Physics and Modeling of Submicron Devices

Supriyo Datta  
Purdue University

Mark S. Lundstrom  
Purdue University

Amitava Das  
Purdue University

Roger Lake  
Purdue University

Michael j. McLennan  
Purdue University

See next page for additional authors

Follow this and additional works at: https://docs.lib.purdue.edu/ecetr

https://docs.lib.purdue.edu/ecetr/683

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Authors
Supriyo Datta, Mark S. Lundstrom, Amitava Das, Roger Lake, Michael J. McLennan, and Mark Stettler
Physics and Modeling of Submicron Devices

Supriyo Datta
Mark S. Lundstrom
Amitava Das
Roger Lake
Michael J. McLennan
Mark Stettler

TR-EE 89-59
Annual Report: August 1, 1988 - July 31, 1989

School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907

Supported by the Semiconductor Research Corporation Contract Number 88-SJ-089
PHYSICS and MODELING

of

SUBMICRON DEVICES

Annual Report: August 1, 1988 - July 31, 1989
Supported by the Semiconductor Research Corporation
Contract Number 88-SJ-089

Supriyo Datta
Mark S. Lundstrom

Amitava Das
Roger K. Lake
Michael J. McLennan
Mark A. Stettler

Purdue University
School of Electrical Engineering
Technical Report: TR-EE 89-59
West Lafayette, Indiana 47907
Motivation

The semiconductor industry is driven by a continuous trend towards miniaturization. However, as device sizes shrink below 0.25 μm, we approach a scaling barrier. To circumvent this scaling barrier, the SRC is funding work on new devices such as the heterojunction bipolar transistor (HBT) and the bipolar inversion channel field-effect transistor (BICFET) which make use of bandgap engineering and ultrathin active layers. The device simulation programs now in use throughout the semiconductor industry lose validity for such devices. New approaches to device simulation are required to model these and future devices.

Conventional devices modeling programs provide self-consistent solutions to the drift-diffusion equations and the Poisson equation subject to the appropriate boundary conditions on the carrier densities and electrostatic potential. This approach has provided an adequate description of electronic devices for the last three decades. However, with the continuing advancement of technology, devices have now shrunk to submicron dimensions, and there is an increasing concern regarding the validity of this approach. The familiar drift-diffusion theory is based on two assumptions:

1. Electrons are particles moving in an external electric field according to Newton's law and are scattered occasionally by phonons and impurities.
2. The electric field changes slowly over the scale of a mean free path, so that an electron is scattered many times before the field changes significantly.

In advanced silicon transistors assumption 2 is violated, leading to transient hot electron effects such as velocity overshoot which are described by the semiclassical Boltzmann Transport Equation. At even smaller dimensions, assumption 1 is violated; the wave nature of carriers becomes important leading to quantum interference and confinement effects. These effects will become increasingly important as device sizes shrink and even offer the possibility of a new class of electronic devices. It seems clear that fundamentally new approaches have to be developed to model devices of the future.

Objectives

Our work is directed at developing the new approaches that will be required to model future devices. Specific program objectives are: 1) to study transport in a bipolar context and to develop new approaches for advanced bipolar simulation, 2) to formulate and demonstrate a viable approach for simulating quantum effects in ultra-small devices, and 3) to apply our evolving simulation tools for exploring advanced post-shrink and bipolar devices.
Relevance to SRC

The benefits of this research to SRC members include: 1) an improved understanding of the device physics of small and ultra-small devices, 2) the demonstration of new, post-shrink device concepts and the identification and assessment of structures for improving the performance of conventional devices, and 3) the development of new device simulation strategies for advanced devices. During the course of this work, several numerical device simulation programs are being developed. The acquisition of this evolving "tool box" of advanced device simulators is, perhaps, the most tangible benefit that SRC-members realize. Copies of these simulation programs have been distributed for several years and are now widely-used by the industrial, government, and academic research communities for advanced device work.

Overview of the Report

This work summarizes our progress during the past year. Chapters 1 and 2 are related to our work on non-stationary transport in bipolar transistors while Chapters 3, 4 and 5 relate to our work on quantum transport in ultrasmall unipolar devices. In Chapter 1, we describe non-stationary transport effects which are beginning to influence advanced Si transistors. The chapter explains why drift-diffusion equations are losing validity as devices shrink in size, and it describes the Monte Carlo technique which treats classical transport very accurately. The development of a Si Monte Carlo program to study electron transport in advanced bipolar transistors is the subject of Chapter 2.

Chapter 3 reviews the recent theoretical and experimental developments in the area of quantum transport and identifies the important problems that need to be solved in order to develop practical tools for quantum device simulation. Chapters 4 and 5 describe the progress that we have made in the past year towards solving these problems.
### TABLE OF CONTENTS

1. CARRIER TRANSPORT IN ADVANCED Si TRANSISTORS .................................................1
2. MONTE CARLO ANALYSIS FOR Si TRANSISTORS .....................................................12
3. QUANTUM TRANSPORT IN ULTRASMALL DEVICES ..............................................29
4. A QUANTUM KINETIC EQUATION ...........................................................................62
5. POTENTIAL VARIATION ACROSS AN OBSTACLE ..................................................104

APPENDIX: SRC Supported Publications and Talks .......................................................142
Aggressive vertical scaling of Si bipolar transistors continues and is pushing $f_T$'s to 40 GHz and above and gate delays below 30 picoseconds. For the most part, these advanced devices are analyzed with numerical device simulators based on drift-diffusion equations, but the drift-diffusion equation is beginning to lose validity as the vertical dimensions of bipolar transistors shrink. The objective of this chapter is to explain in simple, physical terms the basis of the drift-diffusion equation and why it loses validity in very small devices. The chapter also describes the Monte Carlo technique, which will be increasingly used for analyzing small devices. In Chapter 2, we'll describe a silicon Monte Carlo program, S-DEMON, which is being developed at Purdue.
1.1 Introduction

As semiconductor devices continue to evolve, the techniques required to analyze, design, and optimize them have become increasingly sophisticated. Computer simulation programs such as PISCES and MINIMOS are now important tools for the device engineer [1,2]. By simulating the flow of carriers through a semiconductor device under the influence of their self-consistent electrostatic field and the applied bias, such programs accurately predict device performance. They also make it possible to examine the internal workings of a device at a level that is inaccessible experimentally. The simulation process has two components; the first is a solution of Poisson's equation to find the electric field profile for a given distribution of charge carriers, and the second is a solution of the transport problem to find how the carriers move under the influence of the field. As devices shrink from micrometer to nanometer dimensions (Fig. 1.1) the first problem is relatively unaffected; however, new approaches to the second (transport) problem are becoming necessary.

For 40 years now, the familiar drift-diffusion equation,

$$ J_n = n q \mu_n \mathbf{E} + q D_n \frac{dn}{dx} $$

has provided a reliable description of carrier transport in devices. Two key assumptions underlie this equation:

(i) Electrons are viewed as particles that obey Newton's law in an external electric field and are scattered occasionally by phonons, impurities, and by other carriers.

(ii) The electric field varies slowly compared to the mean free path so that an electron is scattered many times before the field changes appreciably.

The second assumption can be violated in advanced sub-micron transistors leading to so-called non-stationary effects such as velocity overshoot [3]. Consequently, standard simulation tools are beginning to lose their validity and new approaches are becoming necessary. When devices shrink even further and become comparable in size to the wavelength of electrons, assumption (i) will also lose validity. Recent experiments have shown that electron transport on a sub-100 nanometer scale is influenced by wave interference effects not unlike those well-known in microwave or optical networks. As we enter this regime of device dimensions, radically new concepts in electronic devices might emerge. To analyze and design such devices it will be necessary to treat the wave nature of electrons explicitly.

This chapter is intended to present a simple, physical description of the carrier transport effects now occurring in advanced Si transistors. It is directed at the users of simulation programs and has two related objectives. The first is to establish the basis for the drift-diffusion equation and to explain why it loses validity when applied to small devices. The second is to introduce the Monte Carlo approach which is finding increasing use for simulating advanced Si transistors. This chapter should help device engineers to appreciate the limitations of present-
Fig. 1.1. Plot of the minimum linewidth versus time for silicon, integrated circuit technology. (Courtesy of James Plummer)

Fig. 1.2. Typical time between collisions for electrons in pure Si plotted versus the electron's kinetic energy. (Room temperature is assumed.)
day simulation programs and to understand the alternative approaches that are available. Quantum effects, which arise on nanometer scale devices and become especially important at low temperatures, are the subject of Chapter 3.

1.2 Drift-Diffusion Equation

Because it is the basis for today’s device simulators, we begin by examining the drift-diffusion equation. In this section, we’ll explain why the drift-diffusion equation is losing validity as device dimensions shrink, and in the following section, we’ll show how devices can be simulated without using drift-diffusion equations. Our discussion will focus on drift currents, but very similar considerations also apply to diffusion currents.

The drift current is the product of the electric field and the mobility which is defined as

$$ \mu_n = \frac{q<\tau>}{m^*}, $$

where q is the electronic charge, <\tau> a specially-defined "average" time between collisions, and m* is the electron’s effective mass. Electrons in silicon frequently collide with impurities and with lattice vibrations, and the typical time between collisions is a strong function of the electron’s kinetic energy as illustrated in Fig. 1.2. In equilibrium, an average electron has a kinetic energy of \((3/2)k_B T = 0.040 \text{ eV}\), so according to Fig. 1.2 such an electron would typically scatter every 0.25 psec. When an electric field is applied, however, the electrons gain kinetic energy. For \(E = 20 \text{ kV/cm}\), the average kinetic energy is \(= 0.1 \text{ eV}\), so Fig. 1.2 shows that an average electron will scatter much more frequently — typically every 0.1 psec.

Equation (2) suggests that the mobility will decrease with electric field because as the field increases, electrons gain energy which reduces the average time between collisions, <\tau>. Because the increase in kinetic energy is frequently the result of an applied electric field, it is usually more convenient to think in terms of a field-dependent mobility, \(\mu_n(E)\), rather than to deal with average kinetic energies and times between collisions. The field-dependent mobility is readily measured and is tabulated in most semiconductor device textbooks, so one can make use of the measured results without needing to consider the underlying scattering physics.

Today’s device simulators model drift transport by using a position-dependent mobility determined by the nonuniform electric field within the device. In modern devices the electric field is large, and it varies rapidly with position. But in the presence of rapid variations in the electric field, the concept of a field-dependent mobility loses validity; to understand why we need to examine the microscopic physics again.

Figure 1.3 illustrates what can happen in an advanced device — in this case a heterojunction bipolar transistor. The abrupt change in the semiconductor’s bandgap at the emitter-base junction produces a "launching ramp" which injects electron’s with high kinetic energy into the
base. The kinetic energy of injected electrons may be 10 times the equilibrium energy — or even more. Clearly the concept of a field-dependent mobility is without validity within the transistor’s base. The electric field within the base is zero, but the carrier's have high kinetic energy, so their mobility must be low. The mobility is closely related to the local average kinetic energy of electrons (which determines the average time between collisions) but not to the local electric field.

More generally, these considerations apply whenever the electric field changes rapidly. Figure 1.4 illustrates what can happen. In the low-field region, well to the left of the transition, the average velocity is simply \( \mu_n(\mathcal{E}_{\text{low}}) \). Just after the low- to high-field transition, however, the mobility remains high because the electrons haven't yet gained much energy from the high field. The result is a very high initial velocity which causes the drift velocity to overshoot its value in bulk Si. Several mean-free-paths after the transition, however, the electrons have achieved a new balance with the high field and their kinetic energy is that of a bulk semiconductor with the field, \( \mathcal{E}_{\text{high}} \). In this region, the electrons simply drift with a velocity of \( \mu_n(\mathcal{E}_{\text{high}}) \). Immediately after the high- to low-field transition, however, the carrier's kinetic energy is still high, so the mobility is lower than that of the corresponding bulk semiconductor, and the velocity undershoots its value in the bulk. A few mean-free-paths after the transition, however, the electrons have dissipated their excess energy, so they move at the velocity, \( \mu_n(\mathcal{E}_{\text{low}}) \) again.

Also plotted in Fig. 1.4 is the velocity versus position profile that results from a drift-diffusion equation that employs a mobility determined by the local electric field. The comparison demonstrates that the drift-diffusion equation fails to describe the over- and undershoot effects which occur when the field changes abruptly. Such non-local, or non-stationary, effects arise whenever the electric field varies rapidly within a mean-free-path. To simulate small devices accurately, these non-local effects must be treated. Conversely, by designing devices to exploit such effects, it might also be possible to enhance the performance of small devices.

One way to simulate non-stationary transport in small devices properly is to use the Monte Carlo approach as described in the following section. But there are simpler approaches that may work if the device is not too small. For example, it is clear that the mobility depends more on the local carrier energy rather than on the local electric field. By viewing the mobility as energy-dependent and adding an energy balance equation to solve for the average carrier energy versus position within the device, many of these non-stationary transport effects can be described. Alternatively, it may be possible to extend the drift-diffusion equation. For example, some researchers have had success with a current equation of the form [4],

\[
J_n = n q \mu_n(\mathcal{E}) + q D_n(\mathcal{E}) \frac{dn}{dx} + W(\mathcal{E}) \frac{d\mathcal{E}}{dx}.
\]  

(1.3)

The first two terms of this current equation are simply the conventional drift-diffusion equation with field-dependent mobility and diffusion coefficient. The last term accounts for the non-stationary transport effects that arise when the electric field varies rapidly. Since \( W(\mathcal{E}) \) is positive for high electric fields, this equation can account for the velocity overshoot observed in Fig.
Fig. 1.3. Energy band diagram of an npn heterojunction bipolar transistor showing an electron "launching ramp" at the emitter-base heterojunction.

Fig. 1.4. Sketch of the average carrier velocity versus position for a model field profile. The dotted line is the average velocity versus position that would be deduced by using a conventional drift-diffusion equation with a field-dependent mobility.
1.4. Either approach, adding an energy balance equation or modifying the drift-diffusion equation, may make it possible to extend the usefulness of present-day device simulators, but as devices continue to shrink, even these approaches will become questionable.

1.3 Monte Carlo Approach

As concepts such as mobility and diffusion coefficient lose validity in shrinking devices, more rigorous approaches are becoming necessary. The most direct approach is to simulate the microscopic motion of several thousand carriers as they travel through a device. By averaging the results, one can obtain the average carrier density, velocity, energy, and other quantities of interest as a function of position within the device. The method is known as the Monte Carlo technique, and although it has been used for more than 20 years, it is just beginning to find wide applications in device engineering.

The Monte Carlo technique is based on the simple ideas illustrated in Fig. 1.5. Between collisions, carriers move as classical particles which obey Newton's laws. Collision times are specified by a random number, $r_1$, and are selected to be consistent with the physical scattering times displayed in Fig. 1.2. At the end of this free-flight, the carrier's position, velocity, and kinetic energy are updated. To identify the flight-terminating collision, a second random number, $r_2$ is then selected. (Again, the random number is selected so that the distribution of computer-generated scattering events approximates the physical distribution.) Because collisions may change the carrier's kinetic energy and its direction of travel, two more random numbers, $r_3$ and $r_4$ are selected to specify the polar and azimuthal angles of the carrier velocity after scattering. A new free-flight then begins and the process continues.

To apply the method to devices, we first sub-divide the device into small boxes, populate each box with a sample of electrons, and initialize their velocities and the electric field. The carriers are then allowed to move for a short time with their trajectories being simulated by Monte Carlo methods. At the end of the short time step, carriers have moved from box to box and in and out of the contacts, so the electric field must be updated to reflect the new carrier profile. As the process proceeds, we obtain a time-dependent simulation for the device. At any time, the average carrier density (or velocity, or energy) versus position is readily obtained by computing averages within each box.

The average velocity versus position profile sketched in Fig. 1.4 was computed for electrons in silicon using S-DEMON, a Monte Carlo device simulation program based on methods similar to those outlined above (see Chapter 2 for a description of S-DEMON). The results, repeated in Fig. 1.6, give some indication of the magnitude of non-stationary transport effects in silicon devices and suggest that they become important when the active region is shorter than 0.25 μm. Also displayed in Fig. 1.6 is the average kinetic energy versus position. Note that immediately after the low- to high-field transition, the average kinetic energy is less than it is in bulk Si with the same applied field. As discussed in Sec. 1.2, the mobility is higher than
Fig. 1.5. Illustration of the basic steps in Monte Carlo simulation.

Fig. 1.6. Monte Carlo simulated average, steady-state velocity and kinetic energy of electrons in silicon. The field profile is the same as that assumed in Fig. 4. (Although the field is not self-consistent with the resulting electron profile, it does illustrate the important effects that occur.) The letters on the plot refer to the corresponding distribution functions as plotted in Fig. 7.
\( \mu_n(\vec{E}_{\text{high}}) \) in this region which results in velocity overshoot.

For small devices, Monte Carlo simulation provides more accurate carrier density and velocity profiles than does the drift-diffusion approach. It also provides much more information than a drift-diffusion simulation can. For example, the carriers don't all move at the average velocity but are, rather, distributed in velocity. The velocity distribution function, \( f(\vec{r}, \vec{v}) \), which is the probability of finding an electron with velocity, \( \vec{v} \), at position, \( \vec{r} \), is the fundamental quantity describing classical transport. Knowing the distribution function, we can compute the average carrier density (from its zeroth moment), average carrier velocity (from its first moment), and the average carrier kinetic energy (from its second moment). Figure 1.7 displays the computed distribution functions at selected locations within the model structure simulated in Figs. 1.4 and 1.6. In equilibrium the distribution function has a Maxwellian (i.e. Gaussian) shape. Figure 1.7 shows that in the low-field regions well away from the field transition (locations "a" and "f") the distribution function is Maxwellian but that it is displaced slightly to the right of zero which results in the net, positive velocity. Near the middle of the high-field region (location "d") the distribution is also Maxwellian-shaped but with a larger spread, which indicates that the electron kinetic energy (or electron temperature) is higher. Many device simulation codes are based on the assumption that the distribution function always retains a Maxwellian shape, but Fig. 1.7 shows that near the field transitions (locations "b," "c," and "e") the distribution is distinctly non-Maxwellian. (The two-peak characteristics result from bandstructure effects in Si; some electrons respond to the field with the light, transverse effective mass while others respond with the heavier, longitudinal effective mass.) Figure 1.7 is an example of the detail that a Monte Carlo simulation provides.

The Monte Carlo method is appealing because it directly mimics carrier motion at a microscopic level, so it provides highly accurate, detailed, realistic simulations of carrier transport in devices. It imposes a heavy computational burden, however, and the noise associated with the use of a relatively small sample of a few thousand electrons sometimes limits its applications. Nevertheless, the need for Monte Carlo simulation continues to increase as device dimensions shrink. Drift-diffusion and extended drift-diffusion approaches will continue to be widely-used, but Monte Carlo simulation will often be necessary for advanced device research and to act as a standard against which the accuracy of simpler approaches can be gauged.

1.4 Conclusions

Although physical device simulation is now an essential tool in advanced device development, the underlying assumptions upon which it is based are beginning to lose validity as device dimensions continue to shrink. This chapter reviewed the physical basis for the drift-diffusion equation, explained its limitations, and described the Monte Carlo approach which provides a more realistic simulation of transport physics. The description of carrier transport used in simulation programs must continue to improve as device technology matures. Future simulation
Fig. 1.7. Distribution functions at selected locations within the model structure. The letters refer to the various locations as specified in Fig. 1.6.
programs will make use of extended drift-diffusion equations, and device designers will make increasing use of advanced techniques such as Monte Carlo simulation. When devices shrink to nanometer dimensions, however, the electron’s wave nature will become important and the Monte Carlo approach, which treats electrons as classical particles, will also lose validity. The new concepts needed to model quantum devices are the subject of the following chapter.

Chapter 1 References


Chapter 2

MONTE CARLO ANALYSIS FOR Si TRANSISTORS

2.1 Introduction

2.2 SIDEMON Program Description

2.3 Results

2.4 Summary

A one dimensional silicon Monte Carlo program S-DEMON (Silicon DEvice MONte Carlo) has been developed using the GaAs version (DEMON, created by previous SRC research) as a framework [1].

This chapter briefly introduces the theoretical basis of the Monte Carlo method (section 2.1), describes the reformulation of the GaAs version made necessary by the more complicated covalent band structure of silicon (section 2.2), and demonstrates the capabilities of the program by determining transport characteristics (velocity-field curves, distribution functions) of bulk Si and model device structures and comparing them with experimental results where applicable.
2.1 Introduction

The Monte Carlo approach to problem-solving is, as the name suggests, one in which random numbers are used to arrive at the solution [2]. By carefully selecting the distribution of random numbers to coincide with the distribution arising from actual physical processes such as the duration of the free time of flight and the scattering mechanism selection for an electron in a semiconductor, the Monte Carlo method can be used to realistically simulate the microscopic path of an electron through a device. Average values of transport quantities such as velocity and kinetic energy can be estimated from the compiled statistics of many individual electron paths. Because the Monte Carlo method as applied to electron transport involves physics on a fundamental, microscopic level, namely that of the single electron, it reveals many physical phenomena that can not be resolved by the more macroscopic drift-diffusion equations.

There are many excellent reviews describing the use of Monte Carlo methods in solving the electronic transport problem [3,4,5]. The Monte Carlo method is considered to be semi-classical because it uses Newtonian physics to describe the motion of the electron due to external forces such as an applied electric field while employing elements of quantum mechanics to determine the probability and outcome of scattering events. The method used in simulation programs at Purdue is called the "incident flux approach." In this approach, a single electron is injected from one contact and its path computed through the device. The statistical average of several thousand of these computed paths, each arising from individually injected electrons, is used to arrive at a steady state solution.

The simulation begins by randomly choosing the initial velocity for each injected electron at the contacts. The velocity is chosen from a positive weighted Maxwellian distribution so that negative velocities (corresponding to the electron leaving the device) cannot be selected. Once this is chosen, the basic simulation sequence of the Monte Carlo method begins and will be repeated until the electron has left the device. First, the time of free flight for the electron is determined by a random number. This random number has a distribution which reflects the probability that a scattering event (an event which terminates the free flight) has not taken place for a certain time interval. The time interval selected by the random number becomes the free time of flight. As the electron travels during free flight its energy, momentum, and position are periodically recomputed according to the influence of the electric field. These new values are then stored for specific regions in the device (called bins). Later, the average value of quantities such as velocity and energy will be computed from the statistics of all the electrons that have passed through that region. At the end of the free flight, a scattering mechanism is chosen by another random number. The probability that a certain scattering mechanism will be chosen is proportional to the scattering rate of that mechanism computed from the electron's energy and position at the end of the flight. The electron's new energy and momentum are then calculated from expressions for the selected mechanism. Depending upon the mechanism chosen, several more random numbers may need to be generated in order to determine the new values for these quantities. After scattering, a new free time of flight for the electron is decided and the Monte
Carlo sequence repeats. The sequence ends when the electron is determined to have crossed the boundary of the device.

2.2 S-DEMON Program Description

This section describes the application of the Monte Carlo method to the solution of electron transport in silicon as implemented in the Purdue simulation program S-DEMON (Silicon DEvice MONte Carlo). Since much of the program structure from DEMON, a GaAs version of the Monte Carlo program created by previous SRC research [1], has been maintained in S-DEMON, the details of the silicon formulation will be given only where significant differences occur.

2.2.1 Features

S-DEMON is a one dimensional silicon device simulator using the Monte Carlo method for the solution of electron transport. The band structure model uses ellipsoidal constant energy surfaces and takes into account nonparabolic conduction bands. Presently it is designed to simulate devices with an electric field in the <100> direction. The capabilities of the program will soon be extended to simulate fields in the <111> direction and allow for self-consistent solutions with the addition of a Poisson solving routine.

In defining the device to be simulated, the user can assign arbitrary doping and field profiles directly in the program or use the results of traditional drift-diffusion simulations as an input (S-DEMON has a built-in interface for the output of FISH1D, a Poisson solver at Purdue, and can readily be adapted for other programs). The number of bins (the positions in the device where data will be collected in order to determine the average transport quantities such as velocity, electron concentration, and kinetic energy) can be specified as well as the positions where velocity histograms will be generated. While the quantities mentioned above are generated in plot format, other quantities such as longitudinal kinetic energy, equivalent valley occupation ratios, scattering rates, and the percentage scattering for each relevant scattering mechanism are listed in the tabular output. The injection velocity distribution of electrons from the contact can be specified Maxwellian, a delta function, or assigned arbitrarily. Periodic boundary conditions can be invoked for the simulation of bulk material and periodic structures.

2.2.2 Band Structure Model Differences

The band structure model used in S-DEMON differs substantially from that of DEMON because of the significant disparity between the band structure of silicon and gallium arsenide. Within the first Brillouin zone, the conduction band minimum of Si is located in the X (<100>) direction whereas the GaAs band minimum lies in the Γ (zone center) direction. Since the multiplicity of conduction band minima per direction is determined by the number of symmetry points or facets of the Wigner-Seitz cell for the crystal in that direction, Si is found to have six equivalent minima (or valleys) corresponding to the direction containing the bandgap, the most
significant for transport, while GaAs has only one. Multiple minima in the bandgap direction make equivalent valley transfer a significant outcome of phonon scattering in Si. In GaAs, this phenomenon applies only to the upper, non-bandgap, valleys, where it becomes important only after significant intervalley scattering has occurred. Other scattering mechanisms included in DEMON that do not apply to covalent semiconductors or are negligible in Si are omitted in S-DEMON. The scattering mechanisms presently included in S-DEMON are discussed in section 2.2.4.

![Si constant energy surfaces](image)

**Fig. 2.1:** Si constant energy surfaces

When viewed as constant energy surfaces in k-space, the valleys in Si form ellipsoids, two positioned along each axis as seen in Figure 2.1 (because valleys oriented on the same axis are virtually identical, S-DEMON uses just three ellipsoids, one on each axis, to represent all six valleys). As the result of the asymmetry of this arrangement in relation to a field applied in the <100> direction, electrons will react differently depending upon the orientation of the valley in which they reside. The reaction of electrons in the valleys parallel to the field (on the k_x axis) is accounted for in the effective mass assigned to those electrons, the longitudinal effective mass (m_l). The reaction of electrons in the valleys perpendicular to the field (on the k_x and k_y axes) is represented in the effective mass of those electrons, the transverse effective mass (m_t). For fields other than those parallel to the axes the reaction of the electrons can be accounted for with some combination of the two. Since the constant energy surface of the single GaAs valley forms a sphere centered at the origin, electrons in it will respond independent of the field
orientation; thus, GaAs has just one effective mass. Because the response of electrons in the Si valleys does depend on field orientation, a different method than the one used in the GaAs version for updating the electron’s momentum (after scattering events and field acceleration) is required.

To summarize, the added complexity of the Si band structure results in two major differences between the model used in the Si Monte Carlo program and the GaAs version: (1) equivalent valley transfer is a significant outcome of phonon scattering in Si, a phenomenon only important in GaAs at high fields, while several other scattering mechanisms in GaAs do not apply or are negligible, and (2) in Si, the electron’s response depends on the valley it occupies and the orientation of the field which substantially complicates the computation of electron dynamics, whereas electron response in the single valley of GaAs is independent of field orientation. The second difference presents the most serious difficulties to simulation and is the subject of the next section.

2.2.3 Electron Dynamics and the Herring and Vogt Transformation

The \( e(k) \) relationship for a general vector \( k \) in rectangular coordinates and assuming parabolic bands is:

\[
e(k) = \frac{\hbar^2}{2} \left( \frac{k_x^2}{m_x^*} + \frac{k_y^2}{m_y^*} + \frac{k_z^2}{m_z^*} \right),
\]

with

\[
k = k_x \hat{x} + k_y \hat{y} + k_z \hat{z}.
\]

where \( m_x^* \), \( m_y^* \), and \( m_z^* \) are the effective masses corresponding to different directions in \( k \)-space. In GaAs, as a result of the spherical symmetry of the constant energy surface, \( m_x^* = m_y^* = m_z^* \) so the energy expression can be written as:

\[
e(k) = \frac{\hbar^2 k^2}{2 m^*}.
\]

with

\[
k = \sqrt{k_x^2 + k_y^2 + k_z^2}.
\]

This form is very convenient for calculations, especially when computing scattering rates which involve integration over all allowable energy transitions.

Constant energy surfaces in Si, however, have ellipsoidal, not spherical symmetry. The \( e-k \) relationship in this case can at most reduce to:
\[ \epsilon(k) = \frac{\hbar^2}{2} \left( \frac{k_l^2}{m_l^*} + \frac{k_t^2}{m_t^*} \right), \]  

where

\[ k_l = k_x \quad \text{and} \quad k_t = \sqrt{k_y^2 + k_z^2}, \]

for an ellipsoid with major axis on the \( \hat{x} \) axis,

\[ k_l = k_y \quad \text{and} \quad k_t = \sqrt{k_x^2 + k_z^2}, \]

for an ellipsoid with major axis on the \( \hat{y} \) axis, and

\[ k_l = k_z \quad \text{and} \quad k_t = \sqrt{k_x^2 + k_y^2}, \]

for an ellipsoid with major axis on the \( \hat{z} \) axis. The subscripts \( l \) and \( t \) refer to longitudinal and transverse directions. Integrating functions of energy (such as scattering rates) over all allowable energy transitions is much more difficult for an \( \epsilon-k \) relationship with the form of equation 2.3 rather than 2.2. In addition, since the electron's response to an electric field will depend upon the valley it occupies, the computation performed to evaluate its velocity will also change with the valley. When non-parabolicity is included in this relationship,

\[ \gamma(k) = \epsilon(k)(1 + \alpha \epsilon(k)) = \frac{\hbar^2 k^2}{2 m^*} \]  

where \( \alpha \) is the band non-parabolicity factor, the above difficulties are compounded.

