7-1-1988

Physics and Modeling of Submicron Devices. Annual Report: August I, 1987 - July 31, 1988

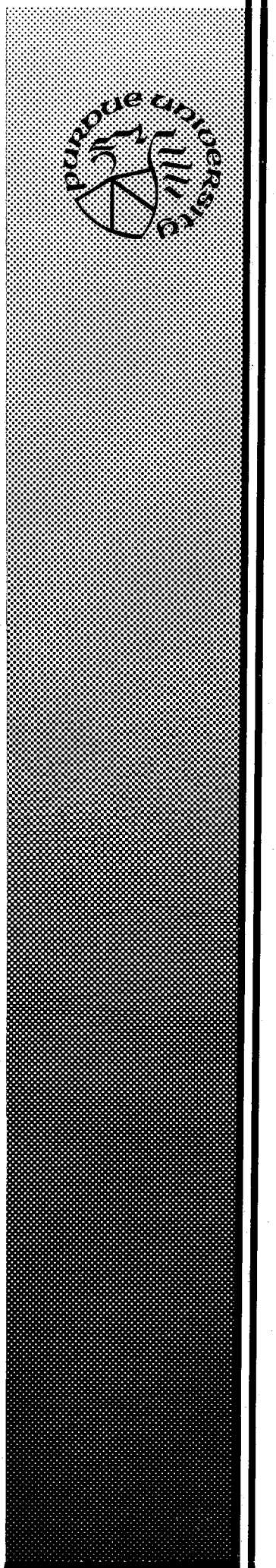
Supriyo Datta
Purdue University

Mark S. Lundstrom
Purdue University

Amitava Das
Purdue University

H. Bob Frohne
Purdue University

Michael J. McLennan
Purdue University



Physics and Modeling of Submicron Devices

Supriyo Datta
Mark S. Lundstrom
Amitava Das
H. Rob Frohne
Michael J. McLennan

TR-EE 88-47
Annual Report: August 1, 1987 - July 31, 1988

School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907

Supported by the Semiconductor Research Corporation Contract Number 87-SJ-089

PREFACE

The work described in this report is directed at understanding quantum transport phenomena in sub-micron heterostructure devices, at developing computational techniques for modeling such devices, and at applying these techniques to develop new device concepts. During the past year we have (1) applied a previously developed collisionless quantum device model (SEQUAL) and Monte Carlo model (DEMON) to the design and study of heterojunction bipolar transistors (Chapter 2); (2) developed a technique for the analysis of arbitrarily shaped quantum devices with elastic scattering (Chapter 3); and (3) developed an approach for incorporating inelastic dissipative processes in quantum transport theory (Chapter 4). As a by-product of the research, several heterostructure device models have been developed: 1- and 2-D equilibrium models, 1- and 2-D drift-diffusion models, a 1-D Monte Carlo simulator and a 1-D collisionless quantum device model. These simulation programs are being applied to advanced device analysis at a number of laboratories and are available to SRC members on request.

TABLE OF CONTENTS

Chapter 1

PROJECT OVERVIEW

1.1 Introduction

This research is directed at the development and application of advanced, physical device simulators. The physical device simulators now being used by the semiconductor industry are losing their ability to accurately describe the physics of small, sophisticated devices. New simulation techniques will have to be developed in order to model devices of the future. Our SRC-supported work is directed at the development of the advanced physical simulation techniques which will be needed to guide the development of future devices. Specific objectives of the program are: 1) initiation of research directed at the development of a new generation of physical device models which account for the wave nature of carriers (quantum effects), 2) the development of a computationally manageable yet physically accurate simulation strategy for treating hot carrier transport in bipolar transistors, and 3) the application of these evolving simulation tools to the exploration of advanced, post-shrink, devices which exploit hot carrier and quantum effects to enhance device performance.

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 widely-used by the industrial, government, and academic research communities for advanced device work.

1.2 Background: Hot Electron and Quantum Effects

Device simulation programs are widely used in industry for the optimization of devices and for exploring new device structures [1]. Such programs will be even more important for the increasingly complex devices of the future. Semiconductor device dimensions are continually shrinking, and device structures are becoming increasingly sophisticated (with the use of heterostructures, for example). Present simulation techniques, however, date back to the 1960's and is not adequate for advanced devices. Conventional device 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 the potential at the contacts. 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 many present day devices assumption 2 is violated, leading to transient *hot electron effects* such as velocity overshoot [2] which are described by the Boltzmann Transport Equation. These effects have been extensively modeled by ensemble Monte Carlo techniques, but little work on engineering these effects to increase the speed of devices has been reported. By contrast, very little work at all has been done in the area of *quantum effects* which arise when assumption 1 is violated. For devices with dimensions comparable to the DeBroglie wavelength of carriers (typically 100-1000 Å), electrons do not behave as particles obeying Newton's law; they must be regarded as waves propagating through the device according to the Schrödinger equation. The relationship between Newton's law and Schrödinger's equation is analogous to that between geometrical and wave optics. A simple ray description is adequate only if the device dimensions are much larger than a wavelength; otherwise a wave description is necessary. It is believed that in future there will be an increasing number of devices that rely on quantum effects for their operation. Resonant tunneling devices pioneered by Tsu and Esaki [3-7] have attracted much attention lately. Device concepts based on the Aharonov-Bohm effect and non-local effects (observed since 1985) are described in a review talk included in Appendix A.

Although the physics of hot carrier transport is now relatively well-understood, there is much work yet to be done in the application of this knowledge to improve device speed and performance. The development of a suitable simulation technique for advanced bipolar devices is also a high priority. The Monte Carlo method is particularly ill-suited to bipolar simulation, and the drift-diffusion approach, though versatile and powerful, does not provide an accurate description of transport in small devices.

We are presently engaged in assessing the speed-limiting factors for bipolar transistors and in engineering hot-electron effects in order to improve speed. This work makes use of our evolving simulation tools. We are also exploring new simulation strategies to accurately describe both transport and recombination in a bipolar context.

Hot electron effects represent one thrust of our work. But our major emphasis at this time is on *quantum mechanical effects.* Compared to hot-electron effects, our understanding of quantum effects is primitive. The challenge here is not just to simulate a well-known set of equations, but to find the proper set of equations that will provide an accurate but tractable description for quantum transport in submicron devices. Our ultimate objective is to develop device engineering models that can be used not only to describe quantum effects in sub-micron devices but also to guide us in the development of new concepts for post-shrink devices. This work is only three years old, but much progress has been achieved. Our group has already developed a simple one-dimensional quantum mechanical device model which treats the collisionless propagation of electron waves in their self-consistent electrostatic potential. This program (SEQUAL) has been released. We have developed a general technique for incorporating elastic scattering processes, such as impurity scattering and boundary scattering. Finally, a promising approach has been identified for handling the most challenging problem in this work, namely to incorporate irreversible dissipative processes into quantum device models.

1.3 Visible By-products of the Research

In the course of this research, several heterostructure device models have been developed, 1-D equilibrium models, 1- and 2-D drift-diffusion models, a 1-D Monte Carlo simulation program, and a 1-D quantum device model. These simulation programs, which are now in use throughout the country, are the most visible benefit of our research. Users have applied these simulation tools to devices in a wide variety of material systems such as Si, Si-Ge, several III-V's, II-VI's, and even combinations of III-V and II-VI semiconductors.

Another important by-product of the research is the development of new device concepts. Two years ago, we proposed a novel collector structure for III-V heterostructure bipolar transistors which was predicted to have a short collector transit time achieved by extending velocity overshoot. This device has recently been built and now holds the record f_T for any transistor [8]. Quantum device concepts developed in our group (Appendix A) have generated much interest in the research community as evidenced by our invited talks at three international research conferences this summer. To complement the SRC-supported theoretical work, we also have an experimental program whose objective is to demonstrate novel device concepts. A summary of the reports, conference presentations, journal articles and talks at SRC member companies which have resulted from this work are contained in Appendix B.

1.4 Overview of the Report

In our proposal we had identified three major steps in the development of a quantum device analysis program: (1) collisionless or ballistic quantum transport, (2) incorporation of elastic scattering due to impurities, defects, boundaries etc. and (3) incorporation of inelastic scattering or dissipative processes. The first step was completed earlier and a 1-D ballistic quantum device model (SEQUAL) was released. During the past year we have applied this model to the design and study of heterojunction bipolar transistors. This work is described in **Chapter 2**. We had made some progress in incorporating elastic scattering processes last year; however, we were limited to devices with rectangular geometries. This year we have developed a technique for analyzing arbitrarily shaped structures. This work is described in **Chapter 3**. Finally in **Chapter 4** we describe a new approach that we have developed for incorporating dissipative processes into quantum device models. This is an important breakthrough, since the incorporation of irreversibility into the description of quantum processes presents major conceptual hurdles. In **Appendix A** we have included a preprint of a review of quantum devices that was presented by us in an invited talk at the Fourth International Conference on Superlattices, Microstructures and Microdevices held at Trieste, Italy in August 1988. A variety of novel quantum device concepts that we have developed are described in this talk. **Appendix B** contains a list of SRC-supported publications and talks. **Appendix C** contains SEQUAL Release 2.1--a document explaining the recent additions to SEQUAL that have been made at the request of various users.

Chapter 1 References

[1] P.J. Zdebel, remarks at the SRC Topical Research Conference on Bipolar Device/Circuit Modeling, University of Florida, Gainesville, May 27-28, 1987.

[2] J. G. Ruch, IEEE Trans. Elec. Dev., ED-19, 652, (1972).

[3] R. Tsu and L. Esaki, Appl. Phys. Letters. 22 (19733) 562-564.

[4] L. L. Chang, L. Esaki and R. Tsu, Appl. Phys. Letters 47(4), (1985) 415-417.

[5] T.J. Shewchuk, P.C. Chapin, P.D. Coleman, W. Kopp, R. Fischer and H. Morkoc Appl. Phys. Letters, 46(5), (1985) 508-510.

[6] E. R. Brown, T. C. L. G. Sollner, W. D. Goodhue and C. D. Parker, Appl. Phys. Lett. 50, 83 (1987).

[7] F. Capasso, S. Sen, A. C. Gossard, A. L. Hutchinson and J. H. English, IEEE Electron Device Letters, EDL-7, 573 (1986).

[8] T. Ishibashi and Y. Yamauchi, "A Possible Near-Ballistic Collection in an AlGaAs/GaAs HBT with Modified Collector Structure," *IEEE Trans. Electron Dev.,* Vol. ED-35, pp. 401-404, 1988.

Chapter 2

CONSEQUENCES OF VALLEY FILTERING ON ABRUPT JUNCTION AlGaAs/GaAs HETEROJUNCTION BIPOLAR TRANSISTORS

Electron transport across the emitter-base junction and across the quasi-neutral base of AlGaAs/GaAs heterojunction bipolar transistors is studied. Transport across abrupt emitter-base heterojunctions is treated quantum mechanically, and the Monte Carlo technique is used to study the transport through the base. The AlGaAs/GaAs heterojunction is found to prevent upper valley electrons from entering the base. This valley filtering enhances device performance by reducing base transit time, but quantum mechanical tunneling lowers the average energy of the injected flux which increases base transit time. The design of a heterojunction bipolar transistor for minimum base transit time involves a trade-off between these competing factors. We examine the influence of varying aluminum fraction and bias on base transport time. The results demonstrate that a moderately doped emitter with high aluminum mole fraction produces the shortest base transit time.

2.1 Introduction

Heterojunction bipolar transistors (HBT's) can broadly be classified into two types according to the nature of emitter-base junction: 1) graded junction HBT's with compositionally graded emitter-base heterojunctions, and 2) abrupt junction HBT's with abrupt emitter-base heterojunctions. Since grading of the emitter-base heterojunction increases electron injection, compositional grading is commonly used to ensure high common emitter current gain [1-3] and a lower turn-on voltage. On the other hand, the abrupt emitter-base heterojunction provides a launching ramp for electrons injected into the base and is therefore expected to improve both the base transit time [4] and base transport factor [5]. Recent experimental [5] and theoretical studies [6] have shown that abrupt junction HBT's can display higher common emitter current gains (β) than graded-junction HBT's when the current gain is limited by the base transport factor (α_t) instead of emitter injection efficiency (γ) [7]. To ensure high common emitter current gain and high speed operation, the design of the emitter and base of an abrupt junction HBT should be optimized to reduce the average base transit time.

For a uniform base HBT, the base transit time primarily depends on the average energy of carriers injected into the base and on the type of scattering carriers undergo during their passage through the base. A higher average injected energy tends to reduce the base transit time, whereas a higher number of momentum randomizing scattering events (such as intervalley scattering) increases the base transit time [8]. The type and frequency of scattering events that minority carriers undergo during their transit through the base depends on the nature of the injected electron flux and on the structure of the base. For example, upper valley electrons present in the injected flux will increase the number of intervalley scattering events, and, in a highly doped base, plasmon scattering will dominate [9]. Issues concerning the design of the base and its impact on the base transit time, have been studied by previous researchers [9-10]. In

this paper we focus on the influence of injected electron flux on the base transit time for HBT's with compositionally abrupt emitter-base junctions.

This paper was motivated by the recent work of Ramberg et. al. [11] who suggested that the base transit time could be improved by filtering out upper valley electrons before they are injected into the base. Previous Monte Carlo studies have demonstrated that a small percentage of upper valley electrons in the injected flux can significantly degrade the base transit time [8]. The filtering effect was to be achieved by properly designing the abrupt emitter-base heterojunction in order to enhance tunneling of Γ-valley electrons through the conduction band spike. Since the Γ-valley electrons are lighter than those in the L or X valley, they have a higher probability of quantum mechanically tunneling through the conduction band spike and, as a result, the electron flux incident on the base will be rich in Γ valley electrons.

We find that the valley filtering is due two separate mechanisms. Firstly, the different band offsets for the Γ, X and L valleys produce different barrier heights for electrons in these valleys which naturally lead to a filtering effect. Consider the energy band diagram for a typical emitter-base heterojunction as displayed in Fig. 2.1(a) (a conduction band discontinuity of 65% was assumed for the Γ valley [12]). This figure shows that the barrier for Γ-valley electrons, $V_{b\Gamma}$ is much smaller than that for the X-valley, V_{bx}. The flux of electrons injected into the base should be correspondingly rich in Γ-valley electrons. The strong tunneling of Γ-valley electrons further reduces their effective barrier height and additionally improves the filtering effect.

The second mechanism for valley filtering is illustrated by the energy band diagram for Γ and L valleys which is displayed in Fig. 2.1(b). The barrier heights for Γ- and L- valley electrons are nearly equal ($V_{b\Gamma} \simeq V_{bL}$), but the strong tunneling of the light Γ-valley electrons reduces the effective barrier for Γ-valley electrons and produces a filtering effect. This is effective mass filtering as described by Ramberg et. al. [11]. We

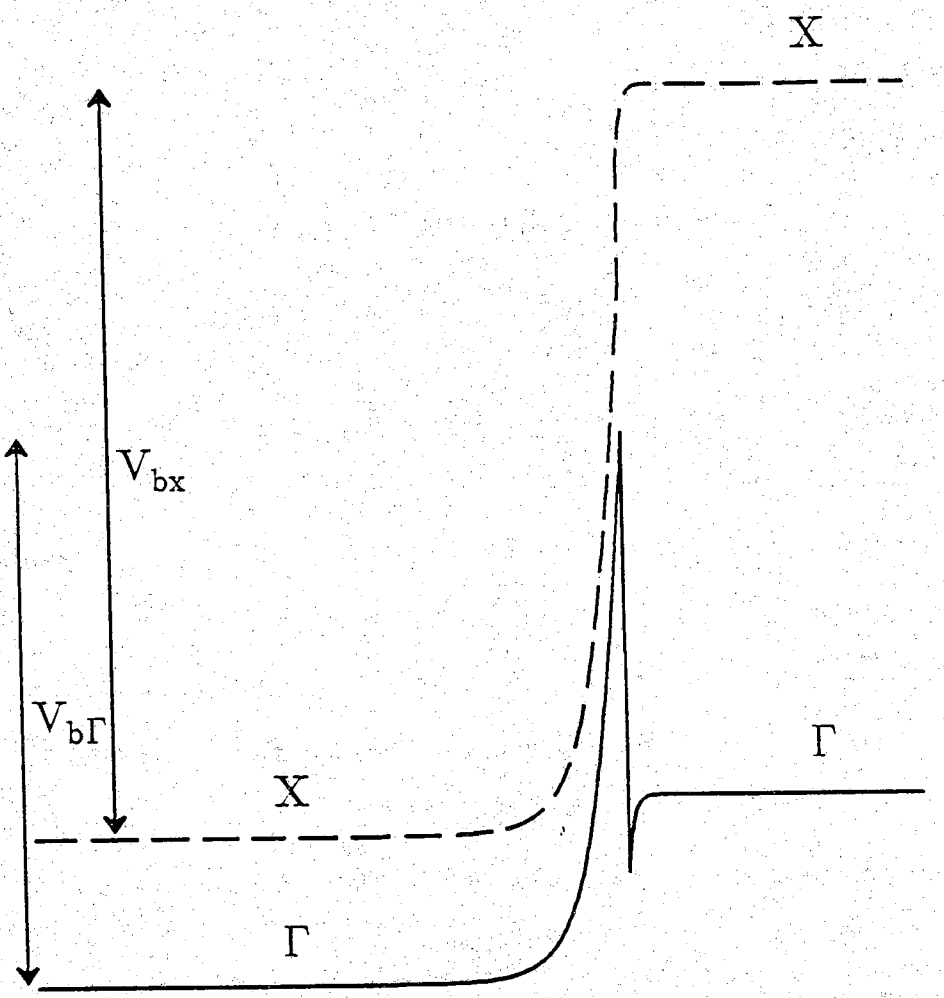


Figure 2.1(a) The band diagrams of Γ and X valley, of an abrupt AlGaAs/GaAs Np heterojunction. The emitter doping is $1.0\times10^{18}/cm^3$. Bias is fixed at 1.2 V. The relevant material parameters are listed in Table 2.1.

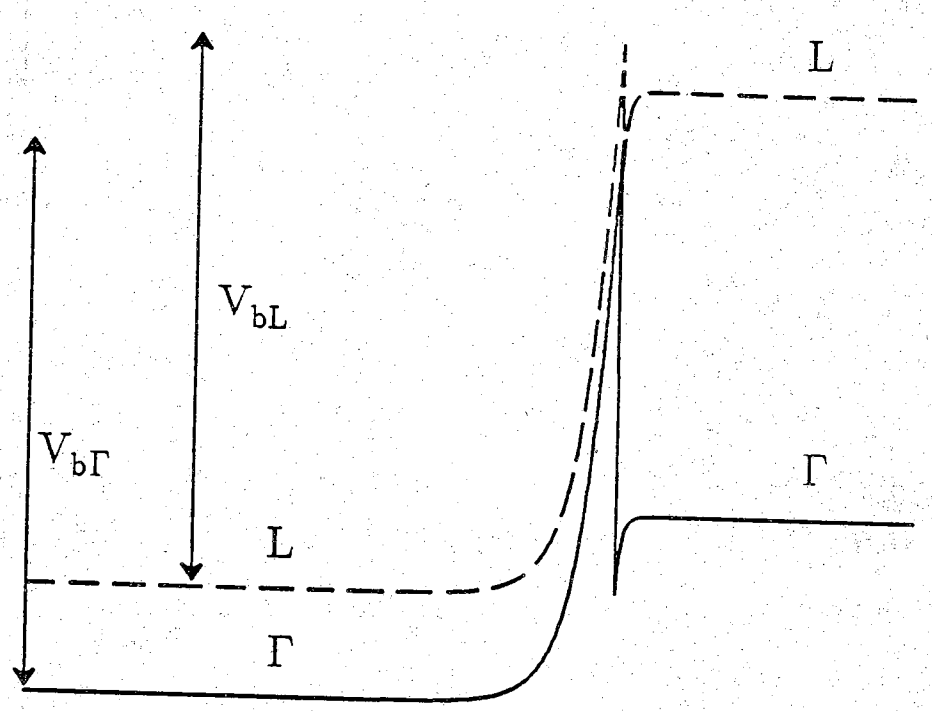


Figure 2.1(b) The band diagrams of Γ and L valley, of an abrupt AlGaAs/GaAs Np heterojunction. The emitter doping is $1.0\times10^{18}/cm^3$. Bias is fixed at 1.2 V. The relevant material parameters are listed in Table 2.1.

should stress that tunneling which enhances filtering in the first case and is responsible for filtering in the second case, also lowers the average energy of injected Γ-valley electrons. The design of abrupt emitter-base heterojunction involves a trade-off; enhanced tunneling improves filtering but reduces the effectiveness of the heterojunction launching ramp. The purpose of this paper is to examine this trade-off quantitatively.

The paper is organized into three sections. In the next section, the simulation techniques are described briefly. In section three, we describe and discuss the results of simulations of various HBT structures. Finally, the paper ends by summarizing the trade-offs involved in designing the emitter-base junction to minimize base transit time.

2.2 The Simulation Approach

To estimate the base transit time, carrier injection across the emitter-base heterointerface has to be considered in series with the transport of injected carriers across the quasi-neutral base. Electron transport across the heterointerface determines both the energy distribution of carriers injected into the base and the composition of the electron flux (the percentage of electrons in different valleys). We measure the extent of valley filtering in terms of the flux ratio which is defined as the ratio of the Γ-valley electron flux to the sum of the electron fluxes in Γ, L and X valleys. Once the energy distribution of the injected electron flux is found, the transport of those carriers through the quasi-neutral base is studied to estimate the average base transit time.

The electron injection (from Γ, L and X valleys of the AlGaAs emitter to the respective Γ, L and X valleys of the GaAs base) across the abrupt emitter-base heterojunction is treated quantum mechanically by numerically solving Schroedinger's equation across the heterojunction as described in [6]. It was necessary to treat all three valleys because high mole fraction $Al_xGa_{1-x}As$ emitters contain significant proportion of Γ, L and X-valley electrons. The energy band profile for each of the three valleys was first

obtained from a conventional numerical simulation program [13]. Across the heterojunction, a Γ-valley discontinuity of 65% of the Γ-valley bandgap difference was assumed [12]. From the resulting conduction band profiles, such as those showed in Figs. 2.1(a) and 2.1(b), we then computed electron current injected into the base by assuming that the emitter contact launched electron waves which propagated without scattering through the structure. Since the probability of electron tunneling from one valley to a different valley across a heterojunction is small [14-15], we treated the process of electron injection across the heterojunction separately for Γ, L and X valleys. The significant assumption underlying the numerical model is the neglect of scattering (scattering is neglected only when the electron injection across the HJ is considered) which is not expected to comprise the accuracy as long as the distance over which the potential varies by $2k_BT$ is greater than a mean free path [16].

After computing the electron flux injected into the base, a standard ensemble Monte Carlo program [17-18] was used to study the steady-state transport of electrons through the base. The initial energy of the electrons was chosen from the quantum mechanically computed energy distribution of the injected flux. The electron trajectories were then followed as they traversed the base under the influence of the scattering potentials. The first Monte Carlo simulations treating the difficult problem of minority carrier electron scattering in p^+ GaAs were recently reported, but a number of uncertainties remain [19]. For our work, we employed a simple treatment of electron-hole scattering which has previously been used for Monte Carlo simulations of HBT's [9]. This approach should serve well to illustrate the nature of the design trade-offs involved.

Scattering of electrons by hole plasmons was calculated after [20] with a cut-off wave vector taken to be the half of inverse of Debye length [21]. Overlap factors and corrections due to non-parabolicity were taken into account appropriately. We

neglected the coupling between hole plasmons and polar optical phonons which should be considered when the plasmon frequency and longitudinal optical phonon frequency are close. For the doping we chose to simulate, the heavy hole plasmon frequency was higher than the optical phonon frequency. So we expect that corrections to the scattering rate due to plasmon-phonon coupling will not significantly change our conclusions [9].

Due to high density of holes in the base, polar optical phonon (POP) scattering should be screened [9]. Figure 2.2 compares the screened POP scattering rate and the plasmon scattering rate. It is clear that plasmon scattering is the dominant mechanism of inelastic scattering for the minority electrons in the p-type base. Scattering of minority electrons by heavy holes was also taken into account by treating the heavy holes fixed in position like ionized impurities [22]. Strictly speaking electron-hole scattering is not purely elastic in nature and energy transfer from electrons to the hole system due to intra and inter-valence-band transitions should be taken into account [19,23]. However the rigorous treatment of electron hole scattering is complicated to implement in Monte Carlo simulations and is by itself a subject of some very recent studies [19]. In our case, the approximate treatment is not expected to change the conclusions of this paper since the primary mechanism of energy loss of minority electrons is through plasmon emission.

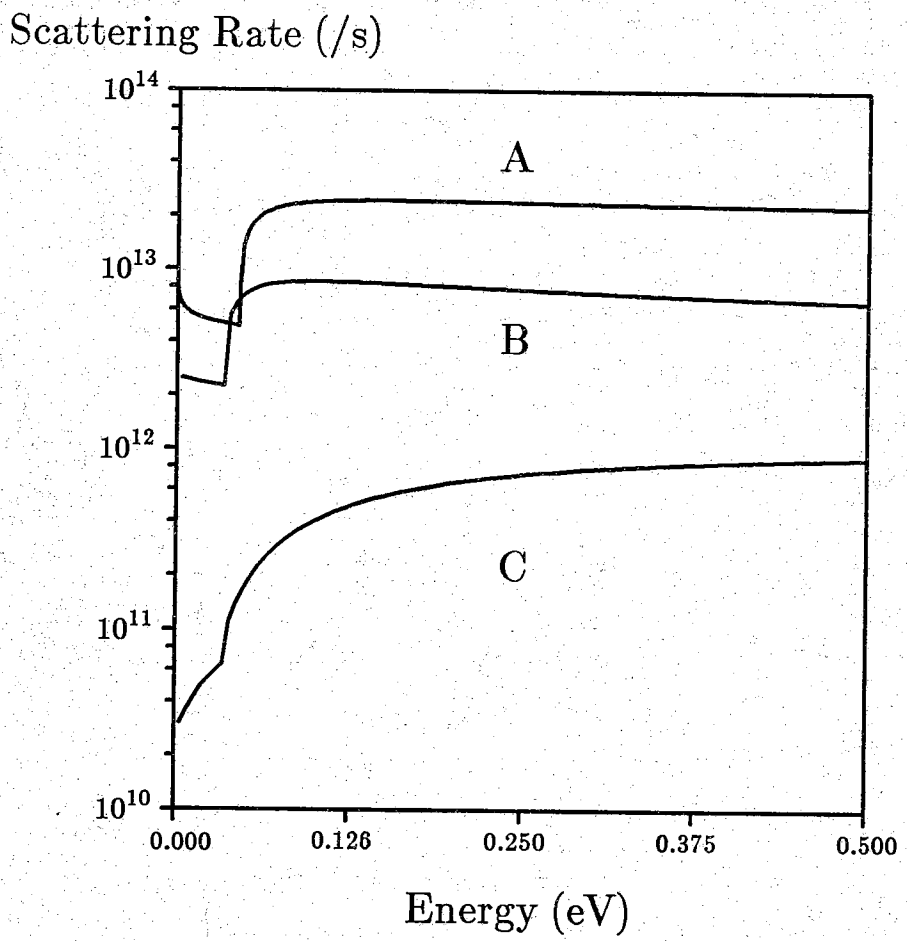


Figure 2.2 Curve A: Plasmon scattering rate *versus* electron energy in the GaAs base. Curve B: Unscreened Polar Optical Phonon scattering rate *versus* electron energy in the GaAs base. Curve C: Screened Polar Optical Phonon scattering rate *versus* electron energy in the GaAs base. The base is doped at $1.0 \times 10^{19}/cm^3$.

2.3 Simulation Results and Discussions

Details of the HBT structures that were simulated are displayed in Table 2.1. The simulations were made with emitter dopings of $1.0\times10^{17}/cm^3$ and $1.0\times10^{18}/cm^3$ at an emitter-base bias of 1.2 V. We begin by discussing the injection of electrons across the emitter-base heterojunction.

	LAYER	THICKNESS	DOPING
		Å	cm^{-3}
Emitter	N $Al_{.35}Ga_{.65}As$	2000	-
Base	p GaAs	500	1.0×10^{19}
Collector	n GaAs	3000	1.0×10^{17}

Table 2.1 The details of the HBT structure used in the simulation.

Figures 2.3(a) and 2.3(b) are plots of the flux of Γ-valley electrons injected into the base *versus* energy of the electrons for two different emitter dopings. The height of the conduction band spike ΔE_c is 0.283 eV. Figures 2.3(a) and 2.3(b) show that tunneling is the dominant mechanism of carrier transport in the two HBT structures we considered, since the average energy of the injected carriers is substantially lower than the height of the conduction band spike, ΔE_c. For the highly doped emitter, the average energy of the injected Γ-valley electrons is about 0.1 eV, whereas for the lightly doped

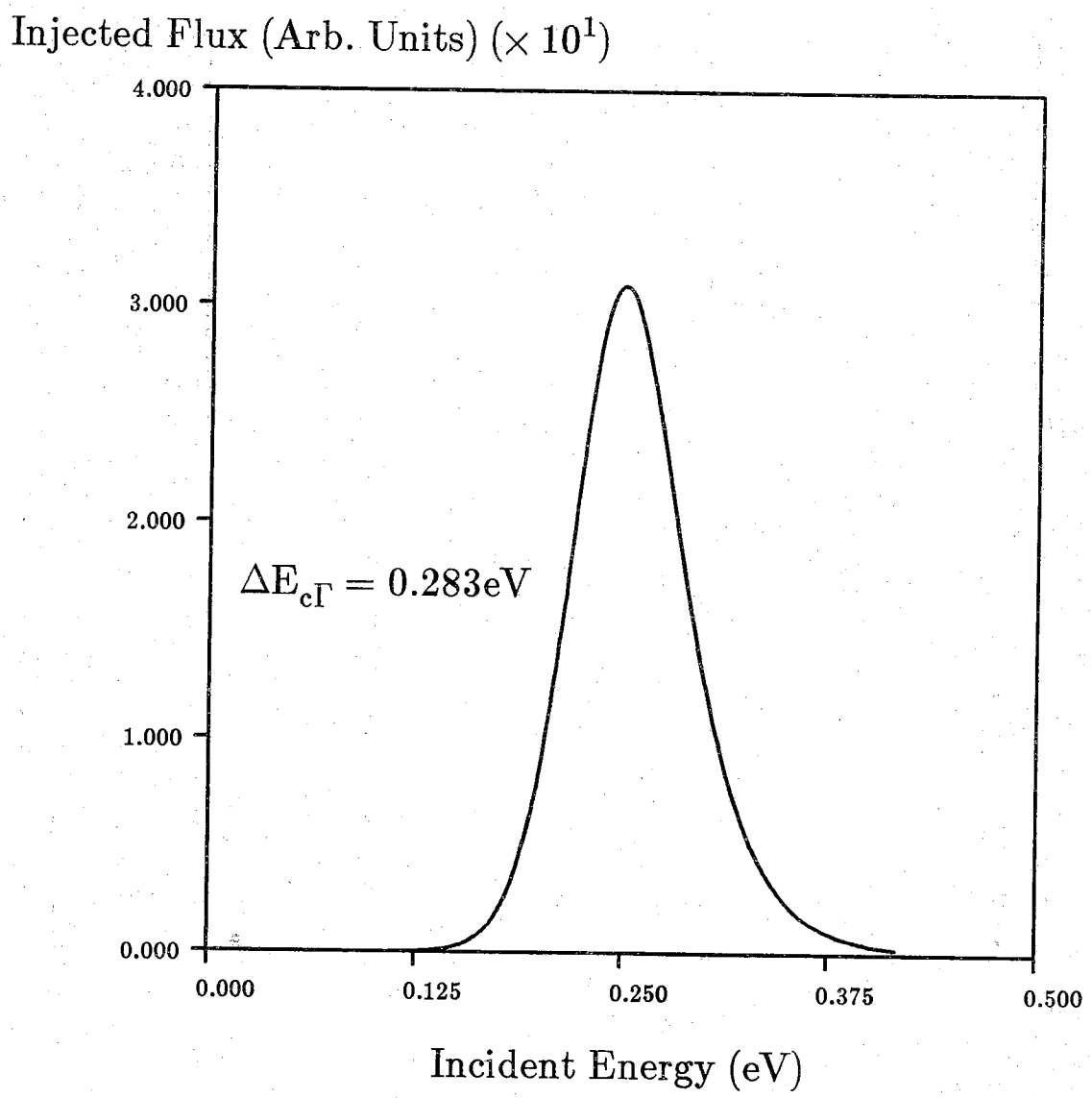


Figure 2.3(a) Injected Flux into the base from the emitter *versus* incident energy with emitter doped at $1.0\times10^{18}/cm^3$.

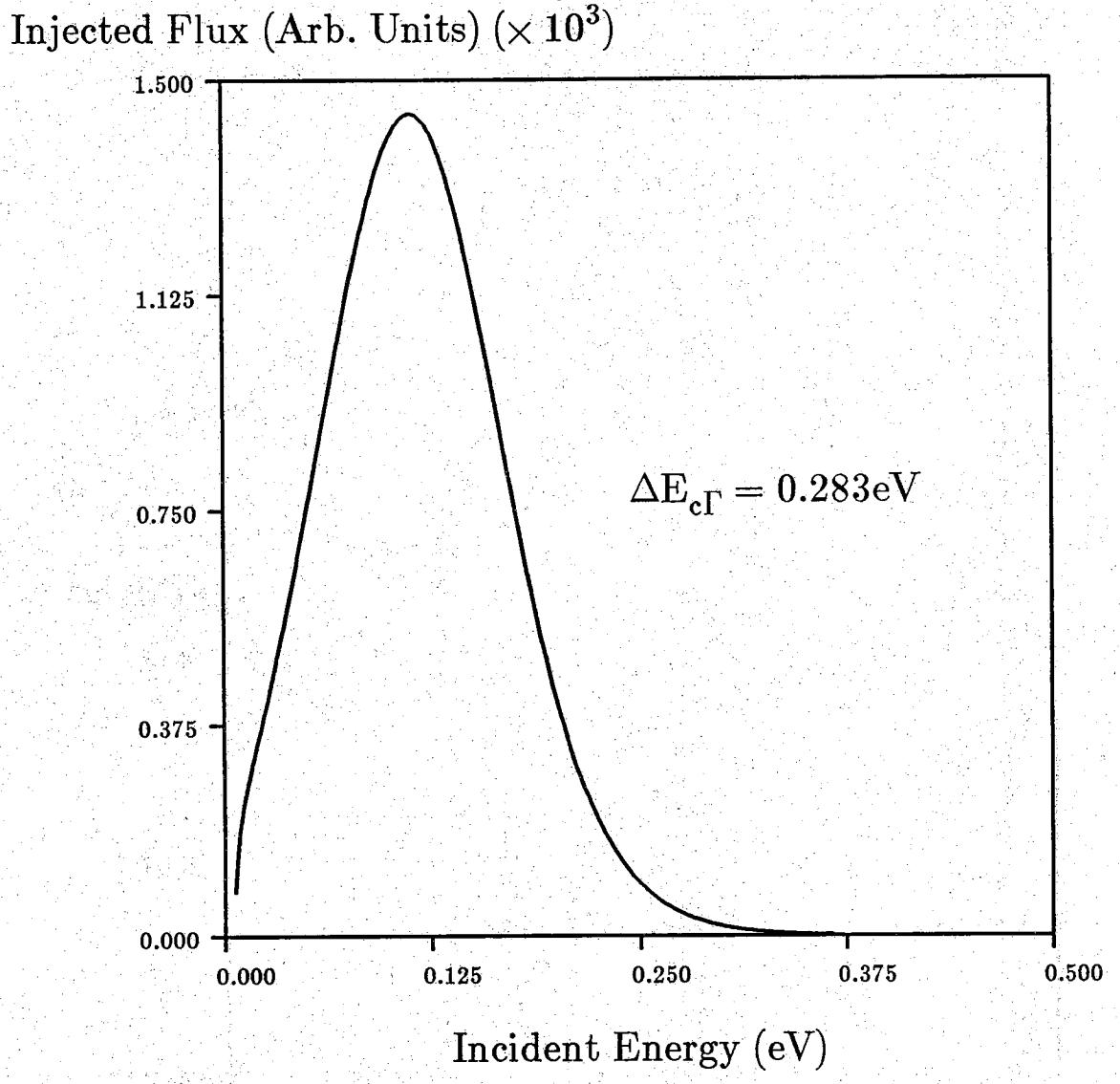


Figure 2.3(b) Injected Flux into the base from the emitter *versus* incident energy with emitter doped at $1.0\times10^{17}/cm^3$.

emitter it is about 0.25 eV. This difference can be explained by the fact that for the highly doped emitter, the width of the band spike is narrower which enhances tunneling and reduces the average energy of the injected flux. On the other hand, the enhanced tunneling produces an injected flux that is richer in Γ-valley electrons which is beneficial for base transport.

Figure 2.4(a) displays the variation of flux ratio with aluminum mole fraction in $Al_xGa_{1-x}As$ emitter. The number of upper valley electrons increases with the mole fraction and reduces the flux ratio in the emitter. Because of increased valley filtering, however, the flux injected into the base is rich in Γ-valley electrons; the injected flux ratio is nearly unity and is almost independent of the emitter composition. A high mole fraction in the emitter is beneficial because it increases the average energy of injection as displayed in Fig. 2.4(b).

Next we examine how the emitter-base bias affects the valley filtering. An increase in the emitter base bias reduces the tunneling current through the conduction band spike because the spike widens with bias. The reduced tunneling current for the Γ-valley electrons degrades the flux ratio after the junction. In Figs. 2.5(a) and 2.5(b), we plot the flux ratio *versus* emitter-base bias for two different emitter dopings. For a highly doped emitter, the decrease in the flux ratio is negligible with bias, but for a lightly doped emitter the proportion of upper valley electron flux is found to increase considerably with bias. The difference in the behavior with bias can be explained by examining the Fig. 2.5(c), which shows the ratio of the thermionic emission component to the total current for the two HBT's. For the lightly doped emitter, the tunneling current, which provides the filtering effect, decreases more rapidly with bias which decreases the population of Γ-valley electrons at high bias.

To illustrate how upper valley electrons increase the base transit time, we conducted the following Monte Carlo simulation. First, the electron flux from an emitter

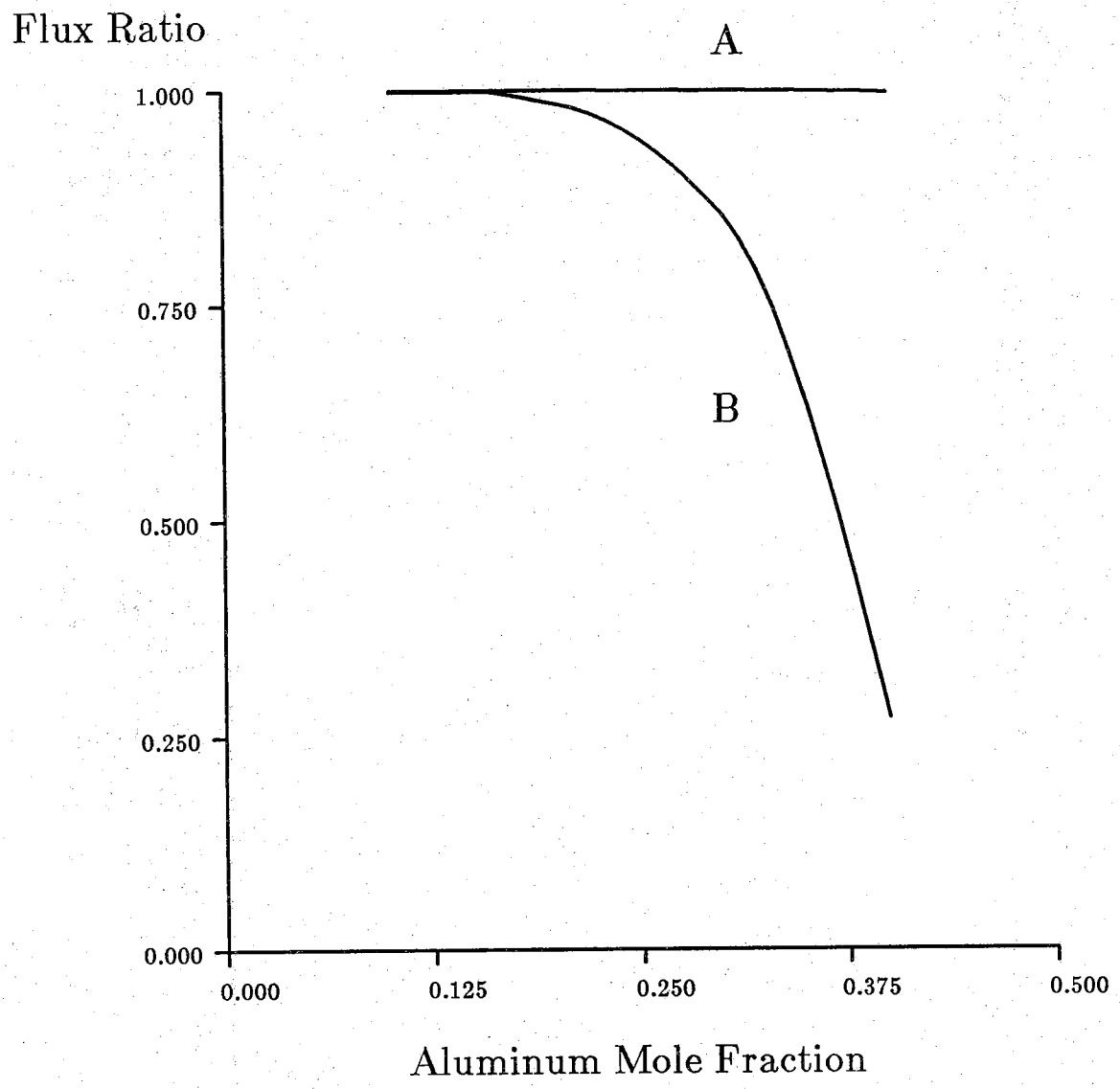


Figure 2.4(a) Dependence of flux ratio on the aluminum mole fraction of the AlGaAs emitter. A: Injected flux; B: Bulk emitter flux. The GaAs base is doped p-type at $1.0\times10^{19}/cm^3$. The AlGaAs emitter is doped n-type at $1.0\times10^{18}/cm^3$.

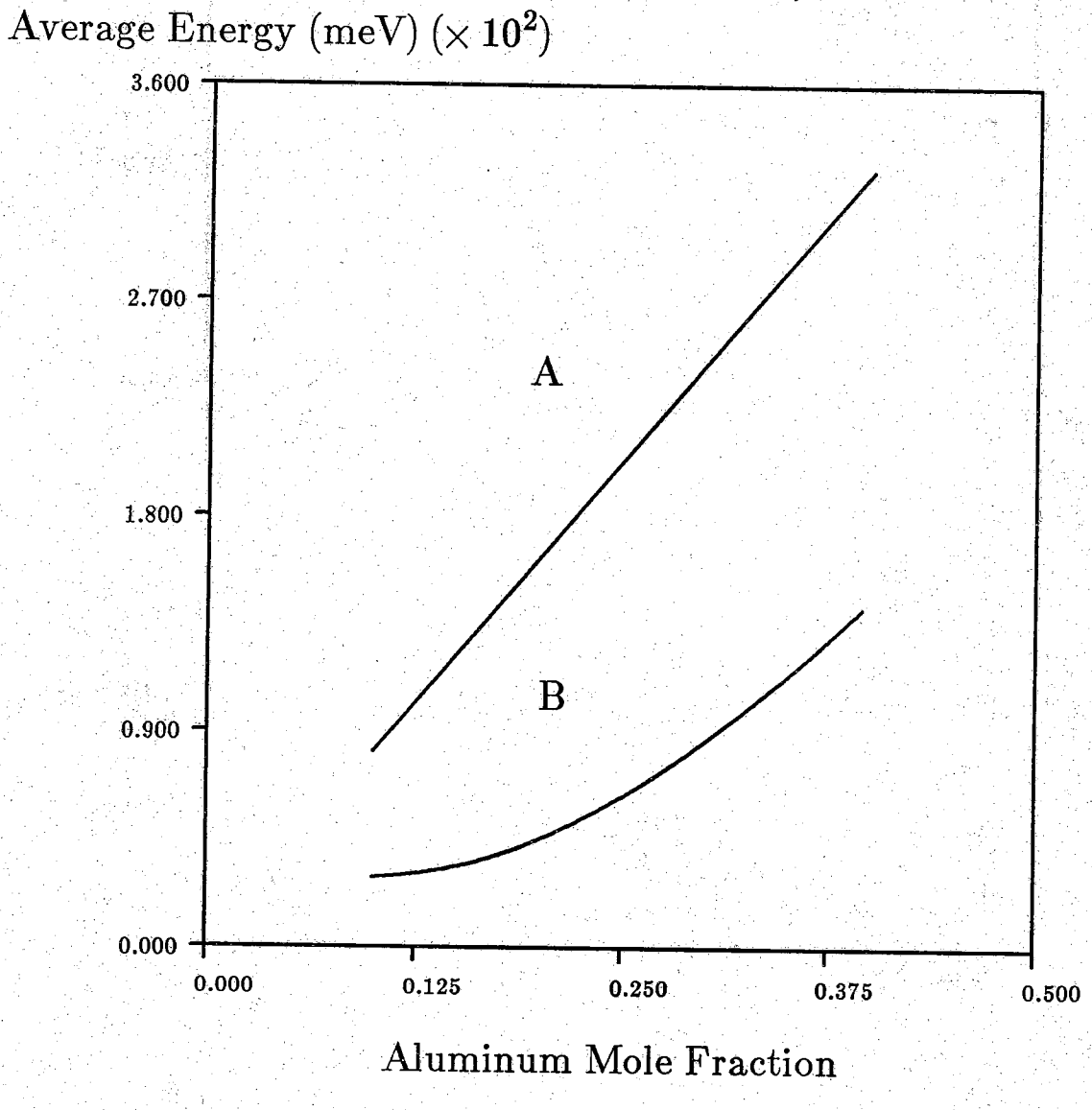


Figure 2.4(b) A: $\Delta E_{c\Gamma}$ *versus* aluminum mole fraction in the emitter. B: Dependence of average energy of the electrons injected into the base from the emitter on the aluminum mole fraction of the emitter. GaAs base is doped p-type at $1.0\times10^{19}/cm^3$. The AlGaAs emitter is doped n-type at $1.0\times10^{18}/cm^3$.

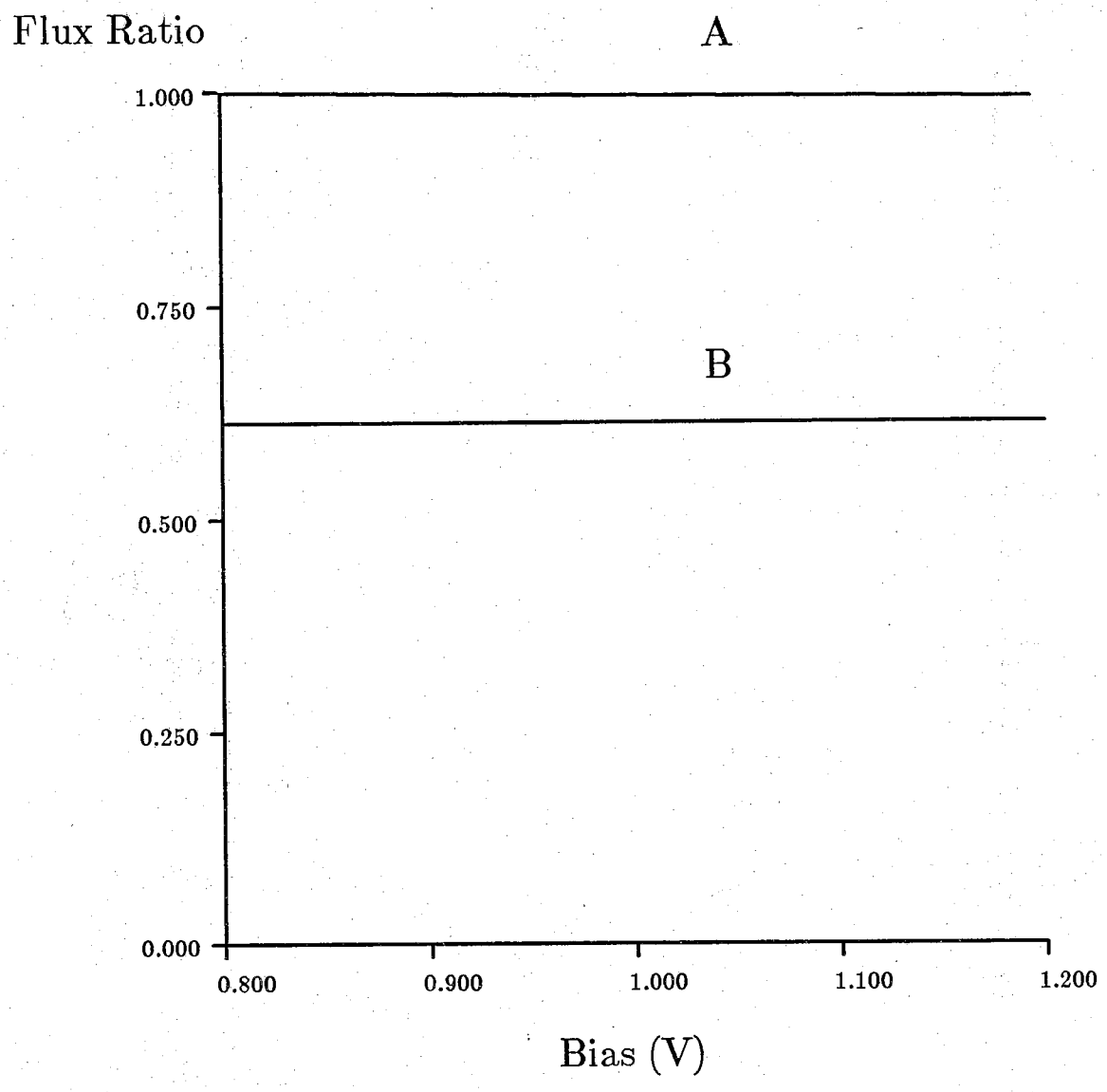


Figure 2.5(a) Dependence of flux ratio on the emitter base bias for an aluminum mole fraction of 35% with an emitter doping of $1.0\times10^{18}/cm^3$. A: Injected flux, B: Bulk emitter flux.

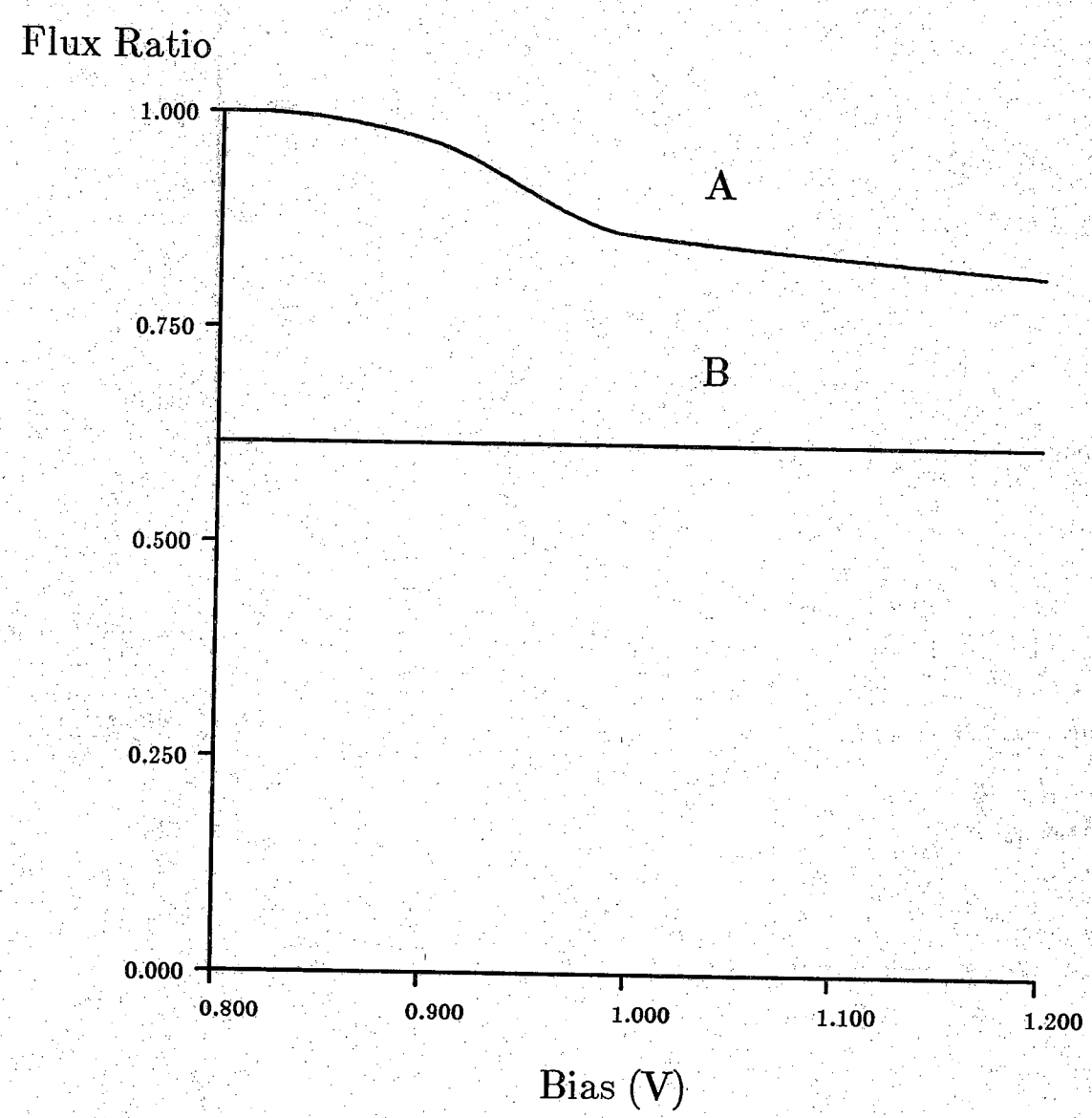


Figure 2.5(b) Dependence of flux ratio on the emitter base bias for an aluminum mole fraction of 35% with an emitter doping of $1.0\times10^{17}/cm^3$. A: Injected flux, B: Bulk emitter flux.

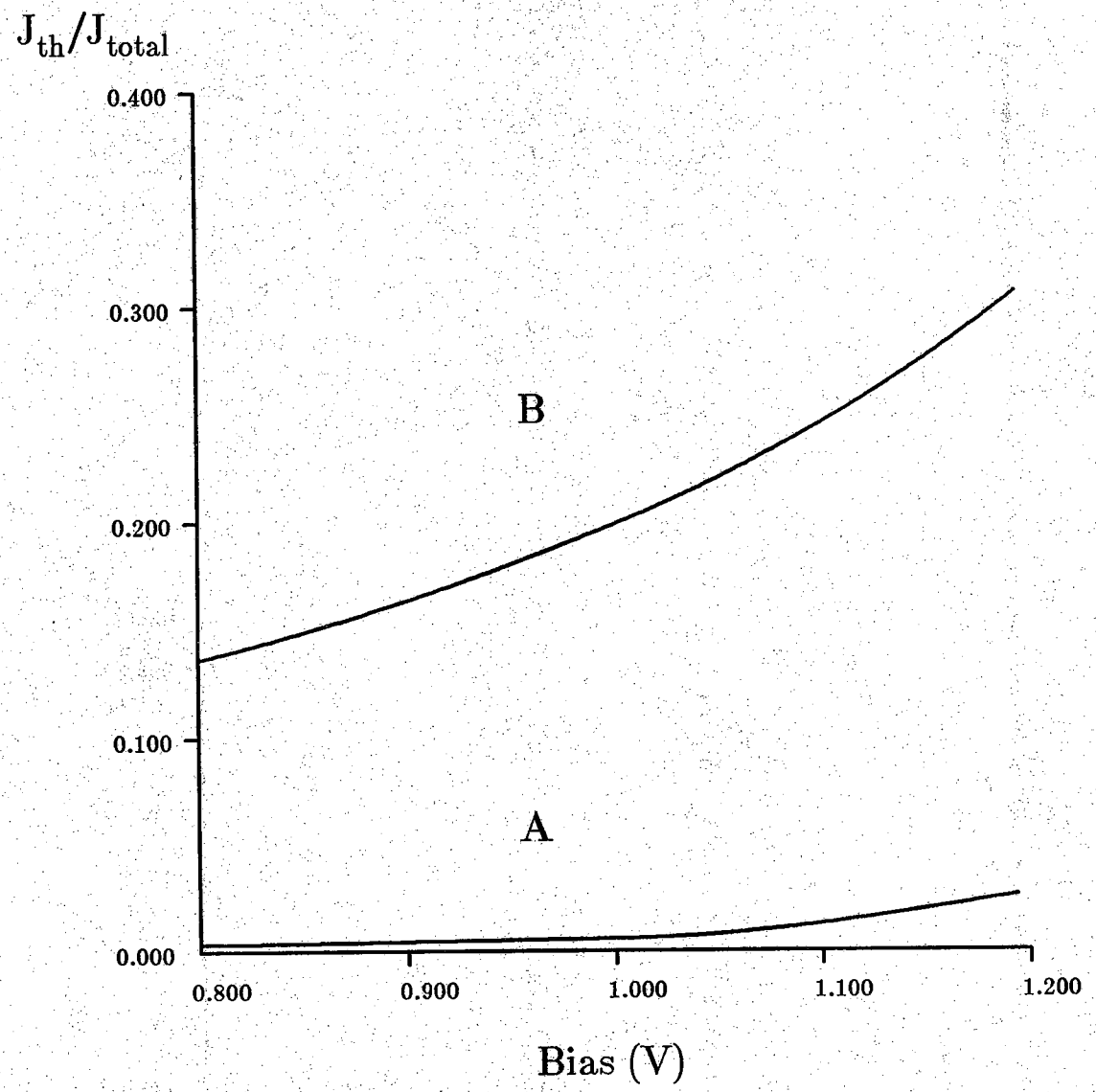


Figure 2.5(c) The proportion of thermionic emission current in the total current across the emitter base junction for A: an emitter doping of $1.0\times10^{18}/cm^3$, B: an emitter doping of $1.0\times10^{17}/cm^3$.

Sl. No.	INJECTION ENERGY	EMITTER DOPING	FILTER RATIO	TRANSIT TIME	INT. VALLEY SCATTERING
	meV	cm^{-3}		pS	%
1	99.0	1.0×10^{18}	0.99	0.36	0.32
2	99.0	1.0×10^{18}	0.76	0.66	8.47
3	250	1.0×10^{17}	0.86	0.28	6.67

Table 2.2 The details of the simulation of carrier transport in the base. The width of the p-type base is 500Å and