-Fermi theory must be seen as an encouragement for attempts to treat *transport* with quantum effects at a macroscopic level, i.e. with densities and current densities as basic variables rather than going through the detailed nature of occupied states. By sacrificing detail in the treatment, one can hope to achieve a description that is numerically easy to implement and can be solved for in computational times that are reasonable for use in engineering design. We would like to be able to describe the lowest order quantum effects of quantization and tunneling within the formulation. Obviously, these ideas must be viewed in the spirit of similar macroscopic theories – e.g. the drift-diffusion formalism. Similar to that theory, several parameters will have to be introduced that model the average effect of energy shifts due to quantization etc. (which cannot be obtained using a description using just the first two moment equations since the energy information is not available). Such models will be the focus of our attention for the rest of the thesis.

1.6 DISSERTATION OVERVIEW

In this dissertation we make a case for macroscopic modeling of transport including quantum effects. The background and the motivation for this have been established in the current chapter which provided a brief description of the various statistical approaches to quantum transport.

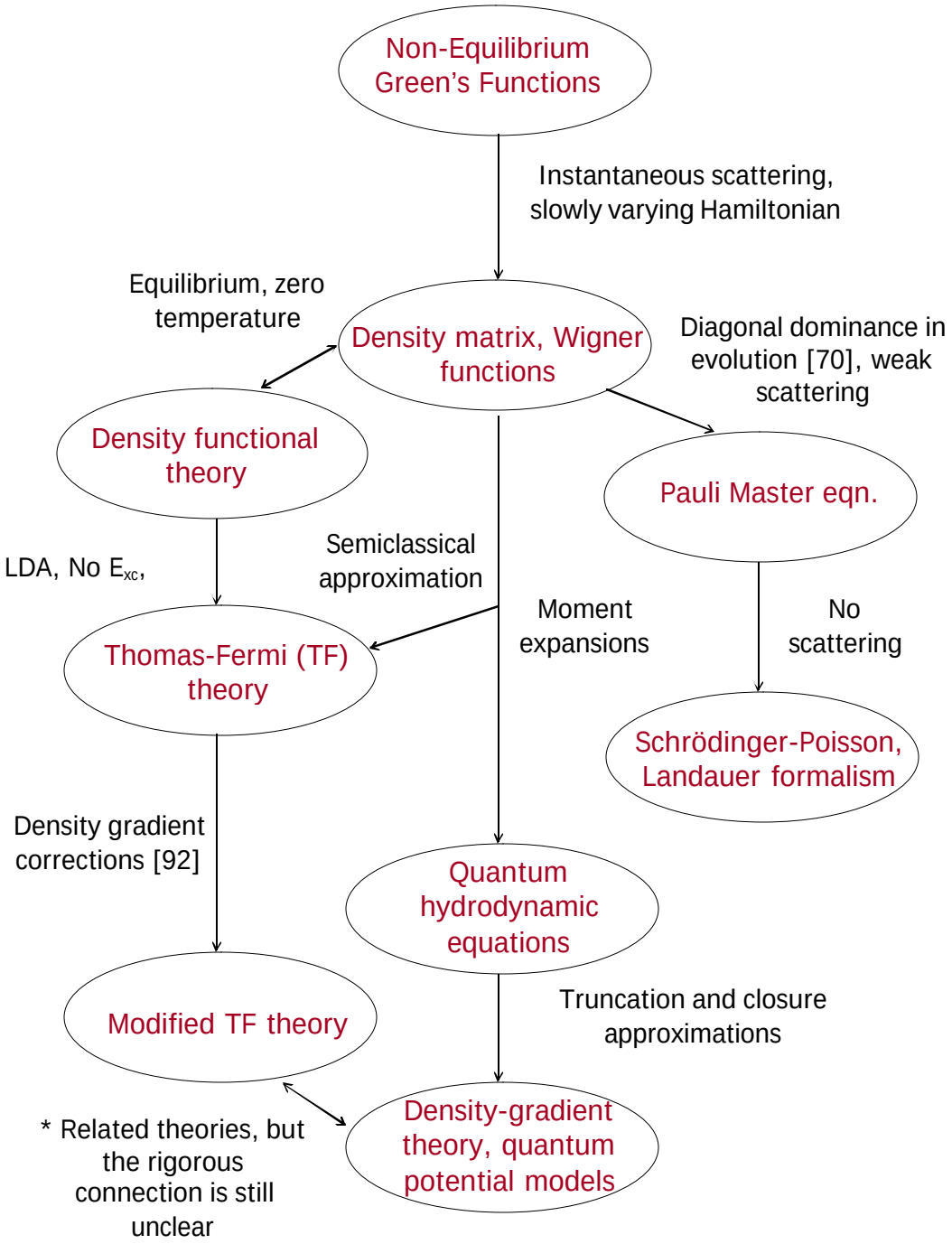

Figure 1.2 A representation of the quantum transport theories at different levels of detail, from the most detailed (NEGF) to the macroscopic, density-gradient and quantum potential models.

Chapter 2 introduces the quantum hydrodynamic equations obtained from a moment expansion of the Wigner function or density matrix equation of motion (1.4.20). The problem of closure of the hydrodynamic hierarchy is first discussed and the necessity of understanding the equilibrium ensembles to aid in this closure is shown. Some approximations for equilibrium and the corresponding transport equations – the density gradient method and the smooth quantum potential model, that they lead to are derived followed by a discussion on their applicability to situations prevalent in common semiconductor devices.

Chapter 3 discusses the description of quantization effects in the inversion layer of a MOSFET and potential wells using the density-gradient method. The boundary layer structure of these equations is examined and comparisons to one-electron effective mass quantum mechanics are made.

Chapter 4 presents a description of the single band tunneling process, a completely quantum mechanical phenomenon in purely macroscopic terms. This represents a significant achievement of this dissertation that shows conclusively that an average description in terms of densities and currents of purely quantum mechanical transport phenomena is possible.

Chapter 5 discusses the main contributions of this dissertation and provides several suggestions for future research that can be based upon this dissertation.

BIBLIOGRAPHY

[1] J.S. Kilby, "Invention of the integrated circuit", *IEEE Trans. Electron Devices,* vol. ED-23, no. 7, pp. 648-654, July 1976.

[2] W. Shockley, "The path to the conception of the junction transistor", *IEEE Trans. Electron Devices,* vol. ED-23, no. 7, pp. 597-620, July 1976.

[3] D. Kahng, "A historical perspective on the development of MOS transistors and related devices", *IEEE Trans. Electron Devices,* vol. ED-23, no. 7, pp. 655-657, July 1976.

[4] F. M. Wanlass and C.T. Sah, "Nanowatt logic using field-effect metal-oxide semiconductor triodes", *IEEE ISSCC Dig. Tech. Pap.*, pp. 32-33, 1963.

[5] G.E. Moore, "Cramming more components onto integrated circuits", *Electronics,* vol. 38, no. 8, pp. 114-117, April 1965.

[6] G. Baccarani, M.R. Wordeman and R.H. Dennard, "Generalized scaling theory and its application to a ¼ micron MOSFET design", *IEEE Trans. Electron Devices,* vol. ED-31, pp. 452-462, April 1984.

[7] M. Song, K.P. MacWilliams, and J. C.S. Woo, "Comparison of NMOS and PMOS hot carrier effects from 300 to 77 K", *IEEE Trans. Electron Devices*, vol. ED-44, no. 2, February 1997.

[8] F. Assederaghi, P.K. Kop and C. Hu, "Observation of velocity overshoot in silicon inversion layers", *IEEE Electron. Dev. Letters,* vol. 14, Issue 10, pp. 484-486, October 1993.

[9] K. Taniguchi, M. Yamaji, K. Sonoda, T. Kunikiyo, C. Hamaguchi, "Monte Carlo study of impact ionization phenomena in small geometry MOSFETs", *IEDM Technical Digest,* , pp. 355-358, 11-14, Dec. 1994.

[10] J. Sune, P. Olivo, B. Ricco, "Quantum-mechanical modeling of accumulation layers in MOS structure", *IEEE Trans. Electron Devices,* vol. ED-39, no. 7, pp. 1732-1739, July 1992.

[11] B. Majkusiak, "Gate tunnel current in MOS transistors", *IEEE Trans. Electron Dev.*, vol. 37, no. 4, pp. 1087-1092, April 1990.

[12] F. Reif, "Statistical and thermal physics", McGraw-Hill publications, 1965.

[13] R. L. Liboff, "Kinetic Theory – Classical, Quantum and Relativistic Descriptions", Springer, 2003.

[14] H. Spohn, "Kinetic equations from Hamiltonian dynamics", *Rev. Mod. Phys,* vol. 53, no. 3, pp. 569-614, 1980.

[15] I. V. Tokatly and O. Pankratov, "Hydrodynamics beyond local equilibrium : Application to an electron gas", *Phys. Rev. B,* vol. 62, no. 4, pp. 2759-2771, July 2000.

[16] C. Wolfe, N. Holonyak, Jr., and G. E. Stillman, "Physical properties of semiconductors", Prentice-Hall Inc., 1989.

[17] N. W. Ashcroft and N.D. Mermin, "Solid state physics", Saunders College Publishing, 1976.

[18] G. H. Wannier, "Structure of electronic excitation levels in insulating crystals", *Physics Rev.*, vol. 52, pp. 191-197, 1937.

[19] J. C. Slater, "Electrons in perturbed periodic lattices", *Phys. Rev,* vol. 76, no. 11, pp. 1592-1601, 1949.

[20] K. Hess, "Advanced theory of semiconductor devices", IEEE Press, 2000.

[21] R. Balescu, "Equilibrium and non-equilibrium statistical mechanics", John Wiley and Sons, Inc. 1975.

[22] E. Merzbacher, "Quantum Mechanics", Wiley, New York, 1998.

[23] M.C. Vecchi, M. Rudan, “Modeling electron and hole transport with full band structure effects, by means of a spherical harmonics expansion of the BTE”, *IEEE Trans. Electron Dev.,* vol. ED-45, no. 1, pp. 230-238, January 1998.

[24] M.V. Fischetti, S. E. Laux, P. M. Solomon and A. Kumar, “Thirty years of Monte Carlo simulations of electronic transport in semiconductors : their relevance to science and to mainstream VLSI technology”, *Journal of Computational Electronics,* Kluwer academic publishers, vol. 4, 2004.

[25] S.E. Laux and M.V. Fischetti, “Monte Carlo simulation of submicrometer Si-n MOSFETs at 77 and 300 K”, *IEEE Electron Dev. Letters,* vol. 9, no. 9, pp. 467-469, September 1988.

[26] M.V. Fischetti, “Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures. I. Homogeneous transport”, *IEEE Trans. Electron Dev.,* vol. 38, no. 3, pp. 634-649, March 1993.

[27] M.V. Fischetti and S.E. Laux, “Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures. II. Submicrometer MOSFETs”, *IEEE Trans. Electron Dev.,* vol. 38, no. 3, pp. 650-660, March 1993.

[28] M.V. Fischetti and S.E. Laux, “Long range Coulomb interactions in small Si devices. Part I : Performance and reliability”, *Journal of Appl. Phys.,* vol. 89, no. 2, pp. 1205-1230, January 2001.

[29] M.V. Fischetti, “Long range Coulomb interactions in small Si devices. Part II : Effective electron mobility in thin-oxide structures”, *Journal of Appl. Phys.,* vol. 89, no. 2, pp. 1231-1250, January 2001.

[30] M. V. Fischetti, D. A. Neumeyer and E. A. Cartier, “Effective electron mobility in silicon inversion layers in metal-oxide-semiconductor systems with a high

kappa insulator : The role of remote phonon scattering", *Journal of Appl. Phys.*, vol. 90, no. 9, pp. 4587-4608, November 2001.

[31] M.V. Fischetti and S.E. Laux, "Band structure, deformation potentials and carrier mobility in strained-Si, Ge and SiGe alloys", *Journal of Appl. Phys.,* vol. 80, no. 4, pp. 2234-2252, May 1996.

[32] F. Gamiz and M.V. Fischetti, "Monte Carlo simulation of silicon-on-insulator inversion layers : The role of volume inversion ", *Journal of Appl. Phys.,* vol. 89, no. 10, 5478-5487, May 2001.

[33] M.V. Fischetti, S.E. Laux, and E. Crabbe, "Understanding hot electron transport in silicon devices : Is there a shortcut?", *Journal of Appl. Phys.,* vol. 78, no. 2, pp. 1058-1087, July 1995.

[34] B. Winstead, U. Ravaioli, "Simulation of Schottky barrier MOSFETs with a coupled quantum injection/Monte Carlo technique", *IEEE Trans. Electron Dev.*, vol. ED-47, no. 6, pp. 1241-1246, June 2000.

[35] E. Kreyszig, "Advanced Engineering Mathematics", Chapter 23, John Wiley and Sons Inc., 1983.

[36] S. Selberherr, "Analysis and simulation of semiconductor devices", Springer-Verlag, 1984.

[37] W. Hansch, "The drift-diffusion equation and its applications in MOSFET modeling", Springer-Verlag, 1991.

[38] T. Grasser, T.-W. Tang, H. Kosina and S. Selberherr," "A review of energy-transport and hydrodynamic models for semiconductor device simulation", *Proc. of the IEEE*, vol. 91, no. 2, February 2003.

[39] T.Grasser, H. Kosina, M. Gritsch and S. Selberherr, "Using six moments of Boltzmann's transport equation for device simulation", *Journal of Appl. Phys.*,vol. 90, no. 5, pp. 2389-2396, September 2001

[40] T. Grasser, A. Gehring and S. Selberherr, “Recent advances in transport modeling for miniaturized CMOS devices”, *Fourth IEEE international Caracas conference on devices, circuits and systems”*, no. D027, pp. 1-8, April 2002.

[41] D.J. Griffiths, “Introduction to Quantum Mechanics”, Prentice-Hall Inc. 1995.

[42] C. Jacoboni, “Quantum dynamics and representations” Chapter 1, in *Quantum Transport in Semiconductors,* Ed. D.K. Ferry and C. Jacoboni, Plenum Press, 1992.

[43] S. Datta, “Electronic Transport in Mesoscopic Systems”, Cambridge University Press, 1999.

[44] A. L. Fetter and J. D. Walecka, “Quantum theory of many-particle systems”, McGraw-Hill publications, 1971.

[45] S. Datta, “Nanoscale device modeling : The Green’s function method”, *Superlattices and Microstructure*, vol. 28, no. 4, pp. 253-278, July 2000.

[46] Z. Ren, R. Venugopal, S. Datta, M. Lundstrom, “Examination of design and manufacturing issues in a 10 nm double gate MOSFET using the nonequilibrium Green’s function simulation”, *IEDM Tech. Dig,* 2001.

[47] R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom, “Quantum mechanical analysis of channel access geometry and series resistance in nanoscale transistors”, *Journal of Appl. Phys,* vol. 95, no. 1, pp. 292-302, 2004.

[48] R. Venugopal, M. Paulsson, S. Goasguen, S. Datta and M. S. Lundstrom, “A simple quantum mechanical treatment of scattering in nanoscale transistors”, *Journal of Appl. Phys,* vol. 93, no. 8, pp. 4628-4633, April 2003.

[49] A. Svizhenko, M.P. Anantram, T. R. Govindan, B. Biegel and R. Venugopal, “Two dimensional quantum mechanical modeling of nanotransistors”, *Journal of Appl. Phys,* vol. 91, no. 4, pp. 2343-2354, February 2002.

[50] R. Venugopal, Z. Ren, S. Datta M. S. Lundstrom and D. Jovanovic “Simulating quantum transport in nanoscale transistors : Real versus mode space approaches”, *Journal of Appl. Phys.,* vol. 92, no. 7, pp. 3730-3739, October 2002.

[51] V. K. Arora, “Nonlinear quantum transport in semiconductors”, *Journal of Appl. Phys.,* vol. 54, no. 2, pp. 824-830, February 1983.

[52] J. B. Krieger and G. J. Iafrate, “ Quantum transport for Bloch electrons in spatially homogeneous electric field”, *Journal of Appl. Phys.,* vol. 35, no. 18, 9644-9658, June 1987.

[53] J. B. Krieger and G. J. Iafrate, “Quantum transport for Bloch electrons in inhomogeneous electric fields”, *Journal of Appl. Phys.,* vol. 40, no. 9, September 1989.

[54] G. J. Iafrate, “Quantum Transport in Solids : The Density Matrix”, Chapter 4, in *Quantum Transport in Semiconductors,* Ed. D.K. Ferry and C. Jacoboni, Plenum Press, 1992.

[55] E. P. Wigner, “On the quantum correction for thermodynamic equilibrium”, *Phys. Rev.,* vol. 40, pp. 749-759, 1932.

[56] J.E. Moyal, “Quantum mechanics as a statistical theory”, *Cambridge Phil. Soc.,* vol. 45, pp. 99, 1949.

[57] W. R. Frensley, “Wigner function modeling of a resonant tunneling semiconductor device”, *Phys. Rev. B,* vol. 36, no. 3, pp. 1570-1580, July 1987.

[58] W. R. Frensley, “Boundary conditions for open quantum systems driven far from equilibrium”, *Rev. Mod. Phys,* vol. 62, pp. 745-791, July 1990.

[59] N. C. Kluksdahl, A. M. Kriman, C. Ringhofer and D. K.Ferry, “Self-consistent study of the resonant tunneling diode”, *Phys. Rev. B,* vol. 39, no. 11, pp. 7720-7735, April 1989.

[60] C. L. Fernando and W. R. Frensley, “Intrinsic high-frequency characteristics of tunneling heterostructures devices”, *Phys. Rev. B,* vol. 52, no. 7. pp. 5092-5104, August 1995.

[61] K. L. Jensen and F. A. Buot, “Numerical simulation of intrinsic bistability and high-frequency current oscillations in resonant tunneling structures”, *Phys. Rev. Lett.*, vol. 66, no. 8., pp. 1078-1081, February 1991.

[62] J. J. Shih, H. C. Huang, and G. Y. Wu, “Effect of mass discontinuity in Wigner theory of resonant-tunneling diodes”, *Phys. Rev. B,* vol. 50, no. 4, pp. 2399-2406, July 1994.

[63] L. Shifren and D. K. Ferry, “Particle Monte-Carlo simulation of Wigner function tunneling”, *Physics Letters A,* no. 285, pp. 217-221, 2001.

[64] M. Nedjalkov, R. Kosik, H. Kosina and S. Selberherr, “Wigner transport through tunneling structures – scattering interpretation of the potential operator”, *Proc. Simulation of Semiconductor Processes. and Devices,* pp. 187-190, Japan 2002.

[65] M. Nedjalkov, H. Kosina, S.Selberherr, C. Ringhofer and D. K. Ferry, “Unified particle approach to Wigner-Boltzmann transport in small semiconductor devices”, *Phys. Rev. B,* vol. 70, 115319, pp. 9-16, 2004.

[66] L. van Hove, “The approach to equilibrium in quantum statistics – A perturbation treatment to general order”, *Proc. International Symposium on transport processes in statistical mechanics,* Part IX, pp. 219-228, August 27-31, 1956.

[67] M.V. Fischetti, “Theory of electron transport in small semiconductor devices using the Pauli Master equation”, *Journal of Appl. Phys,* vol. 83, no. 1, pp. 270-291, January 1998.

[68] M.V. Fischetti, “Master equation approach to the study of electronic transport in small semiconductor devices”, *Journal of appl. Phys.*, vol. 59, no. 7, pp. 4901-4917, February 1999.

[69] M.V. Fischetti, S.E. Laux, A. Kumar, “Simulation of quantum electronic transport in small semiconductor devices : A master equation approach”, *IEDM Tech. Dig.*, pp. 19.3.1-19.3.4, 2003.

[70] S. E. Laux, A. Kumar and M. V. Fischetti, “Analysis of quantum ballistic electron transport in ultra-small silicon devices, including space-charge and geometric effects”, *Journal of Appl. Phys.*, vol. 95, no. 10, pp. 5545-5582, May 2004.

[71] S. E. Laux, A. Kumar and M. V. Fischetti, “Ballistic FET modeling using QDAME : Quantum Device Analysis by Modal Evaluation” *IEEE Trans. on Nanotechnology*, vol. 1, no. 4, pp. 255-260, December 2002.

[72] A. Kumar, J. Kedzierski, S. E. Laux, “Quantum based simulation analysis of scaling in ultrathin body device structures”, *Trans. on Electron Devices*, vol. ED-52, no. 4, pp. 614-617, April 2005.

[73] C. S. Lent and D. J. Kirkner, “The quantum transmitting boundary method”, *Journal of Appl. Phys.*, vol. 67, no. 10, pp. 6353-6359, May 1990.

[74] F. Stern, “Self-consistent results for n-type Si inversion layers” *Phys. Rev B*, vol. 5, no. 12, pp. 4891-4900, June 1972.