To eradicate these complications, the Herring-Vogt transformation is introduced which transforms the ellipsoidal constant energy surfaces into spheres [3]. Since electrons in spheres react with essentially one effective mass \((m_0)\), the calculations for computing scattering rates and updating an electron's velocity as it travels through a Si device become generally no more complex than those for an electron in GaAs. The simulation sequence remains as described before; however, before the dynamic variables of the electron (such as energy and the wave vector) are updated as a result of field acceleration or after a scattering event, the local electric field \( E \) is transformed by matrix \( T \) into the space defined by the Herring and Vogt transformation (the transformed variables are denoted with the superscript *). The calculations are then performed within Herring and Vogt space with the current wave vector (which is always kept within transformed space) and only transport quantities of interest such as velocity are transformed back into the original (actual) space so that statistics can be taken. Energy itself does not need to be transformed since it is the same in both the original and transformed space (the transformation matrix \( T \) is defined such that this occurs). Since the updated position is derived from the energy, it is always kept in terms of the original space and thus does not need to be transformed.
The transformation matrix $T$ along with a table detailing the relationship between quantities in the original space and Herring and Vogt space are listed below and in Table 2.1.

$$
T = \begin{bmatrix}
\left(\frac{m_0}{m_1}\right)^{1/2} & 0 & 0 \\
0 & \left(\frac{m_0}{m_1}\right)^{1/2} & 0 \\
0 & 0 & \left(\frac{m_0}{m_1}\right)^{1/2}
\end{bmatrix}
$$

<table>
<thead>
<tr>
<th>Quantity in Original Space</th>
<th>Transformation Relationship</th>
<th>Quantity in Transformed Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>$k^* = T k$</td>
<td>$k^*$</td>
</tr>
<tr>
<td>$v$</td>
<td>$v = T v^*$</td>
<td>$v^* = \frac{\hbar k^*}{m_0}$</td>
</tr>
<tr>
<td>$E$</td>
<td>$E^* = T E$</td>
<td>$E^*$</td>
</tr>
<tr>
<td>$\epsilon(k) = \frac{\hbar^2}{2} \left( \frac{k_1^2}{m_1^<em>} + \frac{k_2^2}{m_2^</em>} \right)$</td>
<td>$\epsilon(k) = \frac{\hbar^2 k^*}{2 m_0}$</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1 Herring and Vogt transformation relationships. The stars over the effective masses do not denote transformation.

### 2.2.4 Scattering Mechanisms

Because of the dissimilar band structure, the applicable scattering mechanisms in Si vary in type or significance from those in GaAs. Scattering mechanisms included in S-DEMON are intravalley scattering caused by ionized impurities and nonpolar acoustic phonons and intervalley scattering by optical and acoustic phonons. The terms intervalley and intravalley refer to scattering between and within equivalent valleys. Unlike in the GaAs program, non-equivalent intervalley scattering (from the X minima to the $\Gamma$ minimum or the L minima) is not included in S-DEMON because those valleys are far removed in energy and have very small density-of-states effective masses [4]. Also not included in S-DEMON but present in DEMON is polar optical scattering, a mechanism not applicable to covalent semiconductors.

Ionized impurity and nonpolar acoustic phonon intravalley scattering are handled identically in S-DEMON and DEMON (because of the Herring and Vogt transformation), with the appropriate change of scattering constants. Ionized impurity scattering results in intravalley
scattering only because the coulomb cross-section decreases rapidly with increasing cross-
section thus making intervalley scattering very improbable (where substantial momentum
transfer has to occur) [4]. In Si, optical phonon scattering always results in intervalley scattering
because the matrix element of the scattering potential (which appears in the scattering rate
integral) vanishes for reasons of symmetry for transitions within the valley [6]. Intervalley opti-
cal phonon scattering can be divided into two types, \( f \) and \( g \). The \( g \) type scattering occurs
between valleys that share the same axis in k-space while \( f \) type scattering refers to scattering
between valleys with different axes. Scattering constants in S-DEMON where taken from refer-
ence [4].

2.3 Results

This section shows the results of the tests used to verify the accuracy of S-DEMON and
demonstrates its capabilities for device application. The program's accuracy is gauged by com-
paring simulation results for bulk Si with experimental measurements and with other simula-
tions. Device application is demonstrated by simulating electron transport across a model
base-collector field profile. Phenomena resulting from the multivalley band structure of Si are
also investigated.

The first graph (Figure 2.2) shows excellent agreement between the velocity-field charac-
teristic obtained from S-DEMON and the experimental results of Canali et al. [7]. The second
(Figure 2.3) shows the total kinetic energy-field characteristic of electrons in the valleys perpen-
dicular to the direction of the applied field (<100>). The results of S-DEMON are compared to
those from the Monte Carlo program of Zimmerman et al. [8]. Figure 2.4 makes the same
comparison with the energy of electrons in the valleys parallel to the applied field. Since elec-
trons in the perpendicular valleys respond to the field with a lighter, transverse effective mass
than those in the parallel valleys (which have the longitudinal effective mass), their kinetic
energy will be slightly greater. This effect is more pronounced at higher fields and can be seen
if Figures 2.3 and 2.4 are superimposed.

The field profile and doping used by Baccarani et al. [9] to model the base collector junc-
tion of a simple bipolar transistor is portrayed in Figure 2.5. S-DEMON results for the velocity,
normalized electron concentration, and average kinetic energy vs. position are displayed in Fig-
ures 2.6, 2.7, and 2.8 respectively. Slight discontinuities in the figures are caused by statistical
noise and will decrease with an increase in the number of electrons simulated. In Figure 2.6,
the average electron velocity just inside the collector exceeds the bulk steady-state average
velocity. This effect, known as velocity oversadr the rapid increase in the electron
kinetic energy caused by the sudden change from a low to high electric field in that region.
However, the electrons soon scatter and their velocities are redistributed such that the average
becomes the bulk steady-state velocity. The velocity overshoot effect is also present in the
work by Baccarani; however, direct comparison with results from S-DEMON reveals subtle
differences that may be due to disagreement between the scattering constants used or the
Fig. 2.2: Velocity vs. field curve for <100> Si. Experimental curve is from reference [7].

Fig. 2.3: Perpendicular (hot) valley energy vs. field (applied in the <100> direction). Solid curve is from reference [8].
Parallel (cold) valley energy vs. field (applied in the <100> direction). Solid curve is from reference [8].

Electric field vs. position for the bipolar simulation example [9].
Fig. 2.6: Velocity vs. position curve for the bipolar example.

Fig. 2.7: Normalized electron concentration vs. position for the bipolar example.
Fig. 2.8: Average kinetic energy vs. position for the bipolar example.

Fig. 2.9: Velocity vs. position curve for velocity histogram study. Letters denote positions where velocity histograms were taken.
simplified band structure model employed by Baccarani, which assumes parabolic bands and spherical constant energy surfaces.

Phenomena that are not resolvable by the simple model of Baccarani are those arising from a multivalley band structure with nonspherical constant energy surfaces. To study these effects, an undoped structure with constant low field (−5×10^3 V/cm) and high field (−10^6 V/cm) regions was simulated. Velocity histograms were then generated at several points along the high field region for electrons in valleys parallel to the field, in valleys perpendicular to the field, and in all the valleys combined. Figure 2.9 shows the velocity vs. position curve for the structure and roughly indicates the positions where velocity histograms were taken.

At position A (Figure 2.10, left column), the electrons have just reached the low field-high field interface and have a slightly positive velocity shifted Maxwellian distribution for all three cases. At point B (Figure 2.10, right column), the electrons are in the region where velocity overshoot occurs and now begin to exhibit multivalley behavior. Because electrons in the valleys parallel with the field respond with the heavier, longitudinal mass, these electrons react sluggishly to the new field when compared to electrons in valleys perpendicular to the field (which react with the transverse effective mass); as a result, the electrons in the perpendicular valleys have a higher average velocity. In terms of the velocity histograms, this explains the displaced peak of the perpendicular valley histogram and the two peaked (composite) histogram of all the valleys. Also noticeable is the spreading of the distribution that occurs as a result of increased scattering (scattering rate being roughly proportional to energy). This effect is well pronounced in the perpendicular valleys where significant backscattering has occurred as the result of inter-valley scattering between valleys perpendicular to the field and isotropic acoustic phonon scattering events within the valleys (which equally distribute the electrons that scatter between positive and negative momentum).

At C (Figure 2.11, left column), just .03 microns to the right of B, significant scattering (and backscattering) has occurred, smearing out the distributions of both valleys. In the perpendicular valleys the accelerated electrons which have yet to suffer a momentum changing scattering event continue to define the right-most (positive) peak in the histogram. However, by position D (Figure 2.11, right column), sufficient scattering has occurred such that the distribution of all the valleys is returning once more to a Maxwellian centered upon the average bulk steady-state velocity dictated by the field in that region. The distribution has widened from that at A according to the increase in electron temperature caused by the additional energy imparted by the high electric field (which increases the scattering rate and further spreads the electron velocities). The positive hump in the right side of the composite distribution suggests that the electrons in the perpendicular valleys (which have the lighter effective masses) are still traveling with a greater velocity than those of the parallel valleys.
Fig. 2.10  Velocity histograms for position A at .25 microns (left column) and B at .27 microns (right column). The abscissas are velocity with units of $10^8$ cm/sec. The ordinates are the normalized distribution function (normalized for each graph).
Fig. 2.11  Velocity histograms for position C at .30 microns (left column) and D at .42 microns (right column). The abscissas are velocity with units of $10^8$ cm/sec. The ordinates are the normalized distribution function (normalized for each graph).
2.4 Summary

A one dimensional silicon Monte Carlo program has been developed and is now being used at Purdue to study transport in Si bipolar structures. The band structure model utilizes a nonparabolic $E-k$ relationship and ellipsoidal shaped constant energy surfaces. The correct operation of S-DEMON has been verified by comparing its velocity-field characteristic with experimental results and by reproducing the energy-field relationships of other Monte Carlo simulation programs. Transport characteristics were determined for the base-collector junction of a simple bipolar model in order to demonstrate the capabilities of S-DEMON for device simulation. Multivalley effects on the velocity distributions where examined for a model low field-high field region structure.
Chapter 2 References


Chapter 3

QUANTUM TRANSPORT IN ULTRASMALL ELECTRONIC DEVICES

3.1. Introduction

3.2. Background

3.2.3. Current-voltage formula
3.2.2. Linear response
3.2.3. Transmission coefficients

3.3. Quantum Effects in Electron Transport

3.3.3. Plane wave transport
3.3.2. Waveguide transport
3.3.3. Space-charge effects

3.4. Current Theoretical Status

3.5. Open Questions

References

In this chapter we review quantum interference effects that have been observed in ultrasmall structures and their implications for future electronic devices. We also review the current theoretical understanding of such phenomena and discuss some of the unresolved questions that have to be answered in order to develop accurate models for quantum device simulation.
3.1 Introduction

Semiconductor device analysis has traditionally been based on the drift-diffusion equation:

\[ J = e \mu(\vec{E}) n \vec{E} + e D(\vec{E}) \nabla n \]  \hspace{1cm} (3.1)

Here \( J \) is the current density, \( n \) is the electron density, \( \vec{E} \) is the electric field and \( \mu \) and \( D \) are the basic transport parameters called the mobility and the diffusion coefficient respectively. For simplicity, we restrict ourselves to one type of carrier, namely electrons. In deriving eq (3.1) one makes two main assumptions:

1. electrons are particles that move in an external field according to Newton’s law and are scattered occasionally by phonons and impurities, and
2. the electric field changes slowly compared to the mean free path so that an electron is scattered many times before the field changes appreciably.

In many present-day submicron devices, assumption two is violated. An electron may transit through the device with few or no collisions in a manner reminiscent of vacuum tubes. Consequently, the velocity distribution of electrons (and hence transport parameters like \( \mu \) and \( D \)) at any point within the device is not determined uniquely by the local electric field \( \vec{E} \) but is dependent on the boundary conditions as in vacuum tubes. To account for such non-stationary or hot-electron effects as they are often called, new approaches to device simulation are being developed based on the Boltzmann equation, which under steady state conditions can be written as

\[ v(k) \cdot \nabla f + \frac{e}{m} \vec{E} \cdot \nabla_k f = S_{op} f \]  \hspace{1cm} (3.2)

where \( f(r,k) \) is the semiclassical distribution function that tells us the number of electrons at \( r \) having the wave-vector \( k \); \( v \) is the velocity of an electron with wavevector \( k \); \( S_{op} \) is the scattering operator which is usually evaluated by applying Fermi’s golden rule to the individual scatters. In deriving eq. (3.2) one needs only the first of the two assumptions listed after eq. (3.1). Consequently, hot-electron effects are accounted for.

As devices shrink to dimensions comparable to the wavelength of electrons, it is expected that the wave nature of electrons will play an increasingly important role and even the first assumption will not be valid anymore. On such small length scales the semiclassical distribution function is not a valid concept anymore, due to the uncertainty relation between \( r \) and \( k \). To analyze and design devices on a sub-100 nm scale it will be necessary to go beyond the Boltzmann equation (eq. (3.2)) and develop simulation techniques based on kinetic equations (Fig. 3.1). The development of an appropriate kinetic equation is an active topic of current theoretical research, that has recently gained impetus from the surge of experimental activity in the area of quantum transport.

The development of molecular beam epitaxy since the late sixties has made it possible to grow ultrathin (~20 Å) layers of different materials with atomically sharp interfaces. This has led to the development of vertical quantum devices where the current flows perpendicular to
Fig. 3.1. Hierarchy of transport theories

Fig. 3.2. (a) Vertical and (b) lateral quantum devices.
the layers (Fig. 3.2). Some of these have now reached a high level of maturity and have emerged as potentially useful practical devices, such as resonant tunneling diodes and transistors. By contrast lateral quantum devices, with current flowing parallel to the layers, are still in their infancy. They have only recently been made possible by the advances in nanolithographic techniques. It will possibly be many years before useful devices based on such effects become practicable. However, since 1985, there has been a flood of experiments revealing novel quantum effects at low temperatures, causing great excitement among both basic and applied physicists. On the one hand, it opens up new ways to study one of the fundamental questions of physics, namely, the role of dissipation in microscopic phenomena and the microscopic origin of irreversibility. On the other hand, it raises the possibility of radically new electronic devices that operate by controlling the phase of the wavefunction rather than by controlling the carrier density as present-day devices do. The last few years have seen the emergence of a new research area that has been given a variety of names such as "mesoscopic physics," "nanostructure physics" and "nanoelectronics."

The outline of this chapter is as follows. In Section 2, we will briefly discuss the basic conceptual framework that one uses to describe electron transport in ultrasmall devices. We will then review various quantum effects that have been observed and discuss possible device implications in Section 3. We will also discuss the role of space-charge effects in ultrasmall devices and the possibility of taking advantage of such effects to develop novel devices. The recent theoretical work on quantum interference phenomena is briefly reviewed in Section 4. We conclude in Section 5 by discussing some of the unresolved questions associated with the inclusion of dissipative processes in a description of quantum transport.

3.2 Background

All of the phenomena that we will discuss in this article are essentially one-electron phenomena. Although it is possible that exchange and correlation will play a more significant role in the electronic properties of small structures, there is as yet no evidence for such many-body effects. The experimental observations to date are well explained, at least qualitatively, in terms of the simple one-particle picture described below.

3.2.1 Current-voltage formula

An electronic device is typically connected to two contacts across which a voltage is applied (Fig. 3.2.1a). Each of these contacts launches a steady stream of electrons onto the device, of which a fraction is transmitted to the other contact. At equilibrium with both contacts having the same electrochemical potential, the current I(1→2) transmitted from contact "1" to contact "2" is exactly balanced by the current I(2→1) transmitted from contact "2" to contact "3." An applied voltage shifts the local chemical potential μ₁ in contact "1" with respect to the local electro-chemical potential μ₂ in contact "2," making I(1→2) different from I(2→1).
Fig. 3.2.1.  
(a) A device with two contacts 
(b) The two contacts in (a) act as source and detector with the device as the intervening medium.
causing a net current flow through the device. The currents $I(1\rightarrow2)$ and $I(2\rightarrow1)$ may be evaluated as follows.

The incident flux $I_1$ from contact "1" is written as

$$I_1 = -e \sum_n \int \frac{dk}{2\pi} \frac{\hbar k}{m^*} f(E - e\mu_1)$$

where $f(E)$ is the Fermi-Dirac function, $m^*$ is the effective mass, $k$ is the wavenumber in the direction perpendicular to the sample surface and $n$ denotes the transverse modes or subbands, including spin. The energy $E$ is equal to the energy $e_n$ at the bottom ($k = 0$) of subband "n" plus the longitudinal kinetic energy $\hbar^2 k^2 / 2m^*$.

$$E = e_n + (\hbar^2 k^2 / 2m^*)$$

The subband energy $e_n$ is the sum of the potential energy and the transverse kinetic energy. For a large area contact the allowed energies $e_n$ are essentially continuous, while for a small contact they form a discrete set. Using eq. (3.2.2) we may rewrite eq. (3.2.1) as

$$I_1^+ = -\frac{e}{\hbar} \sum_{n} \int dE f(E - e\mu_1)$$

Let $\tau_{mn}^{21}(E)$ be the fraction of electrons incident from contact "1" in subband "n" with energy $E$ that are transmitted to subband "m" in contact "2." We can write

$$I(1\rightarrow2) = -\frac{e}{\hbar} \sum_{n, m} \int dE f(E - e\mu_1) \tau_{mn}^{21}(E)$$

$$= -\frac{e}{\hbar} \int dE f(E - e\mu_1) T_{21}(E)$$

where

$$T_{ij}(E) = \sum_{n, m} \tau_{ij}^{mn}(E)$$

Similarly, we can show that

$$I(2\rightarrow1) = -\frac{e}{\hbar} \int dE f(E - e\mu_2) T_{12}(E)$$

The net current flowing into the device through contact "1" and out through contact "2" is given by

$$I_1 = -I_2 = I(1\rightarrow2) - I(2\rightarrow1)$$

$$= -\frac{e}{\hbar} \int dE \left[ T_{12}(E) f(E - e\mu_2) - T_{21}(E) f(E - e\mu_1) \right]$$

The approach described above has been widely used in tunneling problems (Frenkel 1930, Duke 1969, Tsu and Esaki 1973). In these problems, however, it is usually assumed that there are no
phase-breaking processes within the device so that the transmission coefficients may be obtained from the one-electron Schrödinger equation (this is discussed further in Section 2.3). There is, however, nothing in the above derivation restricting it to phase-coherent transport. Thus, eq. (3.2.7) should be applicable more generally provided one knows how to compute the transmission coefficients in the presence of phase-breaking processes. In the extreme limit of incoherent transport, one could compute the transmission coefficients from a semiclassical Monte Carlo and use them in eq. (3.2.7) to obtain the I-V characteristics.

At equilibrium with \( \mu_1 = \mu_2 = \mu_0 \), the current \( I_1 \) in eq. (3.2.7) must go to zero. Hence, we must have,

\[
\int dE f(E-e\mu_0) \left[ T^{(0)}_{12}(E) - T^{(0)}_{21}(E) \right] = 0
\]  

(3.2.8)

The superscript "0" is added to indicate that the transmission coefficients are evaluated at equilibrium with a constant electrochemical potential \( \mu_0 \) everywhere. In the case of phase-coherent transport, it can be shown from the symmetry properties of the S-matrix that \( T_{12}(E) = T_{21}(E) \). Consequently, the validity of eq. (2.8) is obvious. But in the presence of phase-breaking processes \( T_{12}(E) \neq T_{21}(E) \) in general so that a proof of eq. (3.2.8) is more complicated.

### 3.2.2 Linear response

Eq. (3.2.7) is suitable for computing the current flowing through the circuit in response to an applied potential difference \( \mu_1 - \mu_2 \). If \( \mu_1 - \mu_2 \) is "small," then one can simplify eq (3.2.7) as follows. At equilibrium with \( \mu_1 = \mu_2 = \mu_0 \), \( I_1 \) is zero, as we just discussed. Now, if we assume that the electrochemical potentials \( \mu_1 \) and \( \mu_2 \) deviate only slightly from the equilibrium value \( \mu_0 \) then we can expand the Fermi-Dirac functions in eq (3.2.7) in a Taylor Series about \( e\mu_0 \) as follows.

\[
f(E-e\mu_{1,2}) = f(E-e\mu_0) + e \left[ -\frac{\partial f_0}{\partial E} \right] [\mu_0 - \mu_{1,2}]
\]  

(3.2.9)

Here \( f_0 \) stands for \( f(E-e\mu_0) \). Substituting eq (3.2.9) into eq. (3.2.7), assuming that the transmission coefficients \( T \) are equal to their equilibrium value \( T^{(0)} \) and using eq. (3.2.8), we obtain

\[
I_1 = \frac{e^2}{\hbar} \left[ \tilde{T}_{12} \mu_2 - \tilde{T}_{21} \mu_1 \right]
\]  

(3.2.10)

where

\[
\tilde{T}_{ij} = \int dE \left[ -\frac{\partial f_0}{\partial E} \right] T^{(0)}_{ij}(E)
\]  

(3.2.11)

Since the current in eq. (3.2.10) is zero with \( \mu_2 = \mu_1 \), we must have \( \tilde{T}_{12} = \tilde{T}_{21} \). This can also be proved using eq. (3.2.8) and eqs. (3.2.11a,b). Hence, from eq. (3.2.10),
This is one form of the Landauer formula, which suggests that the effective conductance connecting two contacts is equal to \( (e^2 / h) \tilde{T}_{12} \). However, it should be noted that this is not the conductance of the device itself since we do not know a priori what fraction of the applied potential \((\mu_2 - \mu_1)\) is actually dropped across the device. The question of how the actual conductance of the device can be obtained was raised by Landauer in his pioneering paper in 1957 and has since been addressed by numerous authors. It seems that there is no unique answer to the above question, for it depends on how the potential drop across the device is actually measured. This has led to many different versions of the Landauer formula.

Experimental measurements of the conductance are usually performed using four-probe structures rather than two-probe structures in order to minimize the effect of contacts. The current is fed in through two probes and the voltage is monitored through a pair of probes in the middle (Fig. 3.2.2). For a while it was not clear how this four-probe conductance could be computed theoretically because of the ambiguity regarding what it is that the voltage probes measure. Büttiker found a simple solution to this problem that almost seems obvious in retrospect (Büttiker 1986). He noted that since there is really no qualitative difference between that current probes and the voltage probes in a Hall bridge, one could treat all the probes on an equal footing and simply extend eq. (3.2.10) by summing over all the probes.

\[
I_i = \frac{e^2}{h} \sum_{j \neq i} (\tilde{T}_{ij} \mu_j - \tilde{T}_{ji} \mu_i) \tag{3.2.13}
\]

Since the currents must all be zero when the electro-chemical potentials are all equal, we must have

\[
\sum_{j \neq i} (\tilde{T}_{ij} - \tilde{T}_{ji}) = 0 \tag{3.2.14}
\]

Using eq. (3.2.14) we can rewrite eq. (3.2.13) as

\[
I_i = \frac{e^2}{h} \sum_{j \neq i} \tilde{T}_{ij} (\mu_j - \mu_i) \tag{3.2.15}
\]

If there are no magnetic fields \((B = 0)\) then it can be shown that \(\tilde{T}_{ij} = \tilde{T}_{ji}\). Eq. (3.2.15) is then precisely what one obtains by applying Kirchhoff's laws to a network of resistors formed by connecting each contact "i" and contact "j" through a conductance \(G_{ij}\) given by

\[
G_{ij} = G_{ji} = \frac{e^2}{h} \tilde{T}_{ij} = \frac{e^2}{h} \tilde{T}_{ji} \tag{3.2.16}
\]

Thus, in the absence of magnetic fields, one can visualize mesoscopic systems in terms of an equivalent resistor network as shown in Fig. 3.3.2b.
Fig. 3.2.2.  
(a) A four-probe Hall bridge.
(b) Equivalent resistor network in the absence of magnetic fields.
Finally, we note that one can generalize the Landauer-Büttiker formula (eq. (2.13)) to non-linear response by analogy with eq. (2.7).

\[ I_i = \frac{e}{h} \int dE \sum_{j \neq i} [T_{ij}(E) f(E - e \mu_j) - T_{ji}(E) f(E - e \mu_i)] \]  

(3.2.17)

### 3.2.3 Transmission coefficients

To use any of the equations (eq. (3.2.7) or eq. (3.2.12) or eq. (3.2.13)) or eq. (3.2.17), we need to know the transmission coefficients. The problem of current flow is thus reduced to a scattering problem not unlike those encountered in say, nuclear physics. It is as if the two contacts act as source and detector with the device as the intervening medium (Fig. 3.2.1b). The problem then is to compute the scattering characteristics of this medium.

The procedure for computing the scattering characteristics is quite straightforward if we neglect all "phase-breaking" processes within the device (the precise meaning of phase-breaking will be discussed shortly). The transmission of electrons from the source to the detector is then described by the one-electron time-independent Schrödinger equation.

\[ \left( \frac{p - e A}{2m^*} + e V \right) \Psi(r) = E \Psi(r) \]  

(3.2.18)

Here \( A(r) \) and \( V(r) \) are the vector and scalar potentials within the device. The scalar potential \( V(r) \) includes externally applied fields, *space-charge effects*, band-edge, discontinuities due to heterojunctions and microscopic fields due to elastic scatterers such as defects or impurities. In the absence of magnetic fields, the vector \( A \) potential may be set equal to zero, so that we can simplify eq. (3.2.18) as follows:

\[ \nabla^2 \Psi(r) = -\frac{2m^*}{\hbar^2} (E - V(r)) \Psi(r) \]  

(3.2.19a)

Eq. (3.2.19a) is very similar to Maxwell’s equation used in integrated optics (assuming \( \vec{E} \cdot \nabla \varepsilon = 0 \) for simplicity).

\[ \nabla^2 \vec{E} = -\omega^2 \mu(r) \varepsilon \]  

(3.2.19b)

Here \( \vec{E} \) is the electric field, \( \omega \) is the radian frequency, \( \mu \) is the permeability and \( \varepsilon \) is the spatially varying dielectric constant. Comparing eqs. (3.2.19a) to eq. (3.2.19b), it is evident that electron waves moving through a medium with a spatially varying potential \( V(r) \) is analogous to light waves moving through a medium with a varying dielectric constant (or refractive index). The analogies between electron waves and electromagnetic waves are listed in Table 3.1. Most of the phenomena we discuss in Section 3 (except those involving magnetic fields) have familiar optical analogies which we will mention as we go along.

One question that might bother the reader regarding this analogy is the fact that electrons are fermions while photons are bosons. This difference is not important so long as we are
discussing one-particle phenomena where every electron or photon *interferes with itself*. The electric field in Maxwell's equations can then be viewed as the wavefunction of a single photon and the analogy with the Schrödinger equation seems complete. But if we view the electric field (as we usually do) as the macroscopic field due to a coherent state with billions of photons, then there is no analogous state known with normal electrons. (However, the superconducting state is analogous to the coherent state of light and the Josephson effect is a well-known manifestation of the macroscopic wavefunction of superconducting electrons; we will not discuss this further.)

In view of these analogies between electrons and photons, one might wonder why the Boltzmann picture works at all. Why aren't quantum interference effects more common? One of the chief reasons is the existence of *phase-breaking* scattering processes that destroy interference phenomena. A phase-breaking scattering process is one in which the scatterer changes its internal state. As a result, successive electrons, encountering the scatterer in different states, suffer different phase-shifts thus wiping out any stationary interference patterns. Another way to view a phase-breaking interaction is as a *measurement* process. By monitoring the state of a scatterer, one can gain information regarding the path of the electron between the source and the detector. A well-known principle in quantum interference phenomena is that any process yielding information regarding which of the various interfering alternatives was actually taken, tends to destroy the interference.

---

**Table 3.1. Analogies between electron waves and electromagnetic waves**

<table>
<thead>
<tr>
<th>Electron Waves</th>
<th>↔</th>
<th>E M Waves</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Psi(r,t) = \psi(r) e^{-iEt/\hbar}$</td>
<td>↔</td>
<td>$E(r,t) = \mathcal{E}(r) e^{-i\omega t}$</td>
</tr>
<tr>
<td>Energy $E$</td>
<td>↔</td>
<td>Frequency $\omega$</td>
</tr>
<tr>
<td>Confined Channels</td>
<td>↔</td>
<td>Waveguides</td>
</tr>
<tr>
<td>Subbands</td>
<td>↔</td>
<td>Transverse Modes</td>
</tr>
<tr>
<td>$\Psi$</td>
<td>↔</td>
<td>$E$</td>
</tr>
<tr>
<td>$\nabla \Psi$</td>
<td>↔</td>
<td>$H$</td>
</tr>
<tr>
<td>Charge Density</td>
<td>↔</td>
<td>Energy Density</td>
</tr>
<tr>
<td>Current Density</td>
<td>↔</td>
<td>Poynting Vector</td>
</tr>
</tbody>
</table>

\[ \nabla^2 \psi = -\frac{2m}{\hbar^2} (E - V) \psi \]  \[ \nabla^2 \mathcal{E} = -\omega^2 \mu \varepsilon \mathcal{E} \]
Elastic scattering by the sample boundaries or by defects and impurities plays an important role in determining mobility; but it is not phase-breaking since the scatterer has no internal degree of freedom and remains unaffected by the process. But inelastic scattering by phonons or by other electrons, is phase-breaking. The phase-coherence time $\tau_0$ usually increases significantly as we go to lower temperatures because electron-electron scattering processes are suppressed. Such processes do not contribute to the mobility since the momentum of the electronic ensemble is unchanged; any momentum lost by one electron is picked up by another. Consequently, the mobility is nearly constant at temperatures below say 10$^9$K once the phonons are frozen out. But the phase-breaking time is orders of magnitude larger at 0.1$^9$K than it is at 10$^9$K. Thus, although high mobility films are certainly desirable for quantum devices, the mobility is in general not a good indicator of the phase-coherence time.

A phase-coherence length as long as 1–10 $\mu$m is not uncommon at a temperature of 1$^9$K, but it gets significantly shorter at higher temperatures and for hot electrons. Phase-breaking processes are thus inevitably present in most devices at reasonable temperatures and the assumption of coherent transport (Section 3.2.3) is often not very accurate. However, there is as yet no simple method for computing the transmission coefficients including phase-breaking processes. At the other extreme, if phase-breaking processes are so frequent that one can assume totally incoherent transport, then semiclassical Monte Carlo simulation can be used to compute transmission coefficients; this is equivalent to solving the Boltzmann equation. It is in the middle ground involving partially coherent transport that there are no simple answers, as yet.

### 3.3 Quantum Effects in Electron Transport

In this Section, we will briefly survey various quantum effects that have been observed in semiconductor microstructures. These effects can broadly be divided into two categories: those involving devices whose transverse dimensions are much larger than the phase-coherence length and those involving devices whose transverse dimensions are shorter than the phase-coherence length. In devices belonging to the first category, the subband energies $\epsilon_n$ are nearly continuous and one can describe electron transport in terms of plane waves while, in devices belonging to the second category, it is more appropriate to view transport in terms of discrete waveguide modes.

#### 3.3.1 Plane wave transport

The classic example of a quantum device is the tunnel diode, discovered by Esaki in the late 1950’s. Next to the tunnel diode, the most well-known quantum device is the resonant tunneling diode first demonstrated by Chang, Esaki and Tsu in 1974. This device consists of two barriers in series as shown in Fig. 3.3.1a. It is often compared to the Fabry-Perot interferometer used in optics. The two barriers play the role of partially transparent mirrors that
Fig. 3.3.1. Resonant tunneling diode

(a) Energy band diagram and transmission coefficient versus energy
(b) Band diagram under bias and I-V characteristics
(c) Expected I-V for one barrier (solid line) and for two barriers in series (dotted line).
form a resonant cavity. The transmission coefficient shows sharp peaks as a function of the longitudinal kinetic energy as one would expect for a resonant cavity. The device thus acts as an energy filter that only allows incident electrons with certain discrete values of the longitudinal kinetic energy to go through to the other contact. An applied bias lowers the resonant energy relative to the energy of the incident electrons. When the resonant energy falls below the conduction band edge in the emitter, there is a sharp drop in the current leading to negative differential resistance (NDR) as shown in Fig. 3.3.1b. The current-voltage curve can be computed quantitatively from eq. (3.2.7) using transmission coefficients obtained from the one-electron Schrödinger equation (eq. (3.2.18)).

It will be noted that this NDR is a quantum effect that cannot be understood within a semiclassical framework. It is easy to show that for a single tunnel barrier the current increases monotonically with voltage as shown in Fig. 3.3.1c. If we view electrons as particles, then we would expect a double barrier to act like two single barriers in series. We would thus expect the current to be half that of a single barrier for a given voltage as shown by the dotted curve in Fig. 3.3.1c. Since a single barrier is known not to exhibit NDR, it is hard to explain from a Boltzmann picture why a double barrier exhibits NDR.

The logical extension of a double barrier is a periodic array of barriers or a superlattice (Esaki and Tsu 1970). If the barriers are thin enough that electrons can tunnel from one well to the next, then the energy levels broaden to form minibands as shown in Fig. 3.3.2a. Electrons moving in a superlattice miniband are decelerated by the applied field as they reach the negative mass region at the top of the band resulting in NDR which has been observed in vertical devices (Chang, Esaki and Tsu 1974). NDR has also been reported in lateral devices with a two-dimensional periodic potential imposed through a patterned gate as shown in Fig. 3.3.2b (Bernstein and Ferry 1987; Ismail et al. 1989). Lateral devices can also be operated as transistors where the source-drain current is modulated by the gate which controls the magnitude of the periodic potential; the "washboard transistor" is based on this idea (Tsubaki et al. 1987). Negative transconductances have been observed with both one-dimensional (Tokura and Tsubaki 1989) and two-dimensional (Ismail et al. 1989) periodic potentials. Novel oscillations in the magnetoresistance were recently reported in a lateral structure with a one-dimensional periodic potential (Gerhardt, Weiss and von Klitzing 1989, Winkler, Kotthaus and Ploog 1989).

An interesting question to ask is whether one can observe any quantum interference effects from a random array of elastic scatterers such as an ordinary resistor at low temperatures. The intuitive answer is that due to the randomness, any interference effect would cancel out on the average. This, however, is not true. Quantum interference leads to enhanced backscattering. Fig. 3.3.3 shows the conductance of an array of scatterers computed from the Landauer formula (eq. (3.2.12)) using two different models to obtain the transmission coefficient $\tilde{T}_{12}$ (Cahay, McLennan and Datta 1988). One is the coherent or the quantum model in which the amplitude scattering matrices of the individual scatterers are combined taking phases into account. The
Fig. 3.3.2.  (a) Energy band diagram and miniband structure of a one-dimensional superlattice.

(b) Two-dimensional lateral superlattice.
Fig. 3.3.3. (a) A sample containing an array of static scatterers.
(b) Conductance of the sample computed as a function of the position of the middle scatterers, keeping the rest of the array fixed.
other is the incoherent or the semiclassical model in which the probability scattering matrices are combined taking no account of the phases. Fig. 3.3.3 shows the results obtained for both the quantum and the semiclassical model as the location of one scatterer in the array (the middle one) is changed. The semiclassical result is unaffected by this change, but the quantum result shows fluctuations due to the changing interference patterns. These conductance fluctuations have been observed experimentally in mesoscopic samples. Experimentally it is difficult to move a scatterer as we have done in the theoretical calculation shown in Fig. 3.3.3. Instead, the electron wavelength is changed by changing the magnetic field or the Fermi energy; this too changes the interference pattern and is believed to be equivalent to changing the configuration of scatterers. These fluctuations, discovered in 1985, have come to be known as "magnetofingerprints" and can, in principle, be used to identify mesoscopic samples (Stone 1985, Licini et al. 1985, Kaplan and Hartstein 1986, Skocpol et al. 1986, Lee, Stone and Fukuyama 1987, Taylor et al. 1988). However, in samples with dimensions that are large compared to the phase-coherence length such fluctuations cannot be observed. This is because a large sample is basically an ensemble of uncorrelated units each having dimensions of the order of a phase-coherence length. Due to this self-averaging feature, one measures ensemble-averaged quantities when making measurements on large samples. But the interesting point to note from Fig. 3.3.3 is that the mean (or ensemble-averaged) value of the quantum conductance is less than the semiclassical conductance. This shows that interference causes an enhancement in the average backscattering from an array of scatterers — a phenomenon that has been observed with electromagnetic waves as well (Altshuler and Lee 1988, Bergmann 1984). The enhanced backscattering is destroyed by a small magnetic field of the order of tens of gauss. Consequently, weak localization (as this effect is commonly known) is characterized by a negative magnetoresistance — a magnetic field causes the resistance to decrease from its quantum to its semiclassical value. This magnetoresistance measurement is one of the common techniques for measuring the phase-coherence time.

The spin of an electron, which is analogous to the polarization of electromagnetic waves, usually does not play any significant role in transport processes; it merely doubles the number of states leading to a multiplicative factor of 2. However, non-trivial effects can arise in materials with strong spin-orbit coupling. The phenomenon of weak anti-localization is a well-known example of such an effect (Bergmann 1982). We would now like to describe a novel device concept based on such spin interference phenomena (Datta and Das 1989). The electro-optic light modulator relies on the interference between the two allowed polarizations of electromagnetic waves (Fig. 3.3.4a). A polarizer at the input polarizes the light at 45° to the y-axis (in the y-z plane) which can be represented as a linear combination of z- and y-polarized light.

\[
\begin{bmatrix}
1 \\
1
\end{bmatrix}
\begin{bmatrix}
1 \\
0 \\
1
\end{bmatrix}
(45^\circ - \text{pol.})
(\text{z-pol.})
(y-\text{pol.})
\]

(3.3.1)

As this light passes through the electro-optic material, the two polarizations suffer different
Fig. 3.3.4  (a) An electro-optic light modulator
(b) Proposed electron wave analog of the optical device shown in (a).
phase-shifts $k_1 L$ and $k_2 L$ because the electro-optic effect makes the dielectric constant $\varepsilon_{zz}$ slightly different from $\varepsilon_{yy}$. The light emerging from the electro-optic material is represented as $
abla \begin{bmatrix} e^{ik_1 L} \\ e^{ik_2 L} \end{bmatrix}$. The analyzer at the output lets the component along $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ to pass through. The output power $P_0$ is given by

$$P_0 \propto \left| \begin{bmatrix} e^{ik_1 L} \\ e^{ik_2 L} \end{bmatrix} \right|^2 = \sin^2 \left( \frac{(k_1 - k_2)L}{2} \right)$$

(3.3.2)

The light output is modulated with a gate voltage that controls the differential phase-shift $\Delta \theta \equiv (k_1 - k_2)L$.

The analogous device based on electron waves is shown in Fig. 3.3.4b. The polarizer and analyzer can be implemented using contacts made of a ferromagnetic material like iron. At the Fermi level in such materials the density of states for electrons with one spin greatly exceeds that for the other, so that the contact preferentially injects and detects electrons with a particular spin. Spin current polarization up to ~50% has been experimentally demonstrated utilizing permalloy contacts. Although further work in this area is needed, implementation of the spin polarizer and analyzer seems feasible. A contact magnetized in the x-direction preferentially launches and detects electrons spin-polarized along positive x which is represented as a linear combination of positive z-polarized and negative z-polarized electrons.

$$\begin{bmatrix} 1 \\ (+x \text{ pol.}) \\ (+z \text{ pol.}) \\ (+z \text{ pol.}) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

(3.3.3)

Finally we need the analog of an electro-optic material which will introduce a differential phase-shift between +z polarized and -z polarized electrons, that can be controlled with a gate voltage. Narrow gap semiconductors like InGaAs provide just what we need, as we will describe below.

It has been established both theoretically and experimentally that in 2DEG's in narrow-gap semiconductors there is an energy splitting between up-spin and down-spin electrons even when there is no magnetic field. The dominant mechanism for this "zero-field spin-splitting" is believed to be the Rashba term in the effective mass Hamiltonian (Bychkov and Rashba 1984).

$$H_R = \eta (\sigma_z k_x - \sigma_x k_z)$$

(3.3.4)

This term arises from the perpendicular electric field at heterojunction interfaces. The spin-orbit coupling coefficient $\eta$ can be controlled through a gate voltage. Other mechanisms such as the inversion asymmetry term also contribute to the zero-field spin splitting; however, we will ignore these here as they are usually smaller in narrow-gap semiconductors. It is easy to see that the Rashba term causes +z polarized and -z polarized electrons with the same energy to have different wavevectors $k_1$ and $k_2$. Consider an electron traveling in the x-direction with
$k_z = 0$ and $k_x \neq 0$ (we assume that the electron forms a 2DEG in the x-z plane). The Rashba term $H_R$ is then equal to $\eta \sigma_z k_x$. This raises the energy of $z$-polarized electrons by $\eta k_x$ and lowers that of $-z$ polarized electrons by the same amount. It is as if the electrons feel a magnetic field $B_z$ proportional to $k_x$ ($\eta k_x \rightarrow \mu_B B_z$, $\mu_B$ being the Bohr magneton).

$$E(z \text{ pol.}) = \frac{\hbar^2 k_{x1}^2}{2m^*} - \eta k_{x1}$$

$$E(-z \text{ pol.}) = \frac{\hbar^2 k_{x2}^2}{2m^*} + \eta k_{x2}$$

From eqs. (3.2.11a) and (3.2.11b) we obtain,

$$k_{x1} - k_{x2} = 2m^* \eta / \hbar^2$$

It is apparent that a differential phase-shift

$$\Delta \theta = (k_{x1} - k_{x2}) L = \frac{2m^* \eta L}{\hbar^2}$$

is introduced between up and down spin (or $z$ polarized and $-z$ polarized) electrons, which can be controlled with a gate voltage through the spin-orbit coefficient $\eta$. The extent to which $\eta$ can be controlled by an external gate voltage has not yet been investigated experimentally.

So far, we have considered only electrons traveling along $x$. In practice, of course, we have an angular spectrum of electrons in the x-z plane. As the angle $\theta$ of propagation of the electrons with the x-axis is increased, it can be shown that the effect is reduced gradually to zero at $\theta = 90^\circ$. For larger overall interference effects it seems advisable to restrict the angular spectrum of the electrons. This can be done with a confining potential $V(z)$ that confines the electrons in a waveguide. From an experimental point of view there appear to be at this time two main unknowns: (1) how well the spin polarizer and analyzer can be implemented in a 2DEG with magnetized contacts and (2) to what extent $\eta$ can be controlled with a gate voltage. It is hoped that future experiments will clarify these issues.

### 3.3.2 Waveguide transport

If this article were written before 1985, this section would be absent. The reason is that there were no known techniques for fabricating high-quality electron waveguides with a few propagating transverse modes or subbands. Consider a two-dimensional film with an areal electron density $n_s$. One can estimate the number of transverse modes $M$ in a channel of width $W$ as follows. The Fermi wavevector $k_F$ is related to the electron density $n_s$:

$$k_F = (2\pi n_s)^{1/2}$$

Assuming a rectangular channel, we expect the transverse momentum to be quantized in multiples of $\pi / W$, so that the number of modes below $k_F$ is approximately
Thus a 0.5 \( \mu \text{m} \) wide-channel with an electron density of \( 6.4 \times 10^{11} / \text{cm}^2 \) will have approximately 40 transverse modes. However, wires less than about 0.5 \( \mu \text{m} \) in width usually do not conduct because the Fermi level is pinned near the exposed sidewalls leading to fairly wide depletion layers. It is thus extremely difficult to control the number of modes in a wire without making it totally non-conducting. The depletion layer width can be reduced by using a shallow etch, whereby the sidewalls of the modulation-doped GaAs channel are not exposed; only the top AlGaAs layer containing the dopants is partially etched. The shallow mesa also helps reduce surface effects from degrading the channel mobility (van Houten et al. 1986). This is the technique used by Timp and coworkers in their pioneering work that opened up the field of electron waveguide transport (Timp et al. 1987). Since then, a variety of techniques for channel definition have been used by other groups such as electrostatic confinement with a split gate (van Wees et al. 1988; Wharam et al. 1988) and selective ion etch damage (Roukes et al. 1988). These developments have led to tremendous activity in semiconductors since 1987, though most of the work on mesoscopic systems originated in metals (Webb and Washburn 1988). Magnetoresistance and Hall effect measurements in narrow waveguide structures have revealed unexpected behavior (van Houten et al. 1988a, Roukes et al. 1987) which is not yet understood very well.

A striking demonstration of mode quantization in electron waveguides was provided in a recent experiment that measured the conductance of a ballistic channel whose width was reduced continuously through a split-gate structure (Wharam et al. 1988; van Wees et al. 1988). For a ballistic channel, an incident flux in any mode "n" is completely transmitted \( \left( \tau_{nm}(E) = \delta_{mn} \right) \) so that from eq. (2.5)

\[
T_{21}(E) = T_{12}(E) = 2M
\]

where \( M \) is the number of modes and the factor 2 comes from the two spins. Hence from eq. (2.11)

\[
\tilde{T}_{21} = \tilde{T}_{12} = 2M
\]

The conductance \( G_B \) of a ballistic channel is obtained from eq. (3.2.12) using \( \tilde{T}_{12} \) from eq. (3.3.11).

\[
G_B = \frac{2e^2}{h} M
\]

As the width of the channel is reduced one expects the conductance to decrease linearly in large samples. But eq. (3.3.12) shows that the conductance is quantized since \( M \) is an integer. As the channel width is reduced, the transverse modes are cut off one by one so that the conductance decreases in discrete steps of \( 2e^2 / h \). Experiments have indeed demonstrated this striking result. Such ballistic point contacts have been used to study electron focusing by a magnetic
field (van Houten et al. 1988b).

Now that single mode quantum wires are within technological reach, it seems appropriate to consider the possibility of duplicating with electron waveguides many of the well-known concepts in microwaves or optics. In electromagnetics, one commonly uses monoenergetic beam of electrons (energy plays a role analogous to frequency, see Table 3.2.1). But electron waves in solids commonly have a large spread in energy. This problem can be avoided by using low voltages and low temperatures so that only electrons near the Fermi energy contribute to the conductance (see eq. (3.2.11)).

The Mach-Zender interferometer, used as a modulator in integrated optics, consists of a single input waveguides that subsequently rejoin to form a single output waveguide. The possibility of an analogous device with electron waveguides (Fig. 3.3.5a) was proposed independently by two groups (Fowler 1985; Datta et al. 1985,1986). In the optical interferometer the phase difference between the two arms is controlled by changing the refractive index of one arm through the electro-optic effect. In view of the similarity between eqs. (3.2.18a) and (3.2.18b), we might expect that for electron waves, the phase-difference between the two arms can be changed by changing the potential $V$ of one arm with respect to the other. The applied potential shifts the subband energy $\varepsilon_n$ in one arm with respect to that in the other. Since

$$E = \varepsilon_n + \frac{\hbar^2 k^2}{2m^*}$$

we have

$$|\Delta k| = |\Delta \varepsilon_n| \frac{m^*}{\hbar^2 k}$$

Hence the phase difference $\Delta \theta$ is given by

$$|\Delta \theta| = |\Delta k| L = \frac{|\Delta \varepsilon_n|}{E_k} \frac{k L}{2}$$

$$= |\Delta \varepsilon_n| \frac{\tau_t}{\hbar}$$

where $E_k = \hbar^2 k^2 / 2m^*$ is the kinetic energy of the electron and $\tau_t = m^* L / \hbar k$ is the transit time of the electron across the gate region. We expect the conductance to change periodically with the gate potential as $\Delta \theta$ goes from zero to $\pi$ to $2\pi$ and so on. To obtain a phase-difference of $\pi$ we need

$$|\Delta \varepsilon_n| = \frac{E_k}{L/\lambda} \quad (\Delta \theta = \pi)$$

To deplete the channel would require $|\Delta \varepsilon_n| \approx E_k$. Clearly $L$ should be at least several wavelengths long in order that the device function as a quantum interference transistor rather than as a depletion mode FET.
Fig. 3.3.5. Two possible implementations of an electronic analog of the Mach-Zender interferometer where the two channels are defined (a) lithographically and (b) by film growth.
Recent experiments (deVegvar et al. 1989) showed that a gate over one arm of a ring structure can be used to shift the conductance oscillations in a magnetic field as the gate potential was changed. It will be noted that because the phase-shift is proportional to the transit time (eq. (3.3.15b)), the electrostatic effect is more difficult to observe than the magnetic effect (Timp et al. 1987) where the phase-shift depends only on the flux enclosed and is nearly unique for thin rings. It seems important to design structures that minimize the spread in transit times. An alternative structure has been proposed (Datta et al. 1986) that utilizes film growth rather than lithographic techniques to define the channels (Fig. 3.3.5b). However, this structure has not yet been fabricated. The difficulty lies in embedding a barrier layer in the middle of the conductive channel.

An interesting aspect of mesoscopic structures is the fact that the current in one part is affected by changes made anywhere within a phase-coherence length (Skocpol et al. 1987, Umbach et al. 1987). These non-local effects, as they are called, appear counter-intuitive at first sight but are rather obvious once we get accustomed to viewing electronic circuits as waveguide networks. One implication of such non-locality is that it is not necessary for the gate to be positioned between the source and the drain as we are accustomed to expect in electronic devices. It can be located anywhere within a phase-coherence length. Consider the 3-port network (Figs. 3.6); ohmic contacts are made to two of the ports while a Schottky gate is used to change the phase of the reflection coefficient at the third port (Datta 1989). The transmissivity from the source to the drain is determined primarily by the interference between the two paths shown in Fig. 3.3.6 and the gate controls the phase difference between these paths. One can view this structure too as an interferometer with the T-junction acting like a partially silvered mirror that splits the incoming beam from the source into two. Conductance oscillations as a function of the gate potential have been experimentally observed in this structure, though the underlying mechanisms have not yet been established conclusively (Miller et al. 1989). An alternative structure with the drain and gate interchanged (Fig. 3.3.7a) was proposed independently (Fowler 1988, Sols et al. 1989). Another possibility is a four-port structure shaped like a cross having two gates; the differential voltage between these gates could control the interference in a manner reminiscent of the Michelson interferometer (Fig. 3.3.7b).

It seems that one can come up with new quantum device concepts by looking up a textbook in microwaves or optics. However, the fermionic nature of electrons leads to an important difference with light. A single-mode optical fiber can, in principle, carry any amount of power per unit frequency range. But a single-mode quantum wire can only carry 80 nA of current per meV due to the exclusion principle. An important concern regarding single-mode quantum devices is their low current capability which may make them incompatible, at least with present-day integrated circuits. For this reason, "broadband" structures that are relatively insensitive to wavelength variations may be more suitable as one might be able to obtain significant interference effects even with multiple modes and large bias voltages. Of course, if quantum devices can be fabricated with sufficient precision then it should be possible to enhance the current simply by connecting a number of devices in parallel!
Fig. 3.3.6. A simple T-structure for demonstrating quantum interference effect:
Schematic diagram

Fig. 3.3.7. Two alternative structures for demonstrating quantum interference effects.
3.3.3. Space-charge effects

An important point that often tends to be overlooked is the role of space-charge effects in transport phenomena (Landauer 1987). In fact, one of the limitations of the theoretical approach outlined in Section 3.2 is that it does not tell us how to compute the electron density and hence the space-charge potential within the device. Device engineers are usually quite aware of the fact that the drift-diffusion equation needs to be solved self-consistently with the Poisson equation. Any transport equation that replaces the drift-diffusion equation for ultrasmall devices will also need to be solved likewise. Physicists, however, often view this as a minor detail that has no qualitative effect on transport. A notable exception is the role of electron-electron interactions on localization (Bergmann 1987).

Space-charge effects are known to become more significant in low-dimensional structures. Coulomb blockade in small tunnel junctions is a well-known example (Likharev 1988). Another example is the intrinsic bistability of resonant tunneling diodes (Zaslavsky et al. 1989; Alves et al. 1989), which can be understood as follows. The space-charge build-up within the well gives rise to an electrostatic potential that shifts the resonant energy with respect to Fermi energy in the emitter. This is analogous to the shift in resonant frequency of a non-linear optical Fabry-Perot interferometer similar bistability has recently been demonstrated in the operation of resonant tunneling diodes simultaneously by two different groups. Both groups used asymmetric double-barrier diodes, where the right-hand barrier has a much smaller transmissivity than the left. Consequently, the charge build-up in steady-state is much greater positive bias (when the charge has to leak out through the less transmitting barrier) than under negative bias. A hysteresis in the I-V characteristics is observed in the former case but not in the latter. The hysteresis is understood very simply as follows. As the voltage is increased, one approaches the NDR region with a filled well. The stored charge in the well tends to raise the resonant energy with respect to the emitter, so that a higher applied bias is needed to pull it below the energy of the incident electrons. On the down sweep, however, one approaches the NDR region with an empty well and the transition occurs at a lower applied bias. This result has also been observed in theoretical simulations that solve the Schrödinger equation (eq. (3.2.19a)) self-consistently with the Poisson equation (Mains et al. 1989).

The Poisson equation accounts for electron-electron interactions only in the Hartree approximation. It is possible that exchange and correlation will also play a significant role in small structures as they do in atoms and molecules.

One can view space-charge effects as a large source of non-linearity inherent in electronic transport that one might even be able to take advantage of. Non-linear optics is based on the dependence of the dielectric constant on the light intensity which is a second-order effect. The corresponding phenomenon for electrons is the dependence of the potential on the electron density which is a first-order effect. It may be possible to design novel switching devices based on an interplay between quantum effects and space-charge effects. One example is the possibility of engineering a Mott transition in a lateral superlattice. An analytical treatment of transport in
a non-linear one-dimensional superlattice has recently been published (Hawrylak et al. 1989).

The role of space-charge in electron waveguides transport has so far been neglected. This may be qualitatively different from the role that space-charge plays in large area devices where the electrostatic problem in essentially one-dimensional. By contrast in waveguides one has in general a three-dimensional electrostatic problem and the effect of space-charge may depend on the presence or absence of neighboring ground planes. Clever design may make it possible to design non-linear electron waveguide networks that behave like neural networks! Clearly major breakthroughs are needed before such exotic devices become practicable. But the true power and utility of quantum devices may eventually lie not in making a better transistor, but in the implementation of radically new electronic device concepts.

3.4 Current Theoretical Status

Much of the current theoretical work on quantum transport is based on the Landauer-Büttiker formula (eq. (3.2.15)) derived in Section 3.2; we rewrite it here for convenience.

\[ I_i = \frac{e^2}{\hbar} \sum_j \tilde{T}_{ij} (\mu_j - \mu_i) \]  

(3.4.1)

We have removed the restriction \( j \neq i \) since the term obtained by setting \( j = i \) is clearly zero; \( \tilde{T}_{ii} \) is the probability of an electron reflecting back to the same probe "i" that it was incident from (which is usually written as \( R_i \)). The coefficients \( \tilde{T}_{ij} \) have the following properties (\( B \) being the magnetic field).

\[ \tilde{T}_{ij} |_B = \tilde{T}_{ji} |_{-B} \]  

(3.4.2)

and

\[ \sum_i \tilde{T}_{ij} = \sum_i \tilde{T}_{ji} = 2M \]  

(3.4.3)

where \( M \) is the number of transverse modes in the contacts.

The symmetry property in eq. (3.4.2) is easy to prove if we assume phase-coherent transport through the device. One can then invoke S-matrix reciprocity (based on time reversibility arguments) to write

\[ \tau_{ij}^{nm} (E) |_B = \tau_{ji}^{nm} (E) |_{-B} \]  

(3.4.4)

Using eq. (3.2.5), we obtain

\[ T_{ij} (E) |_B = T_{ji} (E) |_{-B} \]  

(3.4.5)

Eq. (3.4.2) follows readily using eq. (3.2.11). To prove eq. (3.4.3), we note that due to current conservation
Using eq. (3.2.5),
\[
\sum_{i,m} \tau_{ij}^{mn}(E) |_B = \sum_{i,m} \tau_{ij}^{mn}(E) |_{-B} = 1
\]  
(3.4.6)

Using eq. (3.4.5),
\[
\sum_{i} T_{ij}(E) |_B = \sum_{i} T_{ij}(E) |_{-B} = 2M
\]  
(3.4.7)

Using eq. (3.4.5),
\[
\sum_{i} T_{ij}(E) = \sum_{i} T_{ji}(E) = 2M
\]  
(3.4.8)

Again, eq. (3.4.3) follows readily using eq. (3.2.11).

However, these relations are not as easy to prove if we allow phase-breaking processes to occur within the device. In fact, when phase-breaking processes are included, eq. (3.4.5) is no longer valid, though we believe eq. (3.4.2) is still valid.

To see this, consider the simple two-probe device in Fig. 3.4.1 with no magnetic fields \((B = 0)\). We have a single inelastic scatterer on the right of a potential barrier. An electron incident with energy \(E\) from probe 1, crosses the barrier, loses energy to the scatterer and exists into probe 2. But an electron incident with energy \(E\) from probe 2 loses energy to the scatterer before crossing the barrier and cannot cross the barrier into probe 3. Clearly \(T_{21}(E) > T_{12}(E)\) in this case and in general there is no relationship between \(T_{21}(E)\) and \(T_{12}(E)\). A better way to approach the problem is to treat different energies as separate probes and define transmission coefficients \(T_{21}(E_2, E_1)\) and \(T_{12}(E_1, E_2)\) and thereby prove the relations in eq. (3.4.2) and (3.4.3) where the indices \(i\) and \(j\) are now interpreted to include not only individual probes but also individual energies in the same probe. To our knowledge, a proper quantitative demonstration along these lines has not yet been shown using a somewhat different approach that the relations in eqs. (3.4.2) and (3.4.3) are indeed valid even in the presence of phase-breaking processes within the device [Büttiker 1988]. This is an important point since phase-coherent transport is a theorist’s idealization that is never realized precisely. On the other hand, the symmetry property in eq. (3.4.2) can be used to prove general relationships like Onsager reciprocity in mesoscopic systems.

One of the puzzles in the early days of mesoscopic physics was the fact that the conductance measured with a four-probe Hall bridge (see Fig. 3.2.2) was not symmetric in a magnetic field (Benoit et al. 1986). In a large rectangular Hall bridge one measures \(\rho_{xx}\) directly which is symmetric in a magnetic field \((\rho_{xx}(B) = \rho_{xx}(-B), \text{Onsager relation})\). But in a mesoscopic Hall bridge the voltage drop is not uniform and one measures some combination of \(\rho_{xx}\) and \(\rho_{xy}\) which is not symmetric in \(B\). Thus, one should regard mesoscopic samples as inhomogeneous conductors that obey the reciprocity relation.
Fig. 3.4.3. A simple example to show that $T_{12}(E) \neq T_{21}(E)$ in the presence of phase-breaking processes.

$$R_{mn,kl}(B) = R_{kl,mn}(-B) \quad (3.4.9)$$

where $R_{mn,kl}$ is the resistance obtained by feeding a current between terminals $m$ and $n$ and measuring a voltage between terminals $k$ and $l$. Using the symmetry property of the coefficients $\tilde{T}_{ij}$ (eq. (3.4.2)) it can be shown that the four-probe resistances obtained from eq. (3.4.1) indeed obey the reciprocity relation in eq. (3.4.9). (Büttiker 1986).

Eq. (3.4.1) is the starting point for much of the current theoretical work on mesoscopic structures. In numerical calculations, the coefficients $\tilde{T}_{ij}$ are obtained from the Schrödinger equation assuming phase-coherent transport. This approach has been used to study conductance fluctuations (Cahay et al. 1988, Baranger et al. 1988), conductance of a constriction (Szafer and Stone 1989, Avishai and Band 1989b) local and non-local bend resistances (Baranger and Stone 1989a, Avishai and Band 1989a) and the quenching of the Hall effect (Baranger and Stone 1989b). Alternatively, some authors have used diagrammatic techniques to evaluate the non-local conductivity tensor $\sigma(r,r')$ from the Kubo-formulae (Maekawa et al. 1987, Kane et al.
1987, Kane et al. 1988, Hershfield and Ambegaokar 1988, DiVincenzo and Kane 1988); the transmission coefficients $\tilde{T}_{ij}$ are related to the conductivity tensor through the Fisher-Lee relation (Fisher and Lee 1981, Stone and Szafer 1988). The advantage of the diagrammatic technique is that the phase-breaking time $\tau_o$ can be concluded in the computation. However, this approach is better suited to computing ensemble-averaged rather than sample-specific quantities. A totally different approach is to use quantum kinetic equations (Barker 1982, Jauho and Ziep 1989) based on the Wigner distribution function. A number of authors have applied such kinetic equations to the description of resonant tunneling devices (Kriman et al. 1987, Frensley 1986, 1987a, b); dissipative processes are introduced through a phenomenological relaxation time.

### 3.5 Open Questions

Despite the impressive success of eq. (3.4.1) in explaining mesoscopic phenomena a number of unresolved questions remain:

1. How can we compute the transmission coefficients $\tilde{T}_{ij}$ in general, starting from a microscopic model for the phase-breaking scatterers within the device?
2. How can we compute the electron density $n(r)$ within the device so that we can determine the space-charge potential self-consistently from the Poisson equation?
3. Even for phase-coherent or non-dissipative quantum transport, it is not clear how the transmission coefficients $\tilde{T}_{ij}$ can be computed for arbitrarily shaped contacts taking the exclusion principle into account. Usually it is assumed that the contacts are connected to the device through perfectly ordered leads and the scattering matrix is computed from the one-electron Schrödinger equation.

During the last year we have made some progress towards answering these questions. This work is described in the following chapters.

We have derived a steady-state kinetic equation that can be used to describe quantum transport in the presence of phase-breaking processes (Datta 1989b). The description is simplified significantly by the assumption that the phase-breaking scatterers are point-like with no spatial extent. The derivation of the kinetic equation is described in detail in Chapter 4 and some preliminary numerical results are described in Chapter 5. The kinetic equation is simple enough that we believe we can obtain quantitative numerical solutions for practical devices self-consistently with the Poisson equation. We thus feel that within the limitations of our model we can answer the questions listed above. By comparing these with experimental results one can assess the validity of our model and decide if further refinements are needed.
Chapter 3 References

Bychkov Y A and Rashba E I 1984 J. Phys. C 17 6039
Büttiker M 1986 Phys. Rev. Lett. 57 1761
Chang L L, Esaki L and Tsu R 1975 Appl. Phys. Lett. 24 593
Datta S 1989a Superlattices and Microstructures 6 83
Esaki L and Tsu R 1970 IBM J. Res. Dev. 14 61
Fowler A B 1985 US Patent No. 4550330
Fowler A B 1988 Workshop on Quantum Devices (Atlanta)
Frenkel J 1930 Phys. Rev. 36 1604
Frensky W R 1986 Phys. Rev. Lett. 57 2853
Frensky W R 1987a Phys. Rev. B 36 1570
Frensky W R 1987b Appl. Phys. Lett. 51 448
Jauho A P and Ziep O 1989 Physica Scripta T25 329
Landauer R 1957 IBM J. Rev. Dev. 1 223
Landauer R 1988 IBM J. Res. Dev. 32 306
Likharev K K 1988 IBM J. Res. Dev. 32 144
Stone A D 1985 Phys. Rev. Lett. 54 2692
Tokura Y and Tsubaki K 1987 Appl. Phys. Lett. 51 1807
Tsubaki K, Tokura Y, Fukui T, Saito H and Susa N 1989 25 728
Chapter 4

A QUANTUM KINETIC EQUATION

4.1 Introduction
4.2 Self-energy Function
4.3 Transport Equation
4.4 Local Thermodynamic Equilibrium
4.5 Relationship to Classical Brownian Motion
4.6 Summary

Appendix: Derivation of the Kernel from the Kubo Conductivity

References

Starting from the Keldysh formulation of non-equilibrium statistical mechanics we derive a simple kinetic equation for steady-state quantum transport under the simplifying assumption that the inelastic scattering is caused by uncorrelated point scatterers, such as magnetic impurities or impurities with internal degrees of freedom. While this assumption is not always realistic, we believe that the model can be used to describe much of the essential physics of quantum transport in mesoscopic systems. This assumption allows us to write a transport equation that involves only the electron density and not the spatial correlations of the wavefunction; as such it also has a simple physical interpretation. The kernel of this integral equation is calculated from the Schrödinger equation and contains all quantum interference effects. We show that at equilibrium the electron density relaxes to the expected equilibrium value with a constant chemical potential everywhere in the structure. Assuming local thermodynamic equilibrium we then derive a linear-response transport equation which resembles the Landauer-Büttiker formula extended to include a continuous distribution of probes. An alternative derivation is provided in the appendix for the kernel of the linear-response transport equation, starting from the Kubo formula for the conductivity. We discuss the conditions under which this transport equation reduces to the well-known drift-diffusion equations describing classical Brownian motion. In the present work we neglect electron-electron interaction beyond the Hartree term.
4.1 Introduction

Much of our understanding of electron transport in solids is based on the Boltzmann transport equation (BTE).

\[ \frac{\partial}{\partial t} f(r; k; t) + v \cdot \nabla_r f(r; k; t) + \frac{e}{m} \cdot \nabla_k f(r; k; t) = S_{op} f(r; k; t) \quad (4.1.1) \]

Here \( f(r; k; t) \) is the distribution function in phase space that tells us the number of particles at \( r \) with wavevector \( k \) at time \( t \). \( S_{op} \) is the scattering operator which is local in space.

\[ S_{op} f(r; k; t) = \sum_{k'} \left[ S(k, k') f(k') [1 - f(k)] - S(k', k) f(k) [1 - f(k')] \right]_{r, t} \quad (4.1.2) \]

The scattering function \( S(k, k') \) is commonly obtained from Fermi's golden rule. The BTE is based on a simple semiclassical picture of transport: Electrons are particles that obey Newton's law in an external electric field and are scattered occasionally by phonons and impurities. Despite its impressive successes, it suffers from an important limitation; it cannot describe transport phenomena in which the wave nature of electrons plays a crucial role. A variety of such quantum effects have been discovered over the years, such as tunneling [1], resonant tunneling [2], weak and strong localization [3], the quantum Hall effect [4], etc. Since 1985, experiments on mesoscopic structures have revealed a wealth of new effects such as the Aharonov-Bohm effect, conductance fluctuations, non-local effects, quantized conduction in 1D ballistic channels etc. For ultrasmall structures at low temperature, these phenomena have clearly revealed that electron transport is dominated by wave interference effects not unlike those well-known in microwave networks. It has also become clear that in mesoscopic structures, whose dimensions are comparable to the phase-breaking length, it is necessary to distinguish between sample-specific properties and ensemble-averaged properties; solid-state physics in the past had been almost exclusively concerned with the latter.

An important topic of current theoretical research is to develop a quantum transport formalism that can be used to describe the sample-specific properties of mesostructures. Much of the current theoretical work on mesoscopic structures is based on the Landauer-Büttiker formula [13-20] which relates the current \( I_i \) at lead \( i \) to the chemical potential at lead \( j \).

\[ I_i = \frac{e^2}{h} \sum_j \left[ (T_0)_{ij} \mu_j - (T_0)_{ji} \mu_i \right] \quad (4.1.3) \]

where

\[ (T_0)_{ij} = \int dE \left[ - \frac{\partial f_0}{\partial E} \right] T_{ij}(E) \quad (4.1.4) \]

\[ T_{ij}(E) = \text{Tr} \{ t_j^\dagger(E) t_i(E) \} \quad (4.1.5) \]

Here \( f_0 \) is the Fermi-Dirac function and \( t_{ij}(E) \) is the transmission matrix from lead \( j \) to lead \( i \) for electrons with energy \( E \). It is usually assumed that there is no phase-breaking scattering process within the device; all such processes occur in the contacts. The scattering matrix for the device
can then be computed from the time-independent Schrödinger equation

\[ \left( \frac{(p - eA)^2}{2m^*} + eV \right) \psi = E\psi \]  

(4.1.6)

where \( A(r) \) and \( V(r) \) are the vector and scalar potentials within the device. Eq. (4.1.3) has been derived rigorously from the Kubo formalism [21,22]. This approach thus reduces the problem of computing the conductance of a device to that of computing its scattering matrix, much like microwave circuits.

Eq. (4.1.3) has been quite successful in explaining qualitatively many of the recent experimental observations in mesoscopic systems [23-40]. However, in order to make quantitative comparisons, it is necessary to include the phase-breaking processes that are inevitably present in any device. Phase-breaking processes are those that involve a change in the state of the scatterer and thus cannot simply be included in the Schrödinger equation with an appropriate choice of the scattering potential \( V \). For example, if we use a time-varying potential \( V(r,t) \) in the time-dependent Schrödinger equation it will lead to inelastic processes but such processes will be strictly reversible, with absorption and stimulated emission but no spontaneous emission [41]. To include irreversible dissipative processes within the device one has to include the reservoir degrees of freedom explicitly and trace over the states of the reservoir. Diagrammatic techniques based on the Kubo formalism do include phase-breaking processes, but these techniques are more suited to computing ensemble-averaged rather than sample-specific properties [42-54]. Moreover, this approach is usually restricted to linear response.

Quantum kinetic equations provide a powerful approach to including dissipative processes in quantum transport theory for both linear and non-linear response [55-64]. Here the semiclassical distribution function \( f(r,k,t) \) is replaced by the Wigner distribution function \( W(r,k,E,t) \) which is obtained from the Green function

\[ G^<(r_1,r_2 ; t_1,t_2) = \frac{i}{\hbar} \langle \psi^\dagger(r_2,t_2) \psi(r_1,t_1) \rangle \]  

(4.1.7)

\( \psi(r,t) \) being the electron field operator) by transforming to center-of-mass

\[ r = (r_1 + r_2)/2, \quad t = (t_1 + t_2)/2 \]  

(4.1.8a)

and relative coordinates, and then Fourier transforming with respect to the relative coordinate.

\[ r_1 - r_2 \rightarrow k, \quad t_1 - t_2 \rightarrow E \]  

(4.1.8b)

An equation of motion for the Green function (and hence the Wigner distribution function) is derived starting from the Dyson equation [55]. In the Keldysh formulation of non-equilibrium statistical mechanics, the Dyson equation is conveniently written in matrix form.

\[ G(X_1,X_2) = G_0(X_1,X_2) + \int dX_3 \ dX_4 \ G_0(X_1,X_3) \Sigma(X_3,X_4) G(X_4,X_2) \]  

(4.1.9)

where \( X \) stands for \( (r,t) \). \( G \) is a \( (2 \times 2) \) matrix.
\[ G = \begin{bmatrix} G^T - G^c \\ G^\tau - G^{\tau} \end{bmatrix} \] (4.1.10)

whose elements are defined by

\[ G^c(X_1, X_2) = \frac{i}{\hbar} \langle \psi^*(X_2) \psi(X_1) \rangle \] (4.1.11a)

\[ G^\tau(X_1, X_2) = -\frac{i}{\hbar} \langle \psi(X_1) \psi^*(X_2) \rangle \] (4.1.11b)

\[ G^T(X_1, X_2) = \theta(t_1 - t_2) G^\tau(X_1, X_2) + \theta(t_2 - t_1) G^c(X_1, X_2) \] (4.1.11c)

\[ G^{\tau}(X_1, X_2) = \theta(t_1 - t_2) G^c(X_1, X_2) + \theta(t_2 - t_1) G^\tau(X_1, X_2) \] (4.1.11d)

The bracket \( \langle \cdots \rangle \) denotes an average over the available states of the system, that is, a trace over the reservoir states. The self-energy function \( \Sigma \) is also a \((2 \times 2)\) matrix of the same form as \( G \). \( G_0 \) is the unperturbed Green function. In addition to the four functions defined in eqs. (4.1.11a-d) it is convenient to define a retarded and an advanced Green function as follows.

\[ G^R(X_1, X_2) = \theta(t_1 - t_2) [ G^\tau(X_1, X_2) - G^c(X_1, X_2) ] \] (4.1.12a)

\[ G^A(X_1, X_2) = \theta(t_2 - t_1) [ G^c(X_1, X_2) - G^\tau(X_1, X_2) ] \] (4.1.12b)

The retarded and advanced self-energy functions \( \Sigma^R, \Sigma^A \) are also defined accordingly.

Quantum kinetic equations have so far not found much use in the description of transport in mesoscopic structures. The practical difficulty seems to be their complexity as well as the fact that quantum distribution functions, being complex quantities, often have counter-intuitive properties, so that it is difficult to make approximations. In this paper we will present a simple quantum kinetic equation that can be used to describe steady state transport. In general, in quantum transport theory we encounter the independent variables \((r_1, r_2; t_1, t_2)\) or equivalently \((r; k; E; t)\), as we just discussed. Because we restrict ourselves to steady state transport, the time variable \( t = (t_1 + t_2) / 2 \) (eq. (4.1.8a)) is averaged over. This leaves us with functions of the form \( G(r_1, r_2; E) \) or \( G(r; k; E) \). A further simplification is achieved by assuming a special form for the inelastic scattering. We arrive at a transport equation that only involves the diagonal elements \( G(r, r; E) \) of the Green functions. Spatial correlations of the field represented by the off-diagonal elements \( G(r_1, r_2; E), r_1 \neq r_2 \) do not appear in this equation. In order to achieve this simplification, we assume that inelastic scattering is caused by a distribution of independent oscillators, each of which interacts with the electrons through a \textit{delta potential}. We also assume that inelastic scattering processes are weak and infrequent, just as one does in deriving Fermi's golden rule; however, the \textit{elastic scattering processes are treated exactly}. This model closely approximates a laboratory sample with magnetic impurities, or impurities with internal degrees of freedom. For other types of inelastic scattering the model may not be realistic; however, we believe that it should still be possible to describe much of the essential physics of dissipation in quantum transport. Physically, it is easy to see why the above assumption leads to a simple
transport equation that does not involve spatial correlations of the wavefunction. In the "golden rule" approximation, each scatterer acts independently. Since we have assumed a delta interaction potential, an inelastic scattering event only involves the wavefunction at a particular point and is insensitive to spatial correlations.

The simplification described above is important for two reasons. Firstly, the number of independent variables is reduced from \((r_1, r_2; E)\) (or equivalently, \((r; k; E)\)) to \((r; E)\). Secondly, the diagonal elements have simple physical interpretations; for example, the electron density per unit energy \(n(r; E)\) is identified with

\[
n(r; E) = \frac{-i}{2\pi} G^<(r, r; E)
\]

while the hole density per unit energy \(p(r; E)\) is identified with

\[
p(r; E) = \frac{+i}{2\pi} G^>(r, r; E)
\]

We emphasize that the use of \(r\) and \(E\) simultaneously does not violate the uncertainty principle. As shown in eq. (4.1.8b) the energy spectrum is derived from the temporal correlations of the wavefunction at a point \(r\) and bears no relationship to \(k\) which has to do with the spatial correlations. We are not using conjugate variables like \(r\) and \(k\) or \(E\) and \(t\) simultaneously. The transport equation thus involves only positive quantities and can be understood in simple physical terms. This makes it easy to make intuitive approximations. Monte Carlo analysis based on a probabilistic interpretation should also be possible. We believe that this is a consequence of the fact that inelastic scattering events in our model may be viewed as quantum measurements of the position and energy of the electron. Every time an electron is inelastically scattered it leaves one of these oscillators in an excited state, and energy is dissipated into the surroundings as the oscillator relaxes back to its state of thermodynamic equilibrium. An observer who monitors the states of the oscillators will see a series of flashes with different energies from different spatial locations and can, in principle, deduce the electron density \(n(r; E)\) from the observations. We believe that it is for this reason that the transport equation involves only real positive quantities.

Staring from the Dyson equation (eq. (4.1.9)) we have derived an integral transport equation that can be solved to obtain the electron density per unit energy \(n(r; E)\) under steady state conditions [65]. In this chapter we will provide a detailed derivation of the important results. Similar results were derived by us heuristically from a one-electron picture, earlier [66,67]. As we will see, the kernel of this transport equation is computed from the one-electron Schrödinger equation and contains all quantum interference effects due to elastic scatterers, boundaries, etc. Space-charge effects are taken into account by including in the Schrödinger equation the electrostatic Hartree potential obtained self-consistently from the Poisson equation; electron-electron interactions are neglected beyond the Hartree approximation. It will be noted that setting \(r_1 = r_2\) is equivalent to integrating over the Fourier transform variable \(k\). The electron
density \( n(r; E) \) is equal to the Wigner distribution function \( W(r; k; E) \) integrated over all \( k \). Consequently, our integral transport equation allows us to compute the electron density \( n(r; E) \) but not the current density \( J(r; E) \). The detailed distribution of currents throughout the structure is not obtained from this formulation.

In this paper we adopt a microscopic approach starting from a model Hamiltonian for the inelastic scatterers; however, our model is closely related to the Landauer picture. Since the inelastic scattering process is purely local, it can be viewed as an exit into a reservoir followed by reinjection into the main structure \([68,35]\). From this point of view it would seem that distributed inelastic scattering processes can be simulated by connecting a continuous distribution of reservoirs throughout a structure (Fig. 4.1.1). Indeed, when we simplify our transport equation to linear response we obtain what looks like the Landauer-Büttiker formula \((\text{eq. (4.1.3)})\) generalized to include a continuous distribution of probes. Moreover, the kernel of this linear-response transport equation can be derived directly from the Kubo formula for the conductivity using the Lee-Fisher formula; this is shown in the appendix. Therefore, our transport equation reduces to well-known results in the limit of linear response. A direct generalization of the Landauer-Büttiker formula would appear to be a phenomenological approach to simulating inelastic scattering. Our paper thus provides the rigorous justification for such an approach, by deriving the transport equation directly from a model Hamiltonian making certain well-defined assumptions.

This paper also serves to clarify the meaning of the chemical potential \( \mu(r) \) in quantum transport theory. As we mentioned earlier, the transport equation derived in this paper is formulated in terms of the electron density per unit energy \( n(r; E) \). We emphasize that this is a well-defined quantum mechanical quantity. The energy variable \( E \) is derived from the temporal correlations of the wavefunction at a point \( r \), and bears no relationship to \( k \); there is thus no violation of the uncertainty principle since conjugate variables (like \( r \) and \( k \) or \( E \) and \( t \)) are never invoked simultaneously. In order to derive the linear-response transport equation, we assume \textit{local thermodynamic equilibrium} so that we can write the electron density \( n(r; E) \) in terms of a local chemical potential \( \mu(r) \)

\[
n(r; E) = N_0(r; E) \frac{1}{e^{(E - \epsilon \mu(r))/k_B T} + 1} \tag{4.1.14}
\]

where \( N_0(r; E) \) is the electronic density of states. It is with this assumption that our transport equation simplifies to a form resembling the Landauer-Büttiker formula generalized to a continuous distribution of probes. On the other hand, if the driving forces are large enough (or the inelastic scattering weak enough), local thermodynamic equilibrium may not be maintained. It is then not appropriate to talk in terms of a local chemical potential; we should solve for the actual distribution \( n(r; E) \) using the more general transport equation.

The outline of the paper is as follows. In Section 4.2 we describe the model that we use and compute the self-energy function. The transport equation is then derived in Section 4.3, starting from the Dyson equation. The linear-response transport equation is derived in Section
In the limit of linear response, the transport equation can be viewed as a generalization of the Landauer-Büttiker formula to a continuous distribution of reservoirs. Each reservoir simulates the action of an inelastic scatterer.
4.4, assuming local thermodynamic equilibrium. In Section 4.5 we discuss the conditions under which the linear-response transport equation reduces to the drift-diffusion equation that is widely used to describe classical Brownian motion. The diffusion coefficient is computed for a few simple examples (analytically as well as numerically). Finally in Section 4.6 we conclude by summarizing our important results.

4.2 Self-energy Function

In this section we will describe the basic model that we use and evaluate the self-energy function. The self-energy function will be used in the next section to obtain the transport equation from the Dyson equation.

We consider any arbitrary structure in which the propagation of electrons is described by the following one-electron effective-mass Hamiltonian.

$$H_0 = \frac{[p-eA(r)]^2}{2m} + eV(r)$$  \hspace{1cm} (4.2.1)

The vector and scalar potentials $A(r)$ and $V(r)$ include the Hartree potential obtained from a self-consistent solution with the Poisson equation, as well as externally imposed potentials, and all sources of elastic scattering such as impurities, defects, boundaries, etc. For the inelastic scattering we assume a reservoir of independent oscillators labeled by the index $m$,

$$H_p = \sum_m \hbar \omega_m (a_m^+ a_m + \frac{1}{2})$$  \hspace{1cm} (4.2.2)

where $a_m^+$ and $a_m$ are the creation and annihilation operators for oscillator $m$. We assume that each oscillator interacts with the electrons through a delta-potential, so that the interaction Hamiltonian $H'$ can be written as

$$H' = \sum_m U \delta(r-r_m) (a_m^+ + a_m)$$  \hspace{1cm} (4.2.3)

Note that we have assumed the interaction strength $U$ to be constant. There is no loss of generality since the strength of inelastic scattering can be adjusted through the density of scatterers per unit volume per unit energy, described by some function $J_0(r; \hbar \omega)$. The summation over $m$ is eventually replaced by an integral.

$$\sum_m \Rightarrow \int dr \int d(\hbar \omega) J_0(r; \hbar \omega)$$  \hspace{1cm} (4.2.4)

In calculating the self-energy we restrict ourselves to one-phonon processes (Fig. 4.2.1) as one does in deriving Fermi’s golden rule. For one-phonon processes the self-energy function can be written in the form

$$\Sigma^>(X_1,X_2) = G^>(X_1,X_2) D^>(X_1,X_2)$$  \hspace{1cm} (4.2.5a)

$$\Sigma^<(X_1,X_2) = G^<(X_1,X_2) D^<(X_1,X_2)$$  \hspace{1cm} (4.2.5b)

The electron Green functions $G^>$, $G^<$ were defined earlier in Section 4.1 (eqs. (4.1.11a,b)). The
Fig. 4.2.1: One-phonon contribution to the self-energy function.
functions $D^\rho$, $D^\sigma$ are given by

\[
D^\rho(X_1, X_2) = \langle H'(r_1, t_1) \ H'(r_2, t_2) \rangle
\]

(4.2.6a)

\[
D^\sigma(X_1, X_2) = \langle H'(r_2, t_2) \ H'(r_1, t_1) \rangle
\]

(4.2.6b)

Using eq. (4.2.3) for $H'$ we obtain

\[
D^\rho(X_1, X_2) = U^2 \sum_{m,n} \delta(r_1-r_m) \delta(r_2-r_n) \left( (a_{m}^\dagger(t_1) + a_{m}(t_1)) (a_{n}^\dagger(t_2) + a_{n}(t_2)) \right)
\]

(4.2.7)

We assume that the reservoir of oscillators is in a state of thermodynamic equilibrium, so that

\[
\langle a_{m}(t_1) a_{n}(t_2) \rangle = \delta_{mn} N(\hbar \omega_m) e^{i \omega_m (t_1 - t_2)}
\]

(4.2.8a)

\[
\langle a_{m}^\dagger(t_1) a_{n}^\dagger(t_2) \rangle = \delta_{mn} (N(\hbar \omega_m) + 1) e^{-i \omega_m (t_1 - t_2)}
\]

(4.2.8b)

\[
\langle a_{m}(t_1) a_{n}^\dagger(t_2) \rangle = 0
\]

(4.2.8c)

\[
\langle a_{m}^\dagger(t_1) a_{n}^\dagger(t_2) \rangle = 0
\]

(4.2.8d)

where $N(\hbar \omega_m)$ is the average number of "phonons" in a oscillator of frequency $\omega_m$ and is given by the Bose-Einstein factor

\[
N(\hbar \omega) = \frac{1}{e^{\frac{\hbar \omega}{k_B T}} - 1}
\]

(4.2.9)

Using eqs. (4.2.8a-d) we obtain from eq. (4.2.7),

\[
D^\rho(X_1, X_2) = U^2 \sum_{m} \delta(r_1-r_m) \sum_{n} \delta(r_1-r_n) \left[ N(\hbar \omega_m) e^{i \omega_m (t_1 - t_2)} + (N(\hbar \omega_m) + 1) e^{-i \omega_m (t_1 - t_2)} \right]
\]

(4.2.10)

Replacing the sum over $m$ by an integral (eq. (4.2.4)) and Fourier transforming $t_1 - t_2 \rightarrow \epsilon$ we have,

\[
D^\rho(r_1, r_2; \epsilon) = 2\pi \hbar U^2 J_0(\epsilon) \delta(r_1 - r_2)
\]

\[
\begin{cases} 
N(\epsilon) & , \epsilon < 0 \\
N(\epsilon) + 1 & , \epsilon > 0
\end{cases}
\]

(4.2.11a)

Similarly it can be shown that

\[
D^\sigma(r_1, r_2; \epsilon) = 2\pi \hbar U^2 J_0(\epsilon) \delta(r_1 - r_2)
\]

\[
\begin{cases} 
N(\epsilon) + 1 & , \epsilon < 0 \\
N(\epsilon) & , \epsilon > 0
\end{cases}
\]

(4.2.11b)

To calculate the self-energy functions we Fourier transform eqs. (4.2.5a,b)
\[ \Sigma^<(r_1, r_2; E) = \int \frac{dE'}{2\pi \hbar} \; G^<(r_1, r_2; E') \; D^>(r_1, r_2; E-E') \]  
\[ \Sigma^>(r_1, r_2; E) = \int \frac{dE'}{2\pi \hbar} \; G^>(r_1, r_2; E') \; D^<(r_1, r_2; E-E') \]

Using eqs. (4.2.11a,b) and eqs. (4.1.13a,b) we obtain from eqs. (4.2.12a,b),

\[ \Sigma^>(r_1, r_2; E) = -\frac{i\hbar}{\tau^>(r_1; E)} \delta(r_1 - r_2) \]  
\[ \Sigma^<(r_1, r_2; E) = \frac{i\hbar}{\tau^<(r_1; E)} \delta(r_1 - r_2) \]

where

\[ \frac{1}{\tau^>(r; E)} = \frac{2\pi}{\hbar} \int dE' F(r; E'-E) \; p(r; E') \]  
\[ \frac{1}{\tau^<(r; E)} = \frac{2\pi}{\hbar} \int dE' F(r; E-E') \; n(r; E') \]

\[ F(r; \epsilon) = U^2 J_0(r; |\epsilon|) \begin{cases} N(\epsilon), & \epsilon > 0 \\ N(|\epsilon|) + 1, & \epsilon < 0 \end{cases} \]

Physically \(1/\tau^>(r; E)\) is the rate at which electrons are scattered out of energy \(E\) at the point \(r\) assuming that it is initially full. Similarly \(1/\tau^<(r; E)\) is the rate at which holes are scattered out of energy \(E\) at the point \(r\) assuming that it is initially empty. The similarity of eq. (4.2.14a) to Fermi's golden rule will be noted. However, unlike the usual golden rule we are using the position representation and not the energy eigenstates. The simple result is made possible by our assumption of independent point-size inelastic scatterers that only see the electron wavefunction at one point.

Before concluding this section let us evaluate the retarded and advanced self-energy functions, \(\Sigma^R\) and \(\Sigma^A\).

\[ \Sigma^R(X_1, X_2) = \theta(t_1 - t_2) \left[ \Sigma^>(X_1, X_2) - \Sigma^<(X_1, X_2) \right] \]  
\[ \Sigma^A(X_1, X_2) = \theta(t_2 - t_1) \left[ \Sigma^<(X_1, X_2) - \Sigma^>(X_1, X_2) \right] \]

Fourier transforming with respect to \((q - t_2)\) we have

\[ \Sigma^R(r_1, r_2; E) = i \int_{-\infty}^{+\infty} \frac{dE'}{2\pi} \; \frac{\Sigma^>(r_1, r_2; E') - \Sigma^<(r_1, r_2; E')}{E - E' + i\epsilon} \]  
\[ \Sigma^A(r_1, r_2; E) = -i \int_{-\infty}^{+\infty} \frac{dE'}{2\pi} \; \frac{\Sigma^<(r_1, r_2; E') - \Sigma^>(r_1, r_2; E')}{E - E' - i\epsilon} \]

Using eqs. (4.2.13a,b) we obtain from eq. (4.2.17a),
\[ \Sigma^R(r_1, r_2 ; E) = \frac{n}{2\pi} \delta(r_1 - r_2) \int \frac{dE'}{E - E' + i\epsilon} \frac{1}{\tau^R(r_1; E')} \]  

(4.2.18)

where,

\[ \frac{1}{\tau^R(r; E)} = \frac{1}{\tau^\alpha(r; E)} + \frac{1}{\tau^\beta(r; E)} \]  

(4.2.19)

Hence, we have,

\[ \text{Im}\{\Sigma^R(r_1, r_2 ; E)\} = -\frac{n}{2\tau^R(r_1; E)} \delta(r_1 - r_2) \]  

(4.2.20a)

\[ \text{Re}\{\Sigma^R(r_1, r_2 ; E)\} = \sigma^R(r_1 ; E) \delta(r_1 - r_2) \]  

(4.2.20b)

where,

\[ \sigma^R(r ; E) = \frac{n}{2\pi} P \int \frac{dE'}{(E - E') \tau^R(r ; E')} \]  

(4.2.21)

\( P \) represents the principal value of the integral.

The advanced self-energy function can be obtained from the relation

\[ \Sigma^A(r_1, r_2 ; E) = [\Sigma^R(r_2, r_1 ; E)]^* \]  

(4.2.22)

This is a general relationship between advanced and retarded functions that holds for the Green function as well.

\[ G^A(r_1, r_2 ; E) = [G^R(r_2, r_1 ; E)]^* \]  

(4.2.23)

To obtain eq. (4.2.22) or (4.2.23) we note that from the definition of \( G^<(X_1, X_2) \) in eq. (4.1.11a) we have

\[ G^<(X_1, X_2) = \frac{i}{\hbar} \langle \psi^+(X_2) \psi(X_1) \rangle \]

\[ = -\left[ \frac{i}{\hbar} \langle \psi^+(X_1) \psi(X_2) \rangle \right]^* \]

\[ = -[G^<(X_2, X_1)]^* \]  

(4.2.24)

Since the Green functions depend only on the time differences \( t = t_1 - t_2 \), we can write

\[ G^<(r_1, r_2 ; t) = -[G^<(r_2, r_1 ; -t)]^* \]  

(4.2.25)

Hence, on Fourier transforming

\[ G^<(r_1, r_2 ; E) = -[G^<(r_2, r_1 ; E)]^* \]  

(4.2.26a)

The same relation holds for \( G^\alpha \) as well.

\[ G^\alpha(r_1, r_2 ; E) = -[G^\alpha(r_2, r_1 ; E)]^* \]  

(4.2.26b)

Subtracting eq. (4.2.26a) from eq. (4.2.26b) we obtain,
Eq. (4.2.23) is readily obtained using eq. (4.2.27) and noting that $G^R$ and $G^A$ are related to $G^>$ and $G^<$ through relations analogous to eqs. (4.2.17a,b) for the self-energy functions. Eq. (4.2.22) can also be obtained in a similar fashion.

4.3 The Transport Equation

To derive the transport equation we start from eq. (4.1.9), noting that

$$\left[ i\hbar \frac{d}{dt} - H_0(r_1) \right] G_0(X_1, X_2) = \delta^4(X_1 - X_2) \mathbf{I} \tag{4.3.1}$$

where $\mathbf{I}$ is the (2x2) identity matrix. Operating on eq. (4.1.9) with $\left[ i\hbar \frac{d}{dt} - H_0(r_1) \right]$ and using eq. (4.3.1) we obtain

$$\left[ i\hbar \frac{d}{dt} - H_0(r_1) \right] G(X_1, X_2) = \delta^4(X_1 - X_2) \mathbf{I} + \int dX_3 \Sigma(X_1, X_3) G(X_3, X_2) \tag{4.3.2}$$

Each element in eq. (4.3.2) is a (2x2) matrix, so that it is equivalent to four separate equations. We consider only the component involving $G^<$ on the left.

$$\left[ i\hbar \frac{d}{dt} - H_0(r_1) \right] G^<(X_1, X_2) =$$

$$\int dX_3 \left[ \Sigma^R(X_1, X_3) G^<(X_3, X_2) - \Sigma^<(X_1, X_3) G^R(X_3, X_2) \right] \tag{4.3.3}$$

We note that

$$\Sigma^R(X_1, X_3) = \theta(t_1 - t_3) \Sigma^R(X_1, X_3) + \theta(t_3 - t_1) \Sigma^<(X_1, X_3)$$

$$= \Sigma^R(X_1, X_3) + \Sigma^<(X_1, X_3) \tag{4.3.4}$$

where the retarded self-energy function $\Sigma^R$ was defined earlier (eq. (4.2.16a)). Also,

$$G^R(X_3, X_2) = \theta(t_3 - t_2) G^<(X_3, X_2) + \theta(t_2 - t_3) G^>(X_3, X_2)$$

$$= -G^A(X_3, X_2) + G^<(X_3, X_2) \tag{4.3.5}$$

where the advanced Green function $G^A$ defined in the same way as the advanced self-energy function $\Sigma^A$ (eq. (4.2.16b)). Using eqs. (4.3.4) and (4.3.5) in eq. (4.3.3) we obtain,

$$\left[ i\hbar \frac{d}{dt} - H_0(r_1) \right] G^<(X_1, X_2) =$$

$$\int dX_3 \Sigma^<(X_1, X_3) G^A(X_3, X_2) \tag{4.3.6}$$

Fourier transforming with respect to $(t_1 - t_2)$ we have
\[
[E - H_0(r_1)] G^<(r_1, r_2; E) = \int \text{d}r_3 \Sigma^R(r_1, r_3; E) G^<(r_3, r_3; E)
\]

\[
= \int \text{d}r_3 \Sigma^<(r_1, r_3; E) G^A(r_3, r_2; E)
\]

(Eq. 4.3.7)

Here we have assumed that the self-energy functions as well as the Green functions depend only on time differences like \((t_1 - t_3)\), and not on \((t_1 + t_3)\). The integrals then represent convolution products in time whose Fourier transforms are simple products in energy. Substituting for \(\Sigma^R\) from eqs. (4.2.20a,b) and \(\Sigma^<\) from eq. (4.2.13b) we obtain

\[
\left[ E - H_0(r_1) - \sigma^R(r_1; E) + \frac{i\hbar}{2\tau^R(r_1; E)} \right] G^<(r_1, r_2; E) = i\hbar \frac{G^A(r_1, r_2; E)}{\tau^<(r_1; E)}
\]

(Eq. 4.3.8)

It can be shown from eq. (4.3.2) that

\[
[i\hbar \frac{\partial}{\partial t_1} - H_0(r_1)] G^R(X_1, X_2) = \int \text{d}X_3 \Sigma^R(X_1, X_3) G^R(X_3, X_2) = \delta^4(X_1 - X_2)
\]

(Eq. 4.3.9)

Eq. (4.3.9) is obtained by considering the component of eq. (4.3.2) involving \(G^T\) on the left, subtracting eq. (4.3.3) from it and noting that \(G^R = G^T - G^<\). Fourier transforming and substituting for \(\Sigma^R\) from eqs. (4.2.20a,b) we obtain

\[
\left[ E - H_0(r_1) - \sigma^R(r_1; E) + \frac{i\hbar}{2\tau^R(r_1; E)} \right] G^R(r_1, r_2; E) = \delta(r_1 - r_2)
\]

(Eq. 4.3.10)

Using eq. (4.3.10) we can write down the solution to eq. (4.3.8) as

\[
G^<(r_1, r_2; E) = i\hbar \int \text{d}r_3 \frac{G^R(r_1, r_3; E) G^A(r_3, r_2; E)}{\tau^<(r_3; E)}
\]

(Eq. 4.3.11)

We now set \(r_1 = r_2 \equiv r\); using eqs. (4.1.13a) and (4.2.23), we have

\[
n(r; E) = \frac{\pi}{2\pi} \int \text{d}r' \frac{|G^R(r, r'; E)|^2}{\tau^<(r'; E)}
\]

(Eq. 4.3.12)

By considering the component of the matrix equation, eq. (4.3.2), corresponding to \(G^>\) instead of \(G^<\) we could come up with an equation for the hole density \(p(r; E)\) instead of the electron density \(n(r; E)\). Instead of eq. (4.3.12) we obtain

\[
p(r; E) = \frac{\pi}{2\pi} \int \text{d}r' \frac{|G^R(r, r'; E)|^2}{\tau^<(r'; E)}
\]

(Eq. 4.3.13)

Adding eqs. (4.3.12) and (4.3.13) and using eq. (4.2.19) we obtain an important relationship.

\[
N_0(r; E) = \frac{\pi}{2\pi} \int \text{d}r' \frac{|G^R(r, r'; E)|^2}{\tau^<(r'; E)}
\]

(Eq. 4.3.14)

where \(N_0(r; E) = n(r; E) + p(r; E)\) is the electronic density of states. Neglecting any level broadening due to inelastic scattering processes the density of states is given by
\[ N_0(r; E) = -\text{Im}\left\{ G^R(r,r'; E) \right\} / \pi = \sum_M |\phi_M(r)|^2 \delta(E - \epsilon_M) \quad (4.3.15) \]

where \( \phi_M(r) \) are the eigenfunctions of \( H_0 \) (eq. (4.2.1)) with eigenvalues \( \epsilon_M \).

To obtain the transport equation we could start either from eq. (4.3.12) or from eq. (4.3.13) and include the effect of an external current source. However, before proceeding to do this, we will digress slightly and show how the above relationship (eq. (4.3.14)) can be derived directly from eq. (4.3.10). This is an important identity that will be used often. Consider the continuity equation obeyed by the probability density

\[ n = |G^R(r,r'; E)|^2 \quad (4.3.16) \]

and the probability current density

\[ \frac{1}{e} J = \frac{i \hbar}{2m} \left[ (\nabla G^R)^* G^R - G^R* (\nabla G^R) \right] \quad (4.3.17) \]

that we obtain from the solution to eq. (4.3.10). It can be shown from eqs. (4.3.10), (4.3.16) and (4.3.17) that

\[ \frac{1}{e} \nabla \cdot J + \frac{n}{\tau^R} = \frac{i \hbar}{\pi} \delta(r-r') [G^R - G^R*] \quad (4.3.18) \]

Integrating over all volume, using the divergence theorem and assuming that the boundaries are far away so that no current flows out of the surface, we have (using eq. (4.3.15))

\[ \int dr' \frac{|G^R(r',r; E)|^2}{\tau^R(r'; E)} = \frac{2\pi}{\hbar} N_0(r; E) \quad (4.3.19) \]

Eq. (4.3.19) is almost the same as eq. (4.3.14) except that \( r \) and \( r' \) in the argument of \( G^R \) are interchanged. To obtain eq. (4.3.14) consider eq. (4.3.19) with the magnetic field reversed: \( H \rightarrow -H \).

\[ \int_{-H} dr' G^R(r',r; E) \bigg|_{-H} = \frac{2\pi}{\hbar} N_0(r; E) \bigg|_{-H} \quad (4.3.20) \]

Now it is easy to see from eq. (4.3.20) that the density of states is unaffected by the magnetic field reversal which merely replaces each eigenfunction \( \phi_M(r) \) by its complex conjugate. Similarly the inelastic scattering time \( \tau^R(r; E) \) is unaffected. However, the Green function has the property that

\[ G^R(r',r; E) \bigg|_{-H} = G^R(r,r'; E) \bigg|_{H} \quad (4.3.21) \]

Using these results we obtain from eqs. (4.3.19) and (4.3.20),

\[ \int dr' \frac{|G^R(r',r; E)|^2}{\tau^R(r'; E)} = \int dr' \frac{|G^R(r,r'; E)|^2}{\tau^R(r'; E)} = \frac{2\pi}{\hbar} N_0(r; E) \quad (4.3.22) \]

which is the result we sought.
External current: So far we have not considered any external sources. We will now modify eqs. (4.3.12) and (4.3.13), somewhat heuristically to include the external current which is assumed to be injected or extracted incoherently. First we note that since the self-energy is a delta function (eqs. (4.2.13a,b)), the rate at which electrons are inelastically scattered out is given by $n(r; E)/\tau^e(r; E)$ while the rate at which holes are inelastically scattered out is given by $p(r; E)/\tau^e(r; E)$. Eqs. (4.3.12) and (4.3.13) can be written in terms of the inelastic scattering rates as

\[ \frac{e n(r; E)}{\tau^e(r; E)} = I_n(r; E) + \frac{e\hbar}{2\pi} \int \frac{dr'}{\tau^e(r'; E)} \frac{|G^R(r, r'; E)|^2}{\tau^e(r; E)} \]  

\[ \frac{e p(r; E)}{\tau^e(r; E)} = I_p(r; E) + \frac{e\hbar}{2\pi} \int \frac{dr'}{\tau^e(r'; E)} \frac{|G^R(r, r'; E)|^2}{\tau^e(r; E)} \]  

We have added the terms $I_n(r; E)$ and $I_p(r; E)$ to account for the external source; the net current $I(r; E)$ per unit volume per unit energy entering the structure from the external source is equal to their difference.

\[ I(r; E) = I_p(r; E) - I_n(r; E) \]  

To determine $I_n$ and $I_p$ individually we multiply eq. (4.3.23a) by $\tau^e(r; E)$ and eq. (4.3.23b) by $\tau^e(r; E)$ and add them. This yields (using eq. (4.2.19))

\[ e N_0(r; E) = I_n(r; E) \tau^e(r; E) + I_p(r; E) \tau^e(r; E) \]

\[ + \frac{e\hbar}{2\pi} \int \frac{dr'}{\tau^e(r'; E)} \frac{|G^R(r, r'; E)|^2}{\tau^e(r; E)} \]  

Comparing with eq. (4.3.14) we obtain,

\[ I_n(r; E) \tau^e(r; E) + I_p(r; E) \tau^e(r; E) = 0 \]  

From eqs. (4.3.24) and (4.3.26) we obtain (using eq. (4.2.19)),

\[ I_n(r; E) = -I(r; E) \tau^R(r; E)/\tau^e(r; E) \]  

\[ I_p(r; E) = I(r; E) \tau^R(r; E)/\tau^e(r; E) \]  

Substituting eqs. (4.3.27a,b) into eqs. (4.3.23a,b) we obtain the modified versions of eqs. (4.3.12) and (4.3.13).

\[ e n(r; E) = -I(r; E) \tau^R(r; E) + \frac{e\hbar}{2\pi} \int \frac{dr'}{\tau^e(r'; E)} \frac{|G^R(r, r'; E)|^2}{\tau^e(r; E)} \]  

\[ e p(r; E) = I(r; E) \tau^R(r; E) + \frac{e\hbar}{2\pi} \int \frac{dr'}{\tau^e(r'; E)} \frac{|G^R(r, r'; E)|^2}{\tau^e(r; E)} \]  

Substituting for $\tau^e(r'; E)$ from eq. (4.2.14b) and $\tau^e(r; E)$ from eq. (4.2.14a) we obtain the desired transport equation.
\[ n(r;E) = -\frac{1}{e} I(r;E) \tau^R(r;E) + \int dr' \int dE' K_n(r,r';E,E') n(r';E') \]  
(4.3.29a)

\[ p(r;E) = \frac{1}{e} I(r;E) \tau^R(r;E) + \int dr' \int dE' K_p(r,r';E,E') p(r';E') \]  
(4.3.29b)

where

\[ K_n(r,r';E,E') = |G^R(r,r';E)|^2 F(r';E-E') \]  
(4.3.30a)

\[ K_p(r,r';E,E') = |G^R(r,r';E)|^2 F(r';E'-E) \]  
(4.3.30b)

Either eq. (4.3.29a) or eq. (4.3.29b) may be solved to obtain the carrier density per unit energy. The two equations are clearly not independent since, by adding them, we obtain the identity derived earlier (eq. (4.3.14)).

**Physical Interpretation:** The transport equations (eq. (4.3.29a) or (4.3.29b)) can be understood in terms of a simple physical picture, if we specialize to a dilute Boltzmann gas of electrons or holes [66,67]. For a dilute electron gas, \( \tau^c \) is much shorter than \( \tau^x \) since holes are far more numerous than electrons (eqs. (4.2.14a,b)). We can then write, using eq. (4.2.19),

\[ \tau^c = \tau^R \ll \tau^x \]  
(4.3.31)

The transport equation, eq. (4.3.29a) can then be written as

\[ i_S(r;E) = I(r;E) + \int dr' \int dE' K(r,r';E,E') i_S(r';E') \]  
(4.3.32)

where

\[ K(r,r';E,E') = K_n(r,r';E,E') \frac{\tau^x(r';E')}{\tau^x(r;E)} \]  
(4.3.33)

and the scattering current \( i_S(r;E) \) is proportional to the rate at which electrons are scattered out of \( (r;E) \).

\[ i_S(r;E) = -e \frac{n(r;E)}{\tau^x(r;E)} \]  
(4.3.34)

The kernel \( K(r,r';E,E') \) in eq. (4.3.32) can be written as the product of two factors, using eqs. (4.3.30a) and (4.3.33).

\[ K(r,r';E,E') = P(r,r';E) P_S(r';E,E') \]  
(4.3.35)

where

\[ P(r,r';E) = \frac{n}{2\pi} \frac{|G^R(r,r';E)|^2}{N_0(r';E) \tau^x(r;E)} \]  
(4.3.36)

\[ P_S(r';E,E') = \frac{2\pi}{n} F(r';E-E') N_0(r';E) \tau^x(r';E') \]  
(4.3.37)

As we will show shortly these two factors \( P \) and \( P_S \) have simple physical interpretations:

(a) \( P_S(r';E,E') \) is the fraction of electrons inelastically scattered at \( r' \) from an initial energy \( E' \) that acquire a final energy \( E \).
(b) $P(r, r'; E)$ is the fraction of electrons "injected" at $r'$ with energy $E$ (by an inelastic scattering process) that suffer their very next inelastic scattering event at $r$.

This interpretation provides an intuitive picture of quantum transport as a diffusion process in $(r; E)$ (Fig. 4.3.1a). The rate at which electrons are inelastically scattered at $(r'; E')$ is proportional to $i_S(r'; E')$. A fraction $P_S(r'; E, E')$ of $i_S(r'; E')$ acquires a final energy $E$. A fraction $P(r, r'; E)$ of $P_S(r'; E, E')$ reaches position $r$ before inelastically scattering again. The transport equation (eq. (4.3.32)) simply balances the in-flow and out-flow at $(r; E)$, as illustrated in Fig. 4.3.1(b). The kernel $K(r, r'; E, E')$, which is the product of $P$ and $P_S$, can be viewed as a transfer function from one inelastic scattering event at $(r'; E')$ to the next one at $(r; E)$.

Finally let us justify the above interpretations of the functions $P$ and $P_S$. First let us consider $P_S$. The integrand in eq. (4.2.14a) is interpreted as the rate at which an electron initially at $(r; E)$ is inelastically scattered to $(r; E')$.

$$S(r; E', E) = \frac{2\pi}{\hbar} F(r; E'-E) P(r; E')$$

$$= \frac{2\pi}{\hbar} F(r; E'-E) N_0(r; E')$$ (4.3.38)

since for a dilute electron gas $n << p = N_0$. Using eqs. (4.2.14a) and (4.3.38) we can rewrite $P_S$ from eq. (4.3.37) as

$$P_S(r'; E, E') = \frac{S(r'; E, E')}{\int dE S(r'; E, E')}$$ (4.3.39)

Eq. (4.3.39) clearly justifies the interpretation of $P_S$ in the last paragraph. It is also apparent that

$$\int dE P_S(r'; E, E') = 1$$ (4.3.40)

Next we consider the function $P(r, r'; E)$. The retarded Green function $G^R$ is obtained from the "Schrödinger" equation including an optical potential (eq. (4.3.10)). Since for a dilute gas $\tau^R = \tau^\sigma$ we rewrite eq. (4.3.10) as

$$\left[ E - H_0(r) - \sigma^R(r; E) + \frac{i\hbar}{2\tau^\sigma(r; E)} \right] G^R(r, r'; E) = \delta(r-r')$$ (4.3.41)

Eq. (4.3.41) describes the propagation of electrons injected at $r'$. The imaginary potential causes the electrons to disappear due to inelastic scattering to other energies. As we have shown earlier (eq. (4.3.19)),

$$\int dr \frac{|G^R(r, r'; E)|^2}{\tau^\sigma(r; E)} = \frac{2\pi}{\hbar} N_0(r'; E)$$ (4.3.42)

The integrand on the left is the rate at which electrons are inelastically scattered in a volume $dr$. The total rate integrated over all volume equals the rate at which electrons are injected at $r'$ which is the term on the right (Fig. 4.3.2). Hence the fraction of electrons injected at $r'$ (by inelastic scattering) that are inelastically scattered next at $r$ is given by
Fig. 4.3.1:

(a) Physical picture of the transport process.

(b) Schematic diagram illustrating the different terms in eq. (4.3.32).
Fig. 4.3.2: Sketch of the probability density $|G^R(r,r';E)|^2$ calculated from eq. (4.3.41). The index $E$ is dropped for convenience.
which is precisely the function $P(r, r'; E)$ (eq. (4.3.36)). It is also apparent that

$$\int dr \, P(r, r'; E) = 1$$

(4.4.33)

A similar physical picture is obtained for a dilute hole gas with $n << p = N_0$; eq. (4.3.31) is then replaced by $\tau^* = \tau^R << \tau$. For a general Fermi gas it may be possible to develop a simple physical picture in terms of quasiparticles composed of an electron surrounded by its exchange hole; however, the details are not clear to us at this time.

4.4 Local Thermodynamic Equilibrium

In this section, we consider solutions of the transport equation at or near equilibrium. For small perturbations from equilibrium, we reformulate the transport equation, assuming that the distribution of electrons at any point can be characterized by a local chemical potential or quasi Fermi level. In this form, the equation bears a striking similarity to the Landauer-Büttiker formula. Each inelastic scatterer in our model acts as an independent reservoir in the Landauer model, so that our transport equation appears to be a simple extension of the Landauer-Büttiker formula to account for a continuous distribution of probes. A simple expression is derived for the two-probe conductance of any structure.

At equilibrium, the electron density should be given by the product of the density-of-states $N_0(r; E)$ (eq. (4.3.15)) and the Fermi-Dirac factor $f_0(E)$,

$$n_{eq}(r; E) = N_0(r; E) f_0(E)$$

(4.4.1)

Extending this relationship to non-equilibrium situations, we define a "distribution function" $f(r; E)$:

$$n(r; E) = N_0(r; E) f(r; E)$$

(4.4.2a)

The hole density is given by

$$p(r; E) = N_0(r; E) [1 - f(r; E)]$$

(4.4.2b)

The transport equation was formulated above in terms of the electron density $n(r; E)$ (eq. (4.3.29a)) and the hole density $p(r; E)$. Alternatively, we could formulate it in terms of the distribution function $f(r; E)$. This may be more convenient in obtaining numerical solutions since $f(r; E)$, being spatially constant in equilibrium, is expected to be a spatially smoother function than $n(r; E)$, even away from equilibrium.

We emphasize that the distribution function $f(r; E)$ as defined above is not a semiclassical concept but a well-defined quantum mechanical quantity. There is no violation of the uncertainty principle, since a knowledge of the electron's energy is conjugate to the time coordinate, not the position coordinate. This is in contrast to a semiclassical distribution such as $f(r; k)$, used in the Boltzmann transport equation.
Assuming local thermodynamic equilibrium, the distribution function can be written in the form of a Fermi-Dirac function

$$f(r; E) = \frac{1}{e^{(E - \mu(r))/k_B T} + 1}$$  \hspace{1cm} (4.4.3)$$

where $\mu(r)$ is the local chemical potential which is constant everywhere in the structure at equilibrium. We first note that if the distribution function $f(r; E)$ is described by a local chemical potential $\mu(r)$ as shown in eq. (4.4.3) then

$$f(r; E') [1 - f(r; E)] = f(r; E) [1 - f(r; E')] e^{(E - E')/k_B T}$$  \hspace{1cm} (4.4.4)$$

In terms of electron and hole densities we can write (we are dropping the argument $r$ for convenience, since all quantities are evaluated at the same point in space).

$$-n(E') p(E) = n(E) p(E') e^{(E - E')/k_B T}$$  \hspace{1cm} (4.4.5)$$

Again, from eqs. (4.2.15) and (4.2.9) we have,

$$F(E' - E) = F(E - E') e^{(E - E')/k_B T}$$  \hspace{1cm} (4.4.6)$$

From eqs. (4.4.5) and (4.4.6),

$$n(E) F(E' - E) p(E') = p(E) F(E - E') n(E')$$  \hspace{1cm} (4.4.7)$$

Integrating both sides over $E'$ and using eqs. (4.2.14a,b) we obtain (restoring the argument $r$)

$$\frac{n(r; E)}{\tau^c(r; E)} = \frac{p(r; E)}{\tau^o(r; E)}$$  \hspace{1cm} (4.4.8)$$

Using eqs. (4.4.8) and (4.2.19) we can write

$$\tau^R(r; E) = \tau^c(r; E) f(r; E) = \tau^o(r; E) [1 - f(r; E)]$$  \hspace{1cm} (4.4.9)$$

Note that eqs. (4.4.8) and (4.4.9) are only valid under conditions of local thermodynamic equilibrium. As eq. (4.4.8) shows, the rate at which electrons are scattered out inelastically at any point is then exactly balanced by the rate at which holes are scattered out, that is, the rate at which electrons are scattered in. This local balance is not expected to hold in general when the system is driven far from equilibrium.

Subtracting eq. (4.3.23b) from eq. (4.3.23a) and using eqs. (4.4.8) and (4.3.24) we obtain

$$I(r; E) = \frac{e \hbar}{2 \pi} \int dr' \left| G^R(r, r'; E) \right|^2 \left[ \frac{1}{\tau^c(r'; E) \tau^o(r'; E)} - \frac{1}{\tau^c(r; E) \tau^o(r; E)} \right]$$  \hspace{1cm} (4.4.10)$$

Using eq. (4.4.9) we rewrite eq. (4.4.10) as,

$$I(r; E) = \frac{e}{\hbar} \int dr' T(r, r'; E) \{ f(r'; E) - f(r; E) \}$$  \hspace{1cm} (4.4.11)$$

where
satisfies eq. (4.4.11) with the external current \( I(r; E) \) set equal to zero. It is thus fairly straightforward to calculate the equilibrium density of electrons in any structure. We first calculate the eigenfunctions \( \phi_M(r) \) and eigenenergies \( \varepsilon_M \) for the elastic part of the Hamiltonian \( H_q \); these are then used to obtain the density of states \( N_0(r; E) \) from eq. (4.3.15). The chemical potential \( \mu_0 \) appearing in the Fermi-Dirac function \( f_0(E) \) (eq. (4.4.13)) is adjusted to obtain the correct average density of electrons, according to eq. (4.4.1). In general, any uncompensated space-charge must be accounted for by performing an iterative solution for the electron density and the electrostatic potential. The electron density \( n(r) \) should be inserted into the Poisson equation to obtain a corrected potential; the eigenfunctions \( \phi_M(r) \) and the eigenenergies \( \varepsilon_M \) should then be recalculated including this potential, and iteration should continue until the solution is self-consistent.

In linear response theory we assume that the distribution function \( f(r; E) \) deviates only slightly from the equilibrium distribution \( f_0(E) \), so that we can expand \( f(r; E) \) in a Taylor series about \( \mu = \mu_0 \). Noting that \( \partial f_0/\partial E = -e \partial / \partial E \), we obtain

\[
f(r; E) = f_0(E) + \left[ -\frac{\partial f_0}{\partial E} \right] e [\mu(r) - \mu_0]
\]

Substituting eq. (4.4.14) into eq. (4.4.11), we obtain

\[
I(r; E) = \frac{e^2}{h} \int dr' \left[ \frac{-\partial f_0}{\partial E} \right] T(r, r'; E) \{ \mu(r') - \mu(r) \}
\]

Integrating over \( E \),

\[
I(r) = \frac{e^2}{h} \int dr' T_0(r, r') \{ \mu(r') - \mu(r) \}
\]

where

\[
T_0(r, r') = \int dE \left[ \frac{-\partial f_0}{\partial E} \right] T(r, r'; E)
\]

\( I(r) \) is the total external current integrated over all energies. We can rewrite eq. (4.4.16) in the form
by noting that
\[ \int dr' \{ T_0(r', r) - T_0(r, r') \} = 0 \]  
(4.4.19)

Eq. (4.4.19) is obtained from the definition of \( T_0(r, r') \) (eqs. (4.4.12), (4.4.17)) using the result we derived earlier (eq. (4.3.22)).

Eq. (4.4.18) can be viewed as a generalization of the Landauer-Büttiker formula (eq. (4.1.3)) to a continuous distribution of probes. The coefficients \( T_0(r, r') \) have the same symmetry properties in a magnetic field \( H \) as the coefficients \( T_{ij} \) \[19,26\]. Namely,

\[ T_0(r, r') \]  
(4.4.20a)

This is apparent from the definition of \( T_0(r, r') \) (eqs. (4.4.17), (4.4.12)) and the symmetry property of the Green function (eq. (4.3.21)). Also, eq. (4.4.19) is the counterpart of the relationship \( \sum_j \{ (T_0)_{ij} - (T_0)_{ji} \} = 0 \) \[19\].

Space-charge effects: In deriving eq. (4.4.15) from eq. (4.4.11) we have implicitly assumed that when we drive the system slightly away from equilibrium, the distribution function \( f(r; E) \) deviates from the equilibrium value of \( f_0(E) \), but the coefficients \( T(r', r; E) \) remain fixed. Actually, the coefficients \( T(r', r; E) \) will change because corrections to the electrostatic potential will change the Green function \( G^R(r', r; E) \), as well as the inelastic scattering times \( \tau^R(r; E) \). In considering the variation \( \delta I \), we have accounted for one term,

\[ \frac{e}{h} \int dr' \int dE \delta f_0(E) \{ T(r', r; E) - T(r, r'; E) \} \]

It would seem that we should also have a term of the form

\[ \frac{e}{h} \int dr' \int dE \delta f_0(E) \{ \delta T(r', r; E) - \delta T(r, r'; E) \} \]

where \( \delta T \) is the change in the coefficient \( T \). This term is zero, however, because of the relation (eq. (4.4.18b)) that must be satisfied by \( T(r', r; E) \). Consequently, in linear response theory we can use the coefficients \( T(r', r; E) \) obtained (self-consistently) under equilibrium conditions, and ignore corrections due to the modification of the electrostatic potential under an applied bias. This, however, may not be true if there are sharp resonances in \( T \); second-order terms (\( \sim \delta T \delta f_0 \)) may not be negligible in that case.

Power Dissipation and Circulating Currents: In general, we can solve eq. (4.4.18) for the potential distribution \( \mu(r) \) in any structure. At equilibrium, \( \mu(r) \) is equal to a constant \( \mu_0 \), and \( I(r) \) is equal to zero. In the absence of magnetic fields (\( H = 0 \) in eq. (4.4.20b)), \( T_0(r', r) = T_0(r, r') \) so that at equilibrium the integrand in eq. (4.4.19) is zero, and there is
detailed balance between any two points \( r' \) and \( r \). But in the presence of a magnetic field this is not true. There can be circulating currents, even at equilibrium. However, the net current out of any point is zero, as evident from eq. (4.4.19). Any outflow in one direction is balanced by an inflow from another.

At equilibrium, these circulating currents dissipate no power. To show this, we express the total dissipated power as,

\[
P_0 = -\int dr \mu(r) I(r)
\]

\[
= \frac{e^2}{h} \int dr \int dr' \{ T_0(r',r) \mu(r) - T_0(r,r') \mu(r') \} \mu(r)
\]

(4.4.21)

Upon interchanging the roles of \( r \) and \( r' \) in eq. (4.4.21), we have

\[
P_0 = -\frac{e^2}{h} \int dr \int dr' \{ T_0(r',r) \mu(r) - T_0(r,r') \mu(r') \} \mu(r')
\]

(4.4.22)

The total dissipated power can be written as half the sum of eqs. (4.4.21) and (4.4.22),

\[
P_0 = \frac{e^2}{2h} \int dr \int dr' \left[ T_0(r,r') \mu(r) - T_0(r,r') \mu(r') \right] \left[ \mu(r) - \mu(r') \right]
\]

(4.4.23)

Thus, in equilibrium \( \mu(r) \) is a constant \( \mu_0 \), and the dissipated power is zero. We have changed nothing if we write eq. (4.4.23) as,

\[
P_0 = \frac{e^2}{2h} \int dr \int dr' \left[ T_0(r,r') \left[ \mu(r) - \mu(r') \right] + T_A(r,r') \left[ \mu^2(r) - \mu^2(r') \right] \right]
\]

(4.4.24)

where

\[
T_0(r,r') = \frac{1}{2} \left[ T_0(r',r) + T_0(r,r') \right]
\]

(4.4.25a)

\[
T_A(r,r') = \frac{1}{2} \left[ T_0(r',r) - T_0(r,r') \right]
\]

(4.4.25b)

The net power dissipation due to the antisymmetric kernel is zero:

\[
\frac{e^2}{2h} \int dr \int dr' T_A(r',r) \left[ \mu^2(r) - \mu^2(r') \right] = 0
\]

(4.4.26)

This result follows readily if we note that from eqs. (4.4.19) and (4.4.25b)

\[
\int dr' T_A(r,r') = \int dr T_A(r,r') = 0
\]

(4.4.27)

Hence, the power \( P_0 \) dissipated in the structure arises solely from the first term in eq. (4.4.24).

\[
P_0 = \frac{e^2}{2h} \int dr \int dr' T_0(r,r') \left[ \mu(r) - \mu(r') \right]^2
\]

(4.4.28)

From the point of view of power dissipation we can represent any structure by a continuous network of conductors; any two volume elements \( dr' \) and \( dr \) are connected by a conductance equal to \( (e^2/h) T_S(r,r') dr'dr \) (Fig. 4.4.1). If we have two external probes with a potential difference
From the point of view of power dissipation, any structure can be represented by a continuous network of conductors; any two volume elements $dr'$ and $dr$ are connected by a conductance equal to $(e^2/h) T_S(r, r') dr'dr$.
Δμ between them, the conductance g₀ seen from the terminals can be obtained by equating the total power dissipated in the network to g₀ (Δμ)². From eq. (4.4.28) we obtain the following expression for the two-probe conductance g₀.

\[ g₀ = \frac{e^2}{2h} \int dr \int dr' T_S(r,r') \left( \frac{\mu(r) - \mu(r')}{Δμ} \right)^2 \]  (4.4.29)

One may adopt a variational approach to calculating μ(r): choose a trial function and then minimize the power dissipated.

### 4.5 Relationship to Classical Brownian Motion

The transport equation discussed in this paper (eqs. (4.3.29a,b)) can be viewed as describing a random diffusion process in (r; E), where the kernel \( K_n(r,r'; E,E') \) represents the probability of “hopping” from \( (r'; E') \) to \( (r; E) \). Thus, the transport process can be viewed as classical Brownian motion; the only quantum mechanical input is in computing the kernel. In specializing to linear response (Section 4.4), we have integrated over energy, so that we are left with a diffusion process in real space only. In this section, we will show that the linear-response transport equation (eq. (4.4.18)) reduces to a drift-diffusion equation, if we assume slowly varying ensemble-averaged quantities.

To show this we rewrite eq. (4.4.11) in terms of the electron density \( n(r,E) \) and integrate over all energy to obtain

\[ I(r) = e \int dE \int dr' \left\{ v(r,r'; E) n(r'; E) - v(r',r; E) n(r; E) \right\} \]  (4.5.1)

where

\[ v(r,r'; E) = T(r,r'; E) / h N₀(r'; E) \]  (4.5.2)

If we assume that conduction takes place through a narrow band of energies \( E_F - \varepsilon < E < E_F + \varepsilon \) over which \( v(r,r'; E) \) is essentially constant, then we can rewrite eq. (4.5.1) as

\[ I(r) = e \int dr' \left\{ v(r,r') \delta n(r') - v(r',r) \delta n(r) \right\} \]  (4.5.3)

where

\[ v(r,r') = v(r,r') \bigg|_{E=E_F} \]  (4.5.4a)

\[ \delta n(r) = \int_{E_F-\varepsilon}^{E_F+\varepsilon} dE n(r; E) \]  (4.5.4b)

Eq. (4.5.3) has a simple physical interpretation. \( v(r,r') \, dr' \) tells us the fraction of electrons per unit time that “hop” from \( r' \) to \( r \). The first term on the right-hand side of eq. (4.5.3) is the total number of electrons hopping per unit time out of the volume element \( dr' \), while the second term is the number of electrons hopping per unit time into the volume element \( dr' \). The net hopping frequency \( V_q \) is given by
Quantum transport is thus much like classical Brownian motion with a distribution of hopping lengths $v(r, r')$ that is determined quantum mechanically. We can rewrite eq. (4.5.3) in the following form, noting that the current $I(r)$ flowing into the structure through external probes is equal to the negative of the divergence of the current density within the structure.

$$v_0 = \int dr' v(r', r)$$  \hspace{1cm} (4.5.5)

The obvious question to ask is under what conditions does eq. (4.5.6) reduce to the drift-diffusion equation

$$-D_{ij} \nabla_i \nabla_j n + v_d j_i \nabla_j n = \frac{1}{e} \nabla_i J_i$$  \hspace{1cm} (4.5.8)

Here $D$ is the diffusion coefficient (tensor), $v_d$ is the drift velocity (vector) and summation over repeated indices is implied ($x$, $y$ and $z$). Note that in equilibrium, eq. (4.5.8) becomes

$$-D_{ij} \nabla_i n_0 + v_d j_i n_0 = 0$$  \hspace{1cm} (4.5.9)

so that by subtracting eq. (4.5.9) from (4.5.8) we can formulate the drift-diffusion equation in terms of $\delta n$,

$$-D_{ij} \nabla_i \delta n + v_d j_i \delta n = \frac{1}{e} \nabla_i J_i$$  \hspace{1cm} (4.5.10)

To get from eq. (4.5.6) to eq. (4.5.10) we first assume that we are dealing with ensemble-averaged quantities (denoted by a bar on top) so that the coefficient $\zeta(r, r')$ depends only on the difference coordinate.

$$\zeta(r, r') = v_0 \delta(r-r') - v(r, r')$$  \hspace{1cm} (4.5.7)

Next we assume that $\zeta \delta n = \zeta \delta n$ so that eq. (4.5.4) becomes a convolution integral.

$$\int dr' \zeta(r, r') \delta n(r') = -\frac{1}{e} \nabla \cdot J$$  \hspace{1cm} (4.5.11)

Fourier transforming eq. (4.5.12) we obtain

$$\zeta(q) \delta n(q) = -\frac{i}{e} q_i J_i$$  \hspace{1cm} (4.5.13)

Now we expand $\zeta(q)$ in a Taylor series up to the quadratic term.

$$\zeta(q) = \zeta(0) - i q_j v_d j_i - q_j \zeta D_{ij}$$  \hspace{1cm} (4.5.14)

The coefficients in this expansion are obtained readily from the moments of the function $\zeta(r-r')$ in real space.
\( \zeta(0) = \int d\rho \, \zeta(\rho) \)  
(4.5.15)

\( \nu_{d_j} = \int d\rho \, \rho_j \, \overline{v}(\rho) \)  
(4.5.16)

\( D_{ij} = \frac{1}{2} \int d\rho \, \rho_i \, \rho_j \, \overline{v}(\rho) \)  
(4.5.17)

where we have written \( \rho \) for \( r-r' \). Using eq. (4.5.5), (4.5.13) and (4.5.15) it is easy to show that \( \zeta(0) = 0 \). Hence, inserting eq. (4.5.14) in eq. (4.5.13),

\[ (i q_j \, \nu_{d_j} + q_i \, q_j \, D_{ij}) \, \delta n(q) = \frac{1}{e} q_i \, J_i \]  
(4.5.18)

Fourier transforming to real space we obtain the drift-diffusion equation (eq. 4(4.5.10)).

Eqs. (4.5.16) and (4.5.17) may be used to compute the drift velocity and diffusion coefficient from the ensemble-averaged hopping function \( \overline{v}(r-r') \). It should be noted that a number of approximations have been made in deriving eqs. (4.5.16) and (4.5.17). We feel that in general eq. (4.4.15) is a better starting point for the computation of sample-specific properties that vary rapidly in space. However, for slowly varying ensemble-averaged properties eqs. (4.5.16) and (4.5.17) are more convenient.

4.6 Summary

Starting from a model Hamiltonian, we have derived a simple transport equation for the electron density \( n(r; E) \), or for the hole density \( p(r; E) \), in an arbitrary structure.

\[ n(r; E) = - \frac{1}{e} I(r; E) \, \tau^R(r; E) + \int dr' \int dE' \, K_n(r,r'; E,E') \, n(r'; E') \]  
(4.6.1a)

\[ p(r; E) = \frac{1}{e} I(r; E) \, \tau^R(r; E) + \int dr' \int dE' \, K_p(r,r'; E,E') \, p(r'; E') \]  
(4.6.1b)

Our analysis was greatly simplified by assuming that electrons interact with a bath of oscillators through a delta-potential. Each inelastic scattering event involves the wavefunction at a single point, so that spatial correlations of the wavefunction are unnecessary. In our model, inelastic scattering is treated in the "golden rule" approximation. Inelastic events are weak and infrequent, and after each event, an electron’s phase-memory is completely destroyed. However, elastic scattering due to impurities, device geometry, etc., is treated exactly, so that interference effects can influence transport on a scale shorter than the inelastic mean-free path.

Transport can be viewed as a diffusion process in \( (r; E) \). Each inelastic scattering event causes a change in the energy \( E \) of the electron, and the elastic propagation in between two events causes a change in the position \( r \). The kernel \( K_n(r,r'; E,E') \) (or equivalently, \( K_p(r,r'; E,E') \)) simply represents the transfer function between two inelastic scattering events, the first at \( (r'; E') \) and the second at \( (r; E) \).
If we assume that inelastic scattering is strong enough (or that the applied bias is small enough) to maintain local thermodynamic equilibrium everywhere, then eqs. (4.6.1) can be simplified. In this linear response regime, we assume that electrons can be characterized by a Fermi-Dirac distribution with a local chemical potential,

$$n(r; E) = N_0(r; E) \frac{1}{e^{(E - e\mu(r)) / k_B T} + 1}$$

(4.6.2)

Using this form for $n(r; E)$, eq. (4.6.1a) can be rewritten as,

$$I(r) = \frac{e^2}{\hbar} \int dr' \{ T_0(r, r') \mu(r') - T_0(r', r) \mu(r) \}$$

(4.6.3)

This equation bears a striking resemblance to the Landauer-Büttiker formula (eq. (4.1.3)). In fact, it appears that we have simply generalized eq. (4.1.3) to include a continuous distribution of probes. In this respect, our model justifies a very intuitive picture of the transport process: Each inelastic scattering event appears to be an exit into a Landauer-type reservoir, followed by reinjection into the main structure. Using this physical picture and the Kubo formula for the conductivity, we can derive an expression for the kernel $T_0(r, r')$ which agrees with the result obtained by simplifying eq. (4.6.1a) directly; this derivation is presented in the appendix.

If the chemical potential is slowly varying, eq. (4.6.3) can be simplified to a drift-diffusion equation. This was shown by writing eq. (4.6.3) in terms of the change in the electron density away from equilibrium ($\delta n = n - n_0$),

$$I(r) = e \int dr' \{ n(r; E') - n(r', E) \}$$

(4.6.4)

The kernel $v(r, r')$ for this form of the integral equation can be interpreted as the fraction of electrons per unit time, per unit volume, that "hop" from position $r'$ to position $r$. Each "hop" begins and ends with an inelastic scattering event. In an ensemble-average sense, the "hopping distribution" $v(r, r')$ should depend only on the distance between inelastic events $\rho = |r - r'|$. The moments of the ensemble-averaged hopping distribution $\bar{v}(\rho)$ define the coefficients of the drift-diffusion equation:

$$v_{ij} = \int d\rho \rho_j \bar{v}(\rho)$$

(4.6.5a)

$$D_{ij} = \frac{1}{2} \int d\rho \rho_i \rho_j \bar{v}(\rho)$$

(4.6.5b)

Although eqs. (4.6.5a,b) are convenient for obtaining ensemble-averaged properties, we feel that in general eqs. (4.6.1a,b) or eq. (4.6.3) are a better starting point for sample-specific solutions.

The first step in the solution of eqs. (4.6.1a,b) or eq. (4.6.3) is an analysis of the equilibrium state. Because the quantities $n(r; E)$, $V(r)$ and $\tau^R(r; E)$ are interrelated, they must be determined in an iterative manner, as illustrated in Fig. 4.6.1. The process begins with an initial guess for $V(r)$ and $\tau^R(r; E)$, from which the density of states $N_0(r; E)$ is determined. The electron density $n(r)$ is then computed by selecting the proper Fermi level $E_F$ so that the average...
Fig. 4.6.1: Flow chart outlining the process of obtaining a self-consistent solution in equilibrium.
electron density matches the average doping density. New guesses for $\tau^R(r;E)$ and $V(r)$ are calculated based upon the solution for the electron density, and the process continues until each of the quantities converges to a final solution.

The solution of eqs. (4.6.1a,b), outlined in Fig. 4.6.2, is also iterative. An analysis of the equilibrium state provides the initial guesses for both $V(r)$ and $\tau^R(r;E)$. An initial guess of the distribution function $f(r;E)$ is constructed by assuming a Fermi-Dirac distribution (eq. (4.4.3)) at all points, and then guessing the form of the local chemical potential $\mu(r)$. In the contact regions of the device, the distribution function $f(r;E)$ remains fixed, and the external current $I(r;E)$ is determined from the transport equation (eq. (4.6.1a)). At all other positions $I(r;E) = 0$, and instead the distribution function $f(r;E)$ is determined. Using the new solution for $f(r;E)$, guesses for $\tau^R(r;E)$ and $V(r)$ are updated, and the process is repeated until convergence is achieved.

In the linear response regime, the solution of the transport equation (eq. (4.6.3)) is greatly simplified. After performing the equilibrium solution, all quantities (i.e., $\tau^R$, $N_0$ and the kernel $T_0$) are simply evaluated at the Fermi energy $E_F$. What remains is the solution of eq. (4.6.3), which can be set up numerically as a simple matrix equation. At each contact node, the unknown is the external current $I(r)$; at all other nodes, the chemical potential $\mu(r)$ must be determined. For simulations with large numbers of nodes, it may be more convenient to solve the matrix equation in a Gauss-Siedel fashion, as illustrated in Fig. 4.6.3. An initial guess of $\mu(r)$ is improved upon by repeatedly solving the transport equation (eq. (4.6.3)) until the solution for $\mu(r)$ has converged.

In this work, we have restricted ourselves to steady-state transport and neglected many-body effects beyond the Hartree term. We believe that it should be possible in the future to extend the work, to remove these restrictions. However, our assumption of delta interaction potentials is essential in obtaining a simple transport equation that involves only the electron density, and not the spatial correlations of the wavefunction. We believe that the simplicity of the linear-response transport equation (eq. (4.6.3)) will make it feasible to obtain numerical solutions for specific mesostructures, and thereby quantitatively answer some of the fundamental questions of quantum transport [34]. Also, by comparing the predictions of our model with experiment, it should be possible to identify new phenomena arising from correlations between inelastic scatterers, and to shed light on the microscopic origin of irreversibility.

ACKNOWLEDGMENTS

The authors would like to acknowledge helpful discussions with Paul Muzikar. This work was supported by the Semiconductor Research Corporation (Contract no. 88-SJ-089) and the National Science Foundation (Grant no. ECS-83-51-036).
Fig. 4.6.2: Flow chart outlining the general solution of the transport equation.
Appendix: Derivation of the Kernel from the Kubo Conductivity

In Section 4.4 we assumed that, for small perturbations from equilibrium, inelastic scattering was strong enough to maintain local equilibrium everywhere in a sample. In this linear response regime, the transport equation was reduced to a simple form (eq. (4.4.18)) which resembles the Landauer-Büttiker formula (eq. (4.1.3)) generalized to include a continuous distribution of reservoirs. The purpose of this appendix is to reproduce our expression for the kernel $T(r,r'; E)$ of this integral equation starting from the Kubo formula for the conductivity using the Lee-Fisher formula.

In the Kubo formalism, the conductivity tensor $\sigma$ at a frequency $\omega$ is related to the current-current correlation function [13,14],

$$\sigma_{\alpha\beta}(r,r'; \omega) = [C_{\alpha\beta}(r,r'; \omega)]_{\alpha\beta} - \frac{n e^2}{m} \delta(r-r') \delta_{\alpha\beta}$$

where $n$ is the electron density, $m$ is the effective mass, $\delta_{\alpha\beta}$ is the Kronecker delta and the subscripts $\alpha, \beta$ run over $x, y$ and $z$. The current-current correlation function $C_{\alpha\beta}$ is defined as

$$C_{\alpha\beta}(r,r'; \omega) = \frac{i}{\hbar} \int_0^\infty dt e^{i\omega t} \langle J(r,t) J(r',0) - J(r',0) J(r,t) \rangle$$

where $J(r,t)$ is the current density operator in the Heisenberg picture, and $\langle \cdots \rangle$ denotes the ensemble-averaged expectation value. For convenience, we define each of the terms composing $C_{\alpha\beta}$:

$$C_1(r,r'; \omega) = \frac{i}{\hbar} \int_0^\infty dt e^{i\omega t} \langle J(r,t) J(r',0) \rangle$$

$$C_2(r,r'; \omega) = \frac{i}{\hbar} \int_0^\infty dt e^{i\omega t} \langle J(r',0) J(r,t) \rangle$$

The current density operator can be written as

$$J(r,t) = \sum_{N,M} J_{NM}(r) a_N^*(t) a_M(t)$$

where $J_{NM}(r)$ is defined in terms of the eigenfunctions $\phi_N(r)$ of $H_0$ (eq. (4.2.1)),

$$J_{NM}(r) = \frac{i e \hbar}{2m} [\nabla \phi_N]^* \phi_M - \phi_N^* \nabla \phi_M$$

and $a_N^*$, $a_N$ are the creation and annihilation operators for the eigenstate $N$. Substituting eq. (4.4.4) into eq. (4.4.3a),
Flow chart outlining the solution of the transport equation in the linear response regime.
\[
C_1(r,r';\omega) = \sum_{N,M,N',M'} J_{NM}(r) J_{N'M'}(r') \frac{i}{\hbar} \int_0^\infty dt e^{i\omega t} \left\langle a_N^+(t) a_M(t) a_{N'}(0) a_{M'}(0) \right\rangle \tag{4.6}
\]

Since \( N, M, N', M' \) are eigenstates, the expectation value on the right hand side is zero unless \( N' = M \) and \( M' = N \). Hence

\[
C_1(r,r';\omega) = \sum_{N,M} J_{NM}(r) J_{MN}(r') F_1(\omega) \tag{4.7a}
\]

where

\[
- F_1(\omega) = \frac{i}{\hbar} \int_0^\infty dt e^{i\omega t} \left\langle a_N^+(t) a_N(0) \right\rangle \left\langle a_M(t) a_M^+(0) \right\rangle
\]

\[
= -\frac{f_o(\epsilon_N) [1 - f_o(\epsilon_M)]}{\hbar \omega + \epsilon_N - \epsilon_M + i\eta} \tag{4.7b}
\]

\( \eta \) is an infinitesimal positive quantity (\( \eta = 0^+ \)). Similarly it can be shown that

\[
C_2(r,r';\omega) = \sum_{N,M} J_{NM}(r) J_{MN}(r') F_2(\omega) \tag{4.8a}
\]

where

\[
F_2(\omega) = -\frac{f_o(\epsilon_M) [1 - f_o(\epsilon_N)]}{\hbar \omega + \epsilon_N - \epsilon_M + i\eta} \tag{4.8b}
\]

Substituting eqs. (4.7a,b) and (4.8a,b) into eq. (4.2) we have

\[
C_{12}(r,r';\omega) = \sum_{N,M} J_{NM}(r) J_{MN}(r') F_{NM}(\omega) \tag{4.9a}
\]

where

\[
F_{NM}(\omega) = F_1 - F_2 = \frac{f_o(\epsilon_M) - f_o(\epsilon_N)}{\hbar \omega + \epsilon_N - \epsilon_M + i\eta} \tag{4.9b}
\]

We will now rewrite \( F_{NM}(\omega) \) in a somewhat different form by proceeding as follows:

\[
F_{NM}(\omega) = \int \frac{d\epsilon}{2\pi i} \left\{ \frac{1}{\epsilon + \eta} - \frac{1}{\epsilon - \eta} \right\}
\]

\[
= \int \frac{d\epsilon}{2\pi i} \left[ \frac{\delta(\epsilon - \epsilon_M + \hbar \omega)}{\epsilon - \epsilon_N - i\eta} - \frac{\delta(\epsilon - \epsilon_N)}{\epsilon - \epsilon_M + \hbar \omega + i\eta} \right] \tag{4.10}
\]

Using the relation

\[
\delta(x) = \frac{1}{2\pi i} \left[ \frac{1}{x - i\eta} - \frac{1}{x + i\eta} \right]
\]

we obtain from eq. (4.10),
\[ F_{NM}(\omega) = \int \frac{d\varepsilon}{2\pi i} \left[ -f_0(\varepsilon + \hbar \omega) G_N(\varepsilon) [G_M^R(\varepsilon + \hbar \omega) - G_M^R(\varepsilon + \hbar \omega)] \right. \]
\[ \left. - f_0(\varepsilon) G_M^R(\varepsilon + \hbar \omega) [G_N^R(\varepsilon) - G_N^R(\varepsilon)] \right] \quad (4.A.11) \]

where
\[ G_M^R(\varepsilon) = \frac{1}{\varepsilon - E - i\eta} \quad (4.A.12a) \]
\[ G_M^A(\varepsilon) = \frac{1}{\varepsilon - E + i\eta} \quad (4.A.12b) \]

For small \( \omega \), we can write eq. (4.A.11) as
\[ F_{NM}(\omega) = i\omega a_{NM} + b_{NM} \quad (4.A.13a) \]

where
\[ a_{NM} = \frac{\pi}{2\pi i} \int d\varepsilon \left[ -\frac{\partial f_0}{\partial \varepsilon} \right] G_M^R(\varepsilon) G_N^A(\varepsilon) \quad (4.A.13b) \]
\[ b_{NM} = \frac{1}{2\pi i} \int d\varepsilon f_0(\varepsilon) \left[ G_N^R(\varepsilon) G_N^R(\varepsilon) - G_N^R(\varepsilon) G_N^A(\varepsilon) \right] \quad (4.A.13c) \]

Using eqs. (4.A.9a) and (4.A.13a), we obtain from eq. (4.A.1)
\[ \sigma_{i\alpha}(r, r') = A + \frac{1}{n^2} (B - \frac{n e^2}{m} \delta(r - r') \delta_{i\alpha}) \quad (4.A.14a) \]

where
\[ A = \sum_{N,M} \left[ J_{NM}(r) \otimes J_{MN}(r') \right]_{i\alpha} a_{NM} \quad (4.A.14b) \]
\[ B = \sum_{N,M} \left[ J_{NM}(r) \otimes J_{MN}(r') \right]_{i\alpha} b_{NM} \quad (4.A.14c) \]

It can be shown that \( A \) and \( B \) are both real quantities so that the real part of the conductivity is simply equal to \( A \). From eqs. (4.A.13b) and (4.A.14b) we obtain a familiar expression for the Kubo conductivity [32],
\[ \sigma_0(r, r') = \int dE \left( -\frac{\partial f_0}{\partial E} \right) \sigma(r, r'; E) \quad (4.A.15a) \]
\[ \sigma(r, r'; E) = \frac{\pi}{2\pi} \sum_{N,M} \left[ J_{NM}(r) \otimes J_{MN}(r') \right] \left( G_M^R(\varepsilon) G_N^A(\varepsilon) \right) \quad (4.A.15b) \]

So far in this appendix, we have neglected inelastic scattering; the energy \( \eta \) in eqs. (4.A.12) is then a true infinitesimal. As we have seen in Section 4.3, inelastic scattering causes damping of the quasi particle propagator, which is described by including the optical potential
i\hbar/2\tau^R(r;E) in the defining equation for the Green function (eq. (4.3.10)); consequently, we modify eqs. (4.A.12) to

\[ G_M^R(e) = \frac{1}{e - e_M + i\hbar/2\tau_M} \]  
(4.A.16a)

\[ G_M^I(e) = \frac{1}{e - e_M - i\hbar/2\tau_M} \]  
(4.A.16b)

Since the inelastic scattering time \( \tau^R(r;E) \) is not a constant but can vary spatially, we have used different lifetimes \( \tau_M \) for the different eigenstates; in principle, these may be obtained from the imaginary parts of the eigenenergies \( e_M \) calculated using the Hamiltonian \((H_0 - i\hbar/2\tau^R(r;E))\). However, we assume that the imaginary potential is small enough that we can neglect any complication due to the non-orthogonality of the corresponding eigenfunctions \( \phi_M(r) \).

We obtain the conductivity which accounts for inelastic scattering by inserting eqs. (4.A.16a,b) into eq. (4.A.15b),

\[ \sigma_{\alpha\beta}(r,r';E) = \frac{\hbar}{2\pi} \sum_{N,M} \left[ J_{NM}(r) \otimes J_{MN}(r') \right]_{\alpha\beta} \]  
(4.A.17)

We can relate this expression to the kernel \( T(r,r';E) \) by recalling the Landauer interpretation of the linear-response transport equation: The kernel \( T(r,r';E) \) corresponds to the transmission coefficient between reservoirs connected to the infinitesimal volume elements at \( r \) and \( r' \). With this physical picture, we invoke the Lee-Fisher formula which links transmission to conductivity. In the limit of a continuous distribution of probes, each probe has an infinitesimal cross-section, so that from eq. (4.A.17),

\[ \frac{e^2}{\hbar} T(r,r';E) \, dr \, dr' = \frac{\hbar}{2\pi} \sum_{N,M} \frac{(J_{NM}(r) \cdot \hat{n}(r) \, dr) \cdot (J_{MN}(r') \cdot \hat{n}(r') \, dr')}{(E - e_M + i\hbar/2\tau_M)(E - e_N - i\hbar/2\tau_N)} \]  
(4.A.18)

where \( \hat{n}(r) \) is the unit vector normal to the probe at \( r \). But \( \mathbf{J} \cdot \hat{n} \) is the current entering the probe at \( r \) due to inelastic scattering. To determine this current, we rewrite eq. (4.3.29a) in the following form.

\[ \frac{en(r';E)}{\tau^R(r';E)} = -I(r;E) + \int dr' \int dE' \left[ K_n(r,r';E,E') \frac{\tau^R(r';E')}{\tau^R(r';E)} \right] \frac{en(r';E')}{\tau^R(r';E')} \]  
(4.A.19)

The right-hand side of this equation represents the electronic current flowing in to \( r \) from external sources (\(-I\)) and from other points \( r' \) within the structure. Therefore, the left-hand side \( en/\tau^R \) represents the current which exits \( r \) after inelastically scattering; this is precisely the current entering the probe in our physical picture. With this understanding of \( \mathbf{J} \cdot \hat{n} \), we write eq. (4.A.18) as
\[
T(r, r'; E) = \frac{\hbar^2}{e^2} \sum_{N,M} \frac{\rho_{NM}(r) \rho_{MN}(r')}{(E - \varepsilon_M + i\hbar/2\tau_M)(E - \varepsilon_N - i\hbar/2\tau_N)} \tau^R(r; E) \tau^R(r'; E)
\]

where \( \rho_{NM}(r) = \phi_N^*(r) \phi_M(r) \), so that

\[
T(r, r'; E) = \frac{\hbar^2}{\tau^R(r; E) \tau^R(r'; E)} \sum_N \frac{\phi_N^*(r) \phi_N(r')}{(E - \varepsilon_N - i\hbar/2\tau_N)} \sum_M \frac{\phi_M(r) \phi_M^*(r')}{(E - \varepsilon_M + i\hbar/2\tau_M)}
\]

We note that the Green function can be expanded in terms of the eigenstates \( \phi_M(r) \) as

\[
G(r, r'; E) = \sum_M \frac{\phi_M(r) \phi_M^*(r')}{(E - \varepsilon_M + i\hbar/2\tau_M)}
\]

Therefore, we have obtained our previous expression for the kernel (cf. eq. (4.4.12)):

\[
T(r, r'; E) = \frac{\hbar^2 |G(r, r'; E)|^2}{\tau^R(r; E) \tau^R(r'; E)}
\]
Chapter 4 References

Chapter 5

POTENTIAL VARIATION ACROSS AN OBSTACLE

5.1 Introduction

5.2 Theory

5.2.1. Linear Response Transport Equation
5.2.2. Specialization to Layered Geometries
5.2.3. Solution for the Green Function in 1-D

5.2.4. Integration Weights for Integrals Involving \( |G^{1-D}(z,z';E)|^2 \)

5.2.5. Overview of the Solution Technique

5.3 Results and Conclusions

5.4 Future Work

Appendix

References

In this section, we study the distribution of electrochemical potential across an obstacle. From a semiclassical viewpoint, the potential is expected to drop linearly across each resistive region of a device. In the absence of phase-breaking scattering, however, quantum mechanical effects can produce quite a different result: There can be fluctuations in the chemical potential, so that in some regions, the drop in chemical potential can oppose current flow. We present a quantum mechanical transport equation which describes electrons in contact with a thermal bath of inelastic scatterers. After explaining the numerical technique used to solve this equation, we present some solutions of the chemical potential. As the amount of inelastic scattering is increased, we observe a transition from quantum mechanical to semiclassical behavior. A similar transition is observed for high temperatures and for large samples.
5.1 Introduction

Until recently, voltage measurements in semiconductor devices were made with little regard for the detailed nature of the measurement probes. If a device is much larger than the phase-breaking length, the probes can indeed be treated as classical objects. However, in the regime of mesoscopic structures, whose relevant dimensions are on the order of the phase-breaking length, the classical voltage probe no longer exists. Because of its wave nature, an electron "sees" a large area of the sample, including the region up to a phase-breaking length inside of a measurement probe. Experiments have shown that, in the mesoscopic regime, transport can be strongly influenced by regions of the device outside of the classical current path [1]. Although this aspect of quantum transport has led to some interesting device applications [2], it is not at all desirable in the context of making accurate measurements.

The experiments of Benoit et al. [3] illustrate this point. Voltage measurements were made in a four-probe configuration along the length of a wire, with various spacings between voltage probes. In such an arrangement, the measured voltage fluctuates as the strength of an applied magnetic field is increased; this leads to fluctuations in the conductance, a phenomenon that has received widespread attention in the literature [4]. In a classical measurement, one would expect the voltage fluctuations to disappear as the voltage probes are moved closer together. However, the results of Benoit et al. show that the size of the fluctuations instead becomes constant. This result has been verified theoretically by Büttiker [5]. In effect, the separation between the two voltage probes cannot be reduced below a phase-breaking length, since an electron in one probe can "see" this far into another.

Because voltage probes can strongly influence experiments, it is important to understand how a given probe can be modeled. Much of the theoretical work on mesoscopic devices is done in the Landauer picture of transport [6], in which a particular sample is connected to idealized reservoirs of carriers by perfectly ordered leads. The reservoirs act like "black-body" sources of carriers: Electrons entering a perfect lead are completely absorbed by the reservoir, and after being thermalized, they are injected back into the sample according to the energy distribution of the reservoir. Local equilibrium is assumed to exist deep within each reservoir, so that the energy distribution of carriers has the form of the Fermi-Dirac factor characterized by a local chemical potential \( \mu_j \).

In many experiments, the voltage and current leads are formed in exactly the same manner. Büttiker reasoned that they should be treated on an equal theoretical footing, as reservoirs in the Landauer picture. The net current \( I_i \) flowing into the sample through lead \( i \) is determined by the Landauer-Büttiker formula [7],

\[
I_i = \frac{e^2}{h} \sum_j T_{ij} [\mu_j - \mu_i],
\]

where \( \mu_j \) is the chemical potential of reservoir \( j \), and \( T_{ij} \) represents the transmission coefficient from lead \( j \) to lead \( i \). Although this formula is useful for computing terminal characteristics, it says nothing about the interior of the sample. It is desirable, for instance, to have some
knowledge of how electrons are distributed inside the sample, so that the electron density can be
determined; one measure of the distribution of electrons is the local chemical potential. Following
the suggestion of Engquist and Anderson [8], we can measure the chemical potential at any
point within a sample by connecting that point to a reservoir. We require that the current \( I_i \) in
the connecting lead be zero, since the lead is not driven by an external source. The total current
within the lead comes from two sources of electrons: those leaving the sample with some arbit­
rary energy distribution, and those leaving the reservoir with the Fermi-Dirac distribution. In
equating these terms, therefore, we are fitting the energy distribution within the sample to the
Fermi-Dirac distribution. The fitting process underlies all discussions of the chemical potential;
the extent to which such discussions are meaningful is measured by the success of this fitting
process.

Now that we have a means of measuring the chemical potential, we consider how it might
be used to determine the conductance. Suppose that our voltage probes are weakly coupled to
the sample, so that the fraction of current diverted into the probes is small. Furthermore, sup­
pose that only the region between the voltage probes is disordered, and is characterized by some
transmission \( T \) and reflection \( R \). For this arrangement, Büttiker has shown [9] that the conduc­
tance obtained by solving Eq. (5.1.1) is,

\[
G = \frac{e^2}{h} \frac{T}{R},
\]

which is the formula for the conductance originally proposed by Landauer [10]. Clearly, this
formula is but a special solution of the more general Landauer-Büttiker formula (Eq. (5.1.1)); it
assumes that the voltage measurement is made immediately across the disordered region by
weakly coupled measurement probes. On the other hand, if we measure the voltage between the
current contacts, this leads to the conductance,

\[
G = \frac{e^2}{h} T.
\] (5.1.3)

Note that even if there is no scattering within the sample \( (T = 1) \), the conductance obtained from
Eq. (5.1.3) is finite. Imry has shown [6,11] that this arises from a contact resistance that exists
between each reservoir and its perfect lead. Within any sample, therefore, we expect the chemi­
cal potential to drop near the mouth of each reservoir, so that the potential difference across the
disordered interior is less than the potential difference applied at the reservoirs.

Büttiker has used this formalism to study the variation of chemical potential in single­
moded wires [12]. In his calculations, current flowing between two contacts impinged upon a
single potential barrier. The chemical potential was measured at each point by moving a
weakly coupled voltage probe along the length of the wire. Such a measurement has been real­
ized experimentally by Kirtley et al. [13]. A scanning tunneling microscope (STM) probe
served as the weakly coupled voltage probe which is scanned across the sample. The experi­
ments were done at room temperature in amorphous silicon, so that grain boundaries formed an
irregular two-dimensional array of potential barriers. When current was applied to the sample,
the chemical potential dropped in steps, remaining approximately constant across each grain,
and dropping sharply at the grain boundary. This is the expected result [6], since the chemical
potential should drop within each barrier, where the density of states is the least.

Working in the limit of low temperatures and narrow wires, however, Büttiker obtained
quite a different result. His chemical potential showed large oscillations as a function of dis­tance from the barrier. These oscillations arise from the interference of electron waves incident
on and reflected from the barrier. In obtaining this result, Büttiker neglected all phase-breaking
processes, so that the transmission coefficients $T_{ij}$ could be calculated using ordinary wave
mechanics. Of course, in the presence of phase-breaking scattering such oscillations would die­
out after a phase-breaking length. To account for this, Büttiker repeated his calculation in the
incoherent limit, thereby neglecting all interference effects. The result was dramatic: The
chemical potential obtained by averaging over the oscillations was quite different from that
obtained by neglecting interference effects altogether. This indicates that the extent to which
phase-breaking processes are taken into account can affect an analysis in a nontrivial way. It is
not enough to neglect such processes and assume that their only effect is to wash out the
coherent result to its average value.

Strictly speaking, the concept of a chemical potential is not meaningful unless some dissi­
pation exists within the sample itself (not just within the reservoirs). Recall that, in the process
of measuring the potential, we have assumed that the energy distribution of electrons is well
described by a Fermi-Dirac function. Sivan and Imry [14] have improved this fitting process by
allowing temperature to vary spatially, so that the fitted distribution is characterized by two
parameters: the local temperature, and the local chemical potential. In the absence of dissipa­tion,
however, even a small applied bias could heat electrons to the point that their energy distri­
bution is poorly described by any Fermi-Dirac function. If the measurement probes are only
weakly coupled, their reservoirs cannot be counted upon to equilibrate the sample. Even though
a chemical potential can be defined (or physically measured), in this case it is a poor characteri­
zation of the system.

A formalism which accounts for dissipative processes within the sample was recently pro­
posed by Datta [15]. Using the Keldysh formulation of nonequilibrium statistical mechanics,
Datta derived a transport equation which allows electrons to interact with a bath of oscillators
maintained in thermal equilibrium. To simplify the equation, the oscillators were assumed to
have a "point-size" spatial extent, and to have no definite phase relationship with one another.
The first assumption implies that inelastic scattering depends only on local properties; the
second, that an electron's phase memory is completely destroyed after each inelastic event.
Although this model may not represent true phonons (which have a finite spatial extent and
some phase correlations), it should still describe the physics of dissipative transport.

This formalism also leads to a natural definition for the chemical potential. In its most
general form, the transport equation is solved to obtain the energy distribution of electrons at
each point in space. Such a general solution would account for the effects of carrier heating
mentioned above. In the limit of linear response, it is assumed that there is enough dissipation
to maintain local equilibrium everywhere within the sample. The distribution is then fit to the Fermi-Dirac factor, and the transport equation is simplified to a form resembling the Landauer-Büttiker formula with a continuous distribution of probes [15]. Büttiker has shown [16] that each probe acts as an inelastic scatterer; it seems reasonable that the effects of distributed inelastic scattering could be modeled by including a continuous distribution of probes. Such a model, however, needs a more rigorous foundation. Datta’s approach provides this foundation, as well as a method for calculating the transmission coefficients \( T_{ij} \) in the presence of dissipation.

In this report, we study the variation of chemical potential across a single potential barrier in the presence of dissipation. The chemical potential is determined by solving the linear response transport equation proposed by Datta [15]. After reviewing the details of the solution technique, we consider the effects of inelastic scattering time, sample size and temperature on the chemical potential. We summarize by describing considerations for future work.

5.2 Theory

In its most general form, the transport equation derived by Datta [15] is used to determine a function \( f(r; E) \) called the "distribution function." By analogy to semiclassical dynamics, this function describes the probability that an electron at position \( r \) will be in a state with energy \( E \). Unlike semiclassical distribution functions which involve \( r \) and \( k \), this function involves only \( r \) and \( E \). There is no violation of the uncertainty principle, since \( r \) is conjugate to \( k \), not \( E \). The \( (r; E) \) representation arises naturally, since each oscillator has a specific frequency and (point-size) spatial coordinate. Each time an electron interacts with an oscillator, the \( (r; E) \) coordinate of the electron is measured, somewhat like an invisible man walking through a mine field: Between events (or explosions), the man could take any number of paths; however, the start and end of each path is marked by a precise measurement. With a continuous distribution of oscillators, we can form an exact knowledge of an electron’s position, at the expense of any knowledge about its momentum. Because of this, we concern ourselves only with quantities that can be obtained without reference to the momentum. In particular, the electron density at a position \( r \) is defined as,

\[
n(r) = \int dE \, N_0(r; E) \, f(r; E) ,
\]

where \( N_0(r; E) \) is the density of electronic states. By solving the transport equation, we determine \( f(r; E) \), which is then used to define all other quantities of interest.
5.2.1 Linear Response Transport Equation

In equilibrium, electrons are distributed in energy according to the Fermi-Dirac factor,

\[ f_0(E) = \frac{1}{e^{(E-e\mu_0)/k_BT} + 1} \]  

(5.2.2)

where \( \mu_0 \) is the equilibrium chemical potential. Under the application of bias, the distribution function must be determined by solving the general transport equation mentioned above. This task can be greatly simplified if we are merely interested in the linear response of the system. For small applied biases, and with sufficiently strong inelastic scattering, it is reasonable to assume that local equilibrium is maintained everywhere within the structure. If this is true, the distribution function is well described by a Fermi-Dirac function with a local chemical potential \( \mu(r) \),

\[ f(r; E) = \frac{1}{e^{(E-e\mu(r))/k_BT} + 1} \]  

(5.2.3)

Using this form for the distribution function, the transport equation can be greatly simplified. The resulting equation bears a striking similarity to the Landauer-Büttiker formula (Eq. (5.1.1)),

\[ I(r) = \frac{e^2}{h} \int dr' T_0(r, r') [\mu(r') - \mu(r)] \]  

(5.2.4)

where

\[ T_0(r, r') = \int dE \left[ -\frac{\partial f_0}{\partial E} \right] T(r, r'; E) \]  

(5.2.5a)

\[ T(r, r'; E) = \frac{\hbar^2}{\tau(r; E)} \frac{|G(r, r'; E)|^2}{\tau(r'; E)} \]  

(5.2.5b)

and \( G(r, r'; E) \) is the Green function of the Schrödinger equation,

\[ \left[ E + \frac{i\hbar}{2\tau(r; E)} + \frac{\hbar^2}{2m^*} \nabla^2 - eV(r) \right] G(r, r'; E) = \delta(r-r') \]  

(5.2.6)

which includes an imaginary potential \( i\hbar/2\tau \). Such an imaginary potential is common in the definition of Green functions, although it is usually taken to be infinitesimal. This potential, however, is finite. It is the imaginary part of the electron self-energy calculated from the microscopic model for inelastic scatterers. Although it can be computed from parameters in the microscopic model [15], we will treat it as an input parameter for all present calculations. It is interpreted as the lifetime that an electron remains in a coherent state (and is able to produce interference effects) before being inelastically scattered to a different state. The potential \( V(r) \) represents all sources of elastic scattering, such as impurities, defects, and the confining potential of the structure. Strictly speaking, \( V(r) \) also includes the electrostatic potential, which must be determined from a self-consistent solution of Schrödinger's and Poisson's equations. In the examples presented here, all such space-charge effects are ignored.
The analysis of any structure proceeds as follows. Assuming that $\tau(r;E)$ and $V(r)$ are known, we solve Eq. (5.2.6) to determine the Green function for a range of energies. Using these solutions, we compute the kernel $T_0(r,r')$ by performing an integration over energy, as shown in Eq. (5.2.5a). The next step is to solve the integral transport equation (Eq. (5.2.4)). Within the "contact" regions (i.e., those physically connected to external sources), the chemical potential $\mu(r)$ is fixed, and the current $I(r)$ must be determined. In all other regions, $I(r)$ is zero, and $\mu(r)$ must be determined. Thus, if we establish a grid of $N$ nodes, we have $N$ equations (one for each node $r_i$) and $N$ unknowns (which are a mixture of $I(r_i)$ and $\mu(r_i)$). We integrate the final solution of $I(r)$ over the contact regions to determine the terminal currents; this, together with the known applied bias, is used to define the conductance. Finally, we compute the electron density $n(r)$ from the solution for $\mu(r)$ by using Eqs. (5.2.1) and (5.2.3). At this point, our solution is complete. To incorporate the effects of space-charge, we could then solve Poisson's equation for a better guess of the potential $V(r)$, and repeat the entire procedure until the solution converges.

To this point, we have overlooked an important ingredient in the calculation of electron density: the density of states $N_0(r;E)$. This quantity is commonly translated to some simple function based upon the dimensionality of the system (i.e., $\sim \sqrt{E}$ in 3-D, etc.). However, for the structures that we are investigating, it is highly sample-specific. To calculate $N_0(r;E)$ for any arbitrary structure, we use the following well-known relationship [17],

$$N_0(r;E) = -\text{Im}(G(r,r;E))/\pi.$$  

(5.2.7)

This expression requires no extra computation, since it involves the diagonal elements of the same Green function which is used to calculate the kernel. We now have all the theory needed to solve a general problem in three dimensions. For all practical purposes, however, such a calculation is impossible; a mere 100 nodes along four axes (three position-space axes and the energy axis) would require $10^8$ nodes! In order to make the calculation tractable, we specialize to a simple geometry for which the chemical potential varies in only one dimension.

### 5.2.2 Specialization to Layered Geometries

We now consider devices having the layered structure illustrated in Fig. 5.1. At the boundaries the potential $V(r)$ becomes infinite, and thus the entire structure is contained in a quantum box. Within each layer $n$, the potential $V(r)$ is assumed to be a constant $V_n$. Along the longitudinal (z-axis) direction, therefore, the potential varies in a piece-wise constant fashion. In practice, the layers should be made sufficiently thin so that the potential $V_n$ well approximates some $V(z)$. Because the potential is constant in the transverse (x- and y-axis) directions, the eigenfunctions are the usual particle-in-a-box solutions:

$$\phi^i_n(\xi) = \left(2/W_i\right)^{1/2} \sin \frac{\xi}{W_i} \xi,$$  

(5.2.8a)
Fig. 5.1: Devices are composed of a number of layers, each having a constant potential $eV_n$. The entire structure is contained in a quantum box.
Of course, we are neglecting any effects arising from space-charge imbalances in the transverse plane. If you like, we are assuming that the profile of ionized dopants exactly matches the profile of electron density across the transverse plane. In the limit as the widths $W_x$ and $W_y$ tend to infinity, both profiles become uniform.

Without approximation, we can assume that the chemical potential varies only along the longitudinal direction, $\mu(r) = \mu(z)$. At first glance, this statement might appear to be untrue, since by inspection of Eq. (5.2.8a), the electron density goes to zero near the boundaries in the transverse plane. However, this is because the density of states $N_0$ goes to zero—not because of any variation in $\mu$. Indeed, in equilibrium $\mu$ must be constant, regardless of how the electron density might fluctuate. Therefore, the dimensionality of $\mu$ is determined by our boundary conditions: If we assume it to be constant across the transverse plane within each contact region, then it will be constant across the transverse plane everywhere else, independent of the inputs $V(r)$ and $\tau(r; E)$. Any assumptions regarding $V(r)$ and $\tau(r; E)$ merely simplify our calculation of the kernel. Assuming that $\mu(r) = \mu(z)$, Eq. (5.2.4) reduces to:

$$I(z) = \frac{e^2}{h} \int dz' T_0(z, z') [\mu(z') - \mu(z)], \quad (5.2.9)$$

where

$$I(z) = \frac{1}{W_x W_y} \int dx \int dy \, I(x, y, z), \quad (5.2.10)$$

$$T_0(z, z') = \frac{1}{W_x W_y} \int dx \int dy \int dx' \int dy' \int dy'' T_0(x, y, z; x', y', z''). \quad (5.2.11)$$

The bar is a reminder that each quantity is averaged over the cross section. By analogy to the definitions above, we define

$$\overline{N}_0(z; E) = \frac{1}{W_x W_y} \int dx \int dy \, N_0(x, y, z; E), \quad (5.2.12)$$

and therefore the electron density $n(z)$ can be expressed as,

$$n(z) = \int dE \overline{N}_0(z; E) \frac{1}{e^{(E - \mu(z))/k_B T} + 1}. \quad (5.2.13)$$

Note that $n(z)$, like $\overline{N}_0(z; E)$, has dimensions of a 3-D quantity (i.e., cm$^{-3}$).

In order to solve this new transport equation for $\mu(z)$ and $I(z)$, we must evaluate the kernel $T_0(z, z')$. Recalling the definition in Eq. (5.2.5a),

$$T_0(z, z') = \int dE \left[ -\frac{\partial f_0}{\partial E} \right] \overline{\bar{T}}(z, z'; E), \quad (5.2.14a)$$

which shows (as before) that at finite temperatures, the range of energies for transport is
broadened. It is shown in the appendix that, if we assume \( \tau(r;E) = \tau(z;E) \),

\[
\bar{T}(z,z';E) = \frac{1}{W_x W_y} \sum_{\alpha, \beta} T^{1-D}(z,z';E-E_\alpha - E_\beta)
\]

(5.2.14b)

where \( T^{1-D}(z,z';E) \) is the kernel obtained by solving the one-dimensional Schrödinger equation,

\[
T^{1-D}(z,z';E) = \frac{\hbar^2}{\tau(z;E)} \frac{|G^{1-D}(z,z';E)|^2}{\tau(z;E)}
\]

(5.2.15a)

\[
\left[ E + \frac{i\hbar}{2\tau(z;E)} + \frac{\hbar^2}{2m^*} \frac{d^2}{dz^2} - eV(z) \right] G^{1-D}(z,z';E) = \delta(z-z')
\]

(5.2.15b)

This reflects the three-dimensional nature of our structure: For a given total energy \( E \), an electron can have a range of longitudinal energies \( E-E_\alpha - E_\beta \). The proper kernel \( \bar{T}(z,z';E) \) is the sum of one-dimensional kernels \( T^{1-D} \) over all longitudinal energies. Note that, for energies well below the conduction band edge, \( T^{1-D} \) must vanish, since electrons are strongly evanescent in the band gap. In the limit of a narrow cross section, the energies \( E_\alpha \) and \( E_\beta \) will increase so rapidly that only the first term in the sum of Eq. (5.2.14b) will contribute. Therefore, in the limit of a narrow cross section, \( \bar{T} \) reduces to the one-dimensional kernel \( T^{1-D} \), as expected.

Also shown in the appendix is the reduction of Eq. (5.2.12),

\[
\bar{N}_0(z;E) = \frac{1}{W_x W_y} \sum_{\alpha, \beta} N^{1-D}_0(z;E-E_\alpha - E_\beta)
\]

(5.2.16)

where

\[
N^{1-D}_0(z;E) \equiv -\text{Im}(G^{1-D}(z,z;E))/\pi
\]

(5.2.17)

Numerical results presented later will show that this sum over longitudinal energies reproduces the expected results for \( N_0 \) in the limits of 0-D, 1-D, 2-D and 3-D structures.

As the area of the cross section increases, it becomes impractical to perform the sums required by Eqs. (5.2.14b) and (5.2.16). To avoid this problem, we consider the limit as the widths \( W_x \) and \( W_y \) become infinite. In this limit, the summations can be converted into integrals,

\[
\bar{T}(z,z';E) = \frac{m^*}{\pi\hbar^2} \int dE' T^{1-D}(z,z';E')
\]

(5.2.18)

\[
\bar{N}_0(z;E) = \frac{m^*}{\pi\hbar^2} \int dE' N^{1-D}_0(z;E')
\]

(5.2.19)

Again, we are simply summing the one-dimensional results for energies below the desired energy \( E \). These two formulas lead to a substantial savings of computer time by the following trick. Consider the integral for the actual kernel \( \bar{T}_0(z,z') \),
\[ T_0(z, z') = \int_{-\infty}^{+\infty} dE \left[ -\frac{\partial f_0}{\partial E} \right] T(z, z'; E). \] (5.2.20)

This requires two integrals over energy: one to evaluate \( T(z, z'; E) \), and another to evaluate \( T_0(z, z') \). However, we can integrate this expression by parts,

\[ T_0(z, z') = -f_0(E) T(z, z'-E) + \int_{-\infty}^{+\infty} dE f_0(E) \frac{\partial}{\partial E} T(z, z'; E). \] (5.2.21)

The surface terms are zero, since \( f_0 \) vanishes at one limit, and \( T \), at the other. Furthermore, the derivative \( \partial T/\partial E \) is particularly easy to evaluate in light of Eq. (5.2.18). Our final expression involves only a single integral over energy,

\[ T_0(z, z') = \frac{m^*}{\pi \hbar^2} \int_{-\infty}^{+\infty} dE f_0(E) T^{1-D}(z, z'; E). \] (5.2.22)

A similar trick can be performed in the calculation of electron density. Integrating Eq. (5.2.13) by parts, we obtain,

\[ n(z) = \frac{m^* k_B T}{\pi \hbar^2} \int_{-\infty}^{+\infty} dE \ln\left[ 1 + e^{-\left(\frac{E-e\mu(z)}{k_BT}\right)} \right] N_0^{1-D}(z; E). \] (5.2.23)

To this point, we have neglected a problem that arises in any practical calculation. If either of the above integrals is evaluated using the Green function defined in Eq. (5.2.15b), the result will diverge. This difficulty is due to the level broadening caused by our inelastic scatterers. The imaginary energy \( \imath \hbar/2 \tau \) in Eq. (5.2.15b) represents a Lorentzian broadening of each single-electron level. While this is a good approximation for energies close to the level (i.e., times much longer than the inelastic scattering time), the approximation breaks down for energies far from the level. As a result, the tails of each level are greatly exaggerated. This problem has appeared in other contexts of Lorentzian broadening [18]; the usual solution is simply to introduce a cut-off energy \( E_0 \), below which the integral is discarded. Thus, Eqs. (5.2.22) and (5.2.23) are modified to,

\[ T_0(z, z') = \frac{m^*}{\pi \hbar^2} \int_{E_0}^{+\infty} dE f_0(E) T^{1-D}(z, z'; E), \] (5.2.24)

\[ n(z) = \frac{m^* k_B T}{\pi \hbar^2} \int_{E_0}^{+\infty} dE \ln\left[ 1 + e^{-\left(\frac{E-e\mu(z)}{k_BT}\right)} \right] N_0^{1-D}(z; E). \] (5.2.25)

This solution of the problem is particularly convenient, since any numerical integration scheme must be truncated at some point. A more ambitious solution would require a more detailed calculation of the self-energy; we leave such considerations to future work.
5.2.3 Solution for the Green Function in 1-D

It is apparent from the formulas above that our method hinges upon the solution for the Green function $G^{1-D}(z,z';E)$, 

$$
\left[ \frac{d^2}{dz^2} + k^2(z; E) \right] G^{1-D}(z,z';E) = \frac{2m^*}{\hbar^2} \delta(z-z') ,
$$

where 

$$
k(z; E) = \left[ \frac{2m^*}{\hbar^2} (E-eV(z)+i\hbar/2\tau(z;E)) \right]^{1/2} .
$$

We will now describe the numerical method for determining this function. The $\delta$-function in Eq. (5.2.26) divides the $z$-axis into two regions whose solutions are coupled at the injection point $z'$ as follows:

$$
G^{1-D}(z,z';E) \bigg|_{z=(z')} = G^{1-D}(z,z';E) \bigg|_{z=(z')} ,
$$

where 

$$
\frac{d}{dz} G^{1-D}(z,z';E) \bigg|_{z=(z')} = \frac{d}{dz} G^{1-D}(z,z';E) \bigg|_{z=(z')} = \frac{2m^*}{\hbar^2} .
$$

Once we have determined the solutions on either side of the injection point, it is a simple matter to connect them.

Recall our assumption of a layered geometry set forth in the previous section. Within each layer the potential is a constant $V_n$, so that the wavefunction can be expressed as (see Fig. 5.2),

$$
\psi_n(z) = \psi_n^+ e^{ik_n(z-z_n)} + \psi_n^- e^{-ik_n(z-z_n)} \quad \text{for} \quad z_n \leq z \leq z_{n+1} ,
$$

where 

$$
k_n = \left[ \frac{2m^*}{\hbar^2} (E-eV_n+i\hbar/2\tau_n) \right]^{1/2} .
$$

For simplicity, we have neglected any energy dependence of $\tau$. We choose our injection point $z'$ to be slightly to the right of some node $z_m$. Therefore, this point is within a region of constant potential, and the solutions on either side of the injection point are of the form in Eq. (5.2.29). We choose the coefficients $A^\pm$ to denote the solution to the left of the injection point, and $B^\pm$, to denote that to the right. At the injection point, the two solutions are connected according to Eqs. (5.2.28a,b),

$$
A^+ + A^- = B^+ + B^- ,
$$

$$
\text{ik}_m (B^+ - B^- - A^+ + A^-) = \frac{2m^*}{\hbar^2} .
$$

This provides two equations for the solution of our four unknowns. The remaining two equations are obtained as follows. The amplitude $A^-$ represents a plane wave traveling to the
Within each layer, the potential $eV_n$ is constant, and therefore the wavefunction can be expressed as a sum of positively- and negatively- traveling plane waves.
left. This wave encounters a number of steps in the potential, and is therefore scattered in some complicated fashion. Overall, however, the reflection is described by a single coefficient, which relates the incident wave $A^-$ to the scattered wave $A^+$,

$$A^+ = r_L A^-.$$  

(5.2.32a)

A similar relationship exists between $B^+$ and $B^-$,

$$B^- = r_R B^+.$$  

(5.2.32b)

It is a simple matter to determine these reflection coefficients. In an earlier work [19], it was shown that devices having a layered geometry can be described by a number of scattering matrices. Each matrix $S_n$ represents the scattering due to a single layer, by connecting the incoming and outgoing wave amplitudes:

$$(5.2.33)$$

$$\begin{bmatrix} \psi_n^+ \\ \psi_{n+1}^+ \end{bmatrix} = \begin{bmatrix} r_n & t_n' \\ t_n & r_n' \end{bmatrix} \begin{bmatrix} \psi_n^- \\ \psi_{n+1}^- \end{bmatrix},$$

where the elements of $S_n$ are defined as,

$$r_n = \frac{k_n - k_{n+1}}{k_n + k_{n+1}} e^{\pm ik_n (z_{n+1} - z_n)}.$$  

(5.2.34a)

$$t_n = \frac{2k_n}{k_n + k_{n+1}} e^{\pm ik_n (z_{n+1} - z_n)}.$$  

(5.2.34b)

$$t_n' = \frac{2k_{n+1}}{k_n + k_{n+1}} e^{\pm ik_n (z_{n+1} - z_n)}.$$  

(5.2.34c)

$$r_n' = \frac{k_{n+1} - k_n}{k_n + k_{n+1}}.$$  

(5.2.34d)

The overall scattering matrix for a number of successive layers is determined by combining the individual scattering matrices, two at a time, according to the following rule [20],

$$r_1 \oplus_2 \equiv r_1 + t_1 r_2 [1 - r_1' r_2]^{-1} t_1,$$  

(5.2.35a)

$$t_1 \oplus_2 \equiv t_2 [1 - r_1' r_2]^{-1} t_1,$$  

(5.2.35b)

$$t_1 \oplus_2' \equiv t_1' [1 - t_1 r_2']^{-1} t_2',$$  

(5.2.35c)

$$r_1 \oplus_2' \equiv r_2' + t_2 [1 - r_1' r_2']^{-1} r_1' t_2'.$$  

(5.2.35d)

It is convenient to define the composite scattering matrices $S_n^L$ and $S_n^R$ as follows:

$$S_n^L = S_0 \oplus S_1 \oplus \cdots \oplus S_n,$$  

(5.2.36a)
\[ S_n^R = S_{n+1} \oplus S_{n+2} \oplus \cdots \oplus S_{N+1} \, , \quad (5.2.36b) \]

where \( S_0 \) and \( S_{N+1} \) represent scattering matrices at the ends of the device. These matrices are determined by some choice of a boundary condition; in the following calculations, we assume complete reflection at the ends (i.e., infinite potential walls), so that

\[
S_0 = S_{N+1} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} . \quad (5.2.37)
\]

Note that an efficient way of computing the composite matrices is to start at each end and work across to the other side of the device. For example, \( S_1^0 = S_0 \); if we combine this with \( S_1 \), we have \( S_2^0 \); if we combine this with \( S_2 \), we have \( S_2^y \), etc. Therefore, after combining scattering matrices across the entire device, the desired reflection coefficients are simply read out of the composite scattering matrices: At the injection node \( z_m \), they are \( r_m^L \) and \( r_m^R \).

We can now solve the four equations (Eqs. (5.2.31a,b) and (5.2.32a,b)) for the wave amplitudes \( A^\pm \) and \( B^\pm \). The result is,

\[
B_m^+ = \frac{m^*/h^2 (1 + r_m^L)}{ik_m (1 - r_m^L r_m^R)} , \quad (5.2.38a)
\]

\[
A_m^- = \frac{m^*/h^2 (1 + r_m^R)}{ik_m (1 - r_m^L r_m^R)} , \quad (5.2.38b)
\]

where the subscript \( m \) has been added to emphasize that these amplitudes depend on the position of the injection point. Of course, \( A_m^+ \) and \( B_m^- \) are determined from Eqs. (5.2.32a,b). We now have the solution for the Green function at the injection node. While this is sufficient to determine the density of states,

\[
N_0^D(z_m; E) = -\text{Im}(B_m^+ + B_m^-)/\pi , \quad (5.2.39)
\]

we must compute the Green function at all other points to define the kernel \( T_0^D(z, z') \). Now that we have the amplitudes \( A_m^+ \) and \( B_m^\pm \), we can treat this remaining task purely as a scattering problem.

Consider a plane wave impinging on two scatterers from the right, as shown in Fig. 5.3(a). Between the two scatterers, there is an infinite number of multiply reflected paths. The sum of all paths traveling to the left is,

\[
\phi_{l\rightarrow l} = (1 + \sum r_2 r_1^' + \sum r_2 r_1^' r_2 r_1^' + \cdots) t_2^' \psi_{l\rightarrow l} \\
= \left[ 1 - r_2 r_1^' \right] t_2^' \psi_{l\rightarrow l} , \quad (5.2.40a)
\]

while that for all paths traveling to the right is,

\[
\phi_{r\rightarrow l} = r_1^' \left[ 1 - r_2 r_1^' \right] t_2^' \psi_{r\rightarrow l} . \quad (5.2.40b)
\]

Thus, we have found a way to determine the wave amplitudes at any node, given the left-moving wave amplitude at a node further up the line. For the problem at hand, it is convenient
Fig. 5.3: The amplitudes of positively- and negatively-traveling waves between two scatterers are determined by summing the contributions of all multiply reflected paths.
to define the coefficients \( L^\pm_n \), so that for all points to the left of the injection point,

\[
\psi_n^\pm = L_n^\pm \psi_{n+1}^+ ,
\]

(5.2.41)

where

\[
L^+_n = \frac{r_n^L \left[ 1 - r_{n+1}^L \right]^{-1}}{T_{n+1}}, \quad (5.2.42a)
\]

\[
L^-_n = \frac{\left[ 1 - r_{n+1}^L \right]^{-1} r_n^L}{T_{n+1}}. \quad (5.2.42b)
\]

Starting with \( A_m^- \) as the incident wave, we can determine \( \psi_n^\pm \) at each node toward the left, as shown schematically in Fig. 5.4. In general,

\[
\psi_n^\pm = L_n^\pm L_{n+1}^- L_{n+2}^- \cdots L_{m-1}^- A_m^- \quad \text{for} \quad n < m. \quad (5.2.43)
\]

Similar arguments can be presented for all nodes to the right of the injection point. By defining the coefficients \( R^\pm_n \),

\[
R^+_n = \frac{r_n^R \left[ 1 - r_{n+1}^R \right]^{-1}}{T_n}, \quad (5.2.44a)
\]

\[
R^-_n = \frac{r_n^R \left[ 1 - r_{n+1}^R \right]^{-1} r_n^R}{T_n}, \quad (5.2.44b)
\]

we can express the wave amplitudes toward the right as,

\[
\psi_n^\pm = R_n^\pm R_{n-1}^+ R_{n-2}^+ \cdots R_{m+1}^- B_m^- \quad \text{for} \quad n > m. \quad (5.2.45)
\]

Our solution for the Green function \( G^{1-D}(z, z'; E) \) is therefore complete. For each injection point \( z_m \), we determine the wave amplitudes \( A_n^\pm \) and \( B_n^\pm \) at the injection point. Using these amplitudes, we determine \( \psi_n^\pm \) at every other node through Eqs. (5.2.43) and (5.2.45). The Green function at each node is then,

\[
G^{1-D}(z_n, z_m; E) = \psi_n^+ + \psi_n^-. \quad (5.2.46)
\]

### 5.2.4 Integration Weights for Integrals Involving \(|G^{1-D}(z, z'; E)|^2\)

In principle, the method of evaluating \( G^{1-D}(z, z'; E) \) presented in the previous section is all that is needed to solve the transport equation (Eq. (5.2.9)). The kernel \( T_0(z, z') \), which depends on the squared-magnitude of the Green function, can be integrated numerically (say, by trapezoidal rule) over \( z' \). Such a procedure, however, would require too many spatial nodes to be of any practical use. The purpose of this section is to derive analytically the integration weights which will convert integrals of the form,

\[
\int dz' \left| G^{1-D}(z, z'; E) \right|^2 F(z'), \quad (5.2.47)
\]

to summations of the form,

\[
\sum_{m=0}^{N} g(z_n, z_m; E) F(z_m). \quad (5.2.48)
\]

An integral that fits this description is,
Fig. 5.4: Using the amplitudes $A_m^-$ and $B_m^+$, we can deduce the amplitudes $\psi_n^\pm$ in all other layers by solving the scattering problem illustrated in Fig. 5.3.
\[ \int dz' T^{1-D}(z, z'; E) \mu(z') = \int dz' |G^{1-D}(z, z'; E)|^2 \left[ \frac{\hbar^2 \mu(z')}{\tau(z) \tau(z')} \right]. \quad (5.2.49) \]

Of course, the kernel \( T^{1-D}(z, z'; E) \) must be summed or integrated over energy to determine the actual kernel \( T_0(z, z') \), which means that the integration weights \( g(z_n, z_m; E) \) should be summed or integrated in exactly the same manner.

To derive the integration weights, we must first determine \( G^{1-D}(z, z'; E) \) at positions \( z' \) between nodes. As illustrated in Fig. 5.5, we sweep our injection point \( z' \) across the injection layer. This is equivalent, through a change of variables, to sweeping the coordinate \( z'' \) from 0 to \( \Delta z_m = z_{m+1} - z_m \). At the left-hand side of the layer (\( z'' = 0 \)), the solutions for \( A_m^\pm \) and \( B_m^\pm \) are identical to those of the previous section. As we move the injection point across the layer, however, the reflection coefficients \( r_m^L \) and \( r_m^R \) acquire an additional phase,

\[ r_m^L(z'') = e^{2i k_m z''} r_m^L, \quad (5.2.50a) \]
\[ r_m^R(z'') = e^{-2i k_m z''} r_m^R. \quad (5.2.50b) \]

Inserting these reflection coefficients into Eqs. (5.2.38a,b), we obtain,

\[ B_m^+(z'') = \frac{m^* /\hbar^2 (1 + e^{2i k_m z''} r_m^L)}{i k_m (1 - r_m^L/r_m^R)}, \quad (5.2.51a) \]
\[ A_m^-(z'') = \frac{m^* /\hbar^2 (1 + e^{-2i k_m z''} r_m^R)}{i k_m (1 - r_m^L/r_m^R)}. \quad (5.2.51b) \]

A central result of the previous section is that the Green function at all points is simply proportional to the amplitudes \( A_m^\pm \) and \( B_m^\pm \) through the coefficients \( L_m^\pm \) and \( R_m^\pm \). We can therefore express the Green function as,

\[ G^{1-D}(z_n, z_{m+1} + z''; E) = \begin{cases} 
C_{n=m} A_m^-(z'') e^{ik_m z''} & \text{for } n \leq m \\
C_{n>m} B_m^+(z'') e^{-ik_m z''} & \text{for } n > m \end{cases}. \quad (5.2.52) \]

where

\[ C_{n<m} \equiv (L_n^+ + L_n^-) L_{n+1}^- L_{n+2}^- \cdots L_{m-1}^-, \quad (5.2.53a) \]
\[ C_{n=m} \equiv 1 + r_m^L, \quad (5.2.53b) \]
\[ C_{n>m} \equiv (R_n^+ + R_n^-) R_{n-1}^+ R_{n-2}^+ \cdots R_{m+1}^+. \quad (5.2.53c) \]

The phase factors \( e^{i k_m z''} \) in Eq. (5.2.52) account for the phase shift in the wave inputs \( A_m^- \) and \( B_m^+ \) as the injection point is swept across the layer. Using Eqs. (5.2.51a,b) and (5.2.52), we can express the squared magnitude of the Green function to the left of the injection point as,
Fig. 5.5: A schematic representation of the solution for the Green function as the injection point is swept across the layer.
\[ |G^{1-D}(z_n, z_m + z''; E)|^2 \]
\[ = \left| \frac{C_{n=m} m^*/h^2 k_m}{1 - \frac{L_m}{m} R_m} \right|^2 \left[ e^{i(k_m - k_m^*) z''} + \frac{R_m}{m} e^{-i(k_m + k_m^*) z''} + (r_m^*) e^{i(k_m + k_m^*) z''} + |r_m|^2 e^{-i(k_m - k_m^*) z''} \right] \]  \hspace{1cm} (5.2.54a)

and to the right of the injection point as,
\[ |G^{1-D}(z_n, z_m + z''; E)|^2 \]
\[ = \left| \frac{C_{n=m} m^*/h^2 k_m}{1 - \frac{L_m}{m} R_m} \right|^2 \left[ e^{-i(k_m + k_m^*) z''} + \frac{R_m}{m} e^{i(k_m - k_m^*) z''} - (r_m^*) e^{-i(k_m + k_m^*) z''} + |r_m|^2 e^{i(k_m - k_m^*) z''} \right] . \]  \hspace{1cm} (5.2.54b)

Note that, in general, the wavevector \( k_m \) is a complex quantity (Eq. (5.2.30)), so that \( k_m^* \neq k_m \).

Next, we assume that the function \( F(z) \) varies slowly enough that it can be interpolated linearly in the region between node points,
\[ F(z'') = \frac{z''}{\Delta z_m} F_{m+1} + \left[ 1 - \frac{z''}{\Delta z_m} \right] F_m , \]  \hspace{1cm} (5.2.55)

where \( F_n \equiv F(z_n) \) is the value at node \( n \). Integrating the product of Eqs. (5.2.54) and (5.2.55) requires a sum of integrals of the form,
\[ \xi_0^{\Delta z_m}(\alpha, k) \equiv \int_0^{\Delta z_m} dz z^\alpha e^{ikz} . \]  \hspace{1cm} (5.2.56)

It is convenient to define the following sum of integrals,
\[ \chi_0^{\Delta z_m}(\alpha, a, b, k) \equiv |a|^2 \xi_0^{\Delta z_m}(\alpha, k-k^*) + a^* b \xi_0^{\Delta z_m}(\alpha, k+k^*) \]
\[ + a^* b \xi_0^{\Delta z_m}(\alpha, -k-k^*) + |b|^2 \xi_0^{\Delta z_m}(\alpha, k^*-k) . \]  \hspace{1cm} (5.2.57)

The integration weights for a single interval are expressed as, for \( n \leq m \),
\[
\Delta z_m \int_0^\Delta z'' \left| G^{1-D}(z_n, z_m + z'') \right|^2 f(z'')
\]

\[
= \left| \frac{C_n z_m m^* / \hbar^2 k_m}{1 - r_m^{L'}, r_m^R} \right|^2 \left\{ \frac{\chi_0^{\Delta z_m}(1, r_m^L, 1, -k_m)}{\Delta z_m} f_{m+1} \right. \\
\left. + \left[ \chi_0^{\Delta z_m}(0, r_m^R, 1, -k_m) - \frac{\chi_0^{\Delta z_m}(1, r_m^R, 1, -k_m)}{\Delta z_m} \right] f_m \right\} \tag{5.2.58a}
\]

and for \( n > m \),

\[
\Delta z_m \int_0^\Delta z'' \left| G^{1-D}(z_n, z_m + z'') \right|^2 f(z'')
\]

\[
= \left| \frac{C_{n > m} m^* / \hbar^2 k_m}{1 - r_m^{L'}, r_m^R} \right|^2 \left\{ \frac{\chi_0^{\Delta z_m}(1, r_m^L, 1, k_m)}{\Delta z_m} f_{m+1} \right. \\
\left. + \left[ \chi_0^{\Delta z_m}(0, r_m^R, 1, k_m) - \frac{\chi_0^{\Delta z_m}(1, r_m^R, 1, k_m)}{\Delta z_m} \right] f_m \right\} \tag{5.2.58b}
\]

To obtain the integration weights \( g(z_n, z_m) \) for the entire structure, we evaluate the result shown above in all intervals \( m = 0, 1, 2, ..., N-1 \), and add the coefficients of the \( f_m \)'s. At this point, the patient reader deserves some consolation: The pain involved in obtaining this result saves hours of grief, when it comes to performing actual calculations.

### 5.2.5 Overview of the Solution Technique

Amidst the forest of equations which sprang up in the previous sections, one is easily lost. It is useful, therefore, to describe an actual calculation from beginning to end. We begin by evaluating the kernel \( \overline{T}_0 \) in equilibrium. To account for space-charge effects, we would first determine the self-consistent potential \( V_n \) by repeatedly solving for the electron density and inserting this into the Poisson equation. In the following calculations we neglect this correction, and simply assume some form for \( V_n \). We can then compute the integration weights required for integrals involving the kernel \( \overline{T}_0(z, z') \). These weights must be either summed and integrated over energy (for a finite cross section),

\[
(\overline{T}_0)_{nm} = \frac{1}{W_x W_y} \int_{E_0}^{\infty} dE \left\{ - \frac{\partial f_0}{\partial E} \right\} \sum_{\alpha, \beta} \frac{\hbar^2 g(z_n, z_m ; E - E_\alpha^x - E_\beta^x)}{\tau(z_n) \tau(z_m)} \tag{5.2.59}
\]

or simply integrated over energy (for an infinite cross section),
where $E_0$ is the cut-off energy discussed in Section 5.2.2. At each energy, the weights $g(z_n, z_m; E)$ are calculated as described in Section 5.2.4.

The next step is to solve the integral transport equation (Eq. (5.2.9)), which can be rewritten in terms of our node representation as,

$$
(\bar{T}_0)_{nm} = \frac{m^*}{\pi \hbar^2} \int_{E_0}^{+\infty} dE \, f_0(E) \frac{\hbar^2 g(z_n, z_m; E)}{\tau(z_n) \tau(z_m)} ,
$$

(5.2.60)

Within each contact region, the chemical potential $\mu_n$ is fixed, and the external current $I_n$ is determined from Eq. (5.2.61). Between contact regions, the external current $I_n$ is zero, and $\mu_n$ is determined by solving

$$
\mu_n = \frac{\sum_{m=0}^{N} (\bar{T}_0)_{nm} \mu_m}{\sum_{m=0}^{N} (\bar{T}_0)_{nm}} .
$$

(5.2.62)

This leads naturally to an iterative solution, such as Gauss-Siedel iteration. We assume some initial guess for $\mu_n$; namely, that $\mu_n$ is constant within contact regions, and dropped linearly between them. We then solve Eq. (5.2.62) iteratively, continually updating our guess for $\mu_n$ between contact regions until the solution converges.

We can then determine the electron density $n(z_n)$ at each node. For structures with a finite cross section,

$$
n(z_n) = \frac{1}{W_x W_y} \int_{E_0}^{+\infty} dE \, \frac{1}{e^{\left(E-e\mu_n\right)/k_B T} + 1} \sum_{\alpha, \beta} N_0^{1-D}(z_n; E-E^\alpha - E^\beta) ,
$$

(5.2.63)

and for those with an infinite cross section,

$$
n(z_n) = \frac{m^* k_B T}{\pi \hbar^2} \int_{E_0}^{+\infty} dE \ln \left[ 1 + e^{-\left(E-e\mu_n\right)/k_B T} \right] N_0^{1-D}(z_n; E) .
$$

(5.2.64)

At each energy, the density of states $N_0^{1-D}(z_n; E)$ is calculated from Eq. (5.2.39). At this point, the analysis is complete. To obtain the conductance of the structure, we evaluate the terminal currents and divide by the applied bias. We evaluate $I_n$ at each contact node using Eq. (5.2.61). From this, we determine the total current at each contact by integrating $I_n$ over the length of the contact. To perform this integral, it is possible to derive integration weights similar to those obtained in Section 5.2.4; such a procedure, however, is enormously complicated and computationally demanding. For typical spatial grids, it is sufficient to use some simple numerical scheme such as trapezoidal rule or Simpson’s rule.
5.3 Results and Conclusions

Before we consider any variations in the chemical potential, it is instructive to compute the density of states $N_q(z; E)$ in a few simple cases. In particular, we consider an empty quantum box with a modest amount of inelastic scattering ($\tau = 10^{-13}$ s), and expand each of its dimensions one by one. We will show that the sum over energies required by Eq. (5.2.16) does indeed reproduce the correct results in the limits of 0-D, 1-D, 2-D and 3-D structures.

Figure 5.6 presents the density of states as a function of position and energy, for a small quantum box ($100 \times 100 \times 300$ Å). In this limit, only the first term $N_0^{-D}(z; E-E_f-E_i)$ in the sum contributes, since the next highest energy is far below the conduction band edge. This plot nicely illustrates the eigensystem of the 0-D box: For energies near the eigenvalues $E = E_0 + E_i + E_f$ of a rectangular well, $N_0$ peaks up, showing the form $|\phi_n(z)|^2$ of the associated eigenfunction. Of course, each level is homogeneously broadened by the inelastic scattering.

If we increase the length of the box from 300 Å to 3000 Å, we obtain the quasi 1-D result shown in Fig. 5.7. Again, the cross section is narrow, so that only the first term $N_1^{-D}(z; E-E_f-E_i)$ contributes to the sum. Due to the increase in length, however, the eigenenergies for the z-direction are more closely spaced; states at lower energies overlap (because of the level broadening) and therefore contribute strongly. At higher energies, the spacing between levels increases, so the overlap is reduced. The net result is a decrease in $N_0$ resembling $1/\sqrt{E}$.

To form a quasi 2-D box, we extend the cross section to infinity and quantize the z-axis. Hence, the sum in Eq. (5.2.16) becomes an integral, as shown in Eq. (5.2.19). The density of states $N_0$ resulting from this calculation is presented in Fig. 5.8. Viewed along the energy-axis, $N_0$ appears to increase in a stair-step fashion: At each energy $E_n^z$ where an eigenstate becomes allowed, $N_0$ abruptly jumps; between these energies it remains constant, as expected in 2-D.

Finally, we increase the length of the box from 300 Å to 3000 Å to obtain the quasi 3-D result shown in Fig. 5.9. The increase in length causes the eigenenergies to become more closely spaced. In effect, this reduces the step size found in the 2-D example to a very fine grain. One by one, with increasing energy, the eigenstates become allowed. Each new eigenstate adds the same amount to the total $N_0$; however, the states are spaced according to a square law (Eq. (5.2.8b)), so that overall, $N_0$ increases as $\sim \sqrt{E}$.

Now that we have some confidence in the physics that our model predicts, we will examine chemical potential under a number of different conditions. We begin by considering the single-moded wire structure shown at the bottom of Fig. 5.10. To avoid any spurious effects arising from the hard-wall boundaries, "padding regions" were added to each end of the device; these regions merely extend the contacts by a few Fermi wavelengths. Electrons having a longitudinal energy of 50 meV impinged upon a 50 Å barrier with a height of 100 meV. The chemical potential $\mu$, computed for a number of different inelastic scattering times, is shown in Fig. 5.10. Calculations were performed at 100K, so that any thermal broadening in the calculation of
Fig. 5.6: Density of states $N_0(z; E)$ for a 0-D quantum box.

$$N_0(z; E) \text{ (m}^{-3} \text{-eV}^{-1} \times 10^{25})$$
Fig. 5.7: Density of states $N_0(z; E)$ for a quasi 1-D quantum box.
Fig. 5.8: Density of states $N_0(z; E)$ for a quasi 2-D quantum box.
Fig. 5.9: Density of states $N_0(z; E)$ for a quasi 3-D quantum box.
Fig. 5.10: Decreasing the inelastic scattering time $\tau$ destroys oscillations in the chemical potential.
the kernel (Eq. (5.2.5a)) is negligible. When the inelastic scattering time is long, oscillations in
the chemical potential are prominent; these oscillations die out away from the barrier, illustrat­
ing the phase-breaking processes at work in our model. When the inelastic scattering time is
short, the oscillations are all but destroyed, and \( \mu \) drops on either side of the barrier as it would
in an ordinary resistor. Note that in all the curves, \( \mu \) jumps abruptly near the contact regions.
This is an artifact of the contact resistance in the Landauer model which was described in the
introduction.

By increasing the amount of inelastic scattering, we have shown that the oscillations in the
chemical potential are destroyed. But even if transport is perfectly coherent, such oscillations
might not be observed if some averaging over energies takes place. For instance, at finite tem­
peratures the kernel is computed by averaging over energies near the Fermi level (Eq.
(5.2.14a)). As we increase the temperature, we increase the range of energies available for tran­
sport. Each energy gives rise to an oscillation with a slightly different period; however, all
oscillations are in-phase at the barrier. The result, shown in Fig. 5.11, is that only those oscilla­
tions nearest the barrier survive the averaging process.

Another source of energy averaging is the sum over longitudinal energies in samples with
a sizable cross section (Eq. (5.2.14b)). Again, each energy gives rise to an oscillation with a
unique period. If the cross section is well quantized, the energies in the sum are widely
separated, so that the resulting oscillations appear to fluctuate randomly; the calculations
presented in Fig. 5.12 confirm this. In the limit as the cross section becomes infinite, the sum
includes all energies up to the Fermi energy (at zero temperature). Thus, for lightly doped
materials, the oscillations may persist despite the size of the cross section: If we neglect any
thermal spreading, the range of energies for averaging will ultimately be limited by the Fermi
level.

5.4 Future Work

In order to have a well-defined chemical potential, it is necessary to account for dissipation
within a sample. This minimizes carrier heating which would otherwise distort the distribution
function. Despite the presence of dissipation, however, the chemical potential was found to
oscillate in single-moded wires near a potential barrier; furthermore, traces of the oscillations
survived the most extreme conditions: short inelastic scattering times (\( \tau \sim 10^{-13} \) s), high tem­
peratures (77°K), and many transverse modes. This suggests that such oscillations might be a
physical reality, rather than an artifact of a poorly fitted distribution. However, a definitive
answer must be obtained by solving the more general transport equation [15] for the actual dis­
tribution function, and determining how well the result is represented by the Fermi-Dirac func­
tion.

Some interesting results might also be obtained by determining the inelastic scattering
time \( \tau \) microscopically. Datta has presented an expression [15] which shows an inverse rela­tion­ship between \( \tau \) and the density of holes in the conduction band. This suggests that \( \tau \) will be
Fig. 5.11: Increasing the temperature increases the range of energies available for transport; therefore, oscillations having some range of periods are averaged together.
Fig. 5.12: If the cross section is well quantized, the energies in the averaging process are widely separated; the resulting oscillations appear to be noisy.
long wherever the hole density is small. Since the carrier density oscillates near a potential barrier, $\tau$ may oscillate as well, leading to a series of layers in which electrons are in and out of local equilibrium. The investigation of such peculiar situations will further clarify the meaning that we attach to the chemical potential. By understanding the circumstances for which this concept fails, we will better understand the nature of quantum transport.
Chapter 5 Appendix

In Section 5.2 it was argued that the integral transport equation could be written in a one-dimensional form if we averaged $I(r)$ and $T_0(r,r')$ over the cross section. The purpose of this appendix is to evaluate Eq. (5.2.11), to obtain an explicit expression for the kernel $\vec{t}(z,z';E)$. We begin with the simpler task of evaluating the average density of states $\overline{N}_0(z;E)$.

$$\overline{N}_0(z;E) \equiv \frac{1}{W_x W_y} \int dx \int dy \overline{N}_0(x,y,z;E). \quad (5.1)$$

In general, $N_0(r;E)$ is defined by the Green function,

$$N_0(r;E) = \frac{-\text{Im}\{G(r,r;E)\}}{\pi}, \quad (5.2)$$

which can be expanded in terms of the eigenfunctions $\Phi_m(r)$ of the time-independent Schrödinger equation [21],

$$G(r,r';E) = \sum_m \frac{\Phi_m(r) \Phi_m(r')}{E - E_m + i\hbar/2\tau}. \quad (5.3)$$

where

$$H_0 \Phi_m(r) = E_m \Phi_m(r). \quad (5.4)$$

For simplicity, we have assumed that the inelastic scattering time $\tau$ is a constant; we will later relax this constraint. Substituting Eq. (5.3) into Eq. (5.2),

$$N_0(r;E) = \frac{1}{\pi} \sum_m |\Phi_m(r)|^2 \frac{\hbar/2\tau}{(E - E_m)^2 + \hbar^2/4\tau^2}, \quad (5.5)$$

and this result into Eq. (5.1),

$$\overline{N}_0(z;E) = \frac{1}{\pi W_x W_y} \sum_m \frac{\hbar/2\tau}{(E - E_m)^2 + \hbar^2/4\tau^2} \int dx \int dy |\Phi_m(x,y,z)|^2. \quad (5.6)$$

We assume that the Hamiltonian $H_0$ is separable, so that the eigenfunctions $\Phi_m(r)$ can be written as a product of the eigenfunctions for each coordinate,

$$\Phi_m(x,y,z) = \phi_\alpha^x(x) \phi_\beta^y(y) \phi_\gamma^z(z), \quad (5.7a)$$

$$E_m = E_\alpha^x + E_\beta^y + E_\gamma^z \quad \text{where} \quad m \leftrightarrow (\alpha, \beta, \gamma). \quad (5.7b)$$

Therefore, we can integrate Eq. (5.6) explicitly over $x$ and $y,$

$$\overline{N}_0(z;E) = \frac{1}{\pi W_x W_y} \sum_\alpha \sum_\beta \sum_\gamma |\phi_\gamma^z(z)|^2 \frac{\hbar/2\tau}{(E - E_\alpha^x - E_\beta^y - E_\gamma^z)^2 + \hbar^2/4\tau^2} \int dx |\phi_\alpha^x(x)|^2 \int dy |\phi_\beta^y(y)|^2. \quad (5.8)$$

and if we assume that the eigenfunctions are properly normalized, the integrals are trivial. We define the density of states in one dimension to be,
\( N_{0}^{-D}(z; E) = \sum_{\gamma} \frac{1}{\pi} \left| \phi_{\gamma}(z) \right|^2 \frac{\hbar / 2 \tau}{(E - E_{\gamma})^2 + \hbar^2 / 4 \tau^2} \). \hspace{1cm} (5.A.9)

This is simply a one-dimensional statement of Eq. (5.A.5). Using this definition, Eq. (5.A.8) can be reduced to our final result,

\[ \bar{N}_{0}(z; E) = \sum_{\alpha, \beta} N_{0}^{-D}(z; E - E_{\alpha} - E_{\beta}) . \] \hspace{1cm} (5.A.10)

The evaluation of Eq. (5.2.11) proceeds in a similar manner. Since the kernel depends on \(|G(r, r'; E)|^2 \) (Eq. (5.2.5b)), we are interested in evaluating the integral,

\[ \int dx \int dy \int dx' \int dy' \left| G(r, r'; E) \right|^2 . \] \hspace{1cm} (5.A.11)

Using Eq. (5.A.3), we write

\[ |G(r, r'; E)|^2 = \sum_{m, m'} \frac{\Phi_{m}(r) \Phi_{m'}^{*}(r') \Phi_{m'}^{*}(r') \Phi_{m}(r')}{(E - E_{m} + i\hbar / 2\tau)(E - E_{m'} - i\hbar / 2\tau)} . \] \hspace{1cm} (5.A.12)

If we perform the integrals over the primed coordinates, we have a useful identity,

\[ \int dx' \int dy' \Phi_{\alpha \beta \gamma}(x', y', z') \Phi_{\alpha^{'} \beta^{'} \gamma^{'}}(x', y', z') = \delta_{\alpha \alpha'} \delta_{\beta \beta'} \delta_{\gamma \gamma'} . \] \hspace{1cm} (5.A.13)

Therefore, if we integrate Eq. (5.A.12) over \( x' \) and \( y' \), we can perform the sums over \( \alpha' \) and \( \beta' \). The result is,

\[ \sum_{\alpha, \beta} \sum_{\gamma, \gamma'} \frac{\phi_{\gamma}(z) \phi_{\gamma'}^{*}(z) \phi_{\gamma'}^{*}(z') \phi_{\gamma'}^{*}(z') \left| \phi_{\alpha}(x) \right|^2 \left| \phi_{\beta}(y) \right|^2}{(E - E_\alpha - E_\beta - E_{\gamma} + i\hbar / 2\tau)(E - E_{\alpha} - E_{\beta} - E_{\gamma'} - i\hbar / 2\tau)} . \] \hspace{1cm} (5.A.14)

Again, we assume that the eigenfunctions are normalized, so that if we integrate Eq. (5.A.14) over \( x \) and \( y \), we obtain

\[ \sum_{\alpha, \beta} \sum_{\gamma, \gamma'} \frac{\phi_{\gamma}(z) \phi_{\gamma'}^{*}(z) \phi_{\gamma'}^{*}(z') \phi_{\gamma'}^{*}(z')}{(E - E_\alpha - E_\beta - E_{\gamma} + i\hbar / 2\tau)(E - E_{\alpha} - E_{\beta} - E_{\gamma'} - i\hbar / 2\tau)} . \] \hspace{1cm} (5.A.15)

In terms of the squared magnitude of the one-dimensional Green function,

\[ |G^{1-D}(z, z'; E)|^2 = \sum_{\gamma, \gamma'} \frac{\phi_{\gamma}(z) \phi_{\gamma'}^{*}(z) \phi_{\gamma'}^{*}(z') \phi_{\gamma'}^{*}(z')}{(E - E_{\gamma} + i\hbar / 2\tau)(E - E_{\gamma'} - i\hbar / 2\tau)} , \] \hspace{1cm} (5.A.16)

Eq. (5.A.15) reduces to,

\[ \sum_{\alpha, \beta} |G^{1-D}(z, z'; E - E_\alpha - E_\beta)|^2 . \] \hspace{1cm} (5.A.17)

Multiplying through by \( \hbar^2 / \tau(z; E) \tau(z'; E) \), we obtain the desired relationship,
\[ T(z, z'; E) = \frac{1}{W_x W_y} \sum_{\alpha, \beta} T^{1-D}(z, z'; E = E_\alpha^x - E_\beta^y) , \quad (5.A.18) \]

where

\[ T^{1-D}(z, z'; E) = \frac{\hbar^2 |G^{1-D}(z, z'; E)|^2}{\tau(z; E) \tau(z'; E)} . \quad (5.A.19) \]

In writing the Green function (Eq. (5.A.3)), we assumed that \( \tau(r; E) \) was simply a constant \( \tau \). However, this was unnecessary; in general, we can allow the inelastic scattering time to vary as \( \tau(r; E) = \tau(z; E) \). Using lowest order perturbation theory, this would correct the energies as \( E_\gamma^z \rightarrow E_\gamma^z + i\hbar/2\tau_\gamma \), and leave the eigenstates unchanged. In this case, the results of our derivations (Eqs. (5.A.10) and (5.A.18)) are unaffected, since the matrix elements \( \tau_\gamma \) become buried in the definition of \( G^{1-D}(z, z'; E) \). It is not necessary to compute these matrix elements, since \( G^{1-D}(z, z'; E) \) is computed by the method described in Section 5.2.3, which allows for a spatially varying \( \tau \).
Chapter 5 References


Appendix
SRC-SUPPORTED PUBLICATIONS AND TALKS

Reports:


Theses:


Conference Presentations:


Journal Publications:


Invited Talks at SRC Companies:


