A Quantum-Mechanical Charge Transport Simulation Methodology

By Daniel Alexander Wiebe

A thesis submitted to the Faculty of Graduate Studies

at The University of Manitoba

in partial fulfillment of the requirements for the degree of Master of Science

Department of Electrical and Computer Engineering

University of Manitoba

Winnipeg, Canada

April 2012

Abstract

A method was developed for finding charge and current distribution in nanoscale electronic devices such as MOS capacitors and resonant tunneling diodes.

A system of differential equations, comprised of the Poisson and Schrödinger equations, was solved iteratively to find the electric field and charge distribution inside devices under simulation. The proposed solution method was based on the non-equilibrium Green's function approach, but expands on that approach by using spatially varying quasi-Fermi levels to construct density operators.

The proposed method was applied to several example device models. The simulation results are presented. Calculated charge distributions in FET transistors were found to have necessary features: for example, the results showed inversion layer formation. However, the calculated current-voltage curves differed significantly from published experimental results and other simulators.

Other published methods for charge transport simulation are compared to the proposed method.

Acknowledgements

Most of all, I would like to thank my advisor, Dr. Douglas Buchanan, for his incredible patience and his confidence in me; I have tested both of these beyond reason. Sincerely, thank you. The many weaknesses in this work, and the errors, which I hope are few, can be attributed directly to me; but for the insights and the errors that were not made, Dr. Buchanan must also be credited. I would also like to thank Dr. Derek Oliver and Dr. Julien Arino.

I would like to thank all of my family, but especially Carol, Melvin, and Laura; and all of my friends, but especially Pak-Yan Chan, Adam Machynia, and Chris Mazur.

I thank my colleagues and former colleagues at Teshmont Consultants, who have not only been supportive of this endeavor, but have also provided inspiration. In particular I would like to thank the following people directly: Dr. Bob Burton, Galen Lam, Rob Baker, and Dr. S. Rao Atmuri.

This work was supported indirectly by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Contents

List of Tables

List of Figures

List of Symbols

Symbol	*Units*	*Description*
$O(\varepsilon^n)$		Signifies an error term of order n in ε
$\mathbb{C}$		Set of all complex numbers
$\mathbb{R}$		Set of all real numbers
$\delta(\dots)$		Dirac or Kronecker delta function
$\mathrm{Tr}(\dots)$		Trace
$\theta(\dots)$		Step function: $\theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$
$f(\dots)$		Fermi function
L		Generic linear differential operator
Ω		Generic operator representing an observable
k		Generic integer index
Δ	$[m]$	Grid spacing
x or $\boldsymbol{x}$	$[m]$	Position
y, z	$[m]$	Other position co-ordinates
v or $\boldsymbol{v}$	$[m/s]$	Velocity
p or $\boldsymbol{p}$	$[kg \cdot m/s]$	Momentum
m	$[kg]$	Mass
t	$[s]$	Time
$\boldsymbol{F}$	$[N]$	Force
$\boldsymbol{E}$	$[N/C]$	Electric field
$\boldsymbol{D}$	$[C/m^2]$	Electric displacement field
$\boldsymbol{H}$	$[A/m]$	Magnetic field
$\boldsymbol{B}$	$[T]$	Magnetic flux density

$\boldsymbol{M}$	$[A/m]$	Magnetization
$\boldsymbol{P}$	$[C/m^2]$	Polarization
$\boldsymbol{J}$	$[A/m^2]$	Electric current density
$\boldsymbol{J}_f$	$[A/m^2]$	Free electric current density
n	$[C/m^3]$	Charge density
n_f	$[C/m^3]$	Free charge density
c	$[m/s]$	Speed of light
μ_0	$[H/m]$	Permeability of free space
μ	$[H/m]$	Permeability
ε_0	$[F/m]$	Permittivity of free space
ε	$[F/m]$	Permittivity
χ_e		Electric susceptibility
χ_m		Magnetic susceptibility
κ		Relative dielectric constant
ϕ	$[V]$	Scalar electric potential
$\boldsymbol{A}$	$[V \cdot s/m]$	Vector magnetic potential
$\hbar$	$[J \cdot s]$	Planck's constant
E	$[J]$	Energy
T	$[K]$	Temperature
k_B	$[J/K]$	Boltzmann constant
m_e^*	$[kg]$	Effective mass of electrons
m_h^*	$[kg]$	Effective mass of holes
ϑ_G	$[J]$	Band gap
ϑ_{CB}	$[J]$	Conduction band offset
ϑ_{VB}	$[J]$	Valence band offset
N_D, N_D^+	$[C/m^3]$	Density of donor sites (total, ionized)
N_A, N_A^-	$[C/m^3]$	Density of acceptor sites (total, ionized)
z_n, z_p	$[C/m^3]$	Fixed negative and positive charge

N_i	$[C/m^3]$	Intrinsic carrier concentration
n_n	$[C/m^3]$	Density of free electrons
n_p	$[C/m^3]$	Density of free holes
q	$[C]$	Electronic charge
Q^-, Q^+	$[C]$	Total amounts of negative and positive charge
U_x	$[J]$	Transverse potential
U_{yz}	$[J]$	Longitudinal potential
$\xi(x,y,z,t)$	$[m^{-3/2}]$	Total wave-function
$\vartheta(t)$		Time part of separable wave-function
$\chi(y,z)$	$[m^{-1}]$	Transverse part of separable wave-function
$\psi(x)$	$[m^{-1/2}]$	Longitudinal part of separable wave-function
ρ	$[m^{-3}]$	Density matrix
ρ_n	$[m^{-3}]$	Density matrix for electrons
ρ_p	$[m^{-3}]$	Density matrix for holes
H	$[J]$	Hamiltonian
H_0	$[J]$	Un-perturbed Hamiltonian
H_1	$[J]$	Perturbation applied to Hamiltonian
ϖ	$[J]$	Magnitude of perturbation energy
μ_n	$[J]$	Quasi-Fermi energy for electrons
μ_p	$[J]$	Quasi-Fermi energy for holes
U_n	$[J]$	Potential applicable to electrons
U_p	$[J]$	Potential applicable to holes
J	$[m/s]$	Quantum-mechanical “current” operator
j	$[s^{-1}m^{-2}]$	Probability flux

1. Introduction

1.1. Motivation

Increasing the speed of the field effect transistors (FETs) used in digital integrated circuits is a problem at the centre of present-day technology. For the past 50 years, continuous progress has been made by decreasing the size of the individual transistors that comprise integrated circuits.

The structure of a FET is similar to that of a capacitor. One main factor controlling the switching speed of a FET is the charging or discharging time associated with its gate capacitance. Another factor is the time required for charge carriers to move through the doped semiconductor region between the FET's source and drain. Like a parallel plate capacitor, the capacitance of a FET is proportional to the permittivity and area of the gate dielectric and inversely proportional to its thickness. The switching speed expected from a FET design can be increased by adjusting parameters to reduce the gate capacitance, unless the adjustments have adverse effects on the charge carrier transit time or the controllability of source-drain conductance that negate the benefit of reduced capacitance.

Reducing the area of the gate is one way to reduce the gate capacitance of a FET. Reducing the gate area implies reducing the channel length, or the distance between source and drain. If the gate area was reduced without adjusting other design parameters or operating voltages, then the electric field in the channel would be increased. Continued size reduction would eventually lead to dielectric breakdown in the channel when a potential was applied from source to drain. The FET would conduct from source to drain even when a potential was not applied to the gate. To prevent such breakdown, the applied voltages and the gate dielectric thickness are reduced when the gate area is reduced.

The limit of FET scaling that can be achieved using the standard device structures, materials and scaling techniques of the past has nearly been reached. The thickness of gate dielectrics is now counted in atomic layers, and much further reduction is not possible. Until recently silicon dioxide was used as the standard gate dielectric material for (silicon-based) FETs because of good interfacial properties, acceptable electrical properties, and ease of use in manufacturing. Now alternative dielectric materials are being used and considered, with the goal of reducing electrical equivalent thickness while increasing or maintaining physical thickness.

In order to evaluate new materials, structures, and manufacturing processes, simulation tools are invaluable. One important prediction that can be supplied by

simulation tools is the expected relation between applied gate voltage and gate leakage current that is expected from a particular design. If the leakage current is extreme, then the device will not operate correctly. Minimizing gate leakage current is often one of several FET design objectives, because high leakage current has an adverse effect by heating the device. Simulation tools can also be helpful for FET design by predicting how the physical properties of silicon-insulator boundaries in devices with alternative dielectric materials will affect the electrical operation of those devices. For example, the concentration of interfacial traps or defects can influence the gate conductivity, capacitance and threshold voltage.

As gate dielectric thickness approaches several atomic layers, it becomes inaccurate to think in terms of classical statistical mechanics and the properties of bulk matter (such as permittivity constant or effective mass) when designing devices. The fundamental nature of charge transport must now be considered in order to accurately predict device behaviour. Specifically, the quantum-mechanical behaviour of charge carriers is becoming a key aspect of charge transport calculations. Simulators based on quantum mechanics are considered necessary to obtain accurate results for devices sized on the order of 100 nm or smaller [23], [24]. A reliable model of quantum mechanical transport, validated on current devices, would be useful to analyse innovative structures.

1.2. Objective

The primary goal of this work was to develop and implement a model and methodology for predicting the flow and accumulation of electric charge in nanoscale electronic devices. The methodology was implemented in the form of a simulation tool. The intended use of the methodology was to analyze and predict the behaviour of evolutionary and innovative device structures, especially FET transistors and metal-oxide-semiconductor (MOS) capacitors. The quantities calculated in simulations were charge concentrations, voltage profiles, and current-voltage (I-V) curves.

The methodology that was developed was meant to be used to simulate device regions that are 100 nm or smaller. At the intended scale, bulk descriptions of matter are still meaningful but quantum mechanical effects are important. It was considered desirable for the model to incorporate parameters descriptive of bulk matter, such as effective mass and dielectric permittivity, so that device models did not need to be sufficiently detailed to yield these effects from first principles. However, it was also considered desirable that the same model could in principle be extended to calculate effective mass or dielectric permittivity from first principles using the same conceptual framework.

A number of competing methods exist for performing simulations of charge transport in electronic devices, some of which are briefly outlined in this thesis. One widely published method is the nonequilibrium Green's function method. That method, as applied in previously published works such as [14] and [16], constructs quantum-mechanical density operators representing the charge distributions inside devices, using Green's functions of the Schrödinger equation and state occupation probabilities calculated at the device contacts. This work attempts a similar approach using Green's functions, but uses Green's functions and state occupation probabilities at all points interior to the device to construct the density matrices. The intent of this modification is to more accurately account for the effects of particles occupying potential wells inside the devices being simulated.

1.3. Outline

In Chapter 2, a brief overview of relevant mathematical and physical modeling concepts is provided, and notation is established. The numerical solution of differential equations using finite difference methods and Green's functions is discussed. Maxwell's equations for finding the force produced by moving charge distributions are presented. The Schrödinger equation, used to find the probability density function describing the position of a charge carrier, is introduced. The

Fermi-Dirac distribution, applicable to the energy of charge carriers, is also introduced. Readers familiar with these topics need not read Chapter 2.

In Chapter 3, a description is provided of the type of device that was meant to be simulated using the methodology described in this thesis. A statement is given of the mathematical problem solved by the simulator. As Chapter 3 states, the essential problem was to obtain the density matrices describing representative charge carriers present inside electronic devices. The assumptions that were incorporated into the problem statement and the proposed solution methodology are discussed. Approaches that have been used in other works to solve the problem are briefly reviewed.

In Chapter 4, the proposed methodology for obtaining a charge carrier density matrix is presented in detail. Methods for using the density matrix to find charge concentration and flux are described.

In Chapter 5, the results of simulations that were performed for two example devices are presented. The devices discussed in chapter 5 were studied because they were similar to devices studied in other works, and therefore facilitated comparison of alternative simulation methodologies. Calculated charge distributions and current-voltage curves are shown, and compared with results from other works.

In Chapter 6, the results of simulations that were performed for example FET transistors are presented. Calculated charge distributions and current-voltage curves are shown, and compared with other results.

Finally, in Chapter 7, some concluding remarks are given. The successes and failures of the method, as compared with experiment and with other simulators, are discussed. Improvements that would be beneficial to the proposed methodology are noted.

2. Background

2.1. Mathematical Background

2.1.1. Notation

The derivative of a function $f(x)$ is denoted $f'(x)$ or simply f'. For functions that are dependent on a position variable and a time variable, the position derivative is shown as $\frac{\partial f}{\partial x} = f'$ and the time derivative is shown as $\frac{\partial f}{\partial t} = \dot{f}$.

Vector quantities are distinguished from scalar quantities by bold type: x is a scalar and $\boldsymbol{x}$ is a vector. Matrices or operators usually appear in uppercase: for example, M or Λ.

When a semicolon is used in the argument list of a function, the symbols preceding the semicolon are arguments in the usual sense, and the symbols following the semicolon are "parameters" required to construct the function. An example is $f(z;a) = \frac{1}{z-a}$. In the notation that is being used, f is a mapping $f: \mathbb{C} \to \mathbb{C}$, and would be written as $f(z)$ without the "parameter" annotation.

The Laplace operator is denoted by Δ, and the symbol δ is used to mean the Dirac or Kronecker delta function, according to context [1]. The notation $[A, B]$ is used

for the commutator of two operators or matrices, equivalently expressed as $AB - BA$.

In some sections discussing quantum mechanics, the Dirac notation is used. In this notation, the notation $|f\rangle$ is used to indicate a column vector, alternatively shown as $\boldsymbol{f}$. Its transpose conjugate is shown in the Dirac notation as $\langle f|$, or alternatively as $(\boldsymbol{f}^T)^*$.

Where the square root operator is used, it should be interpreted as yielding the principal value.

2.1.2. Numerical Solution of Differential Equations

An approximate value for the first derivative of a differentiable function $f(x)$: $\mathbb{R} \to \mathbb{R}$ can be calculated using:

$$f'_k \cong \frac{f_{k+1} - f_{k-1}}{2\Delta} + O(\Delta^2) \qquad (2.1)$$

In equation (2.1), k is an index referencing a position in a vector containing N regularly spaced samples of f, and Δ is the spacing between adjacent samples, sometimes called the "grid spacing". The index k runs from 0 to $N - 1$. The notation $O(\Delta^2)$ is used to indicate that the error in the approximation, compared to the actual value of the derivative, is proportional to the square of the sample spacing.

An approximate value for the second derivative of the same function is given by:

$$f_k'' \cong \frac{f_{k+1} - 2f_k + f_{k-1}}{\Delta^2} + O(\Delta^2) \tag{2.2}$$

Equations (2.1) and (2.2) are examples of finite-difference approximations for the first and second derivatives. Finite difference approximations for higher-order derivatives can be obtained based on Taylor series expansions, as discussed in [4].

An order-*n* finite difference approximation can be used to construct a square matrix such that the product of that matrix and a vector containing samples of a function is a vector containing samples of the approximate order-*n* derivative of that function. For example, the matrix corresponding to the second derivative operator is given by:

$$D^2 = \frac{1}{\Delta^2}\begin{bmatrix} 1 & -2 & 1 & \dots & 0 \\ & & \ddots & & \\ 0 & \dots & 1 & -2 & 1 \end{bmatrix} \tag{2.3}$$

In equation (2.3), D^2 is a finite difference approximation of the second derivative operator, and Δ is the associated grid spacing. Each row of (2.3) is an instance of equation (2.2), with *k* equal to the row number. The first and last rows cannot be supplied by equation (2.2) because for $k = 0$ and $k = N - 1$, equation (2.2) refers to vector indices that do not exist. To generalize: the finite difference

representation of an order-n derivative must be supplemented by at least n additional equations to obtain an invertible matrix. These additional equations can be supplied by boundary conditions.

The finite difference approximations of differential operators can be used to solve differential equations. In general, a differential equation given by $Ly = f$, where L is a linear differential operator, can be solved using the following procedure:

1. Find a finite difference form of L; it will be a sum of the form $\sum_{k=0}^{n} A_k D^k$, where each A_k is an invertible matrix and each D^k is the finite difference approximation of the order k differential operator.
2. Replace all incompletely specified rows of L with the left-hand sides of equations representing boundary conditions. There will be at least n such rows, but their exact number and their row indices will depend on the specific finite difference approximations used in step (1).
3. For each row in step (2), replace the corresponding row of f with the right-hand side of the same boundary condition.
4. Solve the equation $y = L^{-1} f$, after L and f have been modified, to obtain the solution y of the differential equation.

In this thesis, a finite difference method was used to solve the Poisson equation, and a finite difference method was used to obtain Green's functions for the

Schrödinger equation. The use of finite difference methods for the solution of differential equations is further discussed in [2].

2.1.3. Sparse Matrix Techniques

Sparse matrix techniques are applicable to the numerical solution of differential equations by finite difference methods [3], and were used for that purpose in this thesis.

One straightforward representation of a matrix in computer memory is as a contiguous block of numerical values stored in sub-blocks by row or column. With this matrix representation, the algorithms for basic matrix operations such as matrix-matrix multiplication or matrix-vector multiplication are straightforward. When most matrix entries are zero, many of the individual arithmetic operations performed as part of a matrix-algebraic operation have no effect on the result. For matrices with many entries equal to zero, called “sparse” matrices, a significant decrease in calculation time can be achieved when working with a matrix representation where zeros are not stored in memory.

There are a few common choices for storing the nonzero elements of sparse matrices, notably the linked list and compressed sparse row (CSR) formats [5]. In the linked list format, each nonzero value is allocated individually, and stored in a data structure along with its row, column, and pointers to the next element in its

row and/or column. A separate array stores a pointer to the data structure for the first element in each row. The linked list format is used primarily because of the ease with which elements can be added and removed, but it is not preferred for arithmetic or matrix decomposition operations. In the CSR format, all nonzero elements are stored in a contiguous block, ordered in column and then in row order. A parallel array stores column indices. A third array stores the starting index of every row, including empty ones, with empty rows indicated by a repeated or placeholder value.

The algorithms for matrix-algebraic operations are determined not only by the mathematical result that is to be obtained, but also the matrix storage format. Once a choice of representation in memory has been made, developing a naive algorithm for a sparse matrix arithmetic or decomposition operation based on the corresponding dense algorithm is typically straightforward. However, optimizations can sometimes be made that lead to significant performance improvements.

One important operation involved in the solution of linear systems, including linear systems representing finite-difference differential equations, is the operation of "LU decomposition". The LU decomposition operation, when applied to a matrix, yields an upper triangular matrix and a lower triangular matrix which have the original matrix as their product.

An important optimization is possible when performing LU decomposition on sparse matrices. In this optimization, a transformation is applied to a matrix before it is subject to LU decomposition, to re-order its rows and columns. The objective of the transformation is to re-order rows and columns such that the lower and upper triangular matrices that result from LU decomposition have reduced numbers of nonzero elements. The transformation is stored, so that it can be reversed when performing calculations with the triangular matrices. Matrix re-ordering, with the objective of reducing the number of non-zero elements in the LU decomposition results, can result in a large speedup when solving linear systems [5]. Among the notable matrix reordering algorithms are the column-approximate minimum degree ordering (COLAMD) algorithm [6] and the multifrontal algorithm described in [7].

For this thesis, sparse matrix techniques were used in all program code. Finite difference representations of differential equations were solved using sparse re-ordering and LU decomposition algorithms. These techniques were important, in that they allowed a typical 2010 desktop computer to be used to perform the calculations.

2.1.4. Solution of Differential Equations by Green's Functions

The Green's function of a linear differential operator L is an operator $G(x, x')$ that satisfies the equation [1]:

$$\int L \cdot G(x, x')y(x)dx = \int \delta(x - x')y(x)dx = y(x') \tag{2.4}$$

In equation (2.4), $y(x)$ stands for any function to which the operator L can be applied. The sign of the delta function is a matter of convention. The convention used in this thesis, that the delta function has a positive sign, is also used in works such as [1] or [26]; but the convention that it has a negative sign is sometimes seen in physics texts such as [25].

The Green's function for the operator L may be used to solve the differential equation $Ly = f$.

Since $Gy = Gf$ and L is a linear differential operator, the differential equation immediately gives $L(Gy) = Gf$. Integrating both sides of that equation yields:

$$\int L \cdot G(x, x')y(x)dx = \int G(x, x')f(x)dx$$

The definition (2.4) is applied to obtain:

$$y(x') = \int G(x, x')f(x)dx$$

The differential equation has thus been transformed into an integral equation, and the solution can be obtained by performing the integration, provided that the Green's function $G(x, x')$ is known.

By analogy, the Green's function of a finite, invertible $N \times N$ matrix L is the matrix that satisfies $\langle x|LG|x'\rangle = \langle x|x'\rangle$: that is, L^{-1}. L may represent a differential operator. When it does, the solution of the differential equation $L|y\rangle = |f\rangle$ is $|y\rangle = G|f\rangle$. As noted in Section 2.1.2, a discrete differential operator must be supplemented by boundary conditions in order to be invertible, and therefore, for its Green's function to exist.

The Green's function concept is used in the construction of a multi-particle density operator in Chapter 4.

2.2. Electrodynamics

2.2.1. Maxwell's Equations

Maxwell's equations are a system of differential equations that can be solved to calculate the forces acting on a charge distribution, based on the charge distribution itself. Thus, the time evolution of the positions of mobile elements of the charge distribution can be predicted. Maxwell's equations in free space are [25]:

$$\nabla \cdot \boldsymbol{E} = n(\boldsymbol{x}, t)/\varepsilon_0 \tag{2.5}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{2.6}$$

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{2.7}$$

$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J} + \mu_0 \varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} \tag{2.8}$$

In equations (2.5) to (2.8), n is the distribution of electric charge, $\boldsymbol{E}$ is the electric field, $\boldsymbol{B}$ is the magnetic flux density, and $\boldsymbol{J}$ is current density. Based on these equations, the current conservation law can be derived:

$$\frac{\partial n}{\partial t} = -\nabla \cdot \boldsymbol{J} \tag{2.9}$$

Maxwell's equations as stated in (2.5) to (2.8) apply to charge distributions that exist in vacuum and are free to move. It is convenient to modify the equations to describe a population of charge consisting of some free charge and some fixed charge, located in an isotropic, linear, and homogeneous medium such as a crystalline solid. In this context, it is convenient to supplement the original equations with some auxiliary equations defining other quantities, to arrive at an expanded system of equations serving the same purpose as (2.5) to (2.8):

$$\nabla \cdot \boldsymbol{D} = n_f \tag{2.10.1}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{2.10.2}$$

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{2.10.3}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J}_f + \frac{\partial \boldsymbol{D}}{\partial t} \tag{2.10.4}$$

$$\boldsymbol{D} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P} \tag{2.10.5}$$

$$\boldsymbol{H} = \frac{\boldsymbol{B}}{\mu_0} - \boldsymbol{M} \tag{2.10.6}$$

$$\boldsymbol{P} = \varepsilon_0 \chi_e \boldsymbol{E} \tag{2.10.7}$$

$$\boldsymbol{M} = \chi_m \boldsymbol{H} \tag{2.10.8}$$

In equations (2.10.1-8), $\boldsymbol{D}$ is the electric displacement field, $\boldsymbol{H}$ is the magnetic field, $\boldsymbol{M}$ is magnetization, and $\boldsymbol{P}$ is polarization. The "*f*" subscript attached to the current and charge distributions indicates "free charge" and current arising from its motion. Constants are listed in the table of symbols.

When bulk dielectric matter is considered, the material-specific constant $\varepsilon = \varepsilon_0(1 + \chi_e)$ is defined. Using this definition, $\boldsymbol{D} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P} = \varepsilon \boldsymbol{E}$.

This section has just served to introduce and state Maxwell's equations. An introduction to electrodynamics can be found in Griffiths [27], and a fuller discussion can be found in Jackson [25].

2.2.2. The Poisson Equation: Quasi-static Charge Distributions

The scalar electric potential and the vector magnetic potential are abstract quantities defined for the purpose of increasing the ease with which Maxwell's equations can be solved. The scalar electric and vector magnetic potentials are defined by the system of equations:

$$\boldsymbol{E} = -\nabla\phi - \frac{\partial \boldsymbol{A}}{\partial t} \tag{2.11}$$

$$\boldsymbol{B} = \nabla \times \boldsymbol{A} \tag{2.12}$$

Equations (2.11) and (2.12) are insufficient to uniquely specify the vector magnetic potential. An additional equation, called a "gauge", is required to ensure that $\boldsymbol{A}$ is unique.

If the Coulomb gauge is selected, meaning that the condition $\nabla \cdot \boldsymbol{A} = 0$ is imposed on the potentials, or if the magnetic potential is negligible, then the Poisson equation can be obtained.

From equation (2.10.1),

$$\nabla \cdot \boldsymbol{D} = \nabla \cdot \varepsilon \boldsymbol{E} = n_f$$

Therefore, using equation (2.11),

$$\nabla \cdot \varepsilon \boldsymbol{E} = -\nabla \cdot \left[\varepsilon \left(\nabla\phi + \frac{\partial \boldsymbol{A}}{\partial t}\right)\right] = -\nabla \cdot \varepsilon \nabla\phi - \frac{\partial(\nabla \cdot \boldsymbol{A})}{\partial t} = n_f$$

If $\boldsymbol{A} \cong 0$ or $\nabla \cdot \boldsymbol{A} = 0$, then

$$-\nabla \cdot \varepsilon \nabla \phi = n_f \qquad (2.13)$$

Equation (2.13) is the Poisson equation. If the material to which (2.13) is applied is properly homogeneous, meaning that the dielectric constant has no position dependence, then the Poisson equation can be further simplified to $-\varepsilon \Delta \phi = n_f$. However, when dielectric constant is position-dependent, equation (2.13) is sometimes used as stated.

2.3. Quantum Mechanics

2.3.1. Concepts from Classical Mechanics

The subject of "classical mechanics" encompasses rules and conservation laws that describe the motion of particles or physical objects in terms of differential equations. Quantum mechanics and relativity can be viewed as refinements of classical mechanics that apply at scales of mass, momentum and distance at which precise measurements are not possible in everyday experience.

The Hamiltonian is an important concept in both classical and quantum mechanics. The classical Hamiltonian is a mathematical expression that describes a system of particles: it yields the total energy of the particle system at any instant in time. A Hamiltonian is constructed using two state variables per particle and degree of freedom. The state variables for a particle are typically time-parametric

functions yielding position and momentum. State variables for all particles must yield positions and momenta in a consistent coordinate system, but that coordinate system need not be Cartesian.

For a system comprised of N particles with one degree of freedom, Hamilton's equations are a system of $2N$ equations that can be solved to obtain the positions and momenta of the particles at any instant in time. Hamilton's equations are given by:

$$\frac{dp_k}{dt} + \frac{\partial H}{\partial q_k} = 0 \qquad (2.14)$$

$$\frac{dq_k}{dt} - \frac{\partial H}{\partial p_k} = 0 \qquad (2.15)$$

In equations (2.14) and (2.15), k is a particle index that ranges from 1 to N; and the position and momentum of particle k are denoted by $q_k(t)$ and $p_k(t)$ respectively.

This section contains a minimal introduction to the concept of the Hamiltonian that is required to discuss the Schrödinger equation. The subject of classical mechanics is developed in Goldstein [29] or Landau and Lifshitz [28], and chapter 2 of Shankar [9] introduces the subject in the context of quantum mechanics.

2.3.2. The Schrödinger Equation

Quantum mechanics is a set of rules for transforming a Hamiltonian, constructed according to the principles of classical mechanics, into a differential equation that yields wave-functions when solved.

One feature of quantum mechanics is that exact trajectories and momenta, as found in classical mechanics, cannot be defined. As per the Heisenberg uncertainty principle (e.g. [9], ch. 9), position and momentum cannot simultaneously be determined with arbitrary precision. However, instead of precise trajectories, quantum mechanics provides a means to calculate position and momentum probabilistically.

The square of the magnitude of a single-particle wave-function expressed in the position basis is a probability density function (PDF) describing the position of the particle. Provided that the wave-function has been normalized, the PDF may be integrated over a subspace to find the probability that the particle has a position within that subspace. Similarly, the square of the magnitude of the same wave-function, when the wave-function is expressed in the momentum basis, is a probability density function describing the particle's momentum.

The principles of quantum mechanics allow for operators to be defined that represent observables other than position and momentum, and for probabilistic

calculations to be performed using such operators [9]. This procedure is discussed further in Chapter 4: an operator is introduced that measure electrical current when operating on wave-functions that describe charged particles.

The rules set out in quantum mechanics for transforming a classical Hamiltonian expressed in Cartesian coordinates into a differential equation in position space that yields a wave-function are simple. Before the rules can be applied, the Hamiltonian must be expressed using momenta only, and not velocities. With this condition met, the following substitutions may be made:

$$\begin{aligned} &\text{Position} && x && \rightarrow && x \\ &\text{Momentum} && p && \rightarrow && -i\hbar\frac{\partial}{\partial x} \\ &\text{Total energy} && E && \rightarrow && i\hbar\frac{\partial}{\partial t} \end{aligned}$$

Two operators which have the same effect on the wave-function are obtained: on one hand by making these substitutions in the classical Hamiltonian; and on the other hand by using the total energy operator directly. The equation showing the equality of their results is the Schrödinger equation for the system under consideration.

A typical example of the use of a Schrödinger equation is in the description of a particle with one degree of freedom, moving in an external potential. The classical Hamiltonian describing one such particle is given by:

$$E = \frac{p^2}{2m} + u(x) \tag{2.16}$$

In equation (2.16), the total energy, momentum and mass of the particle are given by E, p, and m respectively, and its potential is given by u. By applying the substitution procedure discussed in this section to equation (2.16), the following Schrödinger equation is obtained:

$$i\hbar\dot{\psi} = -\frac{h^2}{2m}\psi'' + U(X)\psi \tag{2.17}$$

Equation (2.17) is similar to the equation that is proposed in Chapter 3 to model the motion of a charge carrier in an electric field for the purpose of charge transport calculations.

2.3.3. The Density Operator

From any set of quantum numbers describing a particle, the Schrödinger equation yields exactly one wave-function. In the case of a simple Schrödinger equation such as (2.17), in which spin is not considered, only the particle energy is required to obtain the wave-function.

The expected value of an observable Ω of a particle, such as the position or momentum of the particle, is given by $\langle\psi|\Omega|\psi\rangle$, where ψ is the wave-function of the particle [9]. If the particle's precise energy, or state, is known at the time of measurement, then its Schrödinger equation yields its wave-function, and the observable's expected value can be calculated.

Usually, a particle is not known to be in a specific, individual state. More typically, the particle may be described by a probability distribution assigning finite or infinitesimal probabilities to a series or spectrum of possible particle states. In this circumstance, the particle is appropriately described by a density operator. The following table compares wave-functions and density operators.

	Wave-function	*Density operator*
Definition	$\|\psi_E\rangle$	$\rho = \sum_k w_k \|\psi_k\rangle\langle\psi_k\|$
Expected value of an observable	$\langle\psi_E\|\Omega\|\psi_E\rangle$	$\mathrm{Trace}(\rho\Omega)$
Probability density function	$\|\psi_E\|^2$	$\langle x\|\rho\|x\rangle$

Any wave-function ψ must be normalized, by scaling it such that $\langle\psi|\psi\rangle = 1$, before it is used to calculate probabilities. Similarly, density operators must satisfy normalization conditions. Each wave-function ψ_k from which a density operator is constructed must individually satisty the requirement that $\langle\psi_k|\psi_k\rangle =$

1. Additionally, the "weights" must individually satisfy $0 \leq w_k \leq 1$, and they must collectively satisfy $\sum_k w_k = 1$.

In addition to being useful when describing a single particle in an unknown state, the density operator can be used to describe an ensemble of non-interacting particles that obey the same Schrödinger equation. In this case the probability of finding any member of the ensemble in each possible state must be known in order to construct the density operator. The difference between these cases is not in the mathematical representation of the operator, but in the meaning of the coefficients w_k that are the weights for each wave-function.

Further discussion of the density operator is found in chapter 3 of Sakurai [8].

2.4. Solid-State Physics

2.4.1. Maxwell-Boltzmann and Fermi-Dirac Statistics

The simulator developed for this thesis represents charge carrier populations using density matrices. As discussed in the previous section, the density operator representing a particle, such as a charge carrier, can be constructed from wave-functions that are solutions to the Schrödinger equation describing the particle. For each state that contributes to the density operator, a weight must be determined. The magnitude of each weight is proportional to the probability of

finding the particle in the corresponding state. States, or wave-functions, are each associated with a specific value of particle energy.

The Maxwell-Boltzmann and Fermi-Dirac distributions are statistical distributions of particle energies, and therefore they are useful when constructing density matrices.

According to Maxwell-Boltzmann statistics, the distribution of particles in momentum for a particle that is free to move in three dimensions is [30]:

$$f(\boldsymbol{p}) = (2\pi k_B T)^{-1.5} \exp\left(-\frac{\boldsymbol{p}\cdot\boldsymbol{p}}{2mk_B T}\right) \tag{2.18}$$

In equation (2.18), $\boldsymbol{p}$ is momentum, T is temperature, m is particle mass, and k_B is the Boltzmann constant.

As a consequence of (2.18), kinetic energy is distributed according to:

$$f(E) = \sqrt{\frac{E}{\pi k_B T}} \exp\left(-\frac{E}{k_B T}\right) \tag{2.19}$$

In equation (2.19), E is kinetic energy per degree of freedom.

Whereas Maxwell-Boltzmann statistics are applicable to bosons, electrons are Fermions, and should consequently be described by Fermi-Dirac statistics. The reason for this is that the Pauli exclusion principle is not considered in the derivation of the Maxwell-Boltzmann distribution [31].

The Fermi-Dirac distribution in energy, yielding the probability that an energy state E will be occupied, is given by [31]:

$$f(E) = \frac{1}{1 + \exp\left(\frac{E - \mu}{k_B T}\right)} \tag{2.20}$$

In equation (2.20), μ is the Fermi energy or "Fermi level". It is a parameter whose value is "determined by the condition that the number of occupied states at all energies must be equal to the total number of electrons present" [31].

The Maxwell-Boltzmann and Fermi-Dirac distributions are derived in Pathiria [32].

Fermi-Dirac statistics are commonly used for quantum-mechanical transport calculations [33]. They are used in Chapter 4 of this thesis when constructing the density matrices.

3. Modeling Concepts

3.1. Problem Summary

The purpose of this work was to develop a method for calculating the charge and current distributions inside a nanoscale electronic device. The method was designed to be applicable to devices such as field effect transistors (FETs) or metal-oxide-semiconductor (MOS) capacitors. It was considered necessary to use quantum mechanics, rather than classical mechanics, to predict the motion of charge carriers. Other works have shown that predicting the gate tunneling current of a nanoscale FET requires a quantum-mechanical approach.

3.2. Device Model

The purpose of this section is to establish how devices are represented mathematically when formulating calculation procedures that are meant to be applied to them.

A device was conceived of as occupying a simply connected subspace V_D of $\mathbb{R}^3$ bounded by the surface S_D.

Devices were modeled as containing two populations of free charge carriers: electrons, and virtual particles called holes that carry positive charge. The virtual

particle concept is described further in Section 3.4.3. Charge carriers were modeled as moving through the device, thus participating in the flow of electric current.

The construction of a device was described using a series of functions defined on V_D that gave material properties inside the device. The functions required to describe a device are listed in Table 1.

Table 1: Device Material Properties

$m_e^*(x), m_h^*(x)$	Effective masses of electrons and holes
$\varepsilon(\boldsymbol{x}), \mu(\boldsymbol{x})$	Permittivity (dielectric constant) and permeability
$\vartheta_G(\boldsymbol{x})$	Band gap
$\vartheta_{CB}(\boldsymbol{x})$, $\vartheta_{VB}(\boldsymbol{x})$	Conduction and valence band offset, respectively
$N_D(\boldsymbol{x})$, $N_A(\boldsymbol{x})$	Concentration of donor and acceptor dopants
$N_i(\boldsymbol{x})$	Intrinsic carrier concentration

The donors and acceptors referred to in Table 1 are particles, embedded in the substrate material of the device, that can be expected to be ionized. As such, they give rise to a concentration of fixed charge.

The simulator that was implemented was based on a model in which the only forces that acted on charge carriers, thus causing them to move, were the electric and magnetic forces. The force that acts on an individual charge carrier with charge q and velocity $\boldsymbol{v}(\boldsymbol{x})$, in the presence of an electric field $\boldsymbol{E}$ and a magnetic field $\boldsymbol{B}$, is the Lorentz force:

$$\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})$$

The simulator was developed for conditions in which $\boldsymbol{E} \gg \boldsymbol{v} \times \boldsymbol{B}$, and the magnetic field could therefore be neglected.

Devices were modeled as being in contact with one or more reservoirs. Like the device, each reservoir was a simply connected subspace of $\mathbb{R}^3$. Reservoirs were modeled as being in contact with the device only at the device surface, and were not in contact with each other, except indirectly via the device. Every point on a device surface S_D was a point on the surface of exactly one reservoir.

Reservoirs did not have the properties that were defined in the device region. Reservoirs were considered to be conductors, and were therefore equipotential regions. The potentials of the reservoirs were required as input parameters for each calculation. These potentials were required to be time-independent.

3.3. General Problem Formulation

In this section, a general method is formulated for finding the charge distribution inside a device. The general procedure described in this section is the basis for the calculation method developed in this thesis, as well as for a number of methods developed in other works. Specific implementations developed in this thesis and in other works differ in part because of assumptions or details not specified in this general formulation, in part because of the numerical methods used, and in part because of differences in formulation as compared to the general formulation presented here.

The general method followed by this and other works involves the solution of the coupled Poisson and Schrödinger equations. Because the Poisson and Schrödinger equations are coupled, they are often solved iteratively.

The Poisson equation may be solved to find the scalar electric potential inside a device. The Poisson equation cannot be solved unless the charge distribution is known in the solution region. The Poisson equation is given by:

$$\frac{d}{dx}\left\{\varepsilon(x)\frac{d\phi}{dx}\right\} + n(x) = 0 \qquad (3.1)$$

In equation (3.1), the scalar electric potential is denoted ϕ and the total charge distribution is denoted n. The total charge distribution applicable to the Poisson equation is given by:

$$n(x) = n_n + n_p + N_A + N_D \tag{3.2}$$

In equation (3.3), n_n is the concentration of free electrons, n_p is the concentration of free holes, and N_D and N_A are the concentrations of donors and acceptors, which are two of the material properties listed in Table 1 as being part of the device description.

The reservoir potentials, together with the condition that the reservoirs are equipotential and the condition that every point on S_D is a point on the surface of exactly one reservoir, are sufficient information to provide a boundary condition for equation (3.1). The electrostatic potential on the device surface is denoted $\phi_S(x)$ for all $x \in S_D$. The boundary condition used to solve equation (3.1) is given by equation (3.3):

$$\phi(\boldsymbol{x}) = \phi_S(\boldsymbol{x}) \text{ for } \boldsymbol{x} \in S_D \tag{3.3}$$

Since one of the quantities being calculated is the charge distribution inside the device, and the charge distribution is required to solve the Poisson equation, an iterative method may be required to solve equation (3.1). For each iteration, the Poisson equation is solved to find the scalar electric potential based on the charge

distribution, and Schrödinger equations are solved to find the charge distributions of the populations of electrons and holes using the scalar electric potential.

The probability density functions (PDFs) of the positions of charge carriers inside the device are calculated using Schrödinger equations. Two Schrödinger equations need to be solved: one describing holes, and one describing electrons. The Hamiltonian used to construct each Schrödinger equation includes only the kinetic energy and the electrostatic potential energy of a single representative particle. The Schrödinger equations applicable to electrons and holes are given by equations (3.4) and (3.5) respectively:

$$-\frac{\hbar^2}{2m_e^*}\Delta\,\xi_e - \xi_e \cdot q\phi = i\hbar\frac{d\xi_e}{dt} \tag{3.4}$$

$$-\frac{\hbar^2}{2m_h^*}\Delta\,\xi_h + \xi_h \cdot q\phi = i\hbar\frac{d\xi_h}{dt} \tag{3.5}$$

In equations (3.4) and (3.5), the wave-functions for electrons and holes are given by ξ_e and ξ_h respectively, the effective masses of electrons and holes are m_e^* and m_h^* respectively, and the electric potential is ϕ.

Equations (3.4) and (3.5) are partial differential equations with both space and time parts. It was assumed that ϕ is not time-dependent, and therefore that (3.4)

and (3.5) are separable with respect to time. When the Schrödinger equation is separable, the separation constant is interpreted as an energy, denoted E.

The solutions to (3.4) and (3.5) are wave-functions of single particles. However, all charge carriers moving in a device are in the presence of the same electric potential ϕ. Therefore, any individual carrier has a wave-function that is a solution to (3.4) or (3.5). However, as a consequence of the exclusion principle, each solution must correspond to a distinct energy, except inasmuch as degeneracy is permitted.

Each solution to equation (3.4) or (3.5), corresponding to a particular energy, is required to satisfy the continuity equation for probability flux,

$$\frac{d|\xi|^2}{dt} - \nabla \cdot \frac{i\hbar}{2m}(\xi^* \nabla \xi - \xi \nabla \xi^*) = 0 \qquad (3.6)$$

The probability flux continuity equation (3.6) may be applied at the device boundaries to obtain boundary conditions for the Schrödinger equations. In equation (3.6), ξ is a wave-function such as ξ_e or ξ_h, and m is the applicable mass, such as m_e^* or m_h^*.

In this thesis, solutions to the Schrödinger equations (3.4) and (3.5) were obtained by constructing them from Green's functions, rather than directly solving the

Schrödinger equations subject to a condition derived from (3.6). Chapter 4 describes the selected approach in detail.

Wave-functions that are found as solutions to (3.4) and (3.5) may be used to determine the concentrations of free electrons and holes, n_n and n_p. The following general procedure may be used to find n_n:

1. Solve (3.4) for each energy in a series of discrete samples of energy.
2. Normalize each wave-function $\xi_e(x; E)$, where E is the wave-function's energy, such that $\int_{V_D} |\xi|^2 dV = 1$ (where dV is the differential volume element).
3. Calculate a weighted sum of the outer products of wave-functions, where the weight assigned to each wave-function is a function of energy. Obtain a density operator by normalizing the resulting matrix.
4. Multiply the density operator by the total amount of charge in the device's population of free electrons, and take the diagonal to obtain the free electron concentration, n_n.

An analogous approach is used to find the free hole concentration, n_p.

The calculation in step (3) above may be performed using equation (3.7):

$$\rho_e(x) = \sum_{E_k} w(E_k)|\xi_e(x; E_k)\rangle\langle\xi_e(x; E_k)| \qquad (3.7)$$

In equation (3.7), w is the weighting function, E_k refers to an energy in the discrete set, and $\xi_e(x; E_k)$ is the wave-function corresponding to that specific energy. The summation runs over all energies for which the Schrödinger equation is solved.

The approach used to derive the weighting function applicable in (3.7) is one of the key differences between the various methods that have been developed for performing calculations of this type. The weighting function is often considered to be composed of two simpler functions: the density of states, and the probability of state occupation. The density of states is the eigenvalue spectrum of the linear differential operator representing the spatial part of the separated Schrödinger equation that was solved to find the wave-functions being weighted. The probability of state occupation is usually given by the Fermi function, as discussed in Chapter 2.

One difficulty inherent in evaluating the weighting function is that of deriving the Fermi energy, required to evaluate the Fermi function. The device model described in the previous section allows for a device to be in contact with several reservoirs, and each reservoir when considered in isolation may have a Fermi

energy different from those of the other reservoirs. In this circumstance, it is not obvious how to find the Fermi energy applicable inside the device when evaluating the weighting function. This work proposes a method that uses a spatially varying "quasi-Fermi energy", as described in Chapter 4.

Another difficulty inherent in evaluating the weighting function is that of finding the eigenvalue spectrum of the energy operator, or finding a suitable approximation. One common approach is to use the eigenvalue spectrum applicable to a free particle moving in a flat potential.

Once n_n and n_p have been found, they may be fed back into equation (3.2), and the process of solving the coupled Schrödinger-Poisson system may be repeated.

3.4. Modelling Assumptions

The general problem formulation presented in the previous section depended upon a number of assumptions and simplifications that may lead to simulated results that differ from reality. Furthermore, the simulator as implemented embodied some assumptions beyond those implicit in that general problem formulation, such as the assumption that a one-dimensional device representation could be used. The following sub-sections discuss the assumptions and simplifications that were made when developing the general problem formulation, or the specific calculation method presented in this thesis.

3.4.1. The Assumption of a Steady State

The models developed in this work require that the potentials applied to device surfaces are not functions of time. Furthermore, it is assumed that any device being simulated has reached a steady state, and consequently that the electrostatic potential inside the device is not time-dependent.

For time-invariant potentials, the Schrödinger equation is separable and wave-functions that solve it have time dependence of the form $\exp(-iEt/\hbar)$. Because the potential inside the device was assumed to be time-invariant, the requirement to specify an initial condition in the form of a charge distribution was avoided. The initial condition was assumed to be the steady state condition. If the simulator had been designed to allow for time-varying boundary or interior potentials, then the calculation of steady-state current would require an initial charge distribution to be supplied, and the time evolution of this distribution would need to be calculated for a number of discrete time steps to find the eventual steady-state behaviour of the device. The amount of computation effort that this would require was seen as prohibitive.

Counter-examples may be constructed to show that even if the boundary potential applied to a device is time-invariant, the interior potential may not be time-invariant. One simple counter-example is the case of a charge distribution at $t = 0$ consisting of a Gaussian wave-packet representing a single charge carrier

with momentum approaching zero, located near a boundary, and boundary potentials such that the particle experiences a strong repulsive force (approaching infinity) near any boundary. For this example, a time-dependent simulator would yield an oscillatory potential inside the device, but a time-independent simulator may not fail to yield a result. The result that would be obtained from a simulator using the steady-state assumption, if a result were obtained, would not be accurate.

The steady state assumption may have practical consequences, in that some devices studied in other works depend upon the Coulomb blockade effect for their operation. The assumption of a time-independent interior potential makes the simulator unable to correctly model some device behaviours, such as those arising from Coulomb blockade. In such devices, a stable "steady state" does not exist.

3.4.2. Consideration of Edge Effects

Devices in this work were modeled as being planar. Current was modeled as travelling in one dimension. That dimension is called the longitudinal dimension. Devices were modeled as being spatially invariant in the other two dimensions. Those dimensions are called the transverse dimensions. Spatial invariance in the transverse dimensions had the consequence that the modeled devices were infinite in volume. In reality, physical devices can be quite small in the transverse dimensions. The materials beyond the transverse boundaries have different

electrical properties, such as permittivity or dopant concentration, compared to the device region, and these varying properties give rise to a non-uniformity of longitudinal current per unit area. The effects of the materials beyond the device boundaries on current per unit area inside the device are referred to as "edge effects".

Edge effects are neither modeled nor quantitatively estimated for the results presented in this thesis. However, the chosen approach of treating devices as infinite in extent and calculating current per unit area is a common one. It has been used in many similar works.

When a method for calculating current in a device is formulated in a way that allows for the device's material composition to described in one dimension only, the method itself is referred to as one-dimensional. The method developed in this thesis is thus referred to as one-dimensional. However, despite the method being nominally one-dimensional, the process by which a particle exchanges momentum in the transverse and longitudinal directions of travel is actually a central feature of the simulator's operation. This process is discussed further in Chapter 4.

3.4.3. The Use of Virtual Particles

In semiconductors it is usually considered that there are two main mechanisms for charge transport: the motion of holes, and the motion of electrons.

Electrons are point particles, but the principles of quantum mechanics require that calculations involving the positions of electrons describe the electrons' positions using time-parameterized probability density functions (PDFs), rather than time-parameterized positions (trajectories). When calculating the expected value of an observable that depends on the positions of several electrons, two common approaches exist. One approach is to work with a joint probability density function representing multiple electrons. The other approach is to use a density operator formed by summing the probability density functions of all contributing electrons. In this approach, the normalization condition $Tr(\rho) = 1$ that is usually applied to a density operator ρ is modified: $Tr(\rho) = N$, where N is the number of particles represented by the operator, is used instead. The multi-particle density operator approach is used.

Holes are sometimes called virtual particles. A hole is a local maximum in charge density that arises when an electron is removed from a charge-neutral region, causing the region to have a positive charge. The concept of a hole is easiest to understand when the expected position of the hole is near an atom in a conducting or semiconducting solid. When an atom is ionized at the first ionization level, one

of the electrons that is normally bound to the atom has been removed. The region around the ionized atom has a net positive charge because that electron is missing. The amount of charge imbalance is equal but opposite in sign to the charge of an electron. The region of positive charge that arises because the atom is ionized is called a hole.

Although the ionized atoms that contain the positive charge necessary for holes to exist may be stationary, a theory of hole transport can nevertheless be developed. An event in which an electron is transferred to an ionized atom from an un-ionized neighbour in a semiconducting crystal could be described as a hole moving the inter-atomic distance. A series of such events is one process by which hole transport can occur.

Unlike electrons, holes are not point-like. For example, a hole may arise as the result of the contributions of several partially ionized atoms. The mathematical description of a hole can take the form of a charge density function, defined on a region in which a solid exists. This function has the property that when integrated over its domain, the result is the total amount of charge represented by the hole; that is, $+q$. The charge density function may have a time parameter, thus allowing for hole transport to be described. This representation allows for hole transport to be described not only in insulators, where the paradigm of a sequence of electron exchanges between ions and adjacent non-ionized atoms is a

physically reasonable description of hole transport, but also in semiconductors or metals, where that paradigm is unrealistic.

Holes are represented as virtual particles by making an analogy between the charge distribution function describing a hole and the probability density function describing a point-like particle. It is considered that the probability density function, multiplied by the electronic charge, is equal to the charge density function. This PDF is considered to have been generated by a Schrödinger equation, in the same way as the PDF of the position of an electron. The Schrödinger equation contains a kinetic energy term, an element of which is the mass of the particle described by the equation. It could be said that a hole, being a virtual particle, does not have a mass. However, it is common to ascribe a material-dependent "effective mass" to holes, used in place of the mass of a physical particle. The effective mass is phenomenologically determined based on the principle that the same applied field should cause the same time-evolution of the PDF of a hole and the PDF of an electron.

3.4.4. Parabolic Dispersion and Effective Mass

One application of the concept of effective mass, as described in the previous sub-section, is to the kinetics of *virtual* particles described by a Schrödinger equation. However, the same concept has another important application. An effective mass is used to model the difference in rate of acceleration that is observed to result

from a force acting on a physical particle, when the particle in a material versus when the particle is in free space. Effective mass is applicable to an electron being accelerated in a solid by an external electric field. Effective masses of electrons and holes have been measured for many materials by comparing the effect of a field on charge in each material to the effect that the same field would have on charge travelling in empty space. Effective mass can be higher or lower than free-space mass.

The effective mass of a particle moving within a crystalline solid is dependent upon the direction of the charge carrier's momentum relative to the crystal's orientation. The reason for this anisotropy is that due to crystal structure, the probability of collision between charge carriers and substrate atoms depends on the direction of travel of the carriers. In this work, scattering and direction-dependent effective mass was not considered. It was assumed that using a single, average value of effective mass, in the expected direction of travel of the charge carrier, was adequate.

3.4.5. Consideration of the Magnetic Field

There are several effects to consider when developing a model of charge transport in the presence of a magnetic field.

Firstly, any electric charge distribution that is evolving in time gives rise to a magnetic field that acts on free charge carriers in the vicinity through its contribution to the Lorentz force, $\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})$. If a magnetic field were included in the model described in Section 3.3, then the free charge distribution $n_n + n_p$ would not only give rise to a scalar electric potential ϕ, but also to a vector magnetic potential $\boldsymbol{A}$**.** Both of these potentials would appear in the Hamiltonian.

Secondly, an individual moving charge carrier is a contributor to the electromagnetic fields that influence its own motion. In the context of the single-particle model described in Section 3.3, this contribution would be considered separately from the field originated by the entire population of carriers, despite the single particle being a member of that population. The effect of the magnetic field generated by a single particle on the particle itself is called the radiation reaction or Abraham-Lorentz force. The classical theory gives the radiation reaction force as:

$$\boldsymbol{F} = \frac{\mu_0 q^2}{6\pi c} \frac{d^2 \boldsymbol{v}}{dt^2} \tag{3.8}$$

A quantum-mechanical formulation of the radiation reaction force has been proposed by Johnson and Hu in [19] and [20], using a Langevin Equation

approach, but that approach is not usable in the context of the basic formulation presented in Section 3.3.

Finally, the Hamiltonian of an electron moving in a magnetic field properly includes a term arising from interaction between the electron's spin and the magnetic field.

The magnetic field generated by the motion of free charge, the radiation reaction force, and the effect of spin have not been incorporated into the general model of Section 3.3 or the specific model developed in Chapter 4. Magnetic field effects were ignored because of the precedent set and justifications provided by other works, including all of those discussed in Section 3.5.

3.4.6. Approximations Inherent in the Single-Particle Model

The general problem formulation presented in Section 3.3 describes a model in which the charge distributions for holes and electrons are constructed from PDFs obtained by solving Schrödinger equations that represent a single hole and a single electron. An alternative to this model is to solve a single, multi-particle Schrödinger equation. The multi-particle Schrödinger equation would be based on a Hamiltonian that included a kinetic and potential energy contribution from every particle participating in conduction. The result of solving it, for a particular value of total system energy, would be the wave-function corresponding to a joint

probability distribution for the simultaneous positions of all particles. This multi-particle model would have the capability of predicting collisions between charge carriers. Collisions would be events involving the exchange of momentum between the participating particles. The number of collisions predicted in a simulation would depend on the total energy of the system, the applied fields, and the initial positions of each particle.

If a multi-particle model were used, the possibility would exist not only to model collisions between several charge carriers, but also collisions between charge carriers and the particles comprising the substrate through which they travel. Atomic nuclei could be represented as point particles having positive charge equal to an integer multiple of the electronic charge, and mass equal to the appropriate atomic mass. In this way, representation of the nuclear forces could be avoided, and a Schrödinger equation based on a multi-particle Hamiltonian could therefore be used. In order to model carrier-substrate collisions, the Hamiltonian would be expanded to include terms representing the kinetic and potential energies of all atomic nuclei and electrons comprising the substrate, as well as all carriers. In this hypothetical model, details such as the boundary conditions for the electric field would require modification.

A multi-particle model that can predict collisions between individual particles, whether it includes only carriers or carriers and substrate particles, is necessarily a

time-dependent model. The hypothetical multi-particle model described in this section was not used because of the prohibitive calculation time that would be required to obtain results. Other works, including all works mentioned in Section 3.5, have used the single particle model. In this thesis, the single particle model was used because precedents showed it to be acceptable and because the perceived alternative was computationally intractable.

The single-particle model does not capture interactions between individual charge carriers or carriers and ions at fixed positions in the substrate. However, the aggregate effect of many such interactions is indirectly captured by the model, in that the electric field generated by all free carriers and ions acts on the representative carrier that is used to derive the free carrier distribution.

In the single-particle model, the carrier-substrate interactions are modeled not only through the contribution of substrate ions to the electric field, but also by using an effective mass, as described in Section 3.4.4.

3.4.7. Assumptions Regarding Density of States

As stated in Section 3.3, the density of states, in the context of the general problem formulation presented in that section, is the eigenvalue spectrum of the linear differential operator representing the spatial part of the separated

Schrödinger equation. The boundary conditions associated with a linear operator affect the eigenvalue spectrum of the operator.

In this thesis, the density of states inside devices is assumed to be constant. This approximation is made because the probability of state occupation, as a function of energy, is assumed to change much faster than the density of states.

In some works such as [16], [17], [18], [45], and [46], density of states is assumed constant. Usually the constant value is determined based on the eigenvalue spectrum of the Hamiltonian operator for a single free particle free to move in two dimensions, with periodic boundary conditions applied, and is thus called the "2D density of states".

3.4.8. Assumptions Regarding Device Contacts

Devices were modeled as being in contact with reservoirs of charged particles. The simulations found a steady state of device operation involving a fixed rate of transfer of particles from one reservoir to another. It was assumed, as a boundary condition, that the Fermi energy in each reservoir was fixed and time-invariant. It follows from this assumption that the reservoirs must contain infinite numbers of particles.

It was assumed that density of states, as a function of position, was continuous at the device contacts. It follows from this assumption that the addition or removal

of particles from a reservoir must not cause the reservoir's density of states to change.

3.4.9. Consideration of Permittivity

The general problem formulation presented in Section 3.3 included the use of a position-dependent dielectric constant in the Poisson equation, (3.1).

As discussed in Chapter 2, the dielectric constant is applicable to charge carriers moving through bulk matter. The dielectric constant is used in the electromagnetic equations for bulk matter to account for the average effect of the fields of the individual atoms comprising the bulk matter on free charge present within the bulk matter. The dielectric constants of various materials are typically determined experimentally.

The alternative to using the electromagnetic equations for bulk matter in a model, thus requiring dielectric constant to be supplied as an input to the model, is for the model to include the potentials of each atom comprising the bulk matter. This approach would be required for the multi-particle model discussed in Section 3.4.6.

The model of Section 3.3 makes use of the form of the Poisson equation applicable to bulk matter, even though it is intended for the model to be applied to devices with features having a thickness of a few atomic layers. It could be argued

that Maxwell's equations for bulk matter are not meant to be applied to such devices, because the specific forms of the potentials of the atoms comprising the device could affect the calculation results. However, if individual atomic potentials were modeled, then it would no longer be possible to reduce the problem to a one-dimensional form, because no specific device cross-section could be considered representative when the atomic potentials were represented in the model. Instead, calculations would be required to cover three spatial dimensions. The computation time that would be required for this was seen as prohibitive, so Maxwell's equations for bulk matter were used, in conjunction with model representing only one spatial dimension. The use of dielectric constants rather than atomic potentials has been used in many similar works, including the works discussed in Section 3.5. The successes of these works were considered to be justifications for continued use of the model.

3.4.10. Consideration of Relativity

It follows from use of the Schrödinger equation in Section 3.3 that the effects of special relativity are not properly reflected in the model. Special-relativistic effects are commonly neglected in semiconductor device simulations, including those described in Section 3.5. For typical input parameters, these simulations predict carrier speeds much less than the speed of light, thus justifying the use of non-relativistic quantum theory.

3.5. Previous Work

The problem of charge transport in nanoscale electronic devices has been addressed in a considerable number of previous works. In this sub-section, a brief and incomplete review of some notable previous works is presented. Some previous results will be discussed in the context of the general problem formulation presented in section 3.3.

The work of Buchanan et al [23] [38] addressed gate tunneling current in FET transistors. In that work, a coupled Schrödinger-Poisson system was solved iteratively and self-consistently to find charge concentration. The Schrödinger equation was solved with closed boundary conditions rather than conditions derived from equation (3.6). Thus, charge carriers were confined to the interior of the device, and not free to enter or exit via contacts. The density of states applicable to (3.7) was considered to be the product of a constant transverse part, and a longitudinal part obtained as the eigenvalue spectrum of a matrix representing the longitudinal Hamiltonian and the chosen boundary conditions (constructed from a finite-difference second order derivative operator). State occupation probability was calculated using Fermi-Dirac statistics. Once the charge concentration was obtained in this way, the corresponding potential was used to calculate current contributions for each energy in the spectrum of the previously constructed Hamiltonian. For states identified as quasi-bound, current

was calculated with an adaptation of the transverse resonance method used for electromagnetic waveguide analysis [42]. For states identified as free, the transfer matrix approach was used to solve the Schrödinger equation with open boundary conditions. This method yielded close agreement with measured results.

The work of Frensley, reviewed in [35] and originally applied to resonant tunneling diodes, was addressed toward re-formulating the Schrödinger equation to enable intuitive and realistic boundary conditions to be applied. The Liouville-von Neumann equation is a matrix equation for the density operator, given by [8]:

$$i\hbar \frac{d\rho}{dt} = [H, \rho] \tag{3.9}$$

The Liouville-von Neumann equation is equivalent to, and can be derived from, the Schrödinger equation [8]. It can be solved to obtain the elements of the density matrix. Frensley's approach was to take the Wigner-Weyl transform [35] of the Liouville-von Neumann equation, so that the transformed equation would yield a phase-space distribution equivalent to the density matrix. The boundary conditions took the form of values assigned to the phase-space distribution for positions at the device boundaries and all momenta. Once the phase-space distribution was obtained, it was de-transformed to find the density matrix. In this method, the solution of the Wigner-Weyl transformed Liouville-von Neumann equation replaced the solution of the Schrödinger equation (3.4) or (3.5), and the density matrix did not need to be constructed as in (3.7) because the entire density

matrix was obtained at once. Iteration between the Liouville-von Neumann and the Poisson equations remained necessary. Density of states and probability of occupation entered into the calculations via the boundary conditions specified for the phase-space form of the Liouville-von Neumann equation. It was assumed in [35] that density of states was constant, and Fermi-Dirac statistics were used. One issue with this method is that the boundary conditions are only true infinitely far from the contacts [43].

The work of Datta et al, reviewed in [14], made use of non-equilibrium Green's functions to obtain the density operator. In that work, it was proposed to obtain Green's functions using the same method as described in this work, i.e. from equation (4.13). In the method described in [14], the Green's functions for each in a series of energy samples were found at the device boundaries. The contribution to the density matrix from each contact was given by a numerical integral over energy of the outer product of each Green's function with itself, as in equation (3.7). Each Green's function outer product in this integral was multiplied by the Fermi function $f(E, \mu)$, where μ is the Fermi energy in the contact. The total density matrix was given by the sum of the contributions from the contacts. In this method, the Schrödinger equation (3.4) or (3.5) was solved via use of Green's functions, and the boundary conditions applied to the Green's functions were those of outgoing waves at the contacts [14]. Iteration between the Poisson

equation and the density matrix construction process was required to solve the Schrödinger-Poisson system.

One issue with this method is in representing the sub-population of charge carriers that exists in potential wells inside the device under simulation, when the potential wells have energy minima significantly lower than the Fermi energies of all contacts. It was this issue that motivated the present work. The method that will be described in Chapter 4 has many similarities with the method described in [14], but makes use of Green's functions and Fermi energies at all points within the device, not just the boundaries.

The work of Fischetti et al, reviewed in [44], was somewhat more of a departure from the general procedure discussed in section 3.3. In that work, the Liouville-von Neumann equation (3.9) served as the starting point to obtain a Pauli master equation that was solved to find the time evolution of the density operator in the basis of wave-functions. The arguments used to obtain scattering rates for the master equation are too complex to do justice here, but are explained in [44]. The wave-functions themselves must be obtained by solving the Schrödinger equation, and with them, the density operator in the position basis can be obtained. This approach has the advantage that dissipative processes can be described relatively easily using a master equation, but this potentially comes at the cost of departing from the Liouville-von Neumann equation.

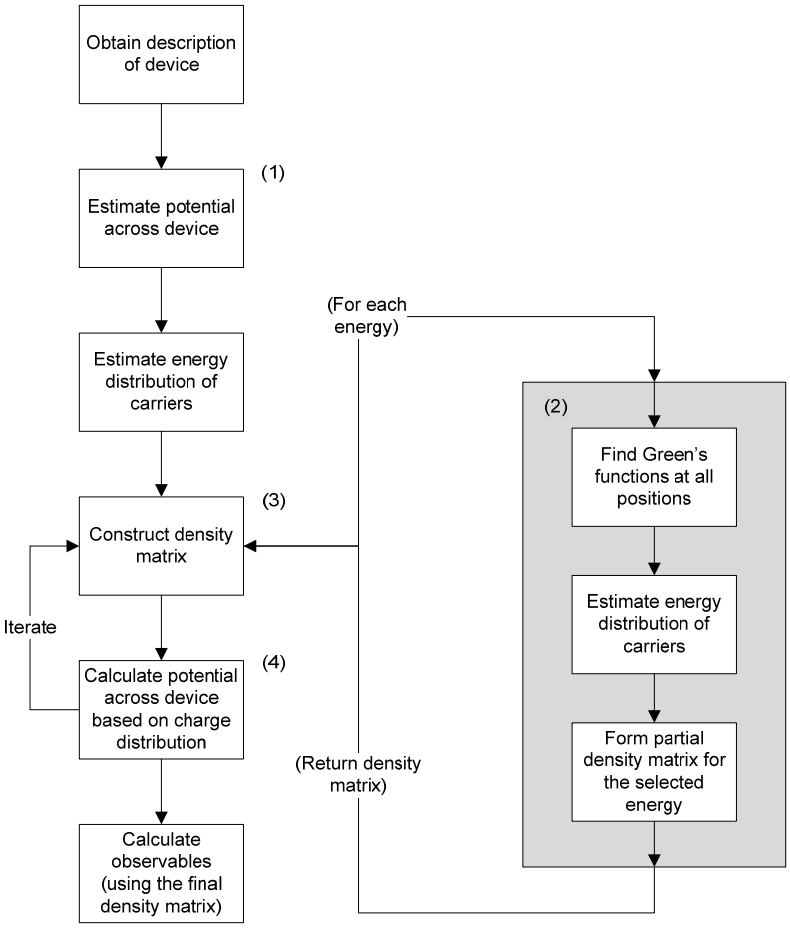
4. Methodology

In this chapter, the method used to calculate charge distribution and current in devices is described. The method was realized in the form of a program. The program's structure and operation are discussed.

4.1. Overview

The flowchart shown in Figure 1 describes the operation of the programs used to obtain the results presented in Chapter 5 and Chapter 6. This section summarises the structure and operation of the calculation programs. Sections 4.2 through 4.5 provide further detail about common elements of the programs.

Figure 1: Program Flowchart



The demonstration programs that implement the method described in this chapter were written in the Python language. One demonstration program was written per device. All of the programs share a common library that implements the calculation routines. Each device-specific program creates a data structure in memory that describes a device, and then calls library routines to perform calculations. No user interface was created, either in the form of an input file

specification and command line interface, or in the form of a graphical interface. However, since the relatively complex parts of the method were implemented in the library, the task of creating demonstration programs for specific devices was simple.

The desired result of some calculations was a charge distribution or the expected value of an observable in a single specific condition. For that purpose, density matrices for electrons and holes were required.

A main purpose of the calculator was to produce current-voltage (I-V) curves. To produce I-V curves, density matrices for electrons and holes were calculated at a series of discrete values of voltage sampled from a curve, and the density matrices were used to find net current.

Density matrices were found iteratively. Iteration was necessary because the density matrices were constructed from the Green's functions for Schrödinger equations in which the potential energy term was dependent on the charge distributions described by the density matrices themselves.

The first step in an iteration of the density matrix calculation was to estimate the minimum kinetic energy that a free charge carrier within the device could have. This energy was a function of position. The model included two populations of free charge carriers, holes and electrons, and two distinct functions were required

to describe their minimum kinetic energies. For electrons, the minimum kinetic energy was given by the conduction band energy. For holes, it was given by the valence band energy. Both the conduction band energy and the valence band energy were found by solving the Poisson equation. For the first iteration, a calculated distribution of free charge carriers was not yet available. Therefore, for the first iteration only, the method described in Section 4.3 was used to find the conduction and valence band energies. The Poisson equation was solved assuming the device contained no free or fixed charge. For iterations subsequent to the first, an approximate free charge distribution was available from the previous iteration. For these iterations, the Poisson equation was solved with approximate free and fixed charge distributions represented as being present within the device. Section 4.5 describes how conduction and valence band energies were calculated for iterations subsequent to the first one. Position-dependent quasi-Fermi distributions were calculated for electrons and holes based on the conduction and valence band energies. The formulae for doing this are given in Section 4.4.2.

The second step in an iteration of the density matrix calculation was performed separately for holes and electrons. For each of these charge carrier populations, iteration over a set of energies was performed. The set of energies was sampled from a continuous range. The lower bound of the energy range was the minimum

kinetic energy that a carrier could have, regardless of its position inside the device. The upper bound of the energy range was the energy such that the probability of finding a carrier with a higher energy anywhere inside the device was negligible. For each of the energies, Green's functions for the Schrödinger equation were calculated. One Green's function was calculated per energy for each discrete position inside the device.

The third step in an iteration of the density matrix calculation was to use all of the Green's functions calculated in step two to construct the density matrix. The density matrix was found as a weighted sum of outer products of Green's functions. The weights were provided by the quasi-Fermi distributions. A mathematical formulation is given in Section 4.4.2.

The final step in an iteration of the density matrix calculation was to use the approximate density matrices for electrons and holes to obtain free charge concentrations to feed back into the iterative process. These were simply the diagonals of the density matrices.

Density matrix calculation iteration was terminated when the difference in free charge distribution between iterations was negligible or when a pre-determined maximum number of iterations was reached. Further details regarding convergence and algorithm termination are given in Section 4.5. When a density

matrix that met one of the termination conditions was found, then observables such as current flow were calculated based on the density matrix.

In the following sub-sections, the steps of the calculation process are discussed in much more detail.

4.2. Device Description

The implementation of the calculation method developed in this work required devices to be described using short programs that built data structures in memory. An implementation of the same method that was constructed for easy re-use would accept data files instead. In any case, a set of data must be collected before calculations can be done.

The data required to begin performing calculations are the following:

- Operating temperature.
- A grid dividing the device into discrete one-dimensional cells. In this work, all such cells were required to be the same size, and that requirement is reflected in the equations stated in this chapter.
- The voltage applied at each device contact.
- Dielectric permittivity as a function of position, throughout the device.
- Effective masses of holes and electrons as functions of position.

- Conduction and valence band offsets as functions of position. In this work, the offsets were referenced to the leftmost discrete cell of the device, but any reference point could be used.
- Concentration of ionized acceptors and donors as a function of position.

The entire device was assumed to be at a uniform and constant temperature.

The macroscopic Maxwell's laws were used to find the field and potential within the device, and therefore, a dielectric constant was a required input. This approximation was discussed in the previous chapter.

The concentrations of ionized acceptors and donors were provided directly. A more accurate calculation would allow a dopant profile to be specified, and use the field, free charge concentration, and temperature within the device to calculate a position-dependent ionization rate. The increased complexity that would result from adding this feature was believed to be inessential to the operation of the basic simulator that was developed. It was envisioned that the simulator would be applied to devices and operating conditions in which the assumption of full ionization was acceptable.

4.3. Initial Calculation of Potential

As explained in Section 4.1, charge density was calculated iteratively, based on a density matrix constructed from the Green's functions of a device Hamiltonian. An estimate of the potential energy of a particle travelling through the device, as a function of position, was required to produce a numerical representation of the Hamiltonian. This potential is necessarily dependent upon the charge distribution, leading to a circularity. Therefore, the potential was calculated by assuming charge neutrality everywhere within the device for the first iteration. With no spatially varying charge distribution present, the one-dimensional potential is given by:

$$\frac{d}{dx}\left\{\varepsilon(x)\frac{d\phi}{dx}\right\} = 0 \qquad (4.1)$$

In equation (4.1): $\varepsilon(x)$ is the dielectric constant; and $\phi(x)$ is the electric potential in Volts. The potential energy of a single hole or electron moving within the device is then given by $\pm q\phi$, where q is the electron charge.

In subsequent iterations, the assumption of no charge being present within the device was removed.

4.4. Construction of the Density Matrix

The density matrix was constructed by summing over a set of energies. For every energy, a Green's function was calculated for each grid point within the device. The density matrix was constructed from all of the Green's functions.

The same method for calculating a density matrix is applicable for both holes and electrons. The calculations must be performed separately for each of these carrier types, however. The reason for this is that distinct potentials and effective masses are applicable to the two carrier types. The balance of the discussion describes the construction of a single density operator, with the understanding that the general procedure must be repeated for both carrier populations.

4.4.1. The Contribution of Particles with a Specific Energy

In this sub-section, a method is described for finding the Green's functions for a particle with a particular total energy. This sub-section also describes how these Green's functions were used to find the component of the density matrix corresponding to the selected energy value. The next sub-section will describe how the energies were selected for which to perform this calculation. It will also describe how the total density matrix was found.

In working toward the desired one-dimensional model, the 3-D case must first be considered. The single dimension to which the calculation was eventually

restricted is referred to as the longitudinal or "x" direction. The calculated current flows were in the longitudinal direction. The other two directions are referred to as the transverse or "y" and "z" directions.

The Schrödinger equation for a single, independent charge-carrying particle moving through a three-dimensional device is given in equation (4.2):

$$\left[-\frac{\hbar^2}{2m}\Delta + U_x(X) + U_{yz}(Y,Z)\right]|\xi\rangle = i\hbar\frac{\partial}{\partial t}|\xi\rangle \qquad (4.2)$$

In (4.2), the total wave-function of the particle, including its dependence on time and all three spatial dimensions, is denoted by $|\xi\rangle = |\xi(x,y,z,t)\rangle$.

The Schrodinger equation has a kinetic energy term and a potential energy term. The reason that the particle is subject to a change in potential energy is that it is in the presence of an electric field. This field has two causes. One cause is the external, applied field. The other cause is charge distribution present within the device. The assumption was made that the potential applicable to the particle can be written as the sum of a part that varies only in the transverse directions, and a part that varies only in the longitudinal direction. In equation (4.2), these two parts are written separately. They are denoted respectively by $U_{yz}(Y,Z)$ and $U_x(X)$.

The assumption was also made that the potentials U_x and U_{yz} are not time-dependent.

Because of the assumptions that were made about the particle's potential energy, the Schrödinger equation (4.2) is a separable partial differential equation. Appendix A shows explicitly how the separation of variables was performed. The part of the particle's wave-function that depends only on t is denoted $\vartheta(t)$. The part that depends only on y and z is denoted $\chi(y,z)$. The part that depends only on x is denoted $\psi(x)$. The total wave-function is given in terms of these by $\xi(x,y,z,t) = \psi(x)\chi(y,z)\vartheta(t)$. The transverse component of energy is denoted E_{xy}. The longitudinal component of kinetic energy is denoted E. Using these symbols, Table 2 summarizes the parts into which equation (4.2) was separated. The right column of Table 2 gives the differential equation that must be solved to obtain the function given in the centre column.

Table 2: Parts of the Separated Schrödinger Equation

Time part	$\vartheta(t)$	$\left(E + E_{yz}\right)\vartheta = i\hbar\frac{\partial\vartheta}{\partial t}$
Transverse part	$\chi(y,z)$	$-\frac{\hbar^2}{2m}\left(\frac{\partial\chi}{\partial y} + \frac{\partial\chi}{\partial z}\right) + U_{yz}\chi = E_{yz}\chi$
Longitudinal part	$\psi(x)$	$-\frac{\hbar^2}{2m}\psi'' + U_x\psi = E\psi$

The direction in which current was calculated is the longitudinal direction. Although it is a slight misuse of terminology, this work refers to the linear differential operator representing the longitudinal part of equation (4.2) (as it appears in the bottom row of Table 2) as "the Hamiltonian". The linear differential operator representing equation (4.2) is called "the total Hamiltonian".

A discrete formulation of the longitudinal Hamiltonian is desired so that $\psi(x)$ can be obtained numerically. The matrix containing the discretized representation of the Hamiltonian is denoted H. It was constructed by starting with the finite difference representation of the second derivative operator as given in Chapter 2, multiplying by a constant, and adding a diagonal matrix with the diagonals equal to discrete samples of the particle's potential energy.

The matrix H was subdivided into nine blocks. The top-left and bottom-right blocks encode the particle's interaction with the left and right reservoirs, respectively. The top-left block is denoted H_L and is square with dimension N_L. The bottom-right block is denoted H_R and is square with dimension N_R. The centre block H_D encodes the particle's interaction with the device. It is square and has dimension N. The other blocks consist entirely of zeros, except for four single elements: one each in four of the blocks τ_L, $\tau_L^\dagger$, τ_R, and $\tau_R^\dagger$. These elements store

the off-diagonal elements of the second derivative operator that are present in H to take derivatives at the boundary positions.

Equation (4.3) gives the precise discrete form of the Hamiltonian. In equation (4.3), the uniform distance between discrete samples of the potential is denoted Δ. This value is called the spatial resolution. The potential $U_x(x)$ has the single, fixed, constant value $U_x(x) = U_L$ inside the left reservoir. Similarly, $U_x(x) = U_R$ inside the right reservoir.

$$H = \begin{bmatrix} H_L & \tau_L & 0 \\ \tau_L^\dagger & H_D & \tau_R \\ 0 & \tau_R^\dagger & H_R \end{bmatrix} \tag{4.3.1}$$

$$a = \hbar^2/2m\Delta^2 \tag{4.3.2}$$

$$\langle k|H_D|j\rangle = (U_x(k) + 2a)\delta(j,k) - a\big(\delta(j,k-1) + \delta(j,k+1)\big) \tag{4.3.3}$$

$$\langle k|\tau_L|j\rangle = -a\delta(j,1)\delta(k,N_L) \tag{4.3.4}$$

$$\langle k|\tau_R|j\rangle = -a\delta(j,N)\delta(k,1) \tag{4.3.5}$$

$$\big\langle k\big|H_{L|R}\big|j\big\rangle = \big(U_{L|R} + 2a\big)\delta(j,k) - a\big(\delta(j,k-1) + \delta(j,k+1)\big) \tag{4.3.6}$$

The density operator for a particle that has the Hamiltonian given in equation (4.3) is denoted ρ. When reference is made in this chapter to the density operator, what is meant is an operator representing an ensemble of a large number of non-

interacting particles, all moving in a common space. The usual normalization condition for a density operator is that $\mathrm{Tr}(\rho) = 1$. It is necessary to scale the density operator by the total amount of charge to obtain a charge concentration.

A block-wise definition for ρ, structured such that the blocks of ρ and H are aligned and have the same size, is given in equation (4.4).

$$\rho = \begin{bmatrix} \rho_L & \rho_{12} & \rho_{13} \\ \rho_{12}^{\dagger} & \rho_D & \rho_{23} \\ \rho_{13}^{\dagger} & \rho_{23}^{\dagger} & \rho_R \end{bmatrix} \tag{4.4}$$

The centre block ρ_D is the solution being sought. The blocks representing particle density in the reservoirs, and the off-diagonal blocks, were not required for any further calculations. They were not calculated.

The reservoirs are infinite in size. Therefore, the components H_L and H_R of the Hamilton operator, and the components ρ_L and ρ_R of the density operator, are infinite in size. However, the infinite size of these matrices does not prevent the numerical value of the centre block of the density operator, ρ_D, from being calculated.

In this work, the starting point for finding ρ_D was the Lippmann-Schwinger equation [10]. The Lippmann-Schwinger equation in the single particle context is applicable to a "system consisting of two interacting parts, which are such that the

interaction energy approaches zero as the two parts are separated spatially" [10]. The Hamiltonian consists of a term that describes the two parts that can interact, and another term that describes their interaction. These parts are denoted H_0 and H_1 respectively. The parts that can interact are the transverse and longitudinal fields and potentials. The interaction that physically occurs is an exchange of kinetic energy between the transverse and longitudinal directions, caused by an interaction between a particle and the device and/or other particles.

The result of solving the Lippmann-Schwinger equation is a basis function; and a set of these basis functions can be used to construct a wave-function describing "the change in the state vector produced by the interaction process" [10]. Lippmann and Schwinger show that such basis functions exist, and show how to find them. These basis functions can be used to construct the wave-function of a particle that is affected by the interaction described by H_1, even though they are time-independent.

In this work, the explicit form of the total interaction potential that causes transfer of kinetic energy between the transverse and longitudinal directions is not given. However, the Green's functions that are found can be used to describe any interaction. Each Green's function corresponds to an interaction in which a particle that has a specific energy exchanges kinetic energy between the transverse and longitudinal directions at a specific position. The form of the

interaction was implicitly defined by the proposed method for constructing the density matrix from the Green's functions, as described in this section and the next.

Here, an abbreviated development of the Lippmann-Schwinger equation is given, following [10] and using similar notation. The full derivation is given in [10]. The Schrödinger equation from which the Lippmann-Schwinger equation is derived is stated as:

$$i\hbar \frac{\partial \psi'}{\partial t} = (H_0 + H_1)\psi' \tag{4.5}$$

In equation (4.5), ψ' is the wavefunction of the particle with the time-independent total energy $H_0 + H_1$. A unitary transformation is made,

$$\psi'(x,t) = \exp(iH_0 t/\hbar) \cdot \psi(x,t) \tag{4.6}$$

Under this transformation the Schrödinger equation for the particle becomes

$$i\hbar \frac{\partial \psi}{\partial t} = \exp(iH_0 t/\hbar) \cdot H_1 \cdot \exp(-iH_0 t/\hbar)\, \psi \tag{4.7}$$

The notation is used:

$$H_1(t) = \exp(iH_0 t/\hbar) \cdot H_1 \cdot \exp(-iH_0 t/\hbar) \tag{4.8}$$

An operator $U_+(t)$ is defined by the equation $\psi(t) = U_+(t)\psi(-\infty)$.

By integrating the Schrödinger equation, an integral equation is produced that can be solved for $U_+(t)$:

$$U_+(t) = 1 - \frac{i}{\hbar}\int_{-\infty}^{t} H_1(t')U_+(t')dt' \tag{4.9}$$

With reference to an initial state $\phi(x,t;E_0)$, Lippmann and Schwinger define

$$\psi^+(x,t;E,E_0) = \lim_{\varepsilon\to\infty}\int_{-\infty}^{\infty} \exp(i(E-H_0)t/\hbar)\exp(-\varepsilon|t| /\hbar)\, U_+(t)\phi(x,t;E_0)dt \tag{4.10}$$

In equation (4.10): the parameter E is the energy of the state ψ^+ with respect to the Hamiltonian $H_0 + H_1$; and the parameter E_0 is the energy of the state ϕ with respect to the Hamiltonian H_0.

Using the definition given in equation (4.10), Lippmann and Schwinger show that:

$$\psi^+ = \phi + \frac{1}{\mathrm{E} + i\varepsilon - H_0} H_1\psi^+ \tag{4.11}$$

Equation (4.11) is called the Lippmann-Schwinger equation.

Appendix B shows that the solutions $(\psi^+ - \phi)$ satisfy the Schrödinger equation for an interaction potential H_1 proportional to the Dirac delta function. In this work, a superposition of the solutions of equation (4.11) with H_1 proportional to

the delta function was used to represent a particle that enters the device, is subject to an interaction, and exits at a later time.

Solutions to equation (4.11) are described in [10] as outgoing scattered waves. In their paper [10], Lippmann and Schwinger develop the time-reversed scenario in parallel. The solutions that result from the time-reversed development are called the incoming scattered waves. As it was argued by Datta in [14], the outgoing solutions must be used for the present application.

The Hamiltonian to which equation (4.11) must be applied is the one stated in equation (4.3). The unperturbed Hamiltonian that must be used when solving equation (4.11) is $H_0 = H$. Equation (4.11) was solved for all energies, with a series of perturbations, to generate the set of Green's functions applicable to that energy. Each of these perturbations was of the form:

$$H_1 = \varpi |x'\rangle\langle x'| \qquad (4.12)$$

In equation (4.12): ϖ is the magnitude of the perturbation energy; and x' is the position at which the perturbed wavefunction is at a maximum. The expected position at which an exchange of transverse and longitudinal kinetic energy takes place is x'. A Green's function was obtained for every possible discrete position x'.

The notation is now adopted that $g(x;x',E) = (\psi^+ - \phi)$, where E is the energy of state ϕ. The functions $g(x;x',E)$ are referred to as Green's functions. It has been shown that each such Green's function is an outgoing solutions to the longitudinal part of the Schrödinger equation (4.2) augmented by a perturbation potential.

Using and the unperturbed and perturbation Hamiltonians from equation (4.3) and equation (4.12), the Lippmann-Schwinger equation that must be solved to obtain the Green's function for the energy E and the position x' is:

$$\big((E + i\varepsilon)I - H\big)|g(x;x',E)\rangle = (\varpi|x'\rangle\langle x'|)|\psi^+\rangle \qquad (4.13)$$

The Green's functions that resulted from solving equation (4.13) were normalized. Therefore, neither the magnitude of the perturbation energy ϖ, nor the value of the perturbation component of the wavefunction ψ^+ at the position x', made any numerical difference to the end result. They can be assumed equal to one, or any other value, with no consequence.

This thesis proposes to construct the component of the density matrix representing particles with energy E using the equation:

$$\rho(;E) = \int |g(x;x',E)\rangle f(x',E)\theta\big(E - U_x(x')\big)\langle g(x;x',E)|\, dx' \qquad (4.14)$$

In equation (4.14), $f(x', E)$ is the Fermi function, given by:

$$f(x', E) = \frac{1}{1 + \exp\left(\left(E - \mu(x')\right)/k_B T\right)} \tag{4.15}$$

Also in equation (4.14): θ is a step function that is zero when its argument is less than zero and one when its argument is greater than or equal to zero; and $U_x(x')$ is the same position-dependent longitudinal potential that was used in the Hamiltonian, equation (4.3).

In equation (4.15): $\mu(x')$ is the position-dependent Fermi energy; k_B is the Boltzmann constant; and T is the temperature, which is assumed to be independent of time and position. Possible procedures for calculating the Fermi energy $\mu(x')$ are discussed in the next sub-section.

The density matrix component $\rho(;E)$ was constructed by weighting the Green's functions for each position and a common energy by the probability of occupation at position x' that a state with the common energy E would be occupied. The step function $\theta\left(E - U_x(x')\right)$ appears in equation (4.14) to ensure that the Green's function at position x' exists. In order to arrive at equation (4.13), it was assumed that the unperturbed and perturbed states at position x' did not have negative kinetic energy. If the integration of equation (4.14) were attempted without the step function present, then wave-functions that do not exist could be included in

the density matrix. Whether this problem actually occurred would depend on the specific potential being considered.

The Hamiltonian that was referred to in equation (4.13) applies to all space, and even a discrete matrix representation of it would have an infinite number of rows and columns. The Hamiltonian was defined block-wise in equation (4.3), and the elements in the block-wise definition of a density matrix with the same matrix structure were assigned symbols in equation (4.4). A procedure for finding the Green's functions block-wise is required so that the centre and only finite block, ρ_D, of the density matrix can be found without calculating the elements of the other blocks. Using the notation of equation (4.3) and equation (4.4), the block-wise statement of equation (4.13) is given as follows:

$$\left\{\frac{E+i\varepsilon}{\varpi}\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} - \frac{1}{\varpi}\begin{bmatrix} H_L & \tau_L & 0 \\ \tau_L^\dagger & H_D & \tau_R \\ 0 & \tau_R^\dagger & H_R \end{bmatrix}\right\}^{-1} \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} = G \qquad (4.16)$$
$$= \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_D & G_{23} \\ G_{31} & G_{32} & G_{33} \end{bmatrix}$$

This system was solved to find G. Each column of G is a Green's function.

As shown in [11], multiplying G by the middle row of $(E + i\varepsilon)I - H$ yields the equations:

$$(E + i\varepsilon)G_D - \begin{bmatrix}\tau_L^\dagger & H_D & \tau_R\end{bmatrix}\begin{bmatrix}G_{12}\\ G_D\\ G_{32}\end{bmatrix} = \varpi I \qquad (4.17.1)$$

$$\big((E + i\varepsilon)I - H_L\big)G_{12} = \tau_L^\dagger G_D \qquad (4.17.2)$$

$$\big((E + i\varepsilon)I - H_R\big)G_{32} = \tau_R^\dagger G_D \qquad (4.17.3)$$

Using the notation $\gamma_L = \big((E + i\varepsilon)I - H_L\big)^{-1}$ and $\gamma_R = \big((E + i\varepsilon)I - H_R\big)^{-1}$, equation (4.17) can be restated as:

$$\big\{\big((E + i\varepsilon)I - H_L\big) - \big(\tau_L^\dagger\, \gamma_L \tau_L + \tau_R^\dagger\, \gamma_R \tau_R\big)\big\}G_D = \varpi I \qquad (4.18)$$

A further notational abbreviation is made: $\Sigma_L = \tau_L^\dagger\, \gamma_L \tau_L$, and $\Sigma_R = \tau_R^\dagger\, \gamma_R \tau_R$.

To proceed further, it was necessary to utilize a means of finding the values of specific elements of the semi-infinite matrices, γ_L and γ_R. The matrices τ_L and τ_R each have only one element, despite being infinite in size. Therefore, the only elements needed are the bottom right element of γ_L and the top left element of γ_R.

For an example device represented by a matrix of size 3×3, this is demonstrated as follows:

$$\tau_L^{\dagger}\,\gamma_L\tau_L = \begin{bmatrix} 0 & & 0 & -a \\ 0 & \dots & 0 & 0 \\ 0 & & 0 & 0 \end{bmatrix} \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1,N_L} \\ \vdots & \ddots & \vdots \\ \gamma_{N_L,1} & \cdots & \gamma_{N_L,N_L} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ & \vdots & \\ 0 & 0 & 0 \\ -a & 0 & 0 \end{bmatrix} = \begin{bmatrix} a^2\gamma_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The matrices of which γ_L and γ_R are the inverses are symmetric and tri-diagonal. Therefore, they are examples of Toeplitz matrices. A procedure for finding individual elements of the inverse of a Toeplitz matrix was found by Southern et al in [12]. That procedure built on the work of Hu and O'Connell in [13]. For the specific physical situation and the matrix formulation that are considered in this work, the applicable Σ_L and Σ_R were found in [14] to be:

$$\langle k|\Sigma_L|j\rangle = a \cdot \delta(0,0)\exp(i \cdot \arccos(1 - [(E - U_x(x_L))/a])) \qquad (4.19.1)$$

$$\langle k|\Sigma_R|j\rangle = a \cdot \delta(N,N)\exp(i \cdot \arccos(1 - [(E - U_x(x_R))/a])) \qquad (4.19.2)$$

In equation (4.19), x_L and x_R are the positions at which the device ends on the left and right sides, respectively. These formulae are based on the dispersion relation for the reservoirs [14], $E_K = 2a(1 - \cos(k\Delta))$ where $E_k = (\hbar k)^2/2m$.

Using the matrix elements for H_D, Σ_L and Σ_R given in equation (4.3) and equation (4.19), the entire process can be summarized using the following expressions:

$$F(E) = \sum_x |x\rangle f(x,E)\theta\big(E - U_x(x')\big)\langle x| \tag{4.20.1}$$

$$Q(;E) = \big((E + i\varepsilon)I - H_D - \Sigma_L - \Sigma_R\big)^{-1}\varpi I \tag{4.20.2}$$

$$P(;E) = F\sum_x Q|x\rangle\langle x|Q^\dagger \tag{4.20.3}$$

$$\rho_D(;E) = \frac{P(;E)}{\mathrm{Tr}\big(P(;E)\big)} \tag{4.20.4}$$

The end result is a partial density matrix that describes a particle with energy E.

4.4.2. The Total Density Matrix

In this section a method is proposed for constructing the total density matrix based on the energy-specific partial density matrices that were constructed in the previous section. It is argued that a density matrix constructed in the proposed manner will naturally incorporate a reasonable density of states and probability of state occupation.

This thesis proposes to construct the total density matrix for the device region, ρ_D, as follows:

$$\begin{aligned}\rho'_D &= \sum_{x'} \sum_{E \in W(x',N)} \rho_D(;E) \\ &= \sum_{x'} \sum_{E \in W(x',N)} |g(x;x',E)\rangle f(x',E)\theta\big(E - U_x(x')\big)\langle g(x;x',E)|\end{aligned} \tag{4.21.1}$$

$$\begin{aligned}W(x',N) = \Big\{ & U_{x;min} \\ & + \frac{k}{N_E}\big(U_{x;max} - U_{x;min} + E_{xm}\big) \quad | \quad k \in \{0,1,\dots N_E\}\Big\}\end{aligned} \tag{4.21.2}$$

$$\rho_D = \rho'_D / \mathrm{Tr}(\rho'_D) \tag{4.21.3}$$

In equation (4.21): the set of positions x' over which the outer summation is performed is the entire set of discrete positions that fall within the device region; N_E is a number of energies over which to perform the summation in energy; $U_{x;max}$ is the maximum value of U_x within the device region; $U_{x;min}$ is the minimum value of U_x within the device region; E_{xm} is a kinetic energy such that the probability of finding a particle with a greater kinetic energy inside the device is negligible; and $f(x',E)$ is the Fermi function, as in equation (4.15).

It is worthwhile to consider the way in which the proposed density operator reflects the density of states.

A typical un-normalized density matrix construction is as follows:

$$\rho = \int_0^\infty f(E_T)a(E_T)\,|\chi_{E_{yz}}\psi_E\rangle\langle\psi_E\chi_{E_{yz}}|\,dE_T \qquad (4.22)$$

In equation (4.22): E_T is the longitudinal plus transverse energy, $E_T = E + E_{yz}$; $f(E_T) = \left(1 + \exp\left(\frac{E_T-\mu}{k_BT}\right)\right)^{-1}$ is the probability of state occupation given by the Fermi function; $a(E_T)$ is the three-dimensional density of states; the transverse part of the wave-function is $\chi_{E_{yz}}$; and the longitudinal part of the wave-function is ψ_E.

Because the Fermi distribution gives the probability of state occupation, a density matrix construction that makes use of it must in some sense utilize a density of states.

In this work, the applicable density of states was assumed to be constant.

Weyl's theorem [15] states that for a Sturm-Liouville operator with one singular endpoint a and one regular endpoint b, there is a unique solution corresponding to an eigenvalue $E \in \mathbb{R}$ of the operator on (a, b). For the problem under consideration in this work, with H_1 given by equation (4.12), each pair (x', E) therefore results in a unique function $g(x; x', E)$.

The density operator given by equation (4.22) can be restated as a double integral:

$$\rho = \int\int f\left(E + E_{yz}\right) a_x(E) a_{yz}\left(E_{yz}\right) |\psi_E\rangle\langle\psi_E| dE \cdot dE_{yz} \qquad (4.23)$$

To arrive at equation (4.23): the contribution from the transverse wave-function $\chi_{E_{yz}} \cdot \overline{\chi_{E_{yz}}}$ is cancelled since the device has no features in the transverse direction; and the partitioning $a = a_x \cdot a_{yz}$ is possible because of the separation of variables. As argued in the previous paragraph, the longitudinal density $a_x(E)$ applicable to Green's functions is constant. Therefore, the density matrix constructed from Green's functions has the proportionality,

$$\rho \propto \int a_{yz}\left(E_{yz}\right)\left\{\int f\left(E + E_{yz}\right) |g(;E)\rangle\langle g(;E)| dE\right\} dE_{yz}$$

As shown above, the assumption of a constant density of states becomes increasingly accurate as E_{yz} becomes small relative to E and $f\left(E + E_{yz}\right) \approx f(E)$. Because the electric field causing net motion of particles participating in conduction is in the longitudinal direction, the assumption that $\langle E\rangle \gg \langle E_{yz}\rangle$ is reasonable.

The idea of a constant density of states has precedent in various other works, such as Lake [16], Rammer and Smith [17], and Jauho and Wilkins [18].

Given an arbitrary $a(E)$, the method that has been presented would allow for calculating ${\rho'}_D$ as $\rho'_D = \int a(E)\rho_D(;E)dE$ rather than using equation (4.21). This work used $a(E) = (const.)$ because no other choice was more obvious.

4.5. Calculation of Potential Using the Density Matrix

The density operators constructed in the previous section were used to calculate charge concentration within the device.

The populations of holes and electrons in a device have different Fermi energies, potentials, and density operators. The following table summarizes the notation used for the populations of electrons and holes.

	Population of electrons	Population of holes
Total amount of charge $[C]$	Q^-	Q^+
Free charge distribution $[C \cdot m^{-3}]$	$n_n(x)$	$n_p(x)$
Ionized dopant charge distribution $[C \cdot m^{-3}]$	$N_D^+(x)$	$N_A^-(x)$
Fixed charge distribution $[C \cdot m^{-3}]$	$z_n(x)$	$z_p(x)$
Fermi energy $[J]$	$\mu_n(x)$	$\mu_p(x)$
Applicable potential $[J]$	$U_n(x)$	$U_p(x)$
Density matrix	ρ_n	ρ_p

Conduction/valence band offset $[J]$	$\vartheta_{CB}(x)$	$\vartheta_{VB}(x)$
Intrinsic carrier density $[C \cdot m^{-3}]$	N_i	
Band gap $[J]$	$\vartheta_G(x)$	
Net charge $[C \cdot m^{-3}]$	$n(x)$	

The total electron charge was found by integrating the ionized donor concentration inside the device region:

$$Q^- = \int N_D^+(x)dx$$

The concentration of free electrons is given by $n_n(x) = Q^-\langle x|\rho_n|x\rangle$.

Similarly, the total hole charge was found by integrating the ionized donor concentration inside the device region:

$$Q^+ = \int N_A^-(x)dx$$

The concentration of free holes is given by $n_p(x) = Q^+\langle x|\rho_p|x\rangle$.

The total charge distribution was found by summing all of the contributions that are included in the model:

$$n(x) = n_n + n_p + N_A^- + N_D^+ + z_n + z_d \qquad (4.24)$$

It is not necessary or expected that 100% of dopant sites will be ionized. However, modeling the interrelationship between Fermi level, field, and ionization rate was not considered central to this work. The flexibility of the

proposed methodology would allow for the proportion of dopant sites that are ionized to be adjusted during the Schrödinger-Poisson iteration process, if an adjustment were deemed necessary.

The net charge concentration may include a population of fixed charge that will not be balanced by free carriers, thus causing the device to have a built-in voltage bias. This population was represented by $z_n + z_d$ in equation (4.24). The fixed charge may be a fixed value, or set to zero.

The voltage profile resulting from the charge distribution given in equation (4.24) was calculated using the Poisson equation:

$$\frac{d}{dx}\left\{\varepsilon(x)\frac{d\phi}{dx}\right\} + n'(x) = 0 \qquad (4.25)$$

In equation (4.25), $n'(x)$ is a charge concentration. The simplest form of the algorithm sets $n'(x) = n(x)$ as given by equation (4.24).

The differential equation, equation (4.25), was discretized and solved numerically. The discrete matrix form of equation (4.25) is:

$$\frac{1}{\Delta^2}\begin{bmatrix} 1 & & & \\ \varepsilon_0 & -2\varepsilon_1 & \varepsilon_2 & \\ & \ddots & \ddots & \ddots \\ & & & 1 \end{bmatrix}\boldsymbol{\phi} = -\boldsymbol{n} + \boldsymbol{b} \qquad (4.26)$$

In equation (4.26): the first and last elements of $\boldsymbol{n}$, the net charge, are set to zero; and the boundary conditions are given by $\boldsymbol{b} = [V_L \quad 0 \quad ... \quad 0 \quad V_R]^T$ with $V_{L|R}$ being the left/right applied voltages.

The potentials applicable to the populations of electrons and holes were calculated based on the solution of the Poisson equation, using the following equations:

$$U_n(x) = -q\phi(x) + \vartheta_{CB}(x) \qquad (4.27.1)$$

$$U_p(x) = q\phi(x) + \vartheta_{VB}(x) \qquad (4.27.2)$$

When the potential calculated from a charge distribution is significantly different from the one that originated it, then consistency between the Schrödinger and Poisson equations requires iteration to be performed. The new potential is fed back into the process by using it to construct a new Hamiltonian. The feedback loop is repeated until the input and output potentials are not significantly different.

There are two degrees of freedom in finding the density of a carrier population corresponding to a potential: the quasi-Fermi level for the population, and the potential itself. The programs developed for this work assigned the quasi-Fermi levels using a very simple scheme. In this scheme, the quasi-Fermi levels were given by:

$$\mu_n = \vartheta_{CB} - \frac{\vartheta_G}{2} + k_B T \ln\left(\frac{N_D^+}{N_i}\right) + \mu_{0,n}(x) \qquad (4.28.1)$$

$$\mu_p = \vartheta_{VB} + \frac{\vartheta_G}{2} - k_B T \ln\left(\frac{N_A^-}{N_i}\right) + \mu_{0,p}(x) \qquad (4.28.2)$$

In equation (4.28), $\mu_{0,n}$ and $\mu_{0,p}$ are position-dependent offsets that relate the quasi-Fermi levels and the electric field inside the device.

A number of different approaches may be used to relate the field and the quasi-Fermi levels. One approach simply offsets the quasi-Fermi levels based on the potential, as follows:

$$\mu_{0,n}(x) = -q\phi(x) \qquad (4.29.1)$$

$$\mu_{0,\mathrm{p}}(x) = q\phi(x) \qquad (4.29.2)$$

Another approach defines $\mu_{0,n}$ and $\mu_{0,\mathrm{p}}$ piecewise in a way that is specific to a device structure. To give one example, the quasi-Fermi level for electrons in a MOS device could be defined as in equation (4.30).

$$\mu_{0,n} = \begin{cases} V_L, & x < x_{ox} \\ \dfrac{V_L + V_R}{2}, & x_{ox} \le x \le x_{ox} + t_{ox} \\ V_R, & x > x_{ox} + t_{ox} \end{cases} \qquad (4.30)$$

In equation (4.30): V_L and V_R are the left and right applied voltages; and x_{ox} and t_{ox} are the position of the left edge and the thickness of the oxide.

Both approaches were used for some of the results in subsequent chapters.

The potentials given in equation (4.27) and the Fermi levels given in equation (4.28) may be used in Schrödinger-Poisson iteration to reconstruct the Hamiltonian and density matrix.

In "self-consistent" calculations, density matrix calculation iteration was terminated when the difference in free charge distribution between iterations became negligible. In practical terms, this meant that charge concentration at any point did not change by more than 0.1% between iterations[1]. When the iteration process terminated because this criterion is met, the solution can be called self-consistent. In order to ensure that the algorithm terminates and to place a limit on the real time that calculations can take, an additional termination criterion was added: The algorithm was terminated if the maximum permitted number of iterations (as chosen by the program operator) was exceeded.

If the density matrix calculation terminated because the iteration limit was reached, the solution could not be called self-consistent. Experience with the calculation programs showed that divergence, in which iterations are progressively farther from the solution than previous ones, was not generally observed. However, this thesis cannot provide a convergence proof.

[1] The 0.1% convergence threshold was selected somewhat arbitrarily, and a different value could be used.

To increase the rate of convergence, it was found beneficial to use a linear combination of the net charge concentration that was used in the previous iteration and the newly calculated net charge concentration. To do this, equation (4.25) may be modified using:

$$n'_{(k)}(x) = (1 - \alpha)n'_{(k-1)}(x) + \alpha n_{(k)}(x) \qquad (4.31)$$

In equation (4.31): a subscript in parentheses indicates the iteration number from which the annotated quantity originates; k is the current iteration; and $0 < \alpha < 1$ is an experientially selected factor meant to accelerate convergence. For example, $\alpha = 0.3$ might be used.

Not all of the results presented in this work are self-consistent. Self-consistent results are noted as such.

4.6. Calculation of Current and Other Observables

The quantum-mechanical procedure for measuring the expected value of a generic observable represented by an operator is to take the trace of the operator applied to the density matrix:

$$\langle \Omega \rangle = \mathrm{Tr}(\rho \Omega) \qquad (4.32)$$

In equation (4.32): Ω is an observable; and ρ is a density matrix.

A position-dependent measurement, given operators in the position basis, is produced by:

$$\omega(x) = \langle x|\rho\Omega|x\rangle$$

If this procedure were to fail, it would suggest the incorrectness of the density operator.

The observable of primary interest in this work is current. The discrete matrix form of the position basis current operator used in the calculations was:

$$J = \frac{\hbar}{m\Delta}\begin{bmatrix} 0 & -i & 0 & \\ i & 0 & -i & \\ 0 & i & 0 & \\ & & & \ddots \end{bmatrix} \tag{4.33}$$

The density operators to which this current operator was applied were scaled as follows:

$$\mathrm{Tr}(\rho_n) = Q^-$$

$$\mathrm{Tr}(\rho_p) = Q^+$$

4.7. Conservation Laws

The conservation laws applicable to charge carrier transport have the following interrelated aspects:

1. Conservation of charge
2. Conservation of particle or quasi-particle count

3. Conservation of probability flux
4. Conservation of momentum
5. Conservation of energy

The probability flux j associated with a wave-function ψ is given by:

$$j = -\frac{i\hbar}{2m}\left(\overline{\boldsymbol{\psi}}\nabla\boldsymbol{\psi} - \psi\nabla\overline{\psi}\right) \tag{4.34}$$

Based on this definition and the Schrödinger equation, the conservation law for probability flux, equation (4.35), may be derived [8].

$$\dot{\rho} + \nabla \cdot j = 0 \tag{4.35}$$

A density operator constructed from valid solutions of the Schrödinger equation must satisfy this conservation law, which implies conservation of momentum.

This work attempted to construct steady-state density operators, which satisfy $\dot{\rho} = 0$. The density operators constructed describe a fixed number of particles. If particle count in a system is conserved, then charge is also conserved. No physical process other than the addition or removal of particles was considered that can could the amount of charge. Particle count is given by the trace of the density matrix; thus the number of particles represented by a steady state density matrix can trivially be enforced by normalization and scaling.

Given a density matrix representing a fixed number of particles, the conservation law (4.35) can easily be verified by finding $\nabla \cdot j$. The numerical results generated by the method described in this chapter do not allow equation (4.35) to be verified, because while the density operators are calculated based on an implicit transfer of energy between the longitudinal and transverse directions, the transverse wave-functions are not calculated. However, provided that the functions from which the operators are constructed are correct solutions of the Schrödinger equation, conservation of momentum and probability flux must be obeyed.

5. Simulation Results

In this section, the results of charge and current calculations for two test devices are presented. The devices were selected with the purpose of testing specific aspects of the calculation programs and the physical model and methodology described in the previous chapter.

The convention that was followed in illustrations was to show devices with one spatial dimension, in which the coordinate increases along the x-axis, from left to right across the page. The y-axis shows the magnitude of position-dependent quantities such as energy, charge concentration, or current. In this chapter and the next, the convention was followed that voltages were applied at the contact where $x = 0$.

5.1. Test Device

The first device that was considered was a test device meant to show the qualitative behaviour of the simulator. The device was not considered to have any purpose other than to test the simulator. It was similar in structure to the n+/n/n+ resistor discussed in [14].

The device model included doping and counter-doping so that the simulations would yield results with substantial concentrations of both majority and minority

carriers. A position-varying dopant concentration was used, with the purpose of testing whether the results would show a correlation between free charge and fixed charge of the opposite sign (produced by ionized dopant atoms). The tendency toward local charge neutrality is expected of a physically realistic model.

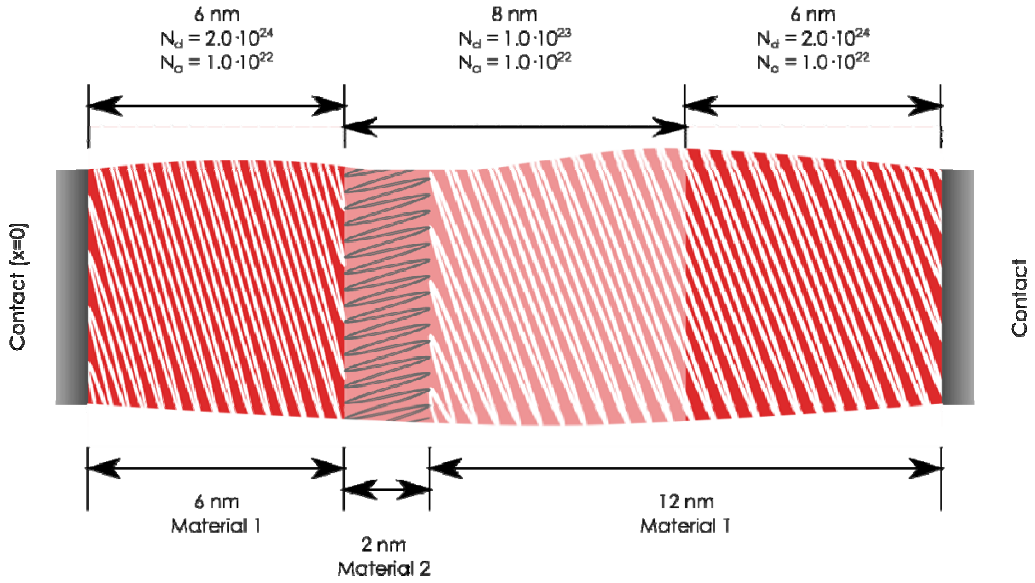
The device had asymmetry in its structure, so that potential errors in the calculation methodology related to symmetry could potentially be revealed.

Self-consistent results were found, with programmatic, independent adjustment of the quasi-Fermi energies for electrons and holes.

A starting potential was used that was far from the correct, self consistent potential, with the purpose of testing whether convergence could be reached in such circumstances. Applied voltages were simulated that were significantly higher than those tested in comparable works, such as [14]. The intention of using high applied voltages was to test the limits of the proposed new methodology.

The device that was simulated is shown in Figure 2. Simulation results showing the concentrations of free electrons and holes are provided in Figure 3. Figure 4 shows the starting potential and the ionized acceptor and donor concentrations that were used for the 1.0 V calculation. The same ionized acceptor and donor concentrations were used for all calculations.

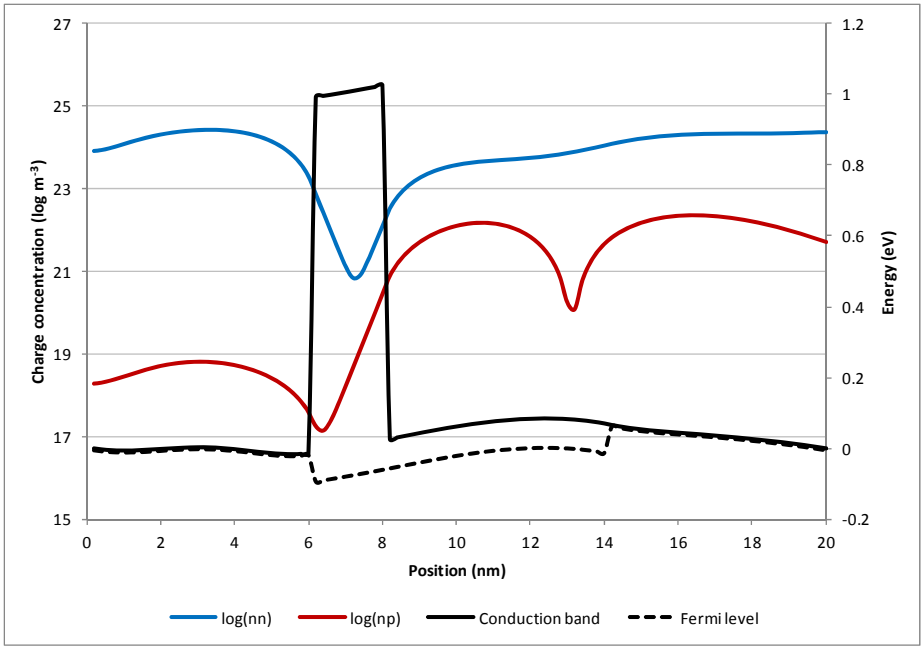
Figure 2: Device Properties – Test Device



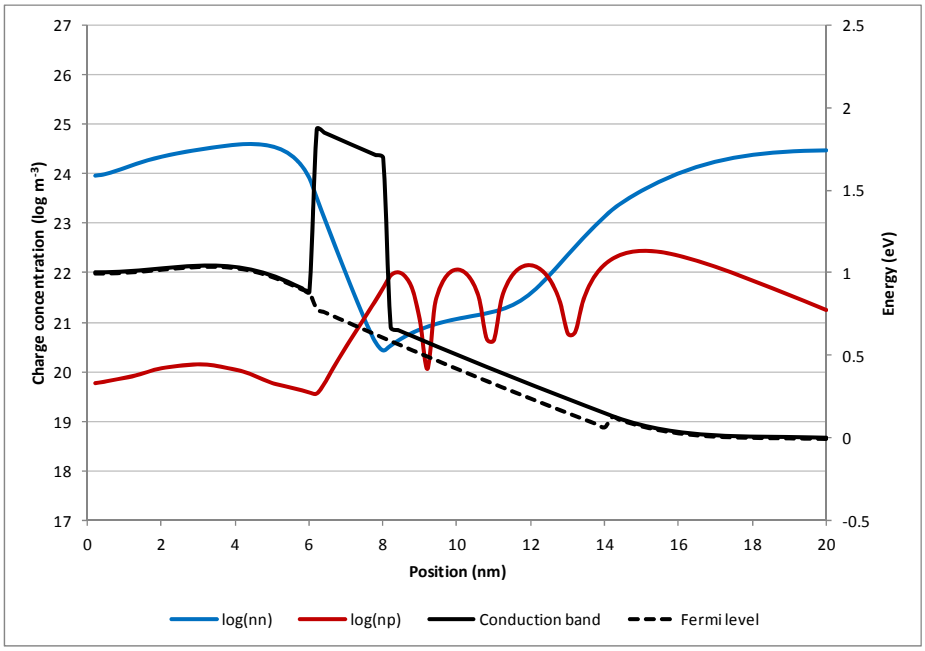
Material 1	**Material 2**
$\Phi_{CB} = 0$	$\Phi_{CB} = 1.0$ eV
$\Phi_{VB} = 0$	$\Phi_{VB} = 1.0$ eV
$\kappa = 1.0$	$\kappa = 1.0$
$E_G = 1.0$ eV	$E_G = 1.0$ eV
$m_e^* = 0.25$	$m_e^* = 0.25$
$m_h^* = 0.25$	$m_h^* = 0.25$

Figure 3: Charge Concentration Results – Test Device

(A) Voltage Applied: 0 V



(B) Voltage Applied: 1.0 V



(C) Voltage Applied: -1.0 V

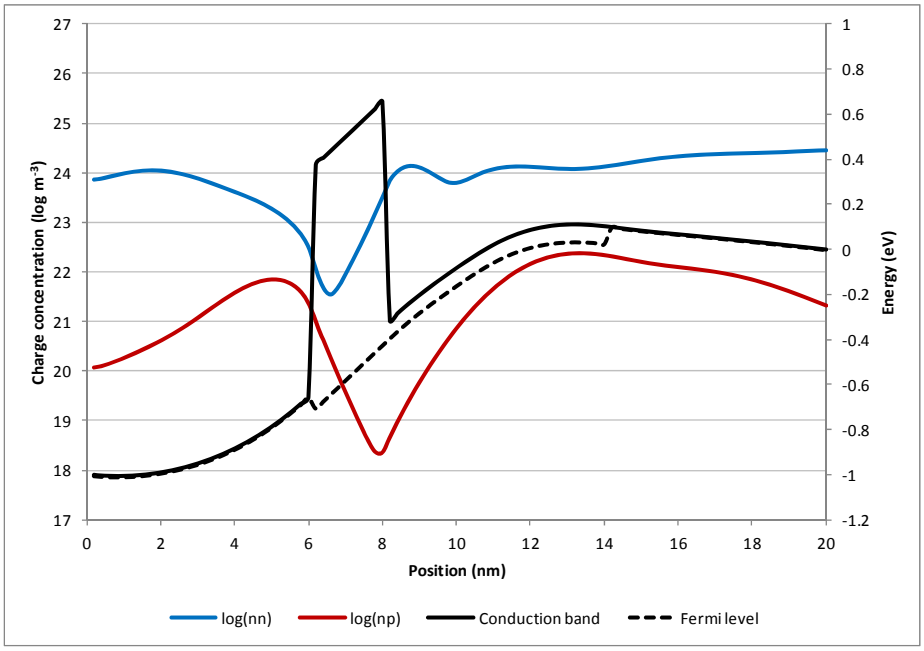
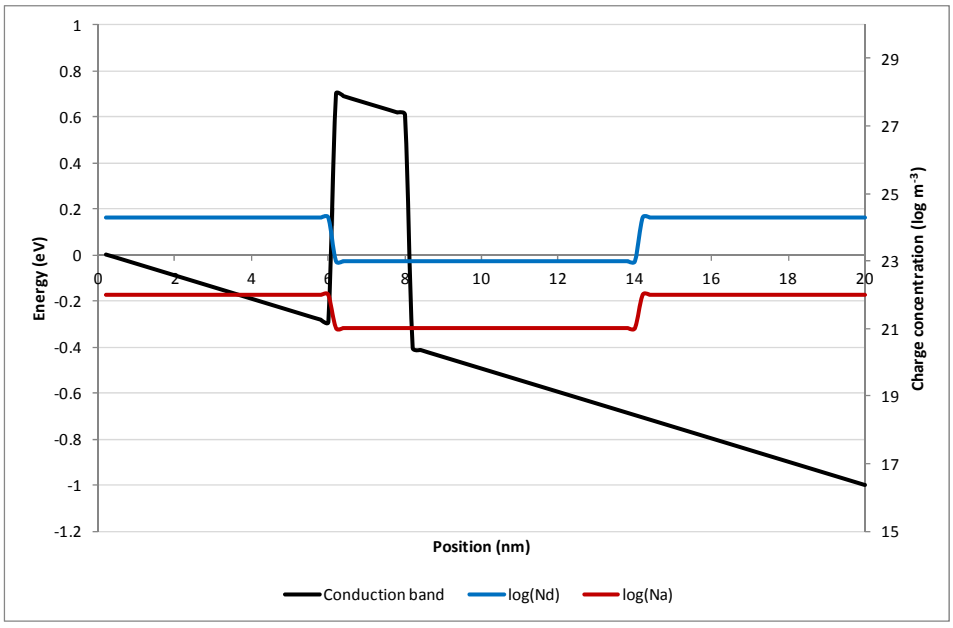


Figure 4: Starting Potential and Charge Concentration – Test Device, 1.0 V applied



The device that was simulated has an 8 nm region with relatively low dopant concentration in its centre, and two 6 nm regions with relatively high dopant concentration adjacent to each contact. There is an insulator region located in the region 6 to 8 nm from the left contact. Electrons are the majority carriers.

Figure 3 (A) shows the charge distributions, the Fermi energy, and the conduction band energy that are calculated when no voltage is applied across the device. The concentrations of free electrons and holes are reduced in the insulator region, as expected. The free electron concentration is reduced by approximately two orders of magnitude inside the potential barrier. Electron concentration is locally correlated with ionized donor concentration. Figure 3 (A) shows that free electron concentration is reduced in the centre of the device, where the concentration of fixed positive ions is lowest. The device has an abrupt changes in donor concentration at the edges of the centre region, but the free charge concentration is continuous, because it is a linear combination of PDFs obtained from continuous wave-functions. Furthermore, the device has an insulator region, inside which the free carrier concentration is reduced. For these reasons, the potential across the device is not expected to be flat. The simulator predicts that to the right of the insulator region, free electron concentration is higher than ionized donor concentration. Therefore, a potential well is created from the perspective of

the minority carriers, and a peak in hole concentration appears at the centre of this potential well.

For Figure 3 (A), (B), and (C), the Fermi levels were calculated using equations (4.28) and (4.29).

Figure 3 (B) and Figure 3 (C) results are shown for +1.0 V and -1.0 V respectively applied at the $x = 0$ contact.

Comparing Figure 3 (B) with Figure 4, it can be seen how the self-consistent Schrödinger-Poisson potential differs from the potential that is calculated when charge neutrality is assumed everywhere. The free electron concentration is reduced to the right of the barrier in Figure 3 (B) for two reasons. Firstly, the dopant concentration is reduced; and secondly, the applied field accelerates free electrons present inside the device toward the contact. In the region to the right of the barrier, there is a potential well for holes. In this potential well, Figure 3 (B) shows the formation of a "quantized inversion layer", where the concentration of free holes exceeds the concentration of free electrons. The pattern of maxima and minima is suggestive of the existence of standing waves in the potential well. Such standing waves could occur if the population of holes in the potential well occupied narrow energy bands. Since this feature of the results was unexpected and precedents were not found in the literature, the simulation was repeated with an increase in resolution in energy, to determine whether the apparent

quantization was the result of a limited number of energy samples having non-negligible probability of occupation at the positions where the potential well existed. It was found that the same results were obtained even when resolution in energy was increased.

Tunnelling from the inversion layer is an important effect co consider when calculating MOSFET leakage current [38], so it is important that inversion layer formation can be observed in the simulation results, and threshold voltage can be extracted.

In Figure 3 (B), the minimum of free electron concentration is at the expected location, which is the right edge of the insulator region. That position is the expected free electron concentration minimum because the direction of electron travel is from left to right, so the only electrons expected to be found there are ones that have tunneled through the insulator region and those that have reached the barrier by travelling from the right contact, in opposition to the direction of net electron current flow.

Figure 3 (B) may be compared with Figure 3 (C) to show the effects of voltage polarity on the charge distributions. As expected, when the voltage polarity is reversed, a potential well for electrons, instead of holes, arises to the right of the barrier. Despite the lower concentration of dopant in the well, electron density still reaches a peak value in this region.

5.2. Resonant Tunneling Diode

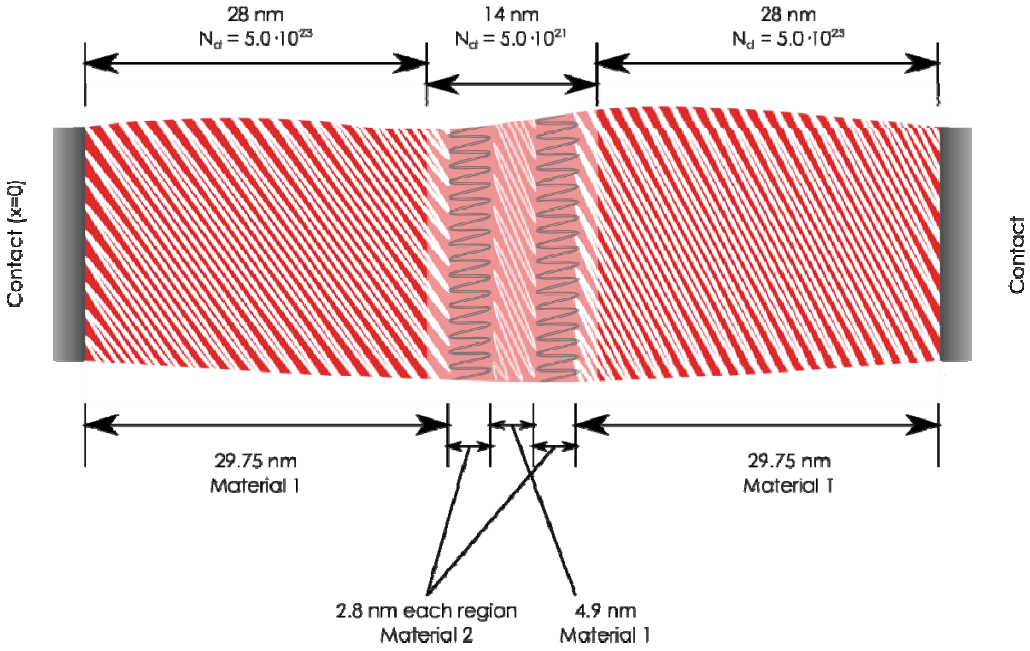
The resonant tunneling diode (RTD) is an often-used device for non-equilibrium quantum-mechanical simulations [35] [36] [37] [22] [39]. In this section, the simulator developed in this thesis is applied to a resonant tunneling diode model, to facilitate comparison with other simulation methodologies.

The resonant tunneling diode subject to simulation is shown in Figure 5. The RTD is meant to be similar to other RTDs simulated in [35], [37], and [22], and tested in [40].

5.2.1. Charge Concentration

The calculated free carrier concentration in the RTD, with 0.25V applied, is shown in Figure 6. The charge concentration shown is self-consistent, meaning that it was produced by a convergent iteration between the coupled Poisson and Schrödinger equations. Figure 6 is provided to facilitate comparison with the results in [22] showing charge concentration for a similar RTD, the same applied voltage, and three alternative simulation methods.

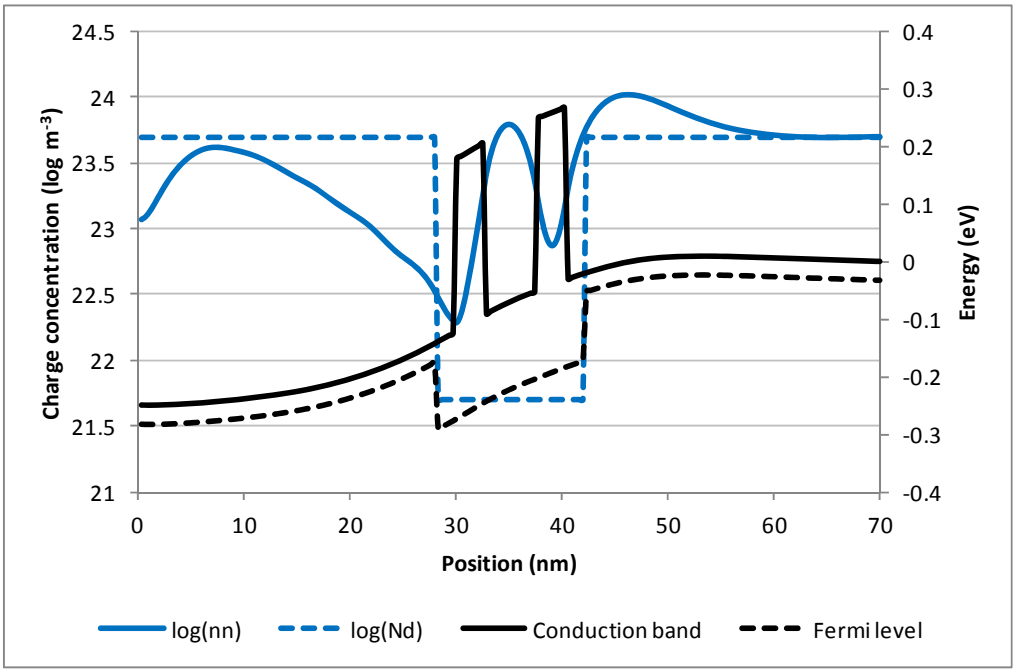
Figure 5: Device Properties – Resonant Tunneling Diode



Material 1	**Material 2**
GaAs	$Al_xGa_{(1-x)}As$[2]
$\Phi_{CB} = 0$	$\Phi_{CB} = 0.3$ eV
$\Phi_{VB} = 0$	$\Phi_{VB} = 0.1$ eV
$\kappa = 13.18$	$\kappa = 12.24$
$m_e^* = 0.067$	$m_e^* = 0.067$
$m_h^* = 0.34$	$m_h^* = 0.34$
$E_G = 1.42$ eV	
$n_i = 2.03e12\ m^{-3}$ [21]	

[2] The conduction band offset and effective mass are from [22] . It was assumed that x ~ 0.3. The dielectric constants are from [34] and the valence band offset is estimated based on [34] and [21].

Figure 6: Charge Concentration – RTD, 0.25 V Applied



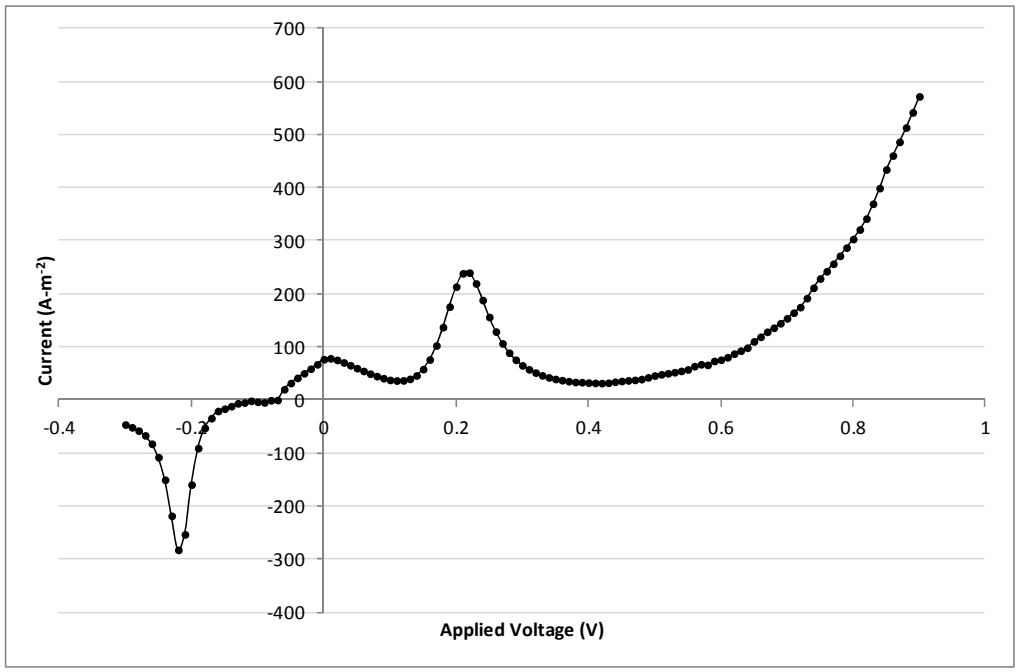
The charge concentrations shown in Figure 6 and calculated in [22] have a few key differences. Foremost among these is that in Figure 6 the free and fixed charge concentrations at the left contact are not equal, whereas [22] has equal free and fixed charge concentrations at the boundaries. The boundary conditions used to solve the Poisson equation were those of fixed voltage at the boundaries. It was assumed that there was no electric field in the contacts, or equivalently, that voltage was spatially invariant in the contacts. The free charge concentration is given by a continuous function. It can be argued that free charge concentration must be continuous because it can be found by adding a series of continuous probability density functions. Since free charge concentration is continuous at the

device/reservoir boundary, it is not equal to the fixed charge concentration, and the voltage in the reservoir is spatially invariant, the result in Figure 6 implies a discontinuity in fixed charge concentration or Fermi energy at the right contact. This result is not inconsistent with any explicit modelling assumption. It was assumed in the model that the Fermi level, as a function of position, could be discontinuous; and that discontinuities could arise at boundaries between different materials. Additionally, the model allows for arbitrary, possibly discontinuous fixed charge concentration. The simulator developed for this thesis did not ensure or guarantee that free and fixed concentrations at the device boundaries are equal, or that free charge concentrations at the device boundaries are equal to specific values. In order for the simulator to produce results with specific charge concentrations at the device boundaries, additional boundary conditions would need to be formulated and incorporated into the simulator.

5.2.2. Current-Voltage Curve

The simulator was used to calculate the current voltage curve for a resonant tunneling diode. The RTD that was simulated was a modified version of the one shown in Figure 5, with a total length of 35 nm. (The regions of "Material 1" adjacent to the contacts, shown in Figure 5 as 29.75 nm long, were shortened to be 12.25 nm). The current-voltage (I-V) curve for the RTD is shown in Figure 7.

Figure 7: Current-Voltage Curve – RTD



A few key features of the resonant tunneling diode's behaviour are the peak current at resonance, the peak/valley current ratio, and the voltage at which the resonance occurs. The peak/valley current ratio is the ratio of the maximum current at the resonance to the minimum value of current reached at a higher voltage. These key features are extracted from other published works, and compared to the results obtained in this thesis, in Table 3. The results provided for comparison are based on simulations of RTDs that similar to, but not exactly the same as, the RTD simulated in this thesis. The full details of the device structures are discussed in the references listed in Table 3.

Table 3: Comparison of RTD Results

Source	Peak Resonant Current [A/m^2]	Peak/Valley Ratio	Resonant Voltage [mV]
Figure 7	$2.4 \cdot 10^2$	8.0	300 [3]
Frensley [35]	$1.8 \cdot 10^8$	4.5	120 [4]
El Ayyadi and Jungel [22]	$1.1 \cdot 10^8$	5.5	260 [5]
Lake and Datta [37]	$7.0 \cdot 10^8$	> 10	150 [6]
Reed et al, experimental [40]	$1.2 \cdot 10^5$	4.5	260

As Table 3 shows, the simulators' predictions for resonant voltage were similar, and were close to the experimental result. The simulators discussed in [22], [35] and [37], made similar predictions for current density, on the order of 10^8 A/m^2. The value measured in [40] was about 3 order of magnitude lower, and the value obtained in this thesis was 3 orders of magnitude lower than the measured value.

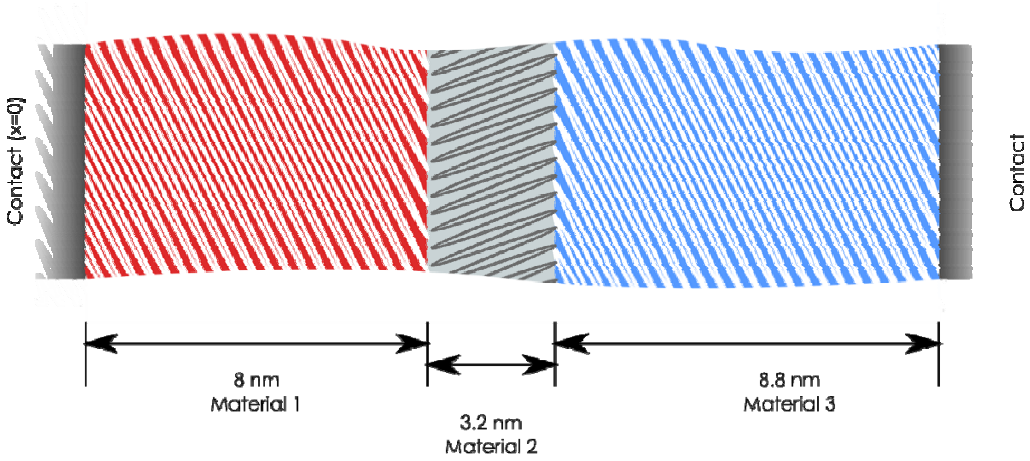
[3] Measured relative to current zero
[4] Based on the Wigner function approach
[5] Based on the QDD-SP approach
[6] Based on the KKB approach

6. Application to Field Effect Transistor Modeling

6.1. Basic Device Model

The basic field effect transistor (FET) device considered in this thesis was similar to the one examined in [23]; it was an n-FET, with an n-type gate, a p-type substrate, and an SiO_2 gate dielectric that varied in width from 1.2 nm to 6.4 nm. The device, with a specific gate dielectric width of 3.2 nm, is shown in Figure 8. Material parameters were obtained from published literature [21] [41]. When dielectric layers thicker or thinner than 3.2 nm were considered, the increase in size of the material 2 (gray) region was compensated by a decrease in size of the material 3 (blue) region.

Figure 8: Device Properties – Field Effect Transistor



Material 1	**Material 2**	**Material 3**
Poly Si	SiO_2	Si, $n_i = 1.0e16\ m^{-3}$
$\Phi = 0$	$\Phi = 3.2$ eV	$\Phi = 0$
$\Phi_{VB} = 0$	$\Phi_{VB} = 4.7$ eV	$\Phi_{VB} = 0$
$E_g = 1.12$ eV	$E_g = 9$ eV	$E_g = 1.12$ eV
$N_d = 5e25\ m^{-3}$		$N_a = 5e23\ m^{-3}$
$\kappa = 11.7$	$\kappa = 3.9$	$\kappa = 11.7$
$m_e^* = 1.08$	$m_e^* = 1.08$	$m_e^* = 1.08$
$m_h^* = 0.48$	$m_h^* = 0.48$	$m_h^* = 0.48$

6.2. Charge Distribution

Two examples of calculated charge distribution and conduction band offset are shown for the simulated device in Figure 9. In Figure 9 (A), the results are shown

for the n-FET in accumulation, with 0.4 V applied. In Figure 9 (B), the results are shown for the n-FET in inversion, with -1.0 V applied. The charge concentrations are self-consistent, in that a convergent solution of the Schrödinger-Poisson system of differential equations was found. The quasi-Fermi energies were defined piecewise, as per equation (4.29). Their values in the gate and body were held fixed to the values in the contacts. The quasi-Fermi energies in the dielectric were the averages of the gate and body values. Section 4.5 discusses alternatives for defining the quasi-Fermi levels and their consequences.

Figure 9: Charge Concentration – FET

(A): Accumulation (+0.4 V applied)

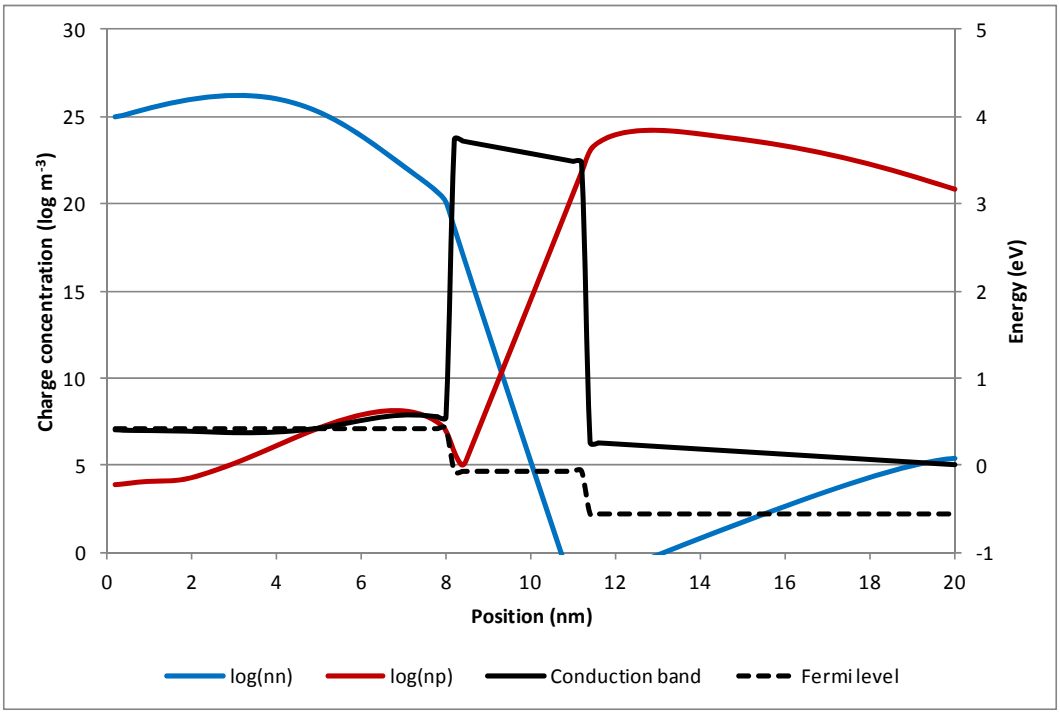
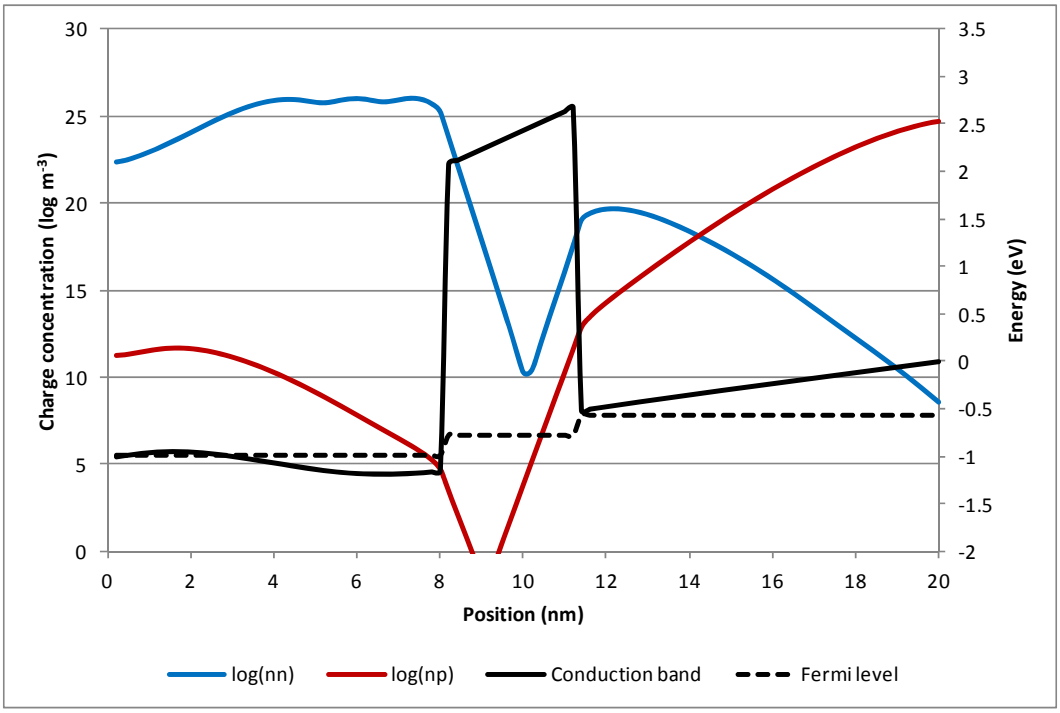


Figure 9 (B): Inversion (-1.0 V applied)



Qualitatively, the charge concentrations had the expected features. Notably, formation of the inversion layer was observed. Without the proposed modifications to the nonequilibrium Green's function method, it seems that this result would not be obtained.

6.3. Gate Leakage Current

The calculated get leakage current, as a function of applied voltage, is shown in Figure 10. As with the results in Section 6.2, the quasi-Fermi levels are defined

piecewise. Figure 10 shows current-voltage curves for three variants of the basic device shown in Figure 8, with gate dielectric thickness varied.

Figure 10: Current-Voltage Curve – FET

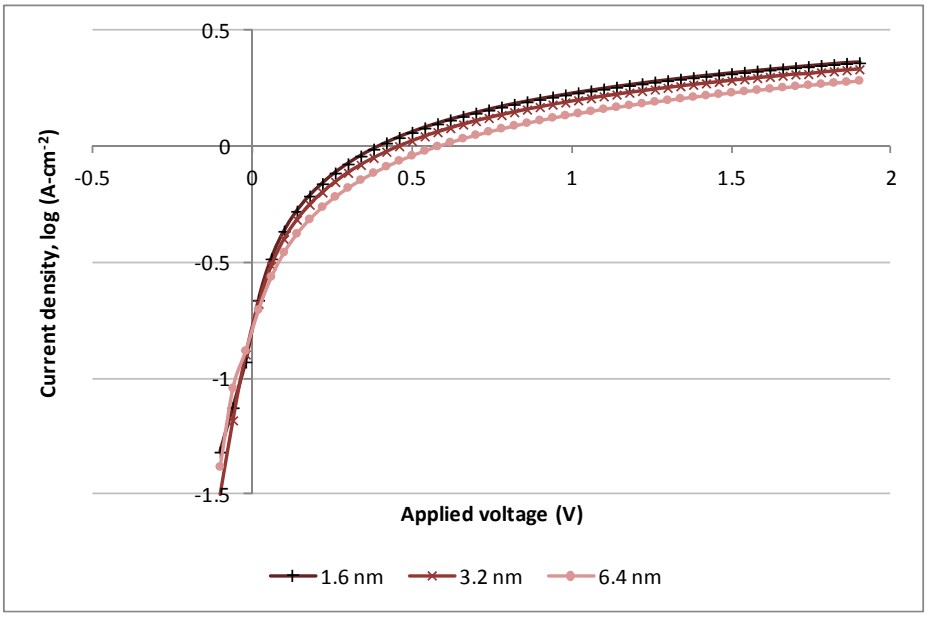


Table 4 shows a comparison with the simulated results in [23] (which match quite closely with experimental results). As Table 4 shows, the results in Figure 10 are not quite realistic.

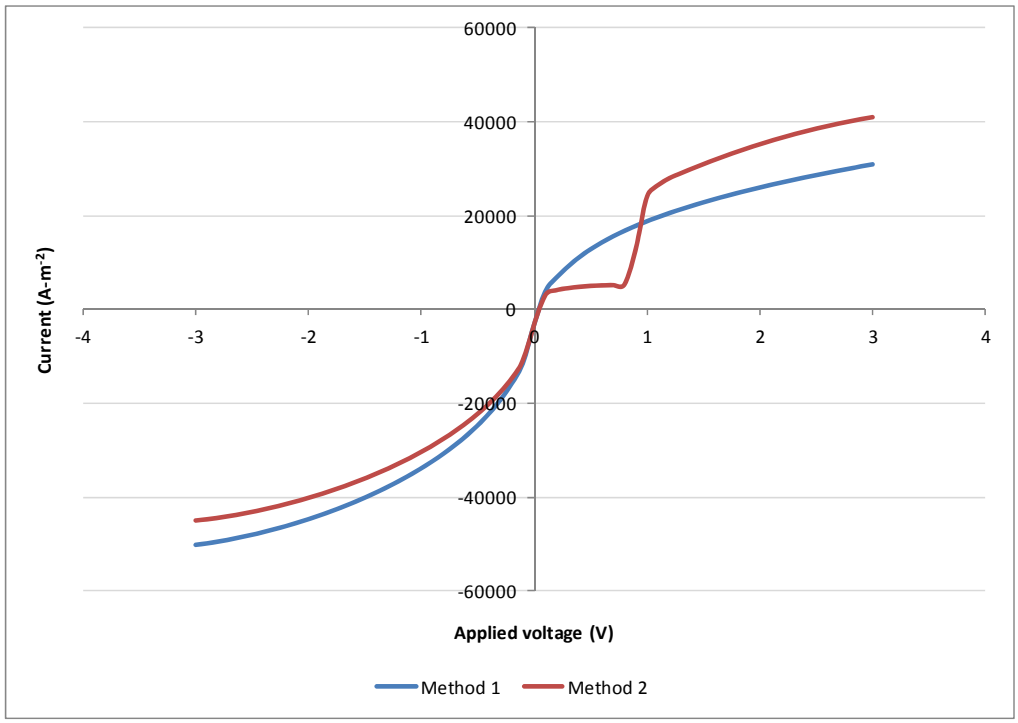
Table 4: Comparison of Tunneling Current Results

Dielectric Thickness	1.6 nm		3.2 nm	
Applied voltage	0.5 V	1.5 V	0.5 V	1.5 V
Figure 10 [A/cm^2]	1.15	2.06	1.05	1.93
Buchanan et al [23] [A/cm^2]	~1	~10	~10^{-8}	~10^{-7}

Calculated results in the two studies give similar I-V curves for the 1.6 nm dielectric, but in this study, increasing the width of the dielectric has a relatively small effect on current. The failure to correctly capture the dependence of current on dielectric thickness is likely attributable in large part to the procedure used to assign quasi-Fermi levels within the device. For example, if there were a discontinuity in Fermi level at either boundary of the dielectric, the I-V curve could be considerably different. A secondary factor is the grid spacing used for the finite difference differential operators; the grid spacing used was 1% of device length. With smaller grid spacing, it is likely that the difference in current as a function of dielectric thickness would increase.

To illustrate the significance of the way in which the quasi-Fermi levels in the device are determined, the limit was considered in which the gate dielectric was removed entirely. Two current-voltage curves, for the device of Figure 8 with the gate dielectric removed, are shown in Figure 11.

Figure 11: Current-Voltage Curves With No Gate Dielectric



For Method 1 shown in Figure 11, smoothly varying quasi-Fermi levels were used, such as given by equation (4.29). For Method 2 shown in Figure 11, piece-wise defined Fermi levels were used, such as given by equation (4.30). When any thickness of gate dielectric is modeled, it blocks current in the "reverse" direction. A qualitative difference in behaviour for no dielectric, versus one that is negligibly thin, suggests assumptions made about the position-dependence of quasi-Fermi levels near material boundaries play an important role in determining simulation outcomes.

7. Conclusion

A general method has been presented for calculating charge concentration and current in nanoscale electronic devices. The proposed simulation methodology has been used to illustrate the qualitative behaviour of a resonant tunneling diode, the gate leakage of a MOS capacitor, and a P-N junction, with modest success.

The calculation method is based on the well-published non-equilibrium Green's function (NEGF) method [14], [16]. However, the standard method was modified by making use of spatially varying quasi-Fermi levels in the construction of density operators.

Additional effort would be needed for the method presented in this thesis to progress from a conceptual demonstration to a practical predictive tool. The most significant improvements that are needed include:

- An adaptation of the method to three dimensions, making it possible to obtain plausible results that obey conservation of energy and momentum
- The development of a convincing procedure for finding uniquely correct position-dependent quasi-Fermi levels. It is possible that some conditions cannot adequately be described by quasi-Fermi levels, as the form of the

distribution could be entirely incorrect. In such cases, a replacement distribution (with position as a parameter) is really what is needed.

- Possible replacement of the condition that devices be strictly charge neutral with the condition that net charge asymptotically approaches zero as the device expands to cover the reservoir regions

The proposed method had a few successes. The shapes of current-voltage curves, though not the magnitudes, could be reproduced for resonant tunneling diodes and field effect transistors. One stated purpose, that of developing a Green's function-based method that predicted charge concentrations sensitive to potential wells inside devices, was achieved. The method was also successful at increasing the range of voltage that could be applied across device contacts as compared to typical NEGF methods. However, the dependence of the I-V curve of a FET on its gate dielectric thickness was not correctly determined by the simulator.

Appendix A: Separation of Variables for the Single Particle Schrödinger Equation

This Appendix begins with equation (4.2), the Schrödinger equation for a single charge-carrying particle moving through a device. Time and three spatial dimensions are considered. It is shown how to separate the partial differential equation, equation (4.2), into the three separate equations listed in Table 2 that can be solved for components of the wave-function.

The Schrödinger equation is:

$$\left[-\frac{\hbar^2}{2m}\Delta + U_x(X) + U_{yz}(Y,Z)\right]|\xi\rangle = i\hbar\frac{\partial}{\partial t}|\xi\rangle$$

It is assumed that the wave-function can be written as:

$$\xi(x,y,z,t) = \psi(x)\chi(y,z)\vartheta(t)$$

Therefore, in the position basis,

$$-\frac{\hbar^2}{2m}\left(\left[\frac{\partial\chi}{\partial y}+\frac{\partial\chi}{\partial z}\right]\psi + \chi\frac{d\psi}{dt}\right)\vartheta + \left(U_x(x)+U_{yz}(y,z)\right)\psi\chi\vartheta = i\hbar\psi\chi\frac{d\vartheta}{dt}$$

Dividing both sides by the total wave-function yields

$$-\frac{\hbar^2}{2m}\left(\frac{1}{\chi}\left[\frac{\partial\chi}{\partial y}+\frac{\partial\chi}{\partial z}\right]+\frac{1}{\psi}\frac{d\psi}{dt}\right) + U_x(x) + U_{yz}(y,z) = \frac{i\hbar}{\vartheta}\frac{d\vartheta}{dt}$$

Since the left-hand side depends only on the variables x, y, and z, and the right-hand side depends only on the variable t, the only way for the two sides to be

equal is for them both to be equal to a constant with an arbitrary value. This constant is denoted C_1. Two equations are obtained:

$$C_1 = \frac{i\hbar}{\vartheta}\frac{d\vartheta}{dt}$$

And,

$$-\frac{\hbar^2}{2m}\frac{1}{\chi}\left[\frac{\partial\chi}{\partial y}+\frac{\partial\chi}{\partial z}\right]+U_{yz}(y,z) = C_1 - U_x(x) + \frac{\hbar^2}{2m}\frac{1}{\psi}\frac{d\psi}{dt}$$

The right and left sides of the latter equation once again have no variables in common. The left side depends only on y and z, and the right side depends only on x. Therefore, both sides must be equal to an arbitrary constant. This constant is denoted C_2. Two equations are obtained:

$$-\frac{\hbar^2}{2m}\frac{1}{\chi}\left[\frac{\partial\chi}{\partial y}+\frac{\partial\chi}{\partial z}\right]+U_{yz}(y,z) = C_2$$

And,

$$-\frac{\hbar^2}{2m}\frac{1}{\psi}\frac{d\psi}{dt}+U_x(x) = C_1 - C_2$$

When the Schrödinger equation is separated, it is usual to interpret the constants of separation as energies. They have units of energy.

The constant C_1 is equal to the total kinetic energy of the particle:

$$E + E_{yz} = C_1$$

The constant C_2 is equal to the transverse kinetic energy of the particle: $C_2 = E_{yz}$. Therefore, the longitudinal kinetic energy E is given by:

$$E = C_1 - C_2$$

In this way, the three separated differential equations that are given in Table 2 as components of the total wave-function are recovered.

Appendix B: Discussion of the Lippmann-Schwinger Equation

This appendix has the goal of demonstrating the sense in which solutions of the Lippmann-Schwinger equation are Green's functions of the Schrödinger equation.

The statement of the Lippmann-Schwinger equation given in Chapter 4 refers to a function ψ^+, which is defined as:

$$\psi^+(x,t;E,E_0) = \lim_{\varepsilon\to 0^+}\int_{-\infty}^{\infty}\exp\left(\frac{i(E-H_0)t}{\hbar}\right)\exp\left(-\frac{\varepsilon|t|}{\hbar}\right)U_+(t)\,\phi(x,t;E_0)dt$$

The function $\phi(x,t;E_0)$ is defined to be a wavefunction that satisfies the time-independent Schrödinger equation $H_0\phi = E_0\phi$. Therefore the time derivative of $\phi(x,t;E_0)$ is:

$$i\hbar\frac{d\phi}{dt} = E_0\phi \qquad (1)$$

The operator $U_+(t)$ for any specific perturbation Hamiltonian H_1 is given by the integral equation,

$$U_+(t) = 1 - \frac{i}{\hbar}\int_{-\infty}^{t} H_1(t')U_+(t')dt'$$

Differentiating the above yields:

$$\frac{dU_+}{dt} = -\frac{i}{\hbar} H_1(t) U_+(t)$$

The time derivative of $\psi^+(x,t;E,E_0)$ is calculated as follows:

$$\begin{aligned}\frac{d\psi^+}{dt} = \lim_{\varepsilon\to 0^+} \int_{-\infty}^{\infty} \Bigg\{ \frac{d}{dt}\Bigg\{ \exp\left(\frac{i(E-H_0)t}{\hbar}\right)\exp\left(-\frac{\varepsilon|t|}{\hbar}\right)\Bigg\} \\ \cdot \big(U_+(t)\phi(x,t;E_0)\big) + \Bigg\{ \exp\left(\frac{i(E-H_0)t}{\hbar}\right)\exp\left(-\frac{\varepsilon|t|}{\hbar}\right)\Bigg\} \\ \cdot \frac{d}{dt}\big(U_+(t)\phi(x,t;E_0)\big)\Bigg\} dt\end{aligned} \tag{2}$$

In the above,

$$\begin{aligned}&\frac{d}{dt}\Bigg\{ \exp\left(\frac{i(E-H_0)t}{\hbar}\right)\exp\left(-\frac{\varepsilon|t|}{\hbar}\right)\Bigg\} \\ &\quad = \left(\frac{i}{\hbar}(E-H_0) - \frac{\varepsilon}{\hbar}\mathrm{sign}(t)\right)\Bigg\{ \exp\left(\frac{i(E-H_0)t}{\hbar}\right)\exp\left(-\frac{\varepsilon|t|}{\hbar}\right)\Bigg\}\end{aligned}$$

and

$$\frac{d}{dt}\big(U_+(t)\phi(x,t;E_0)\big) = -\frac{i}{\hbar}(H_1(t)+E_0)\big(U_+(t)\phi(x,t;E_0)\big)$$

Therefore,

$$\begin{aligned}-\frac{d\psi^+}{dt} = \lim_{\varepsilon\to 0^+} \int_{-\infty}^{\infty} \left(\frac{i}{\hbar}(H_1(t)+H_0) + \frac{\varepsilon}{\hbar}\mathrm{sign}(t) - \frac{i}{\hbar}(E-E_0)\right) \\ \cdot \left(\exp\left(\frac{i(E-H_0)t}{\hbar}\right)\exp\left(-\frac{\varepsilon|t|}{\hbar}\right) U_+(t)\phi(x,t;E_0)\right) dt\end{aligned}$$

The method proposed in this work uses the following H_1:

$$H_1(x;x') = \varpi\delta(x-x')$$

The delta function has the following properties:

$$\left(\frac{d^n}{dx^n}\delta(x-x')\right)f(x) = (-1)^n f^{(n)}(x') = (-1)^n\delta(x-x')\left(\frac{d^n f(x)}{dx^n}\right)$$

$$\delta(x-x')f(x) = f(x')$$

The unperturbed Hamiltonian $H_0(x)$ is of the form

$$H_0 = U_x(x) - \frac{\hbar^2}{2m}\frac{d^2}{dx^2}$$

Therefore,

$$\exp\left(\frac{iH_0 t}{\hbar}\right) = \sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{it}{\hbar}\right)^k\left(U_x(x) - \frac{\hbar^2}{2m}\frac{d^2}{dx^2}\right)^k$$

Using the stated properties of the delta function, and noting that the differential operator only appears in even powers, it can be seen that for any $f(x)$:

$$\int\left[\exp\left(\frac{iH_0(x)t}{\hbar}\right), H_1(x;x')\right]f(x)dx = 0$$

Recalling from equation (4.8) that $H_1(x,t;x') = \exp(iH_0 t/\hbar)\cdot H_1(x;x')\cdot \exp(-iH_0 t/\hbar)$, and using the above, equation (2) becomes:

$$i\hbar\frac{d\psi^+}{dt}(x,t;E,E_0,x')$$

$$=\big(H_0(x)+H_1(x;x')-(E-E_0)\big)\psi^+$$

$$-\lim_{\varepsilon\to 0^+} i\varepsilon\int_{-\infty}^{\infty}\text{sign}(t)\left(\exp\left(\frac{i(E-H_0)t}{\hbar}\right)\exp\left(-\frac{\varepsilon|t|}{\hbar}\right)U_+(t)\phi(x,t;E_0)\right)dt$$

The limit evaluates to zero. The function $\psi^+(x,t;x',E,E_0)$ therefore satisfies the Schrödinger equation,

$$i\hbar\frac{d\psi^+}{dt}=(H_0(x)+H_1(x;x')+E_0-E)\psi^+$$

Making use of equation (1), and the linearity of the time derivative, it is seen that:

$$i\hbar\frac{d}{dt}\{\psi^+-\phi\}=E_0(\psi^+-\phi)+(H_0+H_1-E)\psi^+ \qquad (3)$$

Lippmann and Schwinger show in [10] that $\psi^+(x,t;E,E_0,x')$ satisfies:

$$\psi^+=\phi(x,t;E_0)+\frac{1}{E+i\varepsilon-H_0(x)}H_1(x;x')\psi^+$$

Equivalently,

$$(E+i\varepsilon-H_0)(\psi^+-\phi)=H_1\psi^+$$

Taking the limit $\varepsilon\to 0$ and rearranging yields

$$(E-H_0-H_1)\psi^+ + (E-H_0)\phi=(E-H_0-H_1)\psi^+ + (E-E_0)\phi=0$$

Therefore,

$$i\hbar\frac{d}{dt}(\psi^+-\phi)=E_0(\psi^+-\phi)+(E-E_0)\phi$$

As perturbation magnitude decreases, the following limit applies:

$$\lim_{\varpi \to 0} E = E_0$$

The notation is introduced:

$$g(x; x', E) = \lim_{\varpi \to 0, \varepsilon \to 0^+} \left(\psi^+(x, t; x', E, E_0) - \phi(x, t; E_0) \right)$$

The function $g(x, t; x', E)$ is referred to as a Green's function, and each one satisfies a Schrödinger equation,

$$i\hbar \frac{dg}{dt} = Eg = (H_0 + H_1)g$$

Such Green's functions are obtained by solving the matrix form of the Lippmann-Schwinger equation.

References

[1] G. B. Arfken, H. J. Weber, and F. E. Harris, *Mathematical Methods for Physicists,* 6th ed. Elsevier, 2005.

[2] R. J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady State and Time Dependent Problems*. SIAM, 2007.

[3] A. Iserles, *A First Course in the Numerical Analysis of Differential Equations*, 2nd ed. Cambridge University Press, 2009.

[4] J. Kiusalaas, *Numerical Methods in Engineering with Python*, 2nd ed. Cambridge University Press, 2010.

[5] T. A. Davis, *Direct Methods for Sparse Linear Systems*. SIAM, Philadelphia, Sept. 2006.

[6] T. A. Davis, J. R. Gilbert, S. Larimore, and E. Ng, "A column approximate minimum degree ordering algorithm", ACM Trans. Mathematical Software, vol 30, no 3, pp 353-376, Sept. 2004.

[7] A. Gupta, George Karypis, and Vipin Kumar. "Highly Scalable Parallel Algorithms for Sparse Matrix Factorization". IEEE Trans. Parallel Distrib. Systems, vol. 8, issue 5, pp 502-520, 1997.

[8] J. J. Sakurai, *Modern Quantum Mechanics* (Revised ed.) Addison Wesley, 1993.

[9] R. Shankar, *Principles of Quantum Mechanics*, 2nd ed. Springer, 1994.

[10] B. A. Lippmann and Julian Schwinger, "Variational Principles for Scattering I", Phys. Rev., vol 79, no 3, August 1, 1950.

[11] M. Paulsson, "One Particle NEGF". [Online: http://nanohub.org/resources/1933/download/paulsson_negf_for_dummies.pdf] March 15, 2002.

[12] D. A. Lavis, B. W. Southern, and I. F. Wilde, "The inverse of a semi-infinite symmetric banded matrix". J. Phys. A: Math. Gen., vol 30, no 20, pp 7229, 1997.

[13] G. Y. Hu and R. F. O'Connell, "Analytical inversion of symmetric tridiagonal matrices", J. Phys. A: Math. Gen., vol 29, pp 1511, 1996.

[14] S. Datta, "Nanoscale device modeling: the Green's function method". Superlattices and Microstructures, vol 28, no 4, 2000.

[15] G. W. Hanson and A. B. Yakovlev, *Operator Theory for Electromagnetics: An Introduction*, Springer, 2002.

[16] R. K. Lake, *Application of the Keldysh Formalism to Quantum Device Modeling and Analysis*. Dissertation: Perdue University, 1992.

[17] J. Rammer and H. Smith, "Quantum field-theoretical methods in transport theory of metals", Rev. Mod. Phys., vol 58, no 2, April 1986.

[18] A. P. Jauho and J. W. Wilkins, “Theory of high-electric-field quantum transport for electron-resonant impurity systems”, Phys. Rev. B, vol 29, no 4, February 1984.

[19] P. R. Johnson and B. L. Hu, “Stochastic Theory of Relativistic Particles Moving in a Quantum Field: I. Influence Functional and Langevin Equation”. [preprint: quant-pu/0012137v4], 1 June 2001.

[20] P. R. Johnson and B. L. Hu, “Stochastic Theory of Relativistic Particles Moving in a Quantum Field: II. Scalar Abraham-Lorentz-Dirac-Langevin Equation, Radiation Reaction and Vacuum Fluctuations”. [preprint: quant-ph/0101001v2], 1 June 2001.

[21] B. Van Zeghbroeck, *Principles of Semiconductor Devices*, 2011. [Online: http://ecee.colorado.edu/~bart/book/book/title.htm]

[22] A. El Ayyadi and A. Jungel, “Semiconductor Simulations Using a Coupled Quantum Drift-Diffusion Schrödinger-Poisson Model”. SIAM J. Appl. Math., vol 66, no 2, pp 554-572, 2006.

[23] S-H Lo, D. A. Buchanan, and Y. Taur, “Modeling and characterization of quantization, polysilicon depletion, and direct tunneling effects in MOSFETs with ultrathin oxides”. IBM J Res. Develop., vol 43, no 3, May 1999.

[24] *Z.* Ren, *Nanoscale MOSFETs: Physics, Simulation, and Design*. Dissertation: Perdue University, 2001.

[25] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. John Wiley & Sons, New York, 1998.

[26] C. Lanczos, *Linear Differential Operators*. Dover, 1997.

[27] David J Griffiths, *Introduction to Electrodynamics*, 3rd ed. Prentice Hall, Upper Saddle River, New Jersey. 1999.

[28] L. D. Landau and E. M. Lifshitz. *Mechanics*, 3rd ed. (Course of Theoretical Physics, vol. 1). Butterworth-Heinmann, Linacre House, Jordan Hill, London. 1976.

[29] H. Goldstein, *Classical Mechanics*, 2nd ed. Addison-Wesley, 1980.

[30] Normand M Laurendeau, *Statistical Thermodynamics: Fundamentals and Applications*. Cambridge University Press, New York. 2005.

[31] John S. Blakemore, *Semiconductor Statistics*. Dover, 1987.

[32] R. K. Pathiria and P. D. Beale, *Statistical Mechanics*, 3rd ed. Elsevier, 2011.

[33] M. Lundstrom, *Fundamentals of Carrier Transport*, 2nd ed. Cambridge University Press, 2000.

[34] S. Adachi, "GaAs, AlAs, and $Al_xGa_{(1-x)}As$: Material parameters for use in research and device applications". J Appl. Phys., vol 58, issue 3, R1, August 1985.

[35] W. R. Frensley, “Boundary conditions for open quantum systems driven far from equilibrium”. Rev. Mod. Phys., vol 62, issue 3, pp 745-791, July 1990.

[36] M. A. Reed, W. R. Frensley, R. J. Matyi, J. N. Randall, and A. C. Seabaugh, “Realization of a three-terminal resonant tunneling device: The bipolar quantum resonant tunneling transistor”. Appl. Phys. Lett., vol 54, no 11, pp 1034, March 1989.

[37] R. Lake and S. Datta, “Nonequilibrium Green’s-function method applied to double-barrier resonant-tunneling diodes”. Phys. Rev. B, vol 45, no 12, March 1992.

[38] 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 Device Lett., vol 18, no 5, May 1997.

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

[40] M. A. Reed, W. R. Frensley, W. M. Duncan, R. J. Matyi, A. C. Seabaugh, and H-L. Tsai, “Quantitative resonant tunneling spectroscopy: Current-voltage characteristics of precisely characterized resonant tunneling diodes”. Appl. Phys. Lett., vol 54, no 13, pp 1256, March 1989.

[41] J. Robertson, “High Dielectric Constant Oxides”. Eur. Phys. J. - Appl. Phys., vol 28, pp 265-291, 2004.

[42] B. D. Bates and G. W. Staines, *Transverse Resonance Analysis Technique for Microwave and Millimetre-Wave Circuits*. Tech. Report DSTO-RR-0027. Defence Science and Technology Organisation, Salisbusy, Australia. February 1995.

[43] H. Jiang, W. Cai, and R. Tsu, “Accuracy of the Frensley inflow boundary condition for Wigner equations in simulating resonant tunneling diodes”. J. Comp. Phys., vol 230, pp 2031-2044, December 2010.

[44] M. V. Fischetti, “Master-equation approach to the study of electronic transport in small semiconductor devices”. Phys. Rev. B, vol 59, no 7, February 1999.

[45] Y. Zhao and M. H. White, “Modeling of direct tunneling current through interfacial oxide and high-K gate stacks”. Solid State Electronics, vol 48, pp 1801-1807, 2004.

[46] M. G. Chapline and S. X. Wang, “Analytical formula for the tunneling current versus voltage for multilayer barrier structures”, J. Appl. Phys. 101, 083706 (2007).