[75] C. Moglestue, “Self-consistent calculation of electron and hole inversion charges at silicon-silicon dioxide interfaces”, *Journal of Appl. Phys*, vol. 59, no. 9, pp. 3175-3183, May 1986.

[76] S. Jallepalli, J. Bude, W. K. Shih, M. R. Pinto, C. M. Maziar and A. F. Tasch Jr., “Electron and hole quantization and their impact on deep submicron silicon p

and n MOSFET characteristics", *IEEE Trans. Electron Dev.*, vol. 44, no. 2, pp. 297-303, February 1997.

[77] B. Majkusiak, T. Janik and J. Walczak, "Semiconductor thickness effects in the double gate SOI MOSFET", *IEEE Trans. Electron Dev.*, vol. 45, no. 5, pp. 1127-1134, May 1998.

[78] A. Schenk and A. Wettstein, "Simulation of DGSOI MOSFETs with a Schrödinger-Poisson based mobility model", *SISPAD 2002*, pp. 21-24, September 2002.

[79] A. Pirovano, A. L. Lacaita, and A. S. Spinelli, "Two dimensional quantum effects in nanoscale MOSFETs", *IEEE Trans. Electron Dev.*, vol. 49, no. 1, January 2002.

[80] K. Y. Kim and B. Lee, "Self consistent calculation of Schrodinger-Poisson equation including parallel perpendicular kinetic energy coupling effects in semiconductor quantum wells", *Proc. COMMAD*, pp. 383-386, 2000.

[81] A. S. Spinelli, A. Benvenuti and A. Pacelli, " Self-consistent 2-D model for quantum effects in NMOS transistors", *IEEE Trans. Electron Dev.*, vol. ED-45, pp. 1342-1349, 1998.

[82] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas", *Phys. Rev* vol. 36, no. 3B, 864-871, November 1964.

[83] W. Kohn and L .J. Sham, "Self-consistent equations including exchange and correlation effects", *Phys. Rev.* vol. 140, no. 4A, pp. 1133-1139, November 1965.

[84] L. J. Sham and W. Kohn, "One-particle properties of an inhomogeneous and interacting electron gas,", *Phys. Rev*, vol. 145, no. 2, pp. 561-567, May 1966.

[85] N. H. March and B. M. Deb (Ed), "The single-particle density in physics and chemistry", Academic Press, 1987.

[86] J. C. Slater and H. M. Krutter, “The Thomas-Fermi method for metals”, *Phys. Rev*, vol. 47, pp. 559-568, 1935.

[87] D. M. Newns, “Dielectric response of a semi-infinite degenerate electron gas”,*Phys. Rev. B,* vol. 1, no. 8, April. 1970.

[88] H. Payne, “Screening of a foreign charge by an electron gas”, *Phys. Rev. B*, vol. 1, no. 9, pp. 3645-3648, May 1970.

[89] N. H. March, in ““Theory of the inhomogeneous electron gas”, Edited by S.Lundquist and N. H. March, pp. 189, Plenum press, NY, 1983.

[90] E. Zaremba and H. C. Tso, “Thomas-Fermi-Dirac-von Weiszacker hydrodynamics in parabolic wells”, *Phys. Rev. B*, vol. 49,no. 12,8147-8163, March 1994.

[91] A. Eguiluz, S. C. Ying and J. J. Quinn, “Influence of the electron density profile on surface plasmons in a hydrodynamic model”, *Phys. Rev. B*, vol. 11, no. 6, pp. 2118-2121, March, 1975.

[92] S. Lundquist, “Theory of the inhomogeneous electron gas”, Edited by S.Lundquist and N. H. March, pp. 189, Plenum press, NY, 1983.

[93] M. Brack and R. K. Bhaduri, “Semiclassical Physics”, Addison-Wesley publishing company Inc., 1997.

CHAPTER 2

QUANTUM HYDRODYNAMIC TRANSPORT MODELS

2.1 INTRODUCTION

The previous chapter gave a panoramic view of the various levels of detail at which one can model the transport properties of semiconductor devices. Microscopic approaches starting from distribution functions and statistical mechanics can yield a wealth of information about the device behavior at the cost of significant computational complexity. More often than not however one is interested in modeling the aggregate behavior of devices (the terminal characteristics) with acceptable accuracy for use in engineering design. Macroscopic models derived from moment expansions of the Boltzmann transport equation, such as the drift-diffusion model and the various energy transport and hydrodynamic models have fulfilled this requirement admirably in regimes where transport is predominantly classical. In contrast there are no such general models available that can describe quantum transport fully consistently even in a low energy transport regime in spite of considerable research efforts. This chapter takes a close look at the derivation of the quantum hydrodynamic models that have been suggested to bridge this gap.

There are two major ingredients that are necessary in any consistent microscopic transport theory. The first is an adequate description of equilibrium – i.e. one must have such information about the equilibrium state as will allow us to calculate the average values of the observables that we are interested in. The second is a model for how this equilibrium distribution function evolves under the action of external forces, both deterministic and stochastic (the latter are what lead to irreversibility in the transport process).

2.2 EQUILIBRIUM DISTRIBUTION FUNCTIONS

Since Newton's law is a second order differential equation, the complete description of the classical state of a single particle requires the statement of its position and momentum at some point in time (i.e. the location of the particle in phase space). Hence for an adequate statistical description of the state of an ensemble of particles we require information about their distribution in phase space. In quantum mechanics, the definition of a phase space is not intuitive because of the non-commuting nature of the space and momentum operators and has to be derived based on principles of quantum-classical correspondence. However, one can recognize that for an adequate quantum statistical description, we need to know both the complete set of Hamiltonian eigenstates and the occupation probabilities of these states.

The statistical mechanics of large systems in equilibrium, through notions of the canonical and the grand-canonical ensemble yield the very general concepts of temperature, chemical potential and the partition function. These are as valid for quantum mechanical systems as for classical ones. Through these we get the distributions *in energy* that have to be satisfied by the individual microsystems (e.g. the particles) constituting the large system. For the case of carriers in semiconductors these will specialize to the Boltzmann distribution for a non-degenerate carrier gas and the Fermi-Dirac distribution function for the degenerate case when the exclusion principle is important.

$$f(E_k) = \frac{\exp(-\beta E_k)}{\sum_k \exp(-\beta E_k)} = \frac{\exp(-\beta E_k)}{Z} \quad \text{- Boltzmann Distribution}$$

$$f(E_k) = \frac{1}{1+\exp[\beta(E_F - E_k)]} \quad \text{- Fermi-Dirac Distribution}$$

In the above equations, the quantity β is the inverse temperature $1/k_B T$ and Z represents the partition function given by the usual sum over available energy states.

For the case of classical carriers, the above are readily translated into phase space distributions, by trivially noting that the energy is the sum of the potential and kinetic terms and is essentially a continuum, e.g. the Boltzmann distribution yields the Maxwell distribution

$$f(r,p) = f\left(E = \frac{p^2}{2m} + V(r)\right) = \frac{\exp\left[-\beta\left(\frac{p^2}{2m} + V(r)\right)\right]}{Z} \qquad (2.2.1)$$

This equilibrium phase space distribution (or its Fermi-Dirac analogue) is used, implicitly or explicitly, in the derivation of all the various classical macroscopic transport models.

In quantum mechanics, we can define a system quite generally as an incoherent superposition of states, by providing the probabilities γ_I of the system being in one of a number of pure states $\{\varphi_I, I = 0,1\ldots\}$. Then, to get the expectation value of an observable A in this ensemble of particles we need a *two-stage* averaging process, i.e. we need the expectation value of the observable in each of the pure states (a coherent superposition of eigenvectors u_i in some basis) and then perform a weighted sum of these values over the occupation probabilities of the different states. In other words we will have, for some set of pure states $\{\varphi_I, I = 0,1\ldots\}$,

$$\begin{aligned} \langle A \rangle &= \sum_i \gamma_i \langle A \rangle_i = \sum_i \gamma_i \left\langle \varphi_i \middle| \hat{A} \middle| \varphi_i \right\rangle \\ \varphi_i &= \sum_j c_j^{(i)} u_i \end{aligned} \qquad (2.2.2)$$

where $c_j^{(i)}$ are the weighted coefficients for the eigenstates u_i. By simple manipulation, this can be written as the trace of the product of the operator A with a density operator ρ, which depends only on the state of the system (i.e. the states φ_I and the weights γ_i.)

$$\begin{aligned} \langle A \rangle &= Tr(\rho A) \\ \rho_{jj'} &= \sum_i \gamma_i \left[c_{j'}^{(i)*} . c_j^{(i)} \right] \end{aligned} \tag{2.2.3}$$

In equilibrium, the probability of occupation of the available states is known as a function of energy, and therefore the set of states φ_i can be taken to be the set of many-body Hamiltonian eigenstates $\{\psi_i , i = 0,1\ldots\}$ for which the energies are constant eigenvalues. Therefore, the density matrix in equilibrium becomes diagonal in the Hamiltonian representation and is,

$$\rho(x,x') = \frac{1}{Z} \sum_i f(E_i) \psi_i^*(x') \psi_i(x) = \frac{\exp(-\beta H)}{Tr[\exp(-\beta H)]} \tag{2.2.4}$$

Thus, to describe equilibrium we need the complete set of Hamiltonian eigenstates. The determination of the density matrix is the fundamental problem of equilibrium statistical mechanics [1].

We can now, examine the properties of the Weyl transform of the density matrix, i.e. the Fourier transform with respect to the difference coordinate η, in the difference coordinates defined as under,

$$\begin{aligned} r &\to \frac{x + x'}{2} \qquad \eta \to x - x' \\ f(r, p, t) &= \int_{-\infty}^{\infty} \rho(r, \eta, t) . \exp\left(-\frac{ip\eta}{\hbar} \right) d\eta \end{aligned} \tag{2.2.5}$$

This function was originally defined by Wigner in order to investigate the quantum corrections to the average energy of an inhomogeneous Maxwell gas in thermal equilibrium. One can immediately see from the properties of Fourier transforms that the following hold for the densities *n(R,t)* and *n(p,t)* in *R* and *p* space.

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} f(R,p,t)dp = n(R,t) \qquad (2.2.7)$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} f(R,p,t)dR = n(p,t)$$

The above two relations allow us to interpret the quantity *p* as a parameter corresponding to the classical momentum and the space *(R, p)* as a quantum phase space. The function *f* is called the Wigner function and is a generalization of the classical phase space distribution function to include the effects of quantum mechanics. The reader is referred to [2] for a detailed discussion of the properties of the Wigner function and the quantum phase space.

For a rigorous treatment of a system of charge carriers in semiconductors, the states $\psi_{i,}$ that we use in the equation must be the solutions of the full many-body Schrödinger equation. However, since the solution of this equation is intractable we can use the usual Hartree approximation (or the Hartree-Fock approximation, if we wish to include correlation and exchange effects), and convert the description to a single-particle description, in which case the Hamiltonian eigenstates become the single-particle eigenstates and the coordinates *x* and *x'* become the single particle space coordinates.

It is obvious that even in the above single-particle approximations the density matrix is not known *a priori* as an explicit, exact function of the space coordinates *x* and *x'*. The single-particle Hamiltonian, for instance, is given by the differential operator,

$$\hat{H} = \left[-\frac{\hbar^2}{2m^*}\nabla^2 + V(x) \right] \tag{2.2.8}$$

Therefore the density matrix in equilibrium is a functional of the self-consistent potential. Since the eigenstates are not generally known as explicit functionals of the potential, we cannot write down the density matrix very generally either. Moreover, the potential V in equilibrium is usually not known and has to be obtained self-consistently with the density matrix. This can be done numerically through a self-consistent iteration of the Schrödinger and Poisson equations. For the purposes of deriving macroscopic transport equations, however, we will need analytical forms of the density matrix (or the Wigner function) from which we can assume a linearized response for the non-equilibrium distribution function to calculate averages. We will therefore have to look for approximations to the density matrix in equilibrium under special assumptions on the potential.

For the non-degenerate case, definition of the density matrix in equilibrium can be shown to be equivalent to the following initial value equation (as can be directly verified from the Schrödinger equation) [8]

$$\begin{aligned} \frac{\partial \rho}{\partial \beta} &= -\left[\frac{\hbar^2}{4m}\left(\nabla_x^2 + \nabla_{x'}^2\right) \right]\rho + \frac{\left[V(x)+V(x')\right]}{2}\rho \\ \rho(x,x',\beta = 0) &= \delta^3(x-x') \end{aligned} \tag{2.2.9}$$

This equation (called the Bloch equation for the density matrix henceforth) can be solved numerically for the density matrix, instead of having to solve the Hamiltonian eigenvalue problem and then performing the sum for the density matrix. Using the

definition of the Wigner function we can easily show that the equivalent equation for the Wigner function is,

$$\frac{\partial f_w}{\partial \beta} = \left[\frac{\hbar^2}{8m} \nabla_R^2 - \frac{p^2}{2m} \right] f_w - \theta' f_w \tag{2.2.10}$$

$$\theta' f_w = \cos\left(\frac{\hbar}{2} \nabla_p \nabla_R \right) V(R) f_w$$

The initial condition for the density matrix reduces to the simple initial condition that indicates the absence of any quantum effects at infinite temperature (i.e. all states in phase space are occupied with equal probability),

$$f_w(R, p, \beta = 0) = 1 \tag{2.2.11}$$

The operator θ' is a pseudo-differential operator and should be understood in the sense of a series expansion of the cosine term, with the differential operators with respect to *p* acting only on the Wigner function and the operators with respect to *R* acting only on the potential. When the potential is not differentiable, or varies very sharply, this operator should be replaced by an integral expansion using the Green's function of the differential operator in (2.2.10) [3].

The equations (2.2.9) and (2.2.10) are very useful to derive approximations for the equilibrium Wigner function under particular assumptions on the potential. We will now look at typical expressions for the Wigner function and the density matrix for carrier gases. We will start with a free electron gas to get a feel for its distribution function, and then look at two specific approximations for the equilibrium distribution function that can be usefully employed in deriving transport equations. The first is a method that is based on a small parameter expansion due to Wigner (and later

Kirkwood [4]) while the second is based on the Born approximation of the Bloch equation for the density matrix. In addition to these there are also a separate series of approximations based on the "effective potential" methods [9,10] , which are intended to approximate the carrier density in the same form as in classical transport (i.e. in the Boltzmann form) but with modified expressions for the potential. These are however not directly relevant to what we will discuss in succeeding chapters and are hence not covered here.

2.2.1 Density matrix and Wigner function for a free carrier gas

For a free carrier gas, the Hamitonian eigenfunctions are given by,

$$\psi_k(x) = \frac{1}{\sqrt{(2\pi)^3}} \exp\left(i\vec{k}.\vec{x}\right) \tag{2.2.12}$$

The sum for the density matrix can be explicitly performed as an integral in *k*-space to yield the Gaussian in the difference variable *(x-x')*,

$$\rho(x,x') = \left(\frac{2m}{\hbar^2\beta}\right)^{\frac{3}{2}} \exp\left[-\left(\frac{x-x'}{\lambda_{th}}\right)^2\right] \tag{2.2.13}$$

This is also directly obtainable from the Bloch equation (2.2.9), since for this special case the initial value problem reduces to the heat equation. The Wigner function then becomes the classical Maxwellian distribution function through (2.2.5),

$$f_w(x,p) = \left(\frac{2m}{\hbar^2\beta}\right)^{\frac{3}{2}} \exp\left(-\frac{p^2}{2mk_BT}\right) \tag{2.2.14}$$

In the absence of an external potential the density of the gas is homogeneous in space. A free carrier gas behaves classically as seen from the above distribution function.

Quantum effects only become evident in the presence of variations in the potential, which translate to inhomogeneities in the density.

2.2.2 Wigner function in a slowly varying potential

In pure-state quantum mechanics, the analysis of the Schrödinger equation for a slowly varying potential plays a very important role as it allows one to write down the approximate wavefunctions and energy eigenvalues. For instance, the WKB approximation yields semiclassical wavefunctions that provide information on the near classical dynamics, as well as purely quantum mechanical phenomena (such as tunneling) to the leading order.

In the same spirit one can analyze equation (2.2.10) to get the lowest order deviations from the classical equilibrium phase space distribution function due to quantum mechanics. It is first convenient to cast the equation in a dimensionless form that is amenable to a perturbation series expansion solution. To do this, one can scale the space coordinate with a large length scale *d*, the potential by the thermal voltage, and the momentum by the thermal momentum (i.e. the momentum of a carrier with thermal energy k_BT) to get [5],

$$\frac{\partial f_w(x,p)}{\partial \beta} = \left(\varepsilon^2\nabla^2 - p^2\right) f_w - \theta' f_w \qquad (2.2.15)$$
$$\theta' f_w = \cos\left(\varepsilon \nabla_p \nabla_R\right) V(R) f_w$$

We have changed the label on the space coordinate of the Wigner function to *x* from *R* to be consistent with the labeling of the classical distribution function. Assuming that the parameter ε is small, one can now look for series expansions of the solution of the form,

$$f_w(x,p) = \sum_i \varepsilon^i f_i(x,p) \tag{2.2.16}$$

Since the parameter ε is zero for classical evolution, the first term *(i=0)* in the above expansion should be the classical response. Therefore the initial condition for the first term is identical to (2.2.11) while the rest of the terms in the expansion are zero at infinite temperature.

Substituting now the above expression in the differential equation and equating terms of each order ε, one gets for the zeroth-order solution, i.e., the classical distribution function.

$$\frac{\partial f_0}{\partial \beta} = -\left(p^2 + V\right) f_0 \quad \Rightarrow \quad f_0 = \exp\left[-\beta\left(p^2 + V\right)\right] \tag{2.2.17}$$

Since the cosine series has only even order terms in the expansion, the odd order terms in the expansion (2.2.16) drop out and the lowest-order quantum correction is given by the equation,

$$\frac{\partial f_2}{\partial \beta} = \nabla^2 f_0 - p^2 f_2 - \frac{1}{2}(\nabla_p \cdot \nabla_x)(\nabla_p \cdot \nabla_x) f_0 \tag{2.2.18}$$

This yields as the lowest order, purely quantum correction to the equilibrium phase space distribution,

$$f_2 = f_0\left[-\beta^2 \nabla^2 V + \frac{\beta^3}{3}\left(\left|\nabla V\right|^2 + 2p.\nabla\nabla V.p\right)\right] \tag{2.2.19}$$

Hence the total distribution function to this order in ε is given by

$$f_w(x,p) \approx \exp\left(-\beta\left(\frac{p^2}{2m} + V\right)\right)\left[1 - \frac{\hbar^2\beta^2}{8m}\left(\nabla^2 V - \frac{\beta}{3}\left|\nabla V\right|^2 - \frac{\beta}{3m}\, p.\nabla\nabla V.p\right)\right] \tag{2.2.20}$$

The above distribution function can also be written in a form that is more convenient and illuminating for the purposes of deriving the transport equations. We can also get a different expression for the "quantum potential" that leads naturally to a set of linearized transport equations. We can see that, for the above distribution function, the density is given by straightforward integration as,

$$n(x) \approx \exp(-\beta V)\left[1-\frac{\hbar^2\beta^2}{12m}\left(\nabla^2 V-\frac{\beta}{2}|\nabla V|^2\right)\right] \tag{2.2.21}$$

We therefore have,

$$\log(n) = -\beta V + \log\left[1-\frac{\hbar^2\beta^2}{12m}\left(\nabla^2 V-\frac{\beta}{2}|\nabla V|^2\right)\right] \tag{2.2.22}$$

Using the above relations to eliminate the potential in (2.2.20) by long algebraic division, we can write,

$$f_w(x,p) \approx n\exp\left(-\frac{p^2}{2mk_BT}\right)\left[1+\frac{\hbar^2\beta}{24m}\left(\nabla^2\log(n)-\frac{\beta}{m}\,p.\nabla\nabla\log(n).p\right)\right] \tag{2.2.23}$$

This can now be taken to represent a fundamental relationship between the distribution function and the density, as opposed to the distribution function and the potential. One can then linearize the transport equations about equilibrium by assuming that Eqn (2.2.23) holds locally even when the system is globally off equilibrium.

Eqns (2.2.20) and (2.2.23) yield very close asymptotic approximations to the equilibrium Wigner functions in regions where the potential is smooth. However, they may differ significantly near a large and sharply varying potential perturbation. On the other hand, as opposed to the expansion of the distribution function on the derivatives of the potential, Eqn (2.2.23), in yielding an expansion directly based on the density-gradient is fundamentally free of problems arising from the non-analyticity of the

potential near an abrupt barrier. The functional form of Eqn (2.2.23) simply requires the density to be twice differentiable, a condition which will generally hold, since the density will be smooth even at abrupt potential barriers. The curvature of the potential near such perturbations can be very large, but the curvature of the logarithm of the density will still not be very large.

Instead, what makes Eqn (2.2.23) inaccurate in the vicinity of an abrupt and large potential perturbation is not so much the largeness of the density gradient (or the gradient of the logarithm of the density to be precise), but that the assumed relationship between the potential and the density, i.e. Eqn (2.2.22), can be far off the mark. We need to examine the actual relationship in a more rigorous manner near such abrupt potential perturbations to find the appropriate form of the expansion for the distribution function, and whether Eqn (2.2.22) is sufficiently accurate to second order in Planck's constant.

2.2.3 Density Matrix for Small Potential Perturbations

Another form for the equilibrium distribution function can be obtained under a different assumption on the potential from Section 2.2.2. As mentioned earlier, for the free carrier case, the Bloch equation (2.2.9) for the density matrix reduces to a heat equation with the initial condition given by the uncorrelated density matrix at infinite temperature. This equation has a Gaussian in the difference coordinate as its solution for a finite temperature as derived in Section 2.2.1.

The Green's function of Eqn (2.2.9) in the absence of any potential is the same as the Green's function of the heat equation, i.e. a Gaussian. This Green's function can be used to write an implicit solution for the density matrix in equilibrium.

$$\rho(\beta, x, y) = \rho_0(\beta, x, y) - \frac{1}{2}\int_0^{\beta} d\beta' \int d^3x' d^3y' G(\beta, x, y; \beta', x', y')[V(x') + V(y')]\rho(\beta', x', y') \tag{2.2.24}$$

We can treat the above implicit solution as the basis of an iterative scheme for the density matrix. Then if we use the equilibrium free carrier solution (2.2.13) as the starting point, we can write the first iteration for the density matrix, the *Born approximation* [5], in the above by replacing the full solution ρ by the free-carrier solution ρ_o.

On algebraic simplification, the above solution can be written in terms of a smooth quantum potential V_q as,

$$\rho = \rho_0 \exp(-\beta V_q) \tag{2.2.25}$$

where V_q is a Gaussian smoothened version of the classical self-consistent equilibrium potential [5].

The form (2.2.25) of the density matrix, avoids the problems inherent in the derivation of Eqn (2.2.20) for sharply varying potentials. However, the Born approximation is only good for small absolute values of potential perturbations in Eqn (2.2.9). This requires that the potential energy perturbation must not be much larger than the thermal energy which is not practical for the large abrupt insulating barriers that are encountered in real devices.

2.3 QUANTUM HYDRODYNAMIC EQUATIONS

The quantum hydrodynamic equations (QHD) equations are obtained in a similar manner from the equation of motion of the Wigner function (1.4.22) as the classical hydrodynamic equations are obtained from the BTE, viz. by a method of moments to average out the information in momentum space. The first three moments

of the hierarchy are formally identical to the classical hydrodynamic equations (1.3.3). They are reproduced here for the sake of convenience [8]

$$\begin{aligned}
&\frac{\partial n}{\partial t} + \frac{1}{m}\nabla.\vec{\Pi} = 0 \\
&\frac{\partial \vec{\Pi}}{\partial t} + \nabla.(u\vec{\Pi} - \vec{\vec{P}}) = -n\nabla U - \frac{\Pi}{\tau_m} \\
&\frac{\partial W}{\partial t} + \nabla.(\vec{u}W - \vec{u}.\vec{\vec{P}} - \vec{Q}) = \vec{u}.\nabla U - \frac{(W - W_0)}{\tau_w}
\end{aligned} \tag{2.2.6}$$

The quantities appearing in these equations have the same meaning as in classical hydrodynamic modeling which was described in Section 1.3.2 – the dependent variables are *n* the carrier density, Π, the momentum density and *W,* the energy density. As in the case of the classical equations, the QHD equations are incomplete on truncation at any level and will have to be completed through appropriate closure relations.

The formal equivalence of the first three equations in classical and quantum transport implies that any quantum mechanical properties will have to enter *only* through the closure relations for truncating the hierarchy. As was seen in the context of the classical hydrodynamic equations, the closure will have to be performed assuming a form for the non-equilibrium distribution function.

2.4 DENSITY-GRADIENT THEORY

The density-gradient formalism is a generalization of the drift-diffusion formalism to include quantum effects. It is formally obtained by closing the hierarchy (2.2.6) at the second (momentum density) equation through a local equilibrium assumption for the stress tensor *P* and by ignoring the convective term *u*Π.

As has been pointed out in (1.3.4), the stress tensor can be defined in kinetic terms as an average on the non-equilibrium distribution.

$$\vec{\vec{P}} = \left\langle \frac{(\vec{p} - m^* u)(\vec{p} - m^* u)}{m^*} \right\rangle \tag{2.2.7}$$

The particular form of the stress tensor used in the density-gradient theory is derived from assuming the non-equilibrium distribution to be locally of the form of (2.2.3) with a final drift in the momentum space. This yields the following closure relation [6] for the stress tensor, which can be interpreted as an equation of state for the density-gradient gas.

$$\vec{\vec{P}} = -nk_B T + \frac{\hbar^2}{12m^*}\left[\nabla\nabla n - \frac{\nabla n \nabla n}{n}\right] \tag{2.2.8}$$

On closing the hierarchy with the above relation at the current equation, and after trivial manipulations of the resulting terms we arrive at the fundamental transport equation of the density-gradient theory in steady state.

$$\frac{J_n}{q} = n\mu_n F + D_n \nabla n - n\mu_n \left[2b_n \frac{\nabla^2 \sqrt{n}}{\sqrt{n}}\right] \tag{2.2.9}$$

A similar equation can be derived for holes. These transport equations have to be solved self-consistently with the continuity equation and the Poisson equation for the potential.

2.5 SUMMARY

This chapter examined in detail the derivation of the different equations used in the macroscopic treatment of quantum effects in semiconductor devices. The Bloch equation was introduced and the commonality between the two different formalisms, i.e. the density-gradient theory and the smooth quantum hydrodynamic model in terms of their derivation based on the *free carrier* Wigner function or the density matrix was repeatedly stressed. The next chapter will examine the behavior of density-gradient theory in more detail in situations where quantum effects are important but the conditions of their derivation are explicitly violated.

BIBLIOGRAPHY

[1] N. N. Bogolubov and N. N. Bogolubov Jr, “Introduction to quantum statistical mechanics”, World Scientific Publishing Company, 1982.

[2] J.E. Moyal, “Quantum mechanics as a statistical theory”, *Cambridge Phil. Soc.*, vol. 45, pp. 99, 1949.

[3] I. Oppenheim and J. Ross, “Temperature dependence of distribution functions in quantum statistical mechanics”, *Phys. Rev,* vol. 107, no. 1, pp. 28-32, July 1957.

[4] E. P. Wigner, “On the quantum correction for thermodynamic equilibrium”, *Phys. Rev.* vol. 40, pp. 749-759, June 1932.

[5] J. G. Kirkwood, “Quantum statistics of almost classical assemblies”, *Phys. Rev.* vol. 44, pp. 31-37, July 1933.

[6] D. J. Griffiths, “Introduction to quantum mechanics”, Prentice-Hall, 1995.

[7] M. G. Ancona and G. Iafrate, “Quantum correction to the equation of state of an electron gas in a semiconductor”, *Phys. Rev. B*, vol. 39, no. 13, May 1989.

[8] C. L. Gardner and C. Ringhofer, “Smooth quantum potential for the hydrodynamic model”, *Phys. Rev. E,* vol. 53, no. 1, January 1996.

[9] R.P. Feynmann and A. R. Hibbs, “Quantum mechanics and path integrals”, *McGraw Hill Publishers,* 1965.

[10] D. K. Ferry “The onset of quantization in submicron semiconductor devices”, *Superlattices and Microstructures,* vol. 27, pp. 61-66, 2000.

CHAPTER 3

MACROSCOPIC DESCRIPTION OF BOUNDARY LAYERS AND CONFINEMENT EFFECTS

3.1 INTRODUCTION

The previous chapter gave an account of the derivation of the density-gradient equations and the smooth quantum potential model and discussed the aspects of their derivation near large abrupt potential perturbations. These equations are derived from approximations to the full phase-space distribution function near such perturbations; but nevertheless, do exhibit a boundary layer behavior near such interfaces. Since we are mostly interested in the formulation to yield us aggregate quantities such as threshold voltage shifts and capacitances, they merit further investigation.

The density-gradient equations have been applied in several cases to the simulation of inversion layer physics [1-5] where their derivations make them of dubious validity. This chapter is devoted to assessing aspects of the boundary layer behavior of these equations near insulator interfaces and to gain an understanding of the level to which the macroscopic physics is approximated by these equations.

The methods employed in this chapter are simple, intuitive and semi-analytical in nature, and therefore an insight into the behavior can be obtained. We will first analyze the boundary layer behavior of the equations by expanding the solution to the equations in a singular perturbation series near the interfaces. We will then compare the asymptotic results to what is expected from quantum mechanics in analytically solvable cases. We will next consider the form of the DG solutions for confinement in a potential well. The emphasis here is on getting a clear idea of the *nature* of solutions obtained from the DG equations and their relation to one electron quantum mechanics.

3.2 DENSITY GRADIENT EQUATIONS

We will first briefly reproduce the density-gradient (DG)/Poisson equation system here in the form used in this chapter for the sake of completeness. We will be interested in static ($J_n = J_p = 0$) conditions in this chapter and will hence write down the equation forms relevant for this specialized case.

The general transport equation in DG theory, for electrons is (2.2.9),

$$\frac{J_n}{q} = n\mu_n F + D_n \nabla n - n\mu_n \left[2b_n \frac{\nabla^2 \sqrt{n}}{\sqrt{n}} \right] \qquad (3.2.1)$$

$$\frac{J_p}{q} = p\mu_n F - D_p \nabla p + p\mu_p \left[2b_p \frac{\nabla^2 \sqrt{p}}{\sqrt{p}} \right]$$

The coefficients b_n and b_p refer to the strength of the gradient effects in the electron and hole gases and F is the electric field, the negative gradient of the classical electrostatic potential. The assumed forms of b_n and b_p are,

$$b_n, b_p = \frac{\hbar^2}{4mrq} \qquad (3.2.2)$$

In the above, r can be shown to take the values of 1 for a low-temperature, high-density regime to the value of 3 for a high-temperature and low-density (nearly classical) regime. In the intermediate range, it is usually used as a phenomenological fitting parameter between the two limits [2].

The DG transport equation has to be solved for electrons and holes, self-consistently with the continuity equations and Poisson's equation in the semiconductor,

$$\frac{\nabla \cdot J_n}{q} = (G - R) \qquad (3.2.3)$$

$$\frac{\nabla \cdot J_p}{q} = -(G - R)$$

$$(\varepsilon_r \varepsilon_0)\nabla^2 V_s = q\left(n - p + N_A^- - N_D^+\right) \qquad (3.2.4)$$

For the static conditions that we consider here and for electrons, we will get,

$$nF + \frac{D_n}{\mu_n}\nabla n - n\left[2b_n \frac{\nabla^2 \sqrt{n}}{\sqrt{n}}\right] = 0 \qquad (3.2.5)$$

Assuming that the density is non-zero almost everywhere (except maybe on the boundaries, for strict confinement), we can integrate the above equation directly to yield a fundamental relation for the variation of the density.

$$-V_s + \frac{D_n}{\mu_n}\ln n - 2b_n \frac{\nabla^2 \sqrt{n}}{\sqrt{n}} = C \qquad (3.2.6)$$

where C is an integration constant. The above equation has to be interpreted as representing the constancy of a generalized electrochemical potential for electrons, under static conditions, defined by the left hand side of the expression. The constant C can be obtained from the value of this potential in a nearly classical region of the device where the gradient effects are not important and the density is assumed to be known (for instance at a contact).

The relationship between the diffusion coefficient (D_n) and the mobility (μ_n) is dependent on the conditions (degenerate or non-degenerate) of the electron gas and can be written as a function of the density in one of many available forms in the literature. For instance a particularly convenient one for numerical simulation is [6]

$$\frac{D_n}{\mu_n} = V_t\left[1 + 2a_1 z + \frac{a_2 a_3 (2 + a_3 z)}{(1 + a_3 z)^2}\right] \tag{3.2.7}$$

with, a_1 = 4.8967, a_2 = 0.1334, a_3 = 0.0450 and $z = n/N_c$ obtained from a Pade approximation of the Fermi integral.

A convenient way to write (3.2.6) is in terms of the auxiliary variable *s*, the square root of the density *n*.

$$-V_s + \frac{2D_n}{\mu_n}\ln s - 2b_n\frac{\nabla^2 s}{s} = C \tag{3.2.8}$$

3.3 BOUNDARY LAYERS IN DENSITY-GRADIENT THEORY

For a given potential profile, with an infinite (or large) abrupt barrier as a boundary condition, (3.2.6) will exhibit a boundary-layer behavior [7], as shown in Fig. 3.1. This owes to the fact that the highest order derivative of the density, the DG term, is multiplied by the small parameter b_n given by (3.2.2).

We will first examine the solution of Eqn. (3.2.6) in the vicinity of a simple abrupt potential barrier close to the flat-band condition. This situation is of great practical importance, since it is seen repeatedly in silicon MOS devices (the Si-SiO_2 interface) and in heterostructures. This will give us a good understanding of the leading order physics of barrier repulsion of carriers described macroscopically. For a slowly varying, smooth potential V_s (a more rigorous condition will be made explicit in the following section), the solutions that we discuss below will simply be modulated by the Boltzmann factor of the potential, in the spirit of the WKB approximation for the very similar Schrödinger equation. In fact, it is interesting to note that (3.2.8) is very similar to a Schrödinger equation, with the energy eigenvalue replaced by the non-linear logarithmic term and the kinetic energy operator replaced by the density-gradient coefficient ($2b_n$).

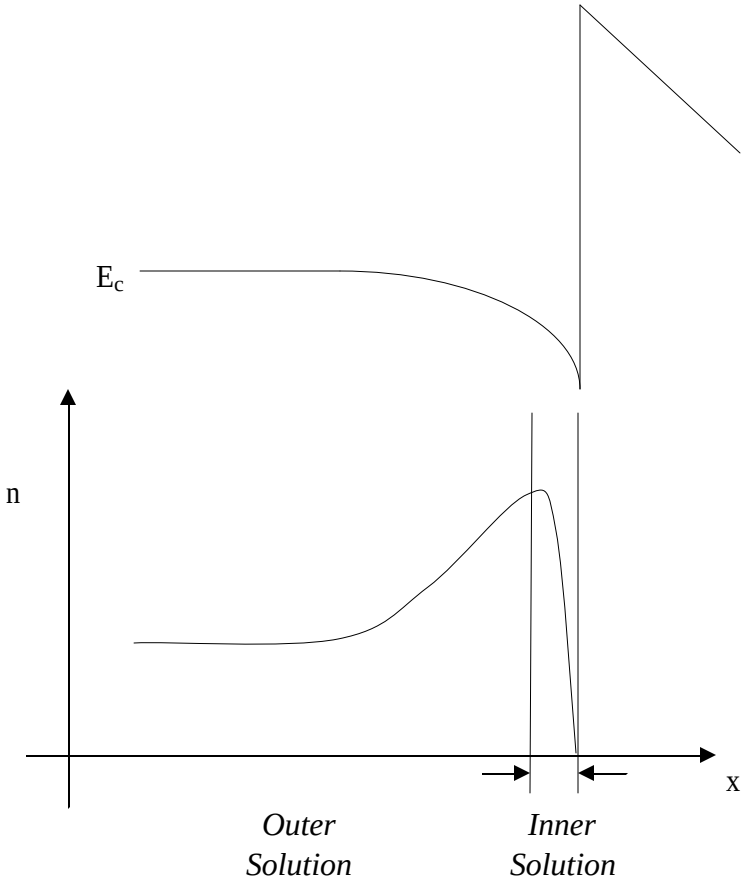

Figure 3.1 A typical DG boundary layer solution showing the regions of validity of the inner solution close to the barrier and the outer solution away from the barrier. The variation of the outer solution is negligible inside the boundary layer if the potential does not vary significantly here.

3.3.1 Quantum-mechanical solution for the boundary layer

The solution of the problem from one-electron effective mass quantum mechanics is particularly simple. Assuming that the potential energy is a constant E_b (the barrier height) for the left half plane and zero in the right, one can write, the solutions of the Schrödinger equation in terms of scattering states as [8],

$$\psi_k(x) = \frac{1}{\sqrt{2\pi}}\left[\exp(-ikx) + r(k)\exp(ikx)\right] \quad , \quad x \geq 0$$
$$= \frac{1}{\sqrt{2\pi}} t(k)\exp(-ik_1 x) \quad , \quad x < 0 \tag{3.3.1}$$

The transmission and reflection coefficients $t(k)$ and $r(k)$ are given by

$$t(k) = \frac{2k}{k + k_1} \quad , \quad r(k) = \frac{k - k_1}{k + k_1} \tag{3.3.2}$$

These wavefunctions can be used to calculate the single-particle density-matrix and the single-particle density through,

$$n = \int_0^\infty |\psi_k(x)|^2 f(E_k) dk = \int_0^\infty |\psi_k(x)|^2 \exp\left(-\beta \frac{\hbar^2 k^2}{2m}\right) dk \tag{3.3.3}$$

Assuming a large barrier, i.e. ($E_b >> k_B T$), we can perform the integration for the density outside the barrier explicitly [9],

$$n = n_0\left(1 - \exp(-\tilde{x}^2)\left[1 + \frac{H_1(\tilde{x})}{\sqrt{\beta E_b}} + \frac{H_2(\tilde{x})}{\beta E_b}\right]\right) \quad , \quad \tilde{x} = \frac{x}{\lambda_{th}} \tag{3.3.4}$$

The $H_i(x)$ in the above expression refer to the Hermite polynomials [10] of degree *i*. The simpler version, corresponding to the case of infinite barrier height, is used in the model proposed by Hänsch [11] for the inversion layer of a MOSFET.

The above expression for the density can be modified in the presence of a non-zero, slowly-varying potential on the right half plane using the WKB approximation for the wavefunctions in (3.3.3). We just provide the resulting intuitive expression from the calculation here,

$$n = n_0 \exp\left(-\beta V_s(x)\right)\left(1-\exp\left(-\tilde{x}^2\right)\left[1+\frac{H_1(\tilde{x})}{\sqrt{\beta E_b}}+\frac{H_2(\tilde{x})}{\beta E_b}\right]\right) \tag{3.3.5}$$

For the non-degenerate case and an infinite barrier, the solution for small x is given from (3.3.4) by

$$n \approx n_0 \exp\left(-\beta V_s(x)\right)\left(\frac{x}{\lambda_{th}}\right)^2 \tag{3.3.6}$$

We also notice an interesting result – the density at a finite barrier interface is suppressed exactly by the factor (βE_b) from the density in the absence of the barrier.

$$n(x=0+) \approx \frac{n_0}{\beta E_b} \tag{3.3.7}$$

Since the typical energies of the scattering states that take part in the summation above are around $k_B T$, we must have the following condition, in order for the assumed WKB approximation to be accurate and to have a quasi-continuous spectrum,

$$\left[V_s(\lambda_{th}) - V_s(0+)\right] << V_t \tag{3.3.8}$$

For the non-degenerate case the condition above will usually be satisfied because of low densities, except at very high doping concentrations (> 10^{19} cm^{-3}). For degenerate conditions and Fermi statistics, the relevant length scale and energy will be the Fermi wavelength and the separation of the Fermi energy from the conduction band edge respectively.

3.3.2 DG density profiles – Non-degenerate conditions

For this situation, the relationship between D_n and μ_n can be taken to be the thermal voltage (V_t), as is done for the usual form of the drift-diffusion equations. Let

us assume that the boundary condition for Eqn (3.2.6) is given by a vanishing density at the barrier interface (assumed at $x = 0$ for convenience), for now thus implying a very large barrier height. Then we get,

$$2V_t s \ln s - 2b_n \nabla^2 s = sC \tag{3.3.9}$$

Far away from the interface, DG effects are negligible and the density is the constant bulk density s_0 and so evaluating the constant here we get,

$$V_t s \ln \frac{s}{s_0} - b_n \nabla^2 s = 0 \tag{3.3.10}$$

The above equation is understood very simply – in the absence of a significant self-consistent potential, there is a balance between the diffusion and "quantum diffusion" given by the gradient term. Since the coefficient of the gradient term b_n is small, we will have a boundary layer near the interface, but the leading order behavior will be classical – i.e. we will see a constant density corresponding to s_0 until we reach very close to the barrier where it will change sharply to comply with the boundary condition at $x=0$.

We would like to facilitate analytical comparisons with the QM solution of the solutions to (3.3.9). Numerical solutions will only yield aggregate information as to the variation and getting some insight as to the *local* behavior of the solutions would require a very fine grid because of the nature of boundary layer behavior. The other important function that will be served by these analytical approximations is that they yield a systematic approach [12] to derive drain current equations when quantum effects are described by the DG theory.

We first cast the equation into a dimensionless form by scaling the length by a typical length term d (may be the device length, or a Debye length or some characteristic potential variation length), and s by s_0. We also notice that the term

$$L_{in} = \sqrt{\frac{b_n}{V_t}} = \sqrt{\frac{\hbar^2}{4mrk_BT}} = \frac{\lambda_{th}}{\sqrt{2r}} \tag{3.3.11}$$

has the dimensions of length and is related to the thermal wavelength λ_{th} (1.4.8) as above. L_{in} is a boundary layer thickness term. This is the typical order of the boundary layer thickness. For a large d we get

$$s \ln s - \varepsilon^2 \nabla^2 s = 0 \tag{3.3.12}$$

$$s(x \to \infty) \approx 1 \quad , \quad s(x = 0) \approx 0 \tag{3.3.13}$$

where the variables are now dimensionless and the small parameter ε is defined naturally as the ratio of L_{in} to d.

For very small x (much smaller than the boundary layer thickness defined by 3.3.11), we have, because of the boundary condition (3.3.13),

$$\lim_{x \to 0} (s \ln s) = 0 \tag{3.3.14}$$

Thus the solution for s yields,

$$s(x \approx 0) = a\, x$$

and is, at least in the sense of having a linear variation(noting that $s \equiv \sqrt{n}$), consistent with the QM solution given by (3.3.6).

The outer solution for the density (s_o) is given by setting ($\varepsilon = 0$) in (3.3.12). This yields as expected, the classical solution for large x.

$$s_o = 1 \tag{3.3.15}$$

The above solution is obviously incorrect when x is of the order of ε or lesser. The inner solution is the solution of (3.3.12) obtained by introducing the new length scale,

$$y = \frac{x}{\varepsilon} \tag{3.3.16}$$

This yields,

$$s_i \ln s_i - \nabla_y^2 s_i = 0 \tag{3.3.17}$$

where the subscript i (for inner) has been explicitly introduced to label the solution. It can be seen that the variable y is the original dimension, scaled by L_{in}. The solution of the above equation is readily written now, in the form of an inverse relation between s_i and y.

$$\int_0^{s_i} \frac{ds}{\sqrt{s^2 \ln s - \frac{s^2}{2} + \frac{1}{2}}} = y \tag{3.3.18}$$

$$\sqrt{2}\int_0^{s_i} \left[1 + \frac{s^2}{2} + \ldots\right] ds = \sqrt{2}\left[s_i + \frac{s_i^3}{6} + \ldots\right] = y \tag{3.3.19}$$

The linear part of the relationship for the density yields,

$$s_i = s_0\left(\frac{rx}{\lambda_{th}}\right) \Rightarrow n = n_0\left(\frac{rx}{\lambda_{th}}\right)^2 \tag{3.3.20}$$

The density (3.3.20) agrees exactly with the calculated QM density (3.3.6) close to the barrier edge for the particular case of r=1. However this is not the usual value used

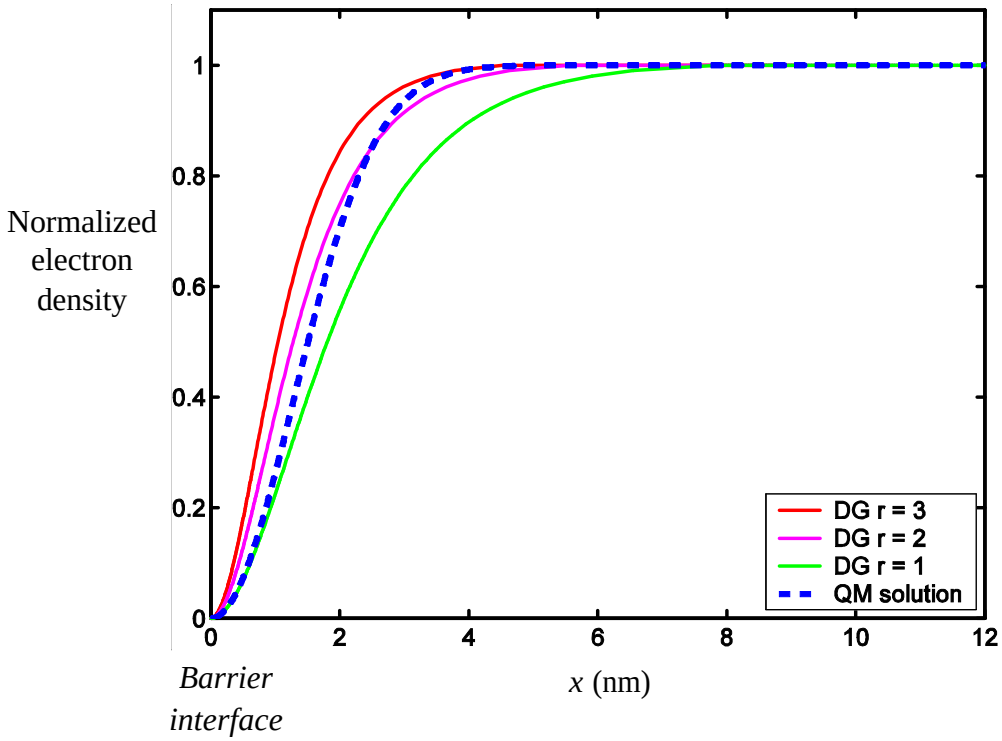

Figure 3.2 A comparison of the boundary layer solutions obtained from QM (dashed line) and DG (solid lines) with different values of the gradient strength from (3.2.2) corresponding to $r = 3$, $r = 2$ and $r = 1$. An average effective mass of 0.45 m_0 corresponding to the silicon conduction band was used.

for the high temperature (300 K) case in density-gradient simulations [1]. For the value of r=3 which is normally used, the densities are different by a factor of nine. This will not be important for capacitance calculations, where the thin boundary layer does not yield to a large charge contribution. However, for situations where the theory is to be used directly for barrier penetration problems (similar to what is discussed in the next chapter), the density at the interface (and therefore inside the classically

forbidden region) would be overestimated. The QM and DG solutions for the electron density near the barrier interface are shown in Fig. 3.2.

The other comparison that we can readily make is that of the width of the boundary layer. This is an important parameter, since it is directly related to the capacitance shifts due to the repulsion of carriers away from the interface. We define here this width, w_b to be the distance from the interface at which the density achieves 99% of the maximum value. This definition is ad-hoc, but very similar results are obtained with other definitions as well.

For the QM solution, we get the equation [11],

$$n(w_b) = 0.99 n_0 = n_0 \left[1 - \exp\left(- \frac{w_b^2}{\lambda_{th}^2} \right) \right] \tag{3.3.21}$$

This readily yields

$$w_{b,QM} \approx 2.146\, \lambda_{th} \tag{3.3.22}$$

For the DG solution (3.3.) we get,

$$\int_0^{0.99} \frac{ds}{\sqrt{s^2 \ln s - \frac{s^2}{2} + \frac{1}{2}}} = \frac{\sqrt{2r}}{\lambda_{th}} w_{b,DG} \tag{3.3.23}$$

Numerically solving the for the above integral we get,

$$w_{b,DG} = 3.427 \frac{\lambda_{th}}{\sqrt{r}} \tag{3.3.24}$$

$$w_{b,DG}(r=1) = 3.427 \lambda_{th} \quad , \quad w_{b,DG}(r=3) = 1.979 \lambda_{th}$$

Thus the assumption $r = 3$, yields a much closer result to that calculated from QM for the width of the barrier layer.

The above results indicate that in general, the values of the DG coefficient b_n (through the parameter r), required to achieve an agreement with the QM solution are different depending on the particular aspect of the boundary layer that we are interested in. It is difficult to achieve a detailed fit to the complete QM density profile using the form (3.2.2) without imposing arbitrary parameterization. However it is obvious that the general trend of the density variation is captured in the DG simulations.

We will make a general comment at this point – the DG coefficient b_n is a measure of the gradient corrections to the statistical free energy of the system. In particular it is related to the extent to which the kinetic energy density of the system is increased locally due to density gradients from the value given by the Thomas-Fermi theory. For a non-degenerate ensemble it has been shown in (2.) that the coefficient takes a value of (3.2.2) with $r = 3$. This, in turn was obtained assuming that the states are free-electron like, an implicit assumption made in the Thomas-Fermi theory.

When we have an abrupt potential barrier, only carriers with a high kinetic energy can penetrate close to the barrier. Hence the local kinetic energy density needs to be *raised* relative to that calculated by assuming a free-electron like ensemble as we explore regions close to the barrier. One can calculate the kinetic energy density explicitly from the scattering states (3.3.1), using the kinetic energy operator. This calculation yields,

$$\left\langle p^2 \right\rangle = \frac{1}{2\pi}\left(\frac{\hbar}{\lambda_{th}}\right)^2 \left[\frac{1}{1-\exp\left(-\tilde{x}^2\right)}\right] \tag{3.3.25}$$

and thus the kinetic energy density is clearly seen to diverge at the barrier edge

The DG theory approximates this by the density dependent expression :

$$\left\langle \frac{p^2}{2m} \right\rangle = \frac{nk_BT}{2} + b_n \frac{\left|\nabla n\right|^2}{n} \tag{3.3.26}$$

This is the fundamental reason for the inaccuracy of the DG solution with a constant r as compared to the QM solution for the boundary layer.

3.3.3 Analytical solutions with a self-consistent potential

The previous sections derived an analytical solution in the inverse form for the DG equations at the flat-band MOS capacitor situation and compared it with the numerical QM solution. We can include the self-consistent potential V_s approximately to derive analytical solutions for the DG equations, provided, that (3.3.7) holds (actually, a simpler version merely requiring the variation of the potential over the width of the inversion layer might be sufficient). This can be applied to the inversion and accumulation conditions in the MOS capacitor. Under this condition, we can ignore the variation of the potential within the boundary layer [12], obtain outer and inner solutions as before, and match them at the maximum charge density n_{max}.

The equation to be solved here is (3.2.8),

$$-sV_s + 2V_t s \ln\frac{s}{s_0} - 2b_n\nabla^2 s = 0 \tag{3.3.25}$$

For simplicity, we have assumed the potential reference at sufficiently large x to be zero. As earlier, we scale the equations into dimensionless form, where additional to the scaling for s and x, the potential is scaled by the thermal voltage V_t.

$$-sV_s + 2s\ln s - 2\varepsilon^2\nabla^2 s = 0 \tag{3.3.26}$$

The outer solution satisfies the equation with $\varepsilon = 0$, yielding the usual Boltzmann variation of the density with potential.

$$s_{outer} = \exp(\beta V_s) \tag{3.3.27}$$

At some $x = w_{max}$ the carrier density achieves a maximum n_{max} (corresponding to s_{max}) and beyond this point, the inner solution is a valid approximation for the density. Because of the assumption on the potential (that it is a constant within the very small inner layer) we can simply use the inverse form (3.3.18) for the solution replacing the quantity s_0 by the maximum s_{max} to scale s. This yields as before,

$$\int_0^{s_i} \frac{ds}{\sqrt{s^2 \ln s - \frac{s^2}{2} + \frac{1}{2}}} = y \tag{3.3.28}$$

for the inner layer density.

3.4 DG BEHAVIOR FOR POTENTIAL WELLS

3.4.1 Analytical DG solutions in a potential well

In deriving approximate analytical solutions for the DG boundary layer, we introduced a small parameter ε that depends on the ratio of a typical length scale in the device d (for instance a depletion layer width) to the length scale of the boundary layer. We also stressed that the solution is reasonable only when the potential variation across the length of the boundary layer is negligible. The parameter ε is no longer negligible when the device dimensions start approaching the length scale L_{in} as for the case of a quantum well, as d will then have to be chosen to be the device dimension. This case is seen in practical cases for ultra thin body SOI and double-gate devices as well as in the bulk MOSFET under strong inversion conditions.

We will look at the behavior of the DG solutions when the carriers are confined in a narrow potential well between $x=0$ and $x=a$. As above, we simplify this problem by ignoring tunneling and setting the boundary condition on density to vanish at the well boundaries.

Physically, the density of states (DOS) in the quantum well is reduced from that of 3-D bulk to a 2-D DOS due to energy quantization. This should lead to a reduction in the total charge as compared to the values expected from a classical situation, which assumes no confinement and a 3D DOS.

One way of physically rationalizing the situation from the DG equations is that there is no "outer" solution in this case, since x everywhere is of the order of the boundary layer thickness. In this case therefore we use the inner solution (3.3.18) everywhere within the quantum well. The maximum value that the density achieves is at the middle of the well (by symmetry). Labeling this value by some ks_0 for $k < 1$ being the suppression factor in the density peak from what it would reach in the absence of confinement (i.e., the bulk density), we can write for s the implicit equation,

$$\int_0^s \frac{dy}{\sqrt{y^2 \ln y - \frac{y^2}{2} + \frac{k^2}{2}}} = \frac{a}{2} \tag{3.3.29}$$

The value of k can be calculated by writing the boundary condition for the center of the well.

$$\int_0^k \frac{dy}{\sqrt{y^2 \ln y - \frac{y^2}{2} + \frac{k^2}{2}}} = \frac{a}{2} \tag{3.3.30}$$

This in turn can be cast into a more convenient form by replacing the integral from the variable y to the variable $q=y/k$. This yields something very much like a quantization condition on k.

$$\int_0^1 \frac{dq}{\sqrt{q^2 \ln q - \frac{q^2}{2} + \frac{1}{2} + q^2 \ln k}} = \frac{a}{2} \tag{3.3.31}$$

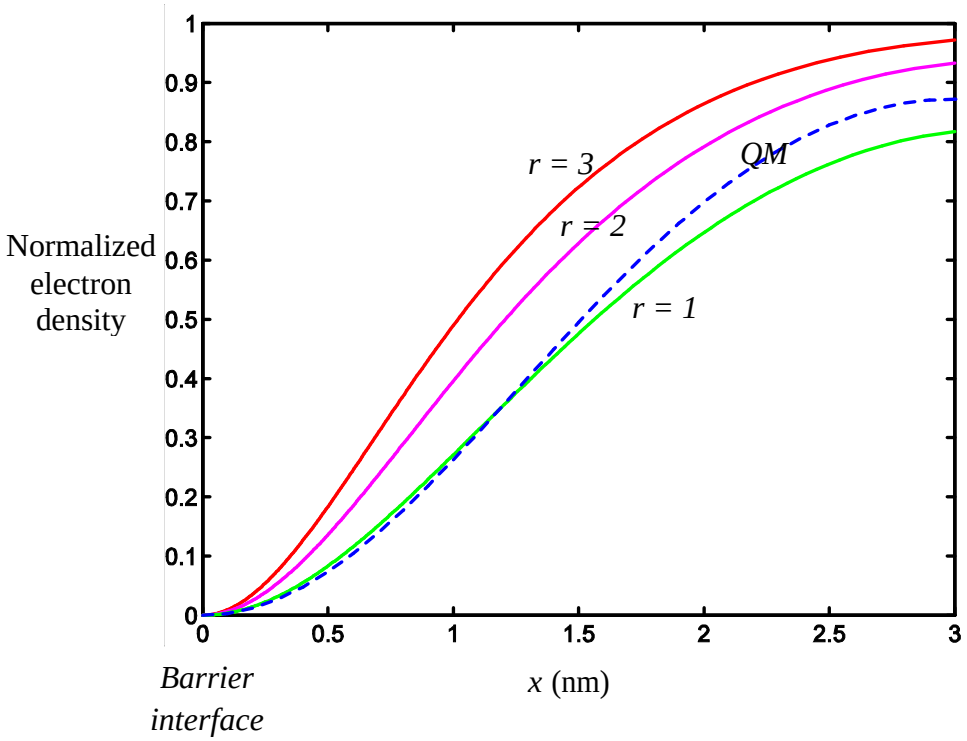

Figure 3.3 (a) A comparison of the solutions for density in a potential well obtained from QM (dashed line) and DG (solid lines) with different values of the gradient strength from (3.2.2) corresponding to $r = 3$, $r = 2$ and $r = 1$. An average effective mass of 0.45 m_0 corresponding to the silicon conduction band was used.

3.4.2 Quantum-mechanical solution in a potential well

The QM solution for the density is given by an explicit sum over the occupied sub-bands (labeled i below). Assuming an isotropic effective mass to make the comparison of the two cases transparent [13]

$$n_{QM} = \frac{mk_BT}{\pi\,\hbar^2}\sum_i \log\left[1+\exp\left(\frac{E_F - E_c - E_i}{k_BT}\right)\right]\left|\Psi_i(x)\right|^2 \qquad (3.3.30)$$

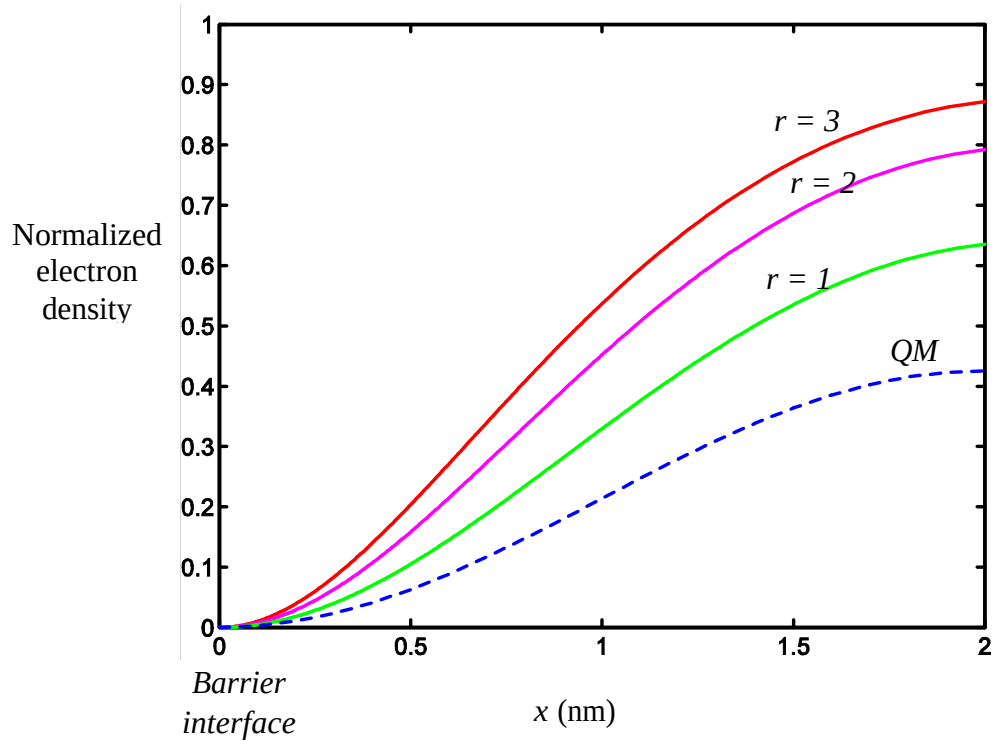

Figure 3.3 (b) Similar plot of the QM and DG densities as in Fig. 3.3 (a) for a potential well width of 4 nm. The discrepancy in the densities is even more pronounced because of confinement not being captured in the DG method.

The mass in the above Eqn. 3.3.30 can be easily related to the effective DOS in the conduction band edge N_c. The classical density expected is obviously

$$n_{CL} = N_c F_{1/2}\left(\frac{E_c - E_F}{k_B T}\right) \tag{3.3.31}$$

3.4.3 Comparison of the DG and QM solutions

Figs. 3.3 (a) and (b) show the trends of these calculations for the two different potential well widths of 6 nm and 4 nm respectively. Since the density is symmetric the density in one half is shown for clarity of comparison of the peak density. It can be

seen clearly that none of the three values of *r* used in the DG equations compares well with the true quantum mechanical density shown in dotted lines. It is also obvious that the discrepancy is significantly worse for the smaller potential well width.

The reason for the poor match between the DG solutions and the full quantum mechanical solutions is easy to understand. As pointed out in the previous chapter the density-gradient theory is derived assuming a perturbation expansion on the free carrier solution of the Bloch equation for the Wigner function. In both the cases considered above the presence of an abrupt potential barrier and a narrow potential well width violate this condition explicitly and therefore the perturbation expansion used for the derivation of DG theory is rendered invalid. Since problems of barrier penetration (tunneling) are in this regime (far from classical transport), it is very questionable whether these equations can be directly used in the treatment of tunneling transport at all.

3.2 SUMMARY

This chapter examined the behavior of the density-gradient theory equations with respect to their behavior in two special cases encountered very often in semiconductor cases. The first was the boundary layer behavior of the equations. It was shown that the equations show similar aggregate behavior to the full quantum mechanical solution near abrupt potential barriers but that the details of the solution for the density do not match very well. Secondly the behavior of the density-gradient theory in modeling confinement and quantization in narrow potential wells was considered and it was once again shown that the full quantum mechanical solution differs quite a lot from the density gradient solution.

BIBLIOGRAPHY

[1] M. G. Ancona and G. Iafrate, "Quantum correction to the equation of state of an electron gas in a semiconductor", *Phys. Rev. B,* vol. 39, no. 13, May 1989.

[2] M. G. Ancona, "Equations of state for silicon inversion layers", *IEEE Trans. Electron Dev.,* vol. ED-47, no. 7, pp. 1449-1456, July 2000.

[3] A. Wettstein, A. Schenk and W. Fichtner, "Quantum device simulation with the density-gradient model on unstructured grids", vol. 48, no. 2, February 2001.

[4] M. G. Ancona, "Diffusion-drift modeling of strong inversion layers", *Int. Journal for Computation and Mathematics in Electrical and Electronic Engg.,* vol. 6, no. 1, March 1987.

[5] M. G. Ancona, "Density gradient theory analysis of electron distributions in heterostructures", *Superlattices and Microstructures,* vol. 7, no. 2, 1990.

[6] V. C. Aguilera-Navarro, G. A. Estevez and A. Kostecki, " A note on the Fermi-Dirac integral function", *Journal of Appl. Phys.,* vol. 63, no. , pp. 2848-, 1988.

[7] J. Kevorkian and J. D. Cole, "Multiple scale and singular perturbation methods", Springer, NY, 1996.

[8] A. M. Kriman, N. C. Kluksdahl and D. K. Ferry, "Scattering states and distribution functions for microstructures", *Phys. Rev. B,* vol. 36, no. 11, pp. 5953-5960, October 1987.

[9] V. Narayanan and E. C. Kan, "A critical examination of macroscopic quantum transport approaches", *Journal of Comp. Elec.,* To be published.

[10] A. Abrahmovich and M. Stegun, "Handbook of mathematical functions",

[11] W. Hänsch, T. H. Vogelsang, R. Kircher and M. Orlowski, "Carrier transport near the Si/SiO_2 interface of a MOSFET", *Solid State Elect.,* vol. 32, no. 10, pp. 839-849, 1989.

[12] H. Abebe and E. Cumberbatch, “Quantum mechanical effects correction models for inversion charge and current-voltage characteristics for the MOSFET device”, *Proc. Nanotech Conf.*, February 2003.

CHAPTER 4

MACROSCOPIC DESCRIPTION OF TUNNELING

4.1 INTRODUCTION

In the previous chapter, macroscopic description of inversion layer quantization and quantum repulsion effects were explored using the density-gradient equations. It was shown that the equations describe the physics to leading order and show most of the qualitative average effects expected from microscopic quantum mechanical calculations.

This chapter extends the macroscopic formulation to include tunneling effects under large barrier potentials, by considering the form that the equation of state of the electron gas must take in those regions. The formulation presented is inherently one dimensional in nature, because a general approximate form of the carrier wavefunction cannot be written down in a multidimensional potential. However the ideas should carry over directly for the case of potentials that allow the Schrödinger equation to be solved by the separation of variables.

Our intention here is not to get the most accurate description of tunneling current for a particular device structure in a particular operating regime, say a MOSFET in the inversion regime, which can be obtained owing to the anisotropy of the MOSFET potential profile in any number of one-dimensional approximations, but rather to demonstrate that a purely macroscopic description of such a tunneling process is possible. Although we demonstrate the ideas for simple one-dimensional calculation here, the real power of these ideas will only be appreciated in multiple dimensions where the calculation of the energy states and the tunneling currents through a direct solution of the Schrödinger equation is impractical, both due to the

computational cost involved as well as due to the inherent difficulties in solving the equation numerically [1].

4.2 CONVENTIONAL CALCULATION OF TUNNELING CURRENTS

Conventionally tunneling currents are calculated using techniques involving the solution of the one-electron effective mass Schrödinger equation. We will restrict

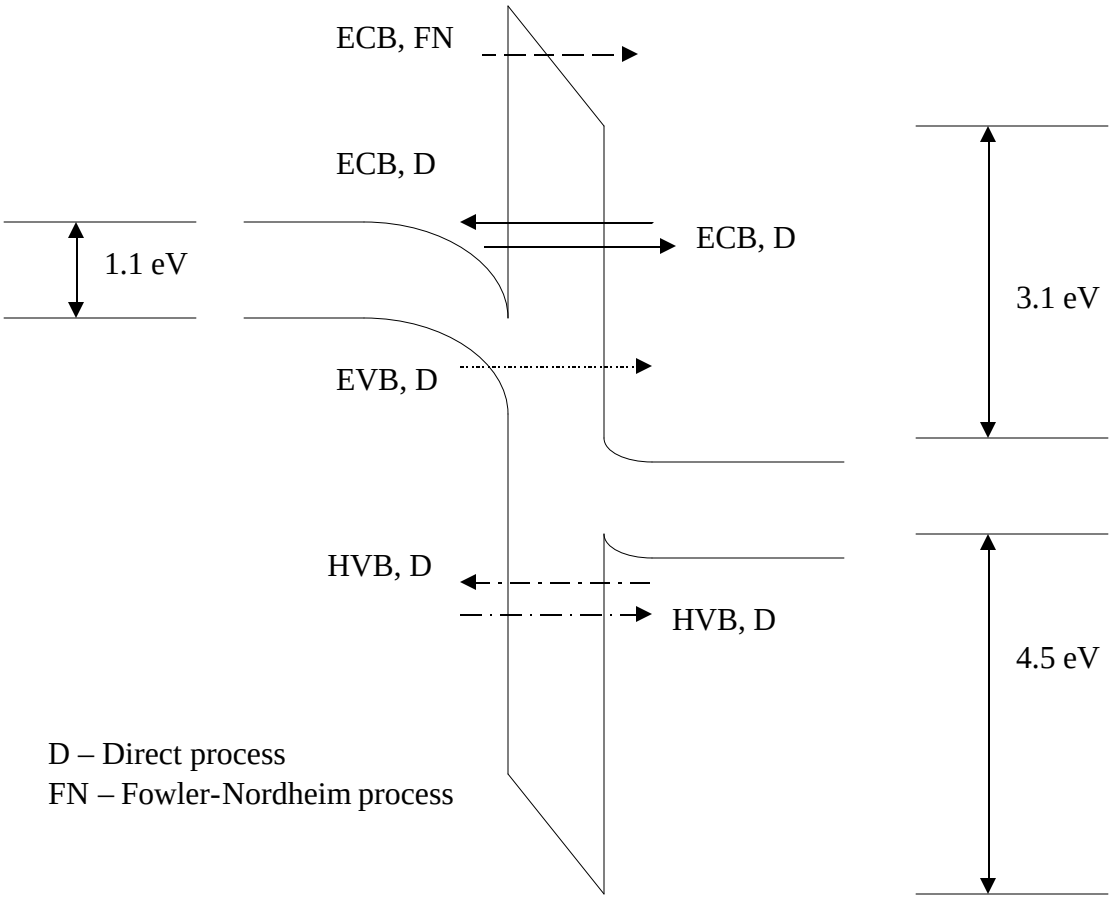

Figure 4.1 (a) Schematic representation of the bands and the direct and Fowler-Nordheim tunneling components for electrons (E) and holes (H) in an NMOS transistor biased in inversion, from the conduction band (CB) and valence band (VB).

ourselves to a discussion of techniques for currents through MIM barriers and the MOSFET gate-oxide barrier in this section. Fig. 4.1 shows a schematic representation of the various direct tunneling and Fowler-Nordheim components of the gate current in a typical silicon MOS capacitor structure, with an oxide insulating layer.

In the simplest treatment of tunneling through an insulating barrier, the mechanism of current is assumed to be from the transmission of free electron like traveling waves (scattering states) impinging on the barrier [2]. The carrier spectrum is assumed to form a continuum in energy from the bottom of the conduction band on one side and the eigenfunctions are treated as traveling waves with a parabolic E-k relation set by the conduction band (valence band for holes) effective mass. A carrier of a particular wavevector *k* is assumed to have a probability of transmission through the barrier given by the WKB transmission function,

$$T_{WKB}(E) = \exp\left[-2\int_0^{t_{ox}} k(x)dx\right]$$
$$k(x) = \sqrt{\frac{2m}{\hbar^2}\left[V(x) - E\right]} \qquad (4.2.1)$$

The current is then given by the difference of transmitted currents [Fig. 4.2 (a)] from either direction, the difference being due to the different populations of carriers on the two sides of the barrier owing to the different Fermi levels, i.e.

$$J_n = e\int \frac{\hbar k_x}{m^*} T_{WKB}(k)\left[f\left(E(k) - E_{fL}\right) - f\left(E(k) - E_{fR}\right)\right]\frac{d^3k}{(2\pi)^3} \qquad (4.2.2)$$

Several minor variations of the above theme, accounting for the reflection of electron waves at the interface, which will yield a prefactor for the above probability

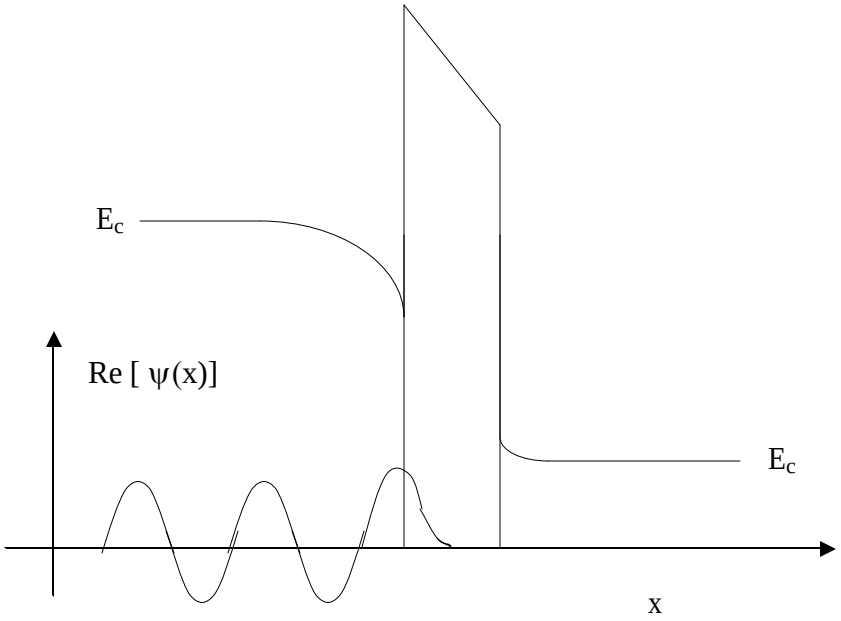

Figure 4.2 (a) Schematic representation of the traveling wave, constant amplitude eigenfunction form assumed in the transmission approach to calculating direct tunneling currents. This picture is accurate only there is no quantization due to the band bending, or for the higher energy states.

[3], or those that take into account the conservation of the momentum parallel to the oxide interface in the tunneling process [4] have also been considered. The BSIM3 compact model [5] is essentially a semi-empirical expression based on the above idea, of course with numerous fitting parameters to capture the variation in current over a large gate voltage range.

A more sophisticated approach for calculating the gate current is to take into account the gradual variation in the surface density of states from a 3-D like situation to a 2-D like situation under inversion [6] or accumulation [7]. This approach calculates the current after inversion as arising from the finite lifetime of carriers in states quasi-bound in the potential well formed by the conduction band and the oxide,

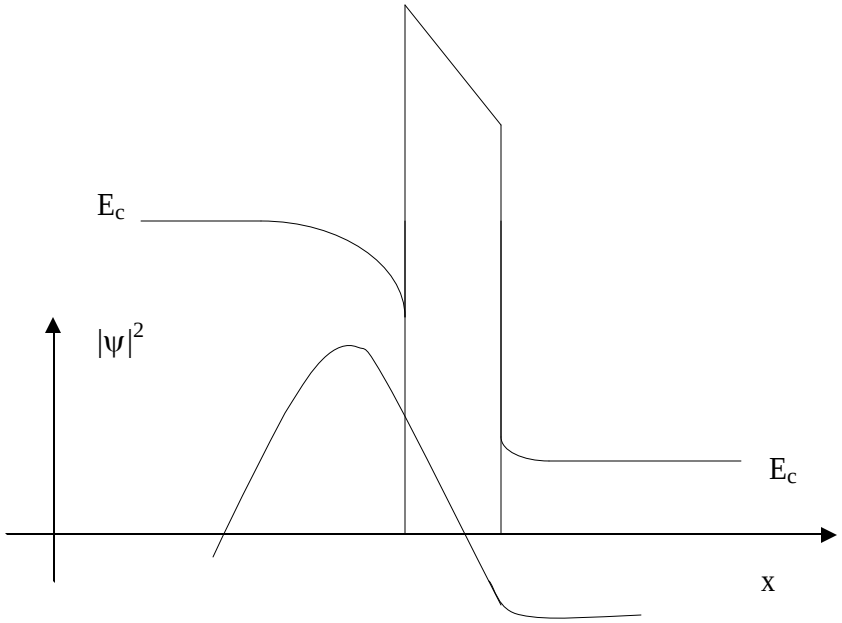

Figure 4.2 (b) Schematic representation of the amplitude of a quasi-bound inversion layer sub-band on a logarithmic axis. The finite amplitude of penetration past the finite width barrier is related to the energy broadening and thus the state's lifetime which determines the tunneling rate from the sub-band.

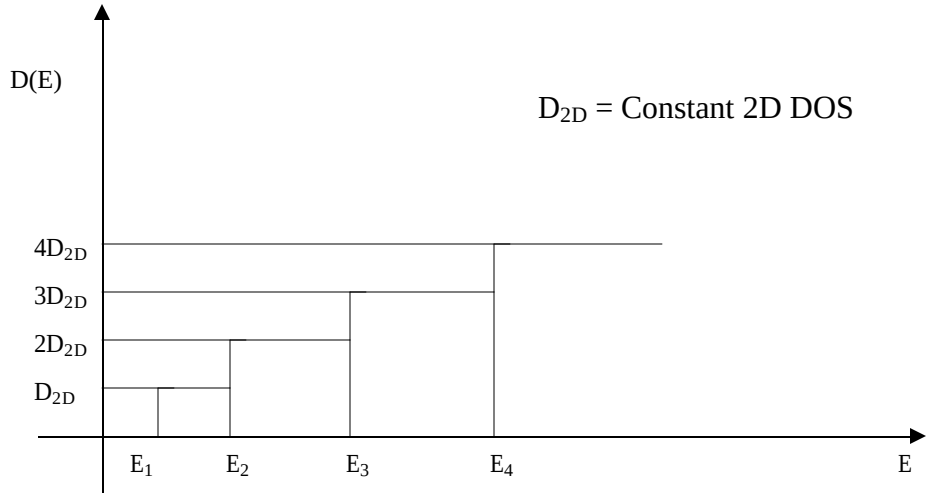

Figure 4.3 (a) Typical density of states (DOS) plot with respect to total energy for a carrier in a confining potential. When states are quasi-bound due to tunneling, the levels are no longer sharp, but this cannot be seen on a linear scale.

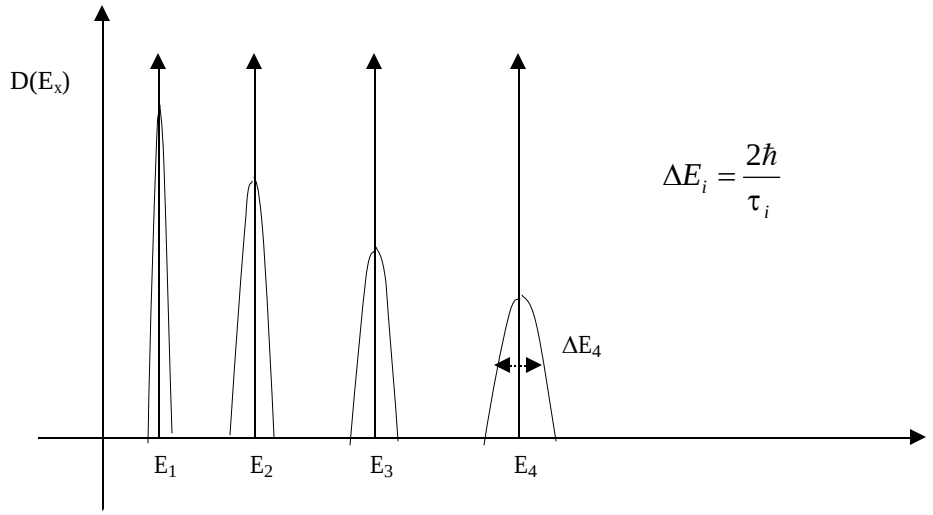

Figure 4.3 (b) Same situation as in (a), but the plot is with respect to the energy due to the motion in the confinement direction only – the delta functions at the quantized levels for bound states are broadened to approximate Lorentzians at slightly shifted energies for quasi-binding, shown exaggerated here

interface [Fig. 4.3 (a)-(b)]. The quasi-bound state lifetime can be related to the imaginary part of the energy eigenvalues, introduced when the Hamiltonian becomes non-Hermitian due to the open boundary conditions [5] (see also Chapter 1), introduced because of tunneling.

These approaches, for the silicon inversion layer, usually also take into account the lifting of the conduction band degeneracy owing to the different effective masses that carriers might have (for instance the transverse and longitudinal masses for electrons, or the heavy hole and light hole masses for the valence band), for motion perpendicular to the interface [6]. Once again several variants of this basic method can be found in the literature, owing to different ways to estimate the quasi-bound state lifetime [8] based on semiclassical or wave-mechanical approaches exact versus approximate treatment of the band splitting etc. Yet another, totally different approach

is to use the transfer-Hamiltonian formalism [9,10] using WKB wavefunctions on the two sides.

There are in addition to these simple one-electron models, a few notable attempts at detailed simulations including the band structure and the short range order in the oxide, using density-functional theory techniques [11].

In all these models, the oxide effective mass is taken as a fitting parameter – the effective mass in the forbidden gap of oxide has been assumed at various widely different values in the literature ranging for electrons from around $0.3\ m_0$ to around $0.6\ m_0$ and for holes over a similar range.

Correspondingly fewer models exist for the treatment of band-to-band tunneling, the efforts in that direction centered mostly on extensions of the Kane model [12] for direct bandgap semiconductors.

4.3 TUNNELING MODELS IN DRIFT-DIFFUSION SIMULATORS

Classical transport simulators, based on the drift-diffusion equations typically assume zero current through insulating barriers, or admit of thermionic emission currents over the barrier only. In applications where it is required to estimate the tunneling currents, they are not obtained self-consistently through a solution of quantum mechanically correct transport equations. Instead the one-electron Schrödinger equation is solved in the classical self-consistent potential and the tunneling currents are calculated and parameterized with respect to, for example the tunneling oxide thickness, the barrier height and the insulator electric field. The pre-calculated currents are then included as generation/recombination terms in the continuity equation for classical carriers at the insulator-semiconductor interfaces. This idea is schematically illustrated in Fig 4.1 below for an idealized MOS transistor geometry.

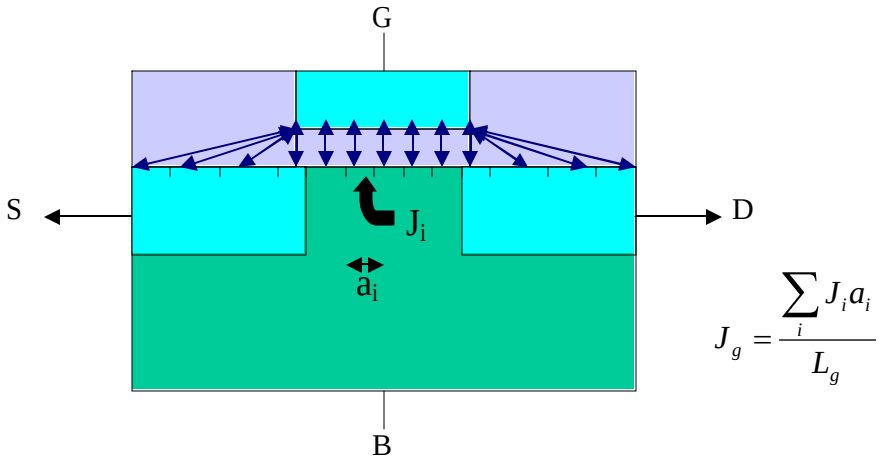

Figure 4.4 Gate current connections, for tunneling simulations using a typical drift-diffusion simulator (e.g. FIELDAY). Generation-recombination terms are introduced at the nodal connections indicated by arrows to simulate carriers lost (recombination) at the surface and created (generation) at the gate.

An approach such as the one above has many undesirable features apart from being physically inelegant – i.e. a compact model is used ad-hoc in a device simulation environment. Analytical forms for the gate current are seldom of consistent use over all the different operating regions of the device [3]. Although theoretically, one can use different current expressions for the different operating regions, this requires several independent parameters similar to, for instance the BSIM model (which itself is calibrated from experiment, is terminal voltage dependent and cannot be used) and is very cumbersome. The carrier density as simulated will retain a predominantly classical form, peaking at the insulator interface which is physically incorrect. Effects of multidimensionality are ignored completely and no attempt is even made to include these in an approximate sense. And finally, the resulting simulations may not converge at all in certain cases where the tunneling currents are large – for instance, charge for tunneling from an inversion layer has to be supplied by the generation process in a

MOS diode, since there are no source-drain contacts to supply carriers. For a thin gate oxide the current will be limited by this generation process and not by the oxide transport, i.e. the inversion layer might not form at all. Assuming a parameterized gate current that is only based on the oxide electric field could lead to convergence problems in the resulting simulations unless the ad-hoc generation terms are accurately reflected in the Jacobian matrix calculation as well.

4.4 SCHRODINGER EQUATION IN MADELUNG-BOHM FORM

For the purposes of the analysis to follow, it is useful to first write the form of the Schrödinger equation in its Madelung-Bohm (hydrodynamic) form [13]. The one single-band effective mass Schrödinger equation is

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(r)\right]\Psi = -i\hbar\frac{\partial\Psi}{\partial t} \tag{4.4.1}$$

We can use a general polar form for the wavefunction in the above equation.

$$\Psi = R.\exp\left(\frac{iS}{\hbar}\right) = \sqrt{P}\exp\left(\frac{iS}{\hbar}\right) \tag{4.4.2}$$

The quantity P can be identified as the probability density of the state Ψ. On substituting the above form for the wavefunction into the Schrödinger equation and equating the real and imaginary parts we get

$$\begin{aligned}\frac{\partial P}{\partial t} &= \nabla.\left(P\frac{\nabla S}{m}\right)\\ \frac{\partial S}{\partial t} &= -\frac{(\nabla S)^2}{2m} - V(x) + \frac{\hbar^2}{2m\sqrt{P}}\nabla^2\sqrt{P}\end{aligned} \tag{4.4.3}$$

These equations have very simple interpretations as a continuity equation for the probability density and a modified form of the Hamilton-Jacobi equation which is familiar in classical mechanics [13]. The potential term which is dependent on the probability density in the above equation is referred to as the Bohm potential or the quantum potential. The probability current is given by,

$$J_P = P\frac{\nabla S}{m} \tag{4.4.4}$$

For an eigenstate at energy E the second of these equations becomes,

$$-\frac{J_P^2}{2mP^2}+\left[E-V(x)\right]+\frac{\hbar^2}{2m}\frac{\nabla^2\sqrt{P}}{\sqrt{P}}=0 \tag{4.4.5}$$

Equations (4.4.3) are identical to the Schrödinger equation in all respects and may be solved for the quantum dynamics of a particle.

Now, the *local* form of equations for the probability density, are bound to be independent of the boundary conditions [14], so we can, without any loss of generality, bound the system in a very large box. This will lead to the following equation for the probability density of a single pure state, bounded in a large box.

$$E-V(x)+\frac{\hbar^2}{2m}\frac{\nabla^2\sqrt{P}}{\sqrt{P}}=0 \tag{4.4.6}$$

4.5 DENSITY EQUATIONS INSIDE A LARGE BARRIER

We are now in a position to consider the forms of the density matrix and therefore the resulting equations for the transport for carriers in a large barrier

potential. We will consider three different cases – one of these is solvable exactly while the others are analyzed approximately using the WKB approximation.

4.5.1 Simple Potential Barrier at Flat Band

Consider a potential barrier as in Fig. 4.5 (a). The form of the potential can be written as,

$$V(x) = E_b \theta(x) \tag{4.5.1}$$

In the above, E_b is the barrier height and $\theta(x)$ is the unit step function.

The solution of the corresponding Schrödinger equation inside the barrier is given by the evanescent forms of the scattering states (3.3.1). For a large barrier ($E_b >> k_B T$) these are given by

$$\begin{aligned} \psi_k(x) &= \frac{1}{\sqrt{2\pi}} t(k).\exp\left(-k_1 x\right) \\ k_1 &= \sqrt{\frac{2mE_b}{\hbar^2} - k^2} = \sqrt{k_b^2 - k^2} \approx k_b\left(1 - \frac{k^2}{2k_b^2}\right) \\ t(k) &= \frac{2k}{k + ik_1} \end{aligned} \tag{4.5.2}$$

For these solutions, we can explicitly write down the form of the density matrix inside the barrier ($x > 0$),

$$\rho(x,x') = \frac{1}{2\pi}\int_0^\infty |t(k)|^2 \exp\left[-k_1\left(x + x'\right)\right]\exp\left(-\beta\frac{\hbar^2 k^2}{2m}\right)dk \tag{4.5.3}$$

On performing the summation explicitly, we get an expression for the density matrix.

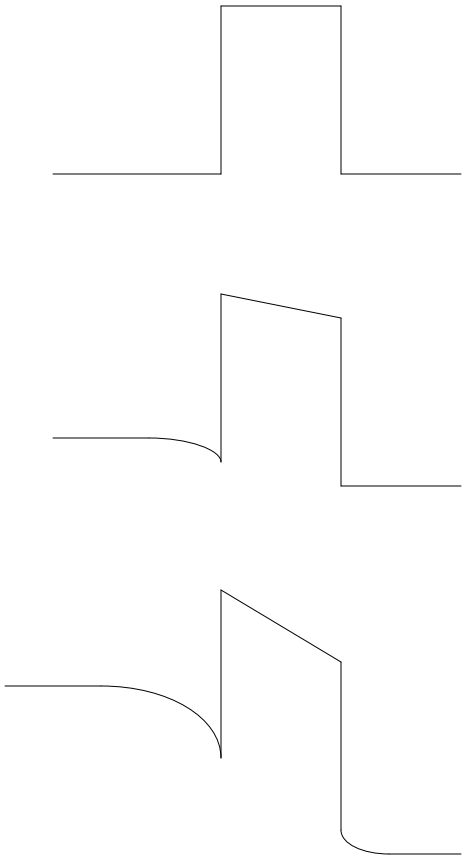

Figure 4.5 Typical potential forms assumed in Sections (4.5.1-3), (a) A simple large potential barrier, (b) Large barrier with a non-confining smooth potential, and (c) Large barrier with a confining potential on one side. The x-axis in all the plots is the coordinate perpendicular to the barrier.

$$\rho(x,x') = \frac{1}{2\sqrt{\pi}\lambda_{th}} \frac{\exp\left[-k_b\left(x+x'\right)\right]}{\sqrt{1-\dfrac{x+x'}{2\sqrt{\beta E_b}\lambda_{th}}}} \approx \frac{1}{2\sqrt{\pi}\lambda_{th}} \exp\left[-k_b\left(x+x'\right)\left(1-\frac{1}{4\beta E_b}\right)\right] \qquad (4.5.4)$$

The density is therefore calculated in a straightforward manner.

$$n(x) \approx \frac{1}{2\sqrt{\pi}\lambda_{th}} \exp\left[-2k_b x \left(1 - \frac{1}{4\beta E_b} \right) \right] \tag{4.5.5}$$

Quite obviously then, the total density is very similar in functional form to each of the evanescent mode densities inside the barrier. It therefore obeys the same differential equation as those densities, i.e. Eqn (4.4.6) with *P* replaced by *n* and with the energy,

$$E = E_b - E_b \left(1 - \frac{1}{4\beta E_b} \right)^2 \approx \frac{1}{2\beta} = \frac{k_B T}{2} \tag{4.5.6}$$

The differential equation that the density obeys is therefore

$$\left(\frac{k_B T}{2} - V(x) \right) + \frac{\hbar^2}{2m} \frac{\nabla^2 \sqrt{n}}{\sqrt{n}} = 0 \tag{4.5.7}$$

Physically, this has the meaning that the electrochemical potential for the gas of tunneling carriers inside the potential barrier is given by

$$\varphi(n) = \frac{k_B T}{2} - V(x) + \frac{\hbar^2}{2m} \frac{\nabla^2 \sqrt{n}}{\sqrt{n}} \tag{4.5.8}$$

The gradient of Eqn (4.5.8) is essentially the second moment equation (i.e. the momentum balance equation) for the distribution function when scattering is neglected in the barrier. On taking the gradient and multiplying by the carrier density, we get an equation that suggests that tunneling carriers move such that their drift due to the external electric field is balanced (or nearly balanced for the case of a non-zero total current) by a “quantum drift” inside the barrier. As a sanity check, the densities calculated from the explicit sum of energy states are shown compared to the density calculated from the above equation in Fig. 4.6 (a)

4.5.2 Potential Barrier in a Non-zero Electric Field

For this case [Fig. 4.5 (b)] we write the potential inside the barrier as a sum of a smoothly varying component V_{sm} and the barrier height.

$$V(x) = V_{sm}(x) + E_b\theta(x) \tag{4.5.9}$$

We will assume that the potential has a quasi-continuous spectrum so that there are a sufficiently large number of available eigenstates within a thermal energy. It is also worth mentioning here, that the spectrum is determined by the form of the potential *outside* the barrier itself. For this case we can write the evanescent mode wavefunctions inside the barrier using the WKB approximation as

$$\begin{aligned} \psi_E(x) &= \frac{C}{\sqrt{p(x)}}\exp\left(-\frac{1}{\hbar}\int_0^x p(x)dx\right) \\ p(x) &= \sqrt{2m(E_b + V_{sm}(x) - E)} \end{aligned} \tag{4.5.10}$$

By definition, this solution approximately satisfies Eqns (4.4.3). For a large barrier in the direct tunneling regime (i.e. $E_b >> | V_{sm} - E |$) we can write the above in a very good approximation as,

$$\begin{aligned} \psi_E(x) &= C_1\exp\left(-k_b\int_0^\infty\sqrt{1+\frac{V_{sm}(x)-E}{E_b}}dx - \frac{1}{4}\ln\left(1+\frac{V_{sm}(x)-E}{E_b}\right)\right) \\ &\approx C_1\exp\left(-k_b x + \frac{k^2 x}{2k_b} - k_b + \frac{k^2}{4k_b^2} - \frac{V_{sm}(x)}{4E_b}\right) \end{aligned} \tag{4.5.11}$$

The replacement of the energy as a parabolic function of the wavevector is possible since we have assumed a quasi-continuum of available states. We can now use this wavefunction to write the density matrix inside the barrier. We will get for the density

$$n(x) = n_0 \int_0^{\infty} \exp\left(-\lambda_{th}^2 k^2\right).\psi_k^*(x).\psi_k(x).dk = n_0 \frac{\exp\left(-2k_b x - 2k_b \int_0^x \frac{V_{sm}(x)}{4E_b} dx - \frac{V_{sm}(x)}{2E_b}\right)}{\sqrt{1 - \frac{x}{\lambda_{th}\sqrt{\beta E_b}} - \frac{1}{2\beta E_b}}} \quad (4.5.12)$$

Writing the density as an exponential by making use of the largeness of the barrier height compared to the thermal energy we get,

$$n(x) = n_0 \exp\left(-2k_b x + \frac{x}{\lambda_{th}\sqrt{\beta E_b}} - 2k_b \int_0^x \frac{V_{sm}(x)}{4E_b} dx + \frac{1}{2\beta E_b} - \frac{V_{sm}(x)}{2E_b}\right) \quad (4.5.13)$$

So we find that once again, the density inside the barrier takes a very similar functional form to each of the wavefunctions and obeys the differential equation

$$\left(\frac{k_B T}{2} - V(x)\right) + \frac{\hbar^2}{2m} \frac{\nabla^2 \sqrt{n}}{\sqrt{n}} = 0 \quad (4.5.14)$$

Numerical solutions are provided in Fig. 4.6 (b) to confirm this result, for a couple of simple barrier potentials, assuming that they yield a quasi-continuous spectrum.

4.5.3 Bound States Leaking into a Potential Barrier

The above argument assumes that the smooth part of the potential is slowly varying outside the barrier and therefore any bound states that exist form a near continuum (i.e. confinement energies are much smaller than the thermal energy and there are several states in an energy range $k_B T$). This allows the sum over energy states that is required to evaluate the density inside the barrier to be converted to the integral which has been evaluated explicitly above.

We now consider the case of bound states, as in a potential well decaying into a neighboring barrier potential. This case is schematically illustrated in Fig. 4.5 (c). There are several interesting practical cases which fall under this category viz. the inversion layer of a MOSFET, ultra thin-body SOI devices, thin body double gate devices etc.

In this case, the Bohr-Sommerfield condition [15] yields us the quantization energies of the carriers as given in the WKB approximation.

$$\int_a^b \sqrt{\frac{2m}{\hbar^2}\left[E_j - V(x)\right]}dx = \left(j + \frac{1}{2}\right)\pi \tag{4.5.15}$$

Here *a* and *b* are the classical turning points and *n* is a quantum number that labels the discrete energy levels. This is known to yield excellent values for the higher levels but is acceptable even for the low lying levels in the spectrum for the case where the potential does not vary too rapidly in between the turning points.

The wavefunctions for this case are given by an expression very similar to the WKB expression given in the previous section inside the barrier with the notable change that the energy is labeled by the level *j* rather than by a wavevector *k* as in a near continuum. We can therefore write the total density, inside the barrier as,

$$n(x) = n_0 \sum_j \exp\left(-\beta E_j\right)\psi^*_{E_j}(x)\psi_{E_j}(x) \tag{4.5.16}$$

We do not know anything about the energies E_j except for the assumed fact that they are large compared to the thermal energy (otherwise the analysis would reduce to that in the previous section), and that they satisfy the above quantization condition. Before proceeding further we notice that the density function for a single state in the WKB approximation can be written from the previous section as,

$$P_{E_j}(x) = C_1 \exp\left(-2k_b x + k_b \frac{E_j x}{E_b} - k_b \int_0^x \frac{V_{sm}(x')}{2E_b} dx' + \frac{E_j}{2E_b} - \frac{V_{sm}(x)}{2E_b} \right) \tag{4.5.17}$$
$$= P(x) \exp\left(k_b \frac{E_j x}{E_b} + \frac{E_j}{2E_b} \right)$$

The probability density, for a single state is hence the product of an energy independent term which contains all the dependence on the external potential and an energy dependent spatial variation which is very slowly varying if the barrier potential is large. Now, for the *significant* few energy levels we can assume that the quantization energies are much smaller than the barrier height and we can write,

$$\sum_j \exp\left(-\beta E_j + k_b \frac{E_j x}{E_b} + \frac{E_j}{2E_b} \right) \approx \sum_j \exp\left(-\beta E_j\right) \left(1 + k_b \frac{E_j x}{E_b} + \frac{E_j}{2E_b} \right) \tag{4.5.18}$$
$$\approx \exp\left(k_b \frac{\langle E \rangle x}{E_b} + \frac{\langle E \rangle}{2E_b} \right)$$

where the averages are self-evident.

Hence the total density inside the barrier is to a very good approximation given by,

$$n(x) \propto P_{\langle E \rangle}(x) = P(x) \exp\left(k_b \frac{\langle E \rangle x}{E_b} + \frac{\langle E \rangle}{2E_b} \right) \tag{4.5.19}$$

Once again, by analogy with the wavefunctions and (4.4.6) the above must satisfy the differential equation,

$$\left[\langle E \rangle - V(x) \right] + \frac{\hbar^2}{2m} \frac{\nabla^2 \sqrt{n}}{\sqrt{n}} = 0 \tag{4.5.20}$$

The only difference between this and the previous cases is therefore the replacement of the average energy from $k_B T/2$ by the appropriate value over the quantization

energies. But as pointed out earlier, the exact value of this energy makes only a marginal contribution to the density variation in the barrier since it is effectively swamped by the potential.

4.5.4 Generalization to Fermi-Dirac Statistics

It is not very hard to generalize the above differential equations for carrier density inside the barrier to the case of Fermi-Dirac weighting of the carrier wavefunctions. Eqn (4.5.17) is independent of the weighting of the states, being for a single probability density that appears in the sum and a procedure similar to that adopted in Eqn (4.5.18) will hold for the case where the energies are weighted by the Fermi distribution as well. In this case, we would have the average energy $<E>$ that appears in Eqn (4.5.20) evaluated over a Fermi distribution as opposed to the Boltzmann distribution.

4.5.5 Generalization to multiple space dimensions

For a general 3D potential barrier, Eqns (4.4.3-6) are still valid being just the Schrödinger equation in a different form. One cannot however construct a general asymptotic proof as above for the differential equation satisfied by the density, since the approximate solution of the Schrödinger equation cannot be written down generally in 3-D coordinate space, unless the Hamiltonian is separable (which requires very specific symmetry conditions imposed on the potential). A semiclassical wavefunction can still be written down, in terms of the classical periodic orbits [16] in the action-angle conjugate variable space, but it is not possible to generally represent it in coordinate space.

The strongly decaying nature of the evanescent modes inside a large potential barrier is however quite general irrespective of dimension and it is expected that the

Eqns (4.5.20) for the density will still hold in such cases, atleast as a good first approximation. The validity of this can only be borne out by extensive numerical experiments and calibration against experimental data, for a truly multidimensional potential barrier profile, such as those obtained with STM probe tips, or for write/erase times in nanocrystal based non-volatile memories [17] and has not been attempted here.

4.6 CURRENT TRANSPORT INSIDE A BARRIER

Elastic tunneling through thin insulating barriers is a transmission phenomenon that is predominantly determined by the characteristics of the barrier itself – i.e. scattering is relatively unimportant, except at the contacts that the carriers tunnel between. One way of saying this is that carriers originating from a particular contact retain the chemical potential of that contact throughout the barrier thickness till they reach the other contact. They are then scattered and relax to the Fermi level of the "downstream" contact.

This is consistently seen in all the different wave mechanical techniques which are employed to calculate the tunneling currents. The idea of calculating the tunneling current as a difference between the transmitted currents in opposing directions, and the calculation of a particular transmitted current using the population at each side and the transmission characteristics of the barrier (e.g. Eqn (4.2.2)) are of course as a result of this implicit assumption.

For the macroscopic description of tunneling, this idea translates to a ballistic transport of the carrier gas between the two contacts. Clearly then the injection must be separated here as well into two carrier populations whose equations of state depend on the contact from which they originate. The Eqns (4.5.20), or their corresponding chemical potential formulation,

$$\nabla \varphi_n = 0 \qquad (4.6.1)$$

yield the equations for the density profile for carriers originating from one of the tunneling contacts. There will be a similar equation for the carriers tunneling from the other contact. The above equation implies that the quasi-Fermi (or the electrochemical potential) of the carriers injected from a particular contact into the barrier region remains a constant. This condition holds for equilibrium (where the current in Eqn (4.4.5) is rigorously zero), but it also holds when the currents themselves are not very large, as then the kinetic energy term due to the current is negligible for most of the barrier.

An analogy with the p-n junction diode is particularly useful here. In a p-n junction diode, on the application of a forward bias across the depletion region, majority carrier electrons from the n-doped side get thermionically injected over the built-in barrier, into the p-doped side of the junction [18].. They then recombine as minority carriers on that side. If the generation/recombination processes in the space charge region are taken to be negligible, the electron current is a constant across the depletion region. The quasi-Fermi level for electrons is then nearly constant all the way across the depletion region (and equal to the Fermi level fixed by the n-doped side as shown in Fig. 4.7 (a). This is because the electron density varies by several orders of magnitude from the value fixed by the doping on the n-side to the value fixed by the doping on the p-side and the applied voltage and the current is given by the product of this density and the gradient of the quasi-Fermi level. At the edge of the depletion region on the p-side the quasi-Fermi level for electrons, E_{fn} relaxes to the bulk Fermi level on the p-doped side across the recombination region over a distance of the order of the diffusion length on that side. The electron current is given by the

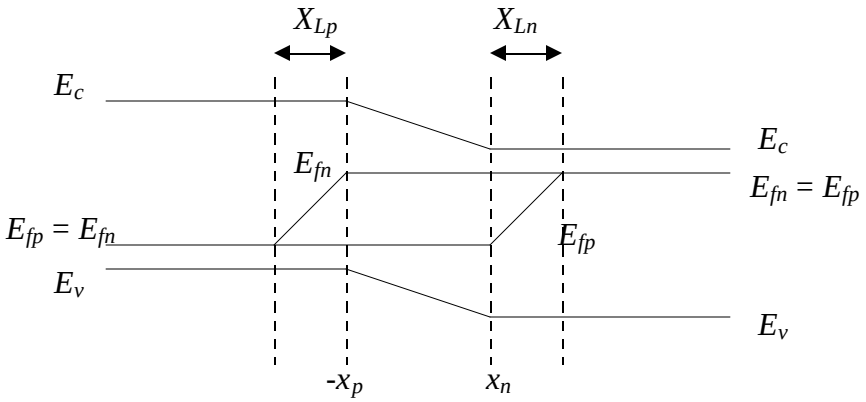

Figure 4.6 (a) Variation of the quasi-Fermi levels for electrons and holes in a typical p-n junction diode. The equalization of Fermi levels occurs on each side due to recombination process and the current depends on the rate of recombination

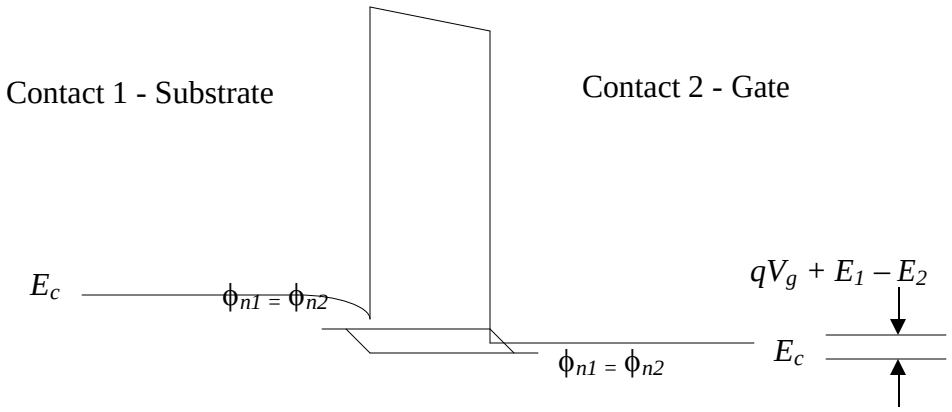

Figure 4.6 (b) Variation of the chemical potentials, for electrons only, injected from the two "contacts" 1 and 2. The equalization of the chemical potentials occurs due to thermalization processes in the contacts. Analogous to (a) the current is decided by this rate and must be modeled accordingly.

rate of recombination of minority carrier electron on the p-doped side of the junction, i.e. the “downstream” contact for electrons injected from the n-doped side.

For the case of two different carriers (electrons and holes) the local driving force for the equalization of the quasi Fermi levels, is recombination. For a single carrier the local driving force that equalizes Fermi levels for carriers moving in different directions is of course scattering and in the absence of this, carriers moving in different directions can have very different Fermi levels.

Eqn (4.5.20) is actually an equation for energy balance in the tunneling carrier gas. For the case of non-equilibrium, we will have to include the “classical” kinetic energy term in Eqn (4.4.5) as well, but this can be ignored in regions where the density is large, similar to what is done for classical carriers with the convective term in the hydrodynamic equations (see Chapter 1). The tunneling carrier density will decay nearly exponentially across the barrier length as given by the typical solutions of Eqns (4.5.20), from large values close to the two “contacts” (substrate and gate). Hence the inertia term (the convective term in the usual form of the hydrodynamic equations (1.3.3)) will be negligible through most of the barrier except very close to the barrier edge downstream, where the density is very small. At this edge, the carriers tunneling from the first contact will have to be scattered to the chemical potential of the downstream contact. In the absence of detailed information about this part of the physics (similar to the recombination process [18] in a p-n junction diode, for instance) we can model this, by assuming a certain recombination velocity γ for these carriers at this contact, i.e.

$$J_{\mathbf{n}} = q\gamma_2 . n_1 \tag{4.6.2}$$

In the above n_1 is the density of carriers tunneling from contact 1 (the substrate) and the quantity γ_2 is a tunneling recombination velocity [19,20], a fundamental macroscopic property of the contact for carriers n_1 at the second contact. This can be calibrated from the experimental results for tunneling current, or estimated from the microscopic calculation from the Schrödinger equation. In the context of microscopic calculations like Eqn (4.2.2), we implicitly assume this to be of the order of the thermal velocity (for instance the rate at which the carriers strike the interface in transmission probability based theories, is given by the average velocity of a hemi-Maxwellian distribution). For the purposes of simulation here, we will assume this to be a constant fitting parameter γ and to be the same for each contact.

The total tunneling current is therefore, given by,

$$J_t = q\gamma_n.[n_1(t_{ox}) - n_2(0)] + q\gamma_p.[p_2(0) - p_1(t_{ox})] \qquad (4.6.3)$$

The tunneling oxide is assumed to be between $x=0$ and $x=t_{ox}$ in the above expression and the subscripts refer to contact from which carriers are injected into the barrier.

4.7 EXAMPLES OF MACROSCOPIC TUNNELING CALCULATIONS

We will consider two sets of examples for a numerical demonstration of the above macroscopic tunneling formalism. The first is the application to the problem of direct tunneling currents in MOSFETs with thin gate oxides. The second concerns the tunneling of carriers between the plates of an MIM capacitor. This provides another easy application to check for the validity of the approach, since the metals can be assumed to be ideal conductors and the barrier repulsion for Fermi statistics in degenerate conductors is negligible and the current is solely determined by how

appropriate the assumed forms of the density equations and the recombination boundary conditions are for the problem.

4.7.1 Numerical Formulation

The Eqns (4.5.20) discretized directly can lead to numerical problems and will require a very fine grid inside the barrier. They can be cast in a more convenient form for the purposes of numerical simulation. Since the density of carriers injected from each contact, is expected to decay nearly exponentially across the barrier, it is better to cast the equations in terms of the logarithms of the densities rather than the absolute values. This can be considered similar to other exponential fitting schemes for easing the requirements on discretization of differential equations in regions where the solutions vary very rapidly [21]. This is easily accomplished by noting that,

$$\frac{\nabla^2\sqrt{n}}{\sqrt{n}} = \frac{1}{2}\left[\nabla^2 \log(n) + \frac{1}{2}\left|\nabla \log(n)\right|^2\right] \tag{4.7.1}$$

Eqn (4.7.1.) then transforms to the nonlinear differential equation, (for $s = log(n)$)

$$\left[\frac{\langle E\rangle}{q} - \frac{E_b}{q} - V(x)\right] + \frac{\hbar^2}{4mq}\left[\nabla^2 s + \frac{1}{2}\left|\nabla s\right|^2\right] = 0 \tag{4.7.2}$$

The above equations have to be solved for each of the carriers (electrons and holes) injected from each contact, self-consistently with the Poisson equation that determines the potential. For electrons injected from each side, we make the assumption that

$$\frac{\langle E\rangle}{q} = \frac{E_{c,1(2)}}{q} + \frac{V_t}{2} \tag{4.7.3}$$

In the above, the subscripts refer to the contact and E_c is the conduction band edge. Basically the assumption is that the mean energy that is relevant in the Eqn (4.7.2) is in excess of the conduction band edge, by the average energy of a Boltzmann averaged continuum. A similar relation is adopted for holes. This assumption is similar to the assumption of a quasi-continuum in energy at the tunneling contact and should be modified explicitly when there is quantization. For instance the mean energy above the conduction band due to quantization can be made a function of the field strength on the substrate side. This will be re-visited later.

4.7.2 Boundary conditions

The boundary condition for the electron density at the upstream contact is obtained from Eqn (2.), the scattering state sum for the density matrix in equilibrium for

$$n_1 = \frac{N_c F_{1/2}(\eta_C)}{\beta E_b} \quad , \quad \eta_C = \frac{E_c - E_F}{k_B T} \tag{4.7.4}$$

Once again the relevant expression is used for holes. Obviously Eqn (4.7.4) could be replaced instead by allowing the density to "float" to the correct value using the relevant macroscopic equations inside the semiconductor as well, although for simplicity this is not done here. The relation above assumes that the density at the interface is fixed simply by the local Fermi level and potential, with the effects of the barrier in reducing the density due to quantum repulsion is taken to be from that for a free electron like situation locally. In reality the effect of the field (through quantization) must also be parameterized in the spirit of the macroscopic approach as with the average carrier energy. This quantity is equivalent to the density of states for the initial state of tunneling carriers.

The boundary condition on the downstream contact is modeled by assuming that the carrier density profile is nearly flat there. i.e. by assuming that

$$\left.\frac{\partial s_1}{\partial x}\right|_{x=t_{ox}} \approx 0 \tag{4.7.5}$$

This is a reasonable guess for the downstream boundary condition that also seems to be borne out by numerical simulations – since the carrier profile decays exponentially, the derivative does as well and for any reasonable barrier length, Eqn (4.7.5) will give an excellent approximation to the carrier density.

The effects of the boundary conditions for the downstream contact are actually rather minimal as far as the results for the tunneling currents are concerned. The exponential nature of the decay of charge density in the barriers is captured by the nature of the differential equation and this largely determines the current. Since the downstream density determines the current through a linear relation (4.6.2), an error in this amounts to an error in the pre-factor for the tunneling current density, while the exponential dependence on voltage is very well captured in the solution of the boundary value problem posed by Eqns (4.7.2-5).

For electrons tunneling from the gate to the substrate under inversion conditions, the low energy electrons will see a substantially wider insulator thickness because of the silicon bandgap (Fig. 4.1). This can be approximately accounted for by modifying the tunneling recombination velocity for these carriers by the Boltzmann factor depending on the oxide voltage

$$\gamma_2 \approx \gamma \min\left[\exp\left(\frac{E_c(t_{ox})-E_c(0)}{k_B T}\right),1\right] \tag{4.7.6}$$

4.7.3 A note on oxide effective mass values

The oxide effective mass, m_{ox} is a fundamental parameter entering into the (4.7.2) applied to a MOSC and will have a particularly strong effect on the tunneling characteristics owing to the exponential variation of the carrier densities inside the barrier. It is worthwhile to investigate what values are reasonable for this parameter. As mentioned briefly in Section 4.2, the values of effective masses in oxide seen in the literature from tunneling models, experimental calibration and numerical band structure simulations vary very widely [22-26]. Values in the range from 0.25 m_0 to as high as 0.7 m_0 have been used.

Since the carriers tunnel through the forbidden gap of the insulator, the E-k relation for the evanescent modes in the SiO_2 is relevant for the effective mass seen by the tunneling carriers. Although SiO_2 is amorphous and it might be suspected that an unambiguous band structure effective mass does not exist [23], local order might be expected and this might lead to an effective mass with minor variations depending on detailed growth. The complex E-k relation (the complex wavevector is labeled κ below) in silicon dioxide is frequently modeled using the Franz E-k dispersion relation with a conduction band effective mass m_c^* (assumed equal to the valence band effective mass)

$$\kappa(E) = \left(\frac{2m_c^*}{\hbar^2}\right)^{1/2} E^{1/2}\left(1-\frac{E}{E_g}\right)^{1/2} \tag{4.7.7}$$

This relation was used for calibrating F-N tunneling currents in [22] and an effective mass of 0.42 m_0 was obtained for the conduction band edge of SiO_2. It was also shown that the rough form of the E(k) relation in the oxide follows (4.7.6). Other band structure calculations obtained 0.55 m_0 [24] for the same value. The above

relation then indicates that the effective mass seen by carriers in the oxide bandgap is energy dependent (Fig. 4.7), roughly according to,

$$\frac{1}{m^*} = \frac{E_g}{m_c(E_g - 2E)} + \frac{4E(E_g - E)E_g}{m(E_g - 2E)^3} \tag{4.7.8}$$

where E is the energy difference between the conduction band edge of oxide and the tunneling carrier energy.

The use of the above Franz relation is equivalent to using a variable effective mass across the oxide as a function of the local potential. However as seen in Fig. 4.7, this leads to a very small effective mass close to the conduction band edge of silicon

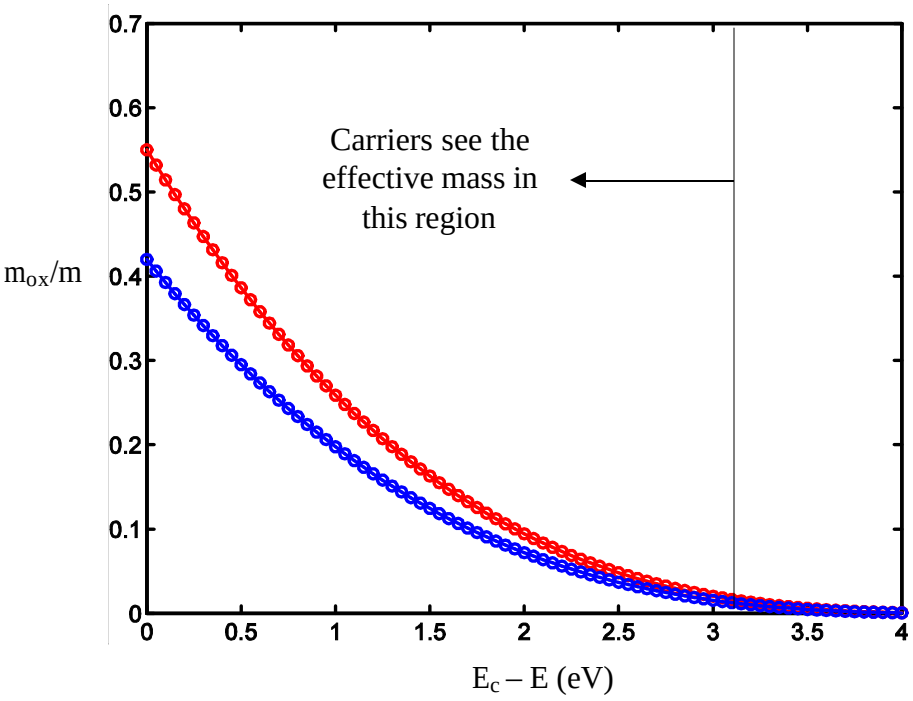

Figure 4.7 The oxide effective masses calculated from the Franz E-k relation (4.7.7) assuming a band edge effective mass of 0.42 m_0 or 0.55 m_0. Both lead to extremely low effective masses close to the Si conduction band edge (dashed line), due to the assumption of equal valence and conduction band masses.

which leads to high tunneling current results and the effective mass at the middle of the oxide bandgap goes to zero. The problem is that the above relation assumes that the conduction band and valence band effective masses in oxide are identical and this is not borne out by numerical simulations of the oxide electronic structure using pseudo potential methods. In fact, the valence band masses are probably very large [24] (5-10 m_0) and hence the valence band states are unlikely to mix with the conduction band states, a condition required for the validity of (4.7.6). In the absence of reliable data or experimental results, we will continue to assume a parabolic relationship with a tunneling effective mass, expected to lie among the reported values of parabolic conduction band effective masses reported in the literature.

4.7.4 Numerical Results

Figs. 4.8 (a) and (b) show the densities inside the barrier, calculated from the macroscopic equations (4.7.2-5) in conjunction with the Poisson equation. The injected densities from each of the contacts decay exponentially across the barrier, as was seen in the analytical forms in Section 4.5, with the self-consistent potential corrections becoming important near the downstream edges. The tails of the densities at near the downstream contacts are much higher for the smaller oxide thickness leading to higher currents in this picture. The substrate p-doping and the n+-poly doping were assumed to be 5×10^{17} /cc and 10^{20} /cc respectively and the bias is 2.0 V.

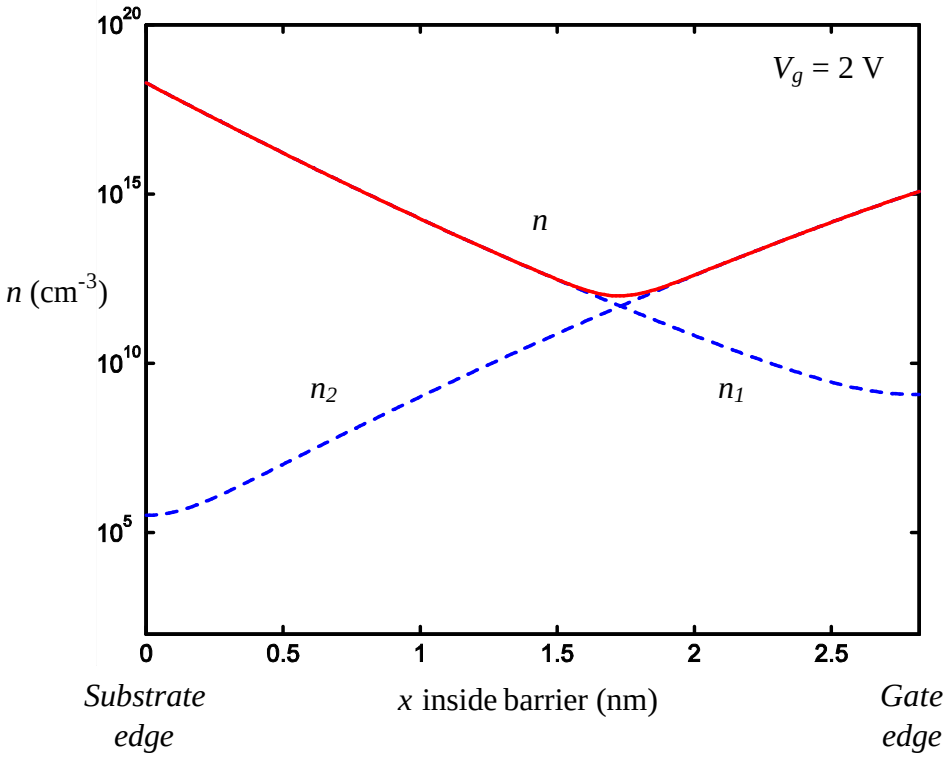

Figure 4.8 (a) The density variation inside the barrier from the solution of (4.7.2-5) for carriers from each contact in conjunction with the Poisson equation. Seen clearly is the exponential decay of n_1 and n_2 as well as the polysilicon depletion effect in the value of n_2 at the gate edge.

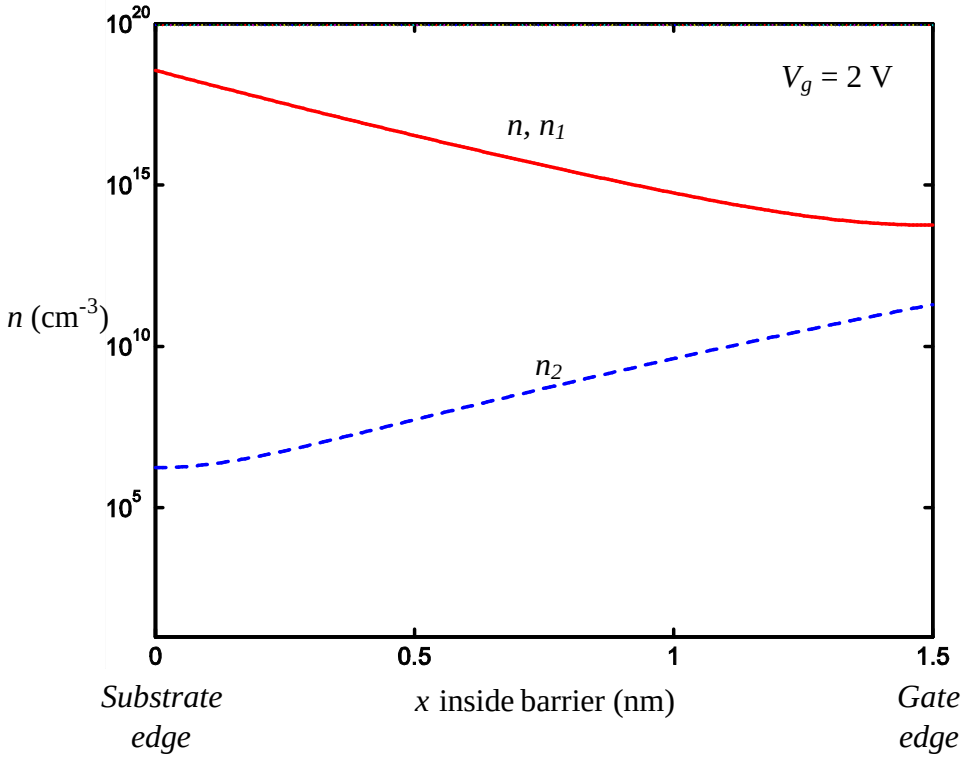

Figure 4.8 (b) Same plot as (a) for a barrier of 1.5 nm. The very high tail of the injected carriers from the substrate is seen clearly and this leads to the high tunneling current. The carrier density is no longer negligible inside the barrier.

Fig. 4.9 (a) and (b) show the tunneling currents for NMOS devices with tunneling oxide thicknesses ranging between 15 to 30 Angstroms, in the inversion regime. The results from the model described here are shown along with data from the devices in [5] and [26], so as to verify the simulation results against independently characterized data. The substrate doping concentrations for these measured devices

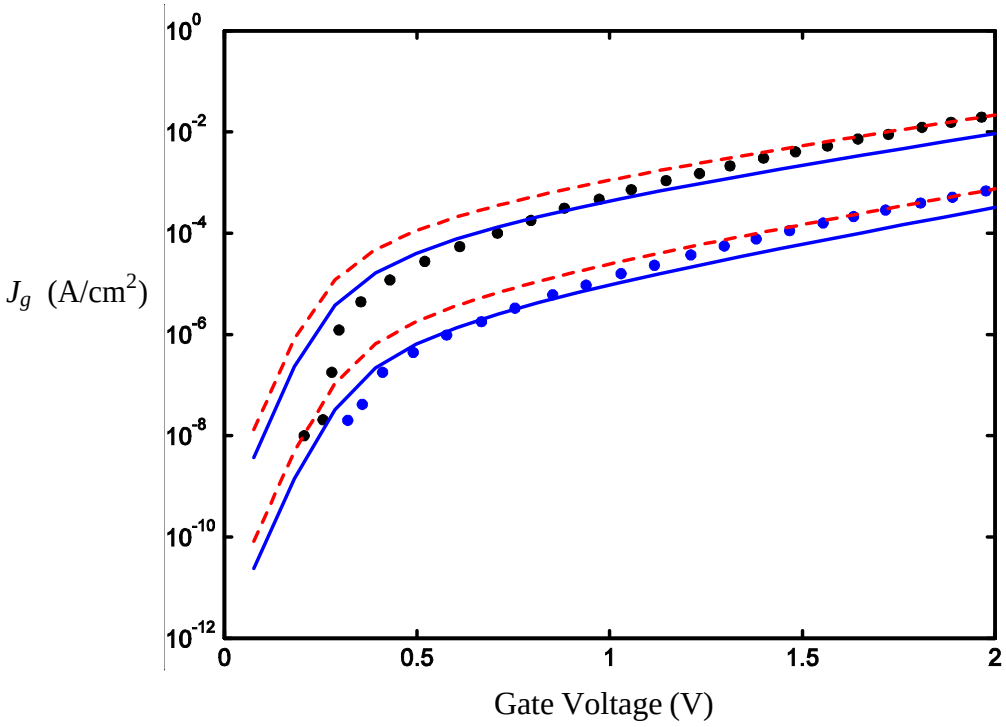

Figure 4.9 (a) Direct tunneling current densities for NMOS transistors from [5] under inversion conditions – Data is shown as dots, the dashed lines represent simulation with m_{ox} = 0.3 m_0 and the solid lines m_{ox} = 0.34 m_0. The oxide thicknesses are indicated next to the plots.

were 5×10^{17} /cc and 4.7×10^{17} /cc respectively while the corresponding gate poly dopings were 1×10^{20} /cc and 9×10^{19} /cc. The latter data suffers from oxide charge and the low voltage regime has unreliable data probably because of traps. , The plots are shown for effective masses of 0.3 m_0 and 0.34 m_0 which lie within the bounds seen in the literature [22-26].

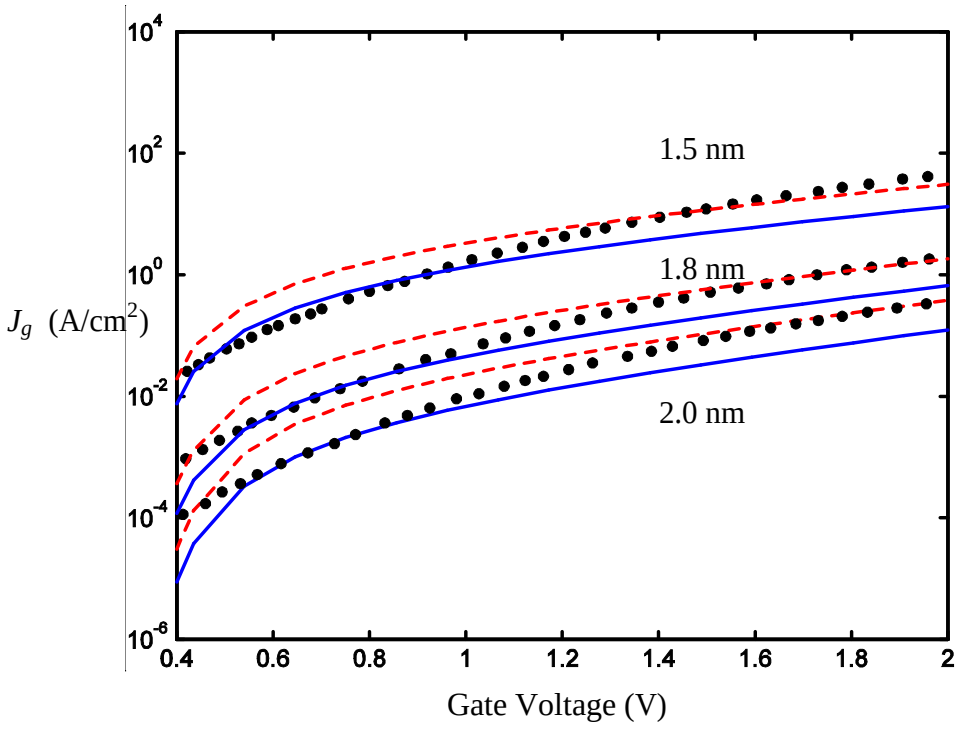

Figure 4.9 (b) Direct tunneling current densities for NMOS transistors from [26] under inversion conditions – Data is shown as dots, the dashed lines represent simulation with $m_{ox} = 0.3\ m_0$ and the solid lines $m_{ox} = 0.34\ m_0$. The oxide thicknesses are indicated next to the plots.

As is immediately seen, even the simple macroscopic approximation employed here is able to explain the tunneling data fairly well, for these devices over the entire range – this is even without taking into account explicitly the effect of quantization on increasing the average energy term in 4.7.2. The trends are also very reasonable, since the quantization energies are bound to increase at higher bias conditions and Eqn (4.7.4) the upstream boundary condition will also have to be modified to account for the field. This approximate treatment that we have shown here is somewhat similar in spirit to calculating the current as the transmission of traveling waves – similar to that method, reasonable results are obtained at either the low voltage regime, or the high

voltage regime by adjusting the tunneling effective mass [3], but not both because the effects of barrier repulsion and quantization are not included. This is typical - This also suggests one reason for the spread among the effective masses in the literature, since they depend on the particular voltage range and oxide thicknesses that simple calculations based on transmission probability approaches are fit to.

4.7.5 Modeling Inversion layer quantization effects on tunneling

We will now explore how to improve the fit for the tunneling simulations to data by including quantization effects in (4.7.2). It is obvious from (4.5.18) how one must do this. The quantization effects can be included by making the average energy of the tunneling carriers in this equation field dependent within the macroscopic approach.

If we are interested in a “local” field approximation for the quantization effect, the field at the interface alone is considered and so the quantization energies for the bound states leaking into the barrier are those for a linear potential. The average of this for a Fermi distribution must be used, as is evident from (4.5.18), but we ignore this refinement and simply use the first energy level [23] for now, similar to the correction proposed for capacitance modeling by van Dort [28]. This can easily be extended to include higher order corrections

$$\langle E \rangle \approx 1.857 \left(\frac{q\hbar}{\sqrt{m_{Si}}} F_{surf} \right)^{2/3} \tag{4.7.9}$$

The above is an estimate that has to be adjusted a bit in practice, since the effective mass normal to the interface is not a constant with bias, because of the lifting of the degeneracy of the conduction band valleys in silicon. We will simply use above value

calculated for the longitudinal effective mass for electrons and the heavy hole mass in this work, since this will lead to the lowest sub-band quantization energies.

The macroscopic tunneling calculations, performed with the simple quantization treatment above are shown in Fig. 4.10. It is evident that the fit to data is improved significantly especially in the higher-field regime, where the slope difference with bias that is not captured well in Figs. 4.9 is better reflected in the calculation with an effective mass of 0.34 m_0.

Even better models can be constructed by allowing the tunneling recombination velocity to be field dependent, as it is likely to be and by treating the

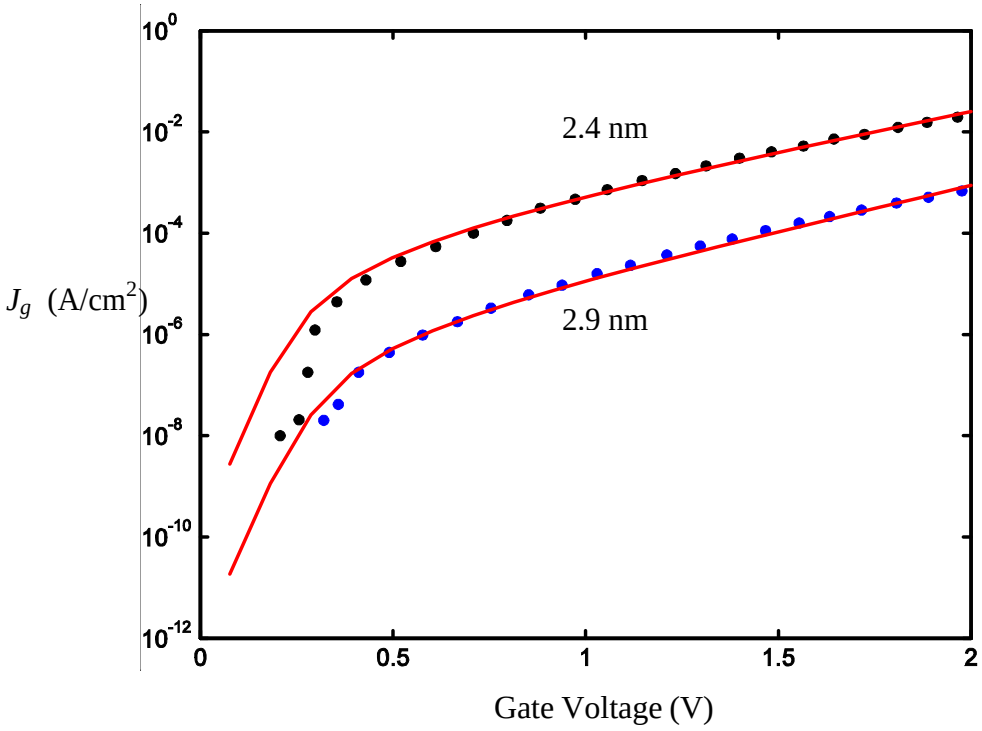

Figure 4.10 (a) Direct tunneling current densities, calculated including the quantization condition (4.7.9) in Eqn. (4.7.2), compared with data from [5]. The oxide thicknesses are indicated next to the plots.

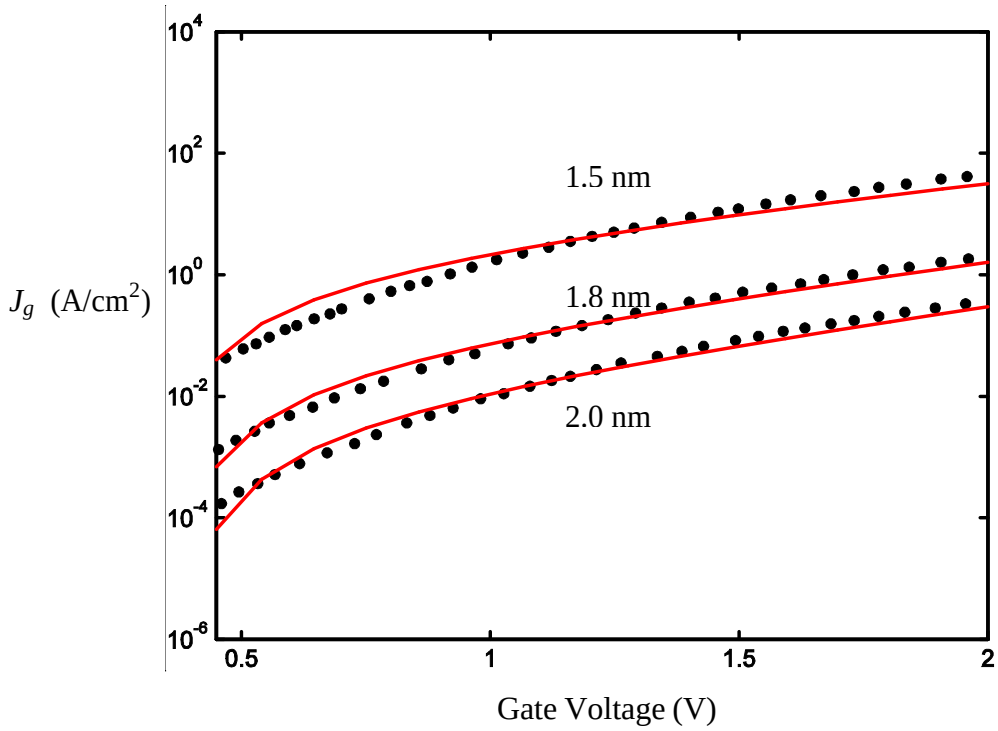

Figure 4.10 (b) Direct tunneling current densities, calculated including the quantization condition (4.7.9) in Eqn. (4.7.2), compared with data from [26]. The oxide thicknesses are indicated next to the plots.

barrier repulsion condition (4.7.2) through a better approximation based on the interface field and Fermi statistics.

4.8 SUMMARY

Tunneling is an entirely quantum mechanical phenomenon and cannot be explained by extensions of classical methods. Hence the excellent agreement to tunneling data of the very simple macroscopic simulations presented in this chapter, with a simple local field approximation to model quantization must be considered as very encouraging results for the application of such models to treat quantum effects in

semiconductors and more generally for many body systems under a mean-field approximation.

The methods presented here also hint at another important idea – the idea of non-equilibrium closures. In the microscopic sense, elastic tunneling currents between two different terminals are carried by extended states originating from each which do not mix incoherently due to scattering. This process has been modeled here by assuming that an average energy (the chemical potential) is conserved for each of the two populations across the barrier. We can think of this as a non-equilibrium closure of the hydrodynamic equations – i.e. since we know *a priori* that the process is elastic we have separated the two populations, which amounts to each half of the distribution function due to each contact, being allowed to evolve separately.

It will be worthwhile to see if this idea carries over for regions of a device where coherent effects are important - i.e. we have already seen in Chapter 1, that the Landauer formalism is equivalent to assuming that the states are completely extended in the device and are populated only because of scattering processes at the contacts, which themselves yield equilibrium conditions there. This would imply that inside a device with many terminals, as many different non-interacting gases, as there are terminals would exist, *each* with an equilibrium closure, i.e. a constant chemical potential across the device (as was done here explicitly considering the Schrödinger equation) in this chapter. We might then be able to formulate equations like (4.5.20) for each of them and solve boundary value problems to get the currents and densities.

BIBLIOGRAPHY

[1] S. E. Laux, A. Kumar and M. V. Fischetti, "Analysis of quantum ballistic electron transport in ultra-small silicon devices, including space-charge and geometric effects", *Journal of Appl. Phys.*, vol. 95, no. 10, pp. 5545-5582, May 2004.

[2] C. B. Duke, "Tunneling in solids" in Solid State Physics, Supplement 10, Academic Press Inc., 1969.

[3] L. F. Register, E. Rosenbaum, K. Yang, "Analytic model for direct tunneling current in polycrystalline silicon gate, metal-oxide-semiconductor devices", *Appl. Phys. Lett.*, vol. 74, no. 3, pp. 457-459, January 1999.

[4] M.V. Fischetti, S.E. Laux, and E. Crabbe, "Understanding hot electron transport in silicon devices : Is there a shortcut?", *Journal of Appl. Phys.*, vol. 78, no. 2, pp. 1058-1087, July 1995.

[5] W.-C. Lee and C. Hu, "Modeling CMOS tunneling currents through ultra-thin gate oxide due to conduction and valence band electron and hole tunneling", *Trans. on Electron Dev.*, vol. ED-48, no. 7, pp. 1366-1373, July 2001.

[6] S.-H. Lo, D. A. Buchanan, Y. Taur and W. Wang, "Quantum mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide NMOSFETs", *IEEE Electron. Dev. Letters*, vol. 18, no. 5, pp 209-211, May 1997.

[7] F. Rana, S. Tiwari and D. A. Buchanan, "Self-consistent modeling of accumulation layers and tunneling currents through very thin oxides", *Appl. Phys. Lett.*, vol. 69, no. 8, pp. 1105-1107, August 1996.

[8] A. Della Serra, A. Abramo, P. Palestri, F. Selmi and F. Widdershoven, " Closed and open boundary conditions for gate-current calculation in n-MOSFETs", *IEEE Trans. on Electron Dev.,* vol. ED-48, no. 8, pp. 1811-1815, August 2001.

[9] K. Hess, "Advanced theory of semiconductor devices", IEEE Press, 2000.

[10] J. Cai and C.-T. Sah, "Gate tunneling currents in ultrathin oxide metal-oxide-silicon transistors", *Journal of Appl. Phys.,* vol. 89, no. 4, pp. 2272-2285, February 2001.

[11] M.Stadele, B. Tuttle, B. Fischet and K. Hess, "Tunneling through thin oxides – new insights from microscopic calculations", *Journal of Computational Electronics,* vol. 1, no. 1-2, pp. 153-160, October 2001.

[12] E. O. Kane, "Zener tunneling in semiconductors", *Journal of Phys. Chem. Solids,* vol. 12, pp. 181-188, 1959.

[13] D. Bohm, "A suggested interpretation of quantum theory in terms of 'hidden' variables", *Phys. Rev.* , vol. 85, no. 2, pp. 166-179, January 1952.

[14] W. Kohn and L. J. Sham, "Quantum density oscillations in an inhomogeneous electron gas", *Phys. Rev.,* vol. 137, no. 6A, pp. A 1697-1706, March 1965.

[15] D. J. Griffiths, "Introduction to Quantum Mechanics",

[16] M. Brack and R. K. Bhaduri, Chapter 3, "Semiclassical Physics", Addison-Wesley Publishing Company, Inc., 1997.

[17] C. Lee, "Charge storage in nanocrystals", Ph.D Dissertation, Cornell University, May 2004.

[18] Y. Taur and T. H. Ning, "Fundamentals of modern VLSI devices", Cambridge University Press, 1999.

[19] M. G. Ancona, "Macroscopic description of quantum-mechanical tunneling", *Phys. Rev. B,* vol. 42, no. 2, pp. 1222-1233 July 1990.

[20] M. G. Ancona, Z. Yu, R. W. Dutton, P. J. Vande Voorde, M. Cao and D. Vook, "Density-gradient analysis of MOS tunneling", *IEEE Trans. Electron Dev.*, vol. ED-47, no. 12, pp. 2310-2319, December 2000.

[21] D. L. Scharfetter and H. K. Gummel, "Large signal analysis of a silicon read diode oscillator", *IEEE Trans. Electron Dev.,* vol. ED-16, pp. 64-77, January 1969.

[22] J. Maserjian and G. P. Petersson, "Tunneling through thin MOS structures : Dependence on energy E-κ", *Appl. Phys. Lett.*, vol. 25, no. 1, pp. 50-52July 1974.

[23] Z. A. Weinberg, "On tunneling in metal-oxide-silicon structures", *Journal of Appl. Phys.*, vol. 53, no. 7 pp. 5052-5058, July 1982.

[24] J. R. Chelikowski and M. Schluter, "Electronic states in α−quartz – a self-consistent pseudopotential calculation", *Phys. Rev. B*, vol. 15, no. 8, pp. 4020-4029, April 1977.

[25] G. Krieger and R. M. Swanson, "Fowler-Nordheim electron tunneling in thin Si-SiO_2-Al structures", *Journal of Appl. Phys.*, vol 52, no. 9, pp. 5710-5717, September 1981.

[26] M. Lenzlinger and E. H. Snow, "Fowler-Nordheim tunneling into thermally grown SiO_2", *Journal of Appl. Phys.*, vol. 40, no. 1, pp. 278-283, January 1969.

[27] N. Yang, W. K. Henson, J. R. Hauser and J. Wortman, " Modeling study of ultrathin gate oxides using tunneling current and capacitance-voltage measurements in MOS devices", *IEEE Trans. Electron Dev.*, vol. ED-46, no. 7, pp. 1464-1472, July 1999.

[28] M. J. van Dort, P. H. Woerlee, and A. J. Walker, " A simple model for quantization effects in heavily-doped silicon MOSFETs at inversion conditions", *Solid State Electronics,* vol. 32, no. 3, pp. 411-414, 1994.

CHAPTER 5

CONCLUSION

5.1 SUMMARY OF CONTRIBUTIONS

- A clear review of various classical and quantum transport approaches has been presented.
- The lack of explicit models for equilibrium quantum mechanical distributions has been identified as the primary reason for the ambiguity in macroscopic quantum transport approaches. A derivation of the common approximations and their limitations has been discussed.
- An analytical treatment of the boundary-layers in DG theory has been presented and a comparison to one-electron QM has been made.
- A DG description of confinement effects has been examined using analytical approximations.
- A clear derivation of the equation of state for a tunneling carrier gas has been presented.
- Tunneling, a completely quantum mechanical phenomenon has been satisfactorily described in purely macroscopic terms

5.2 SUGGESTIONS FOR FUTURE RESEARCH

Several interesting problems can be explored in the light of the work presented in this thesis on macroscopic models for tunneling.

- Extension of the tunneling formalism to include interband tunneling transitions

- A theoretical proof for the multidimensional applicability of Eqn (4.5.8) can be attempted.
- Computational experiments along the line of the tunneling model presented here in multiple dimensions, possibly on STM tips or nanocrystal memory devices.
- Non-equilibrium closure can be explored for the hydrodynamic equations, along the lines discussed in Section (4.8).
- Compact models for quantum effects in DG/UTB SOI devices based on the analytical charge profiles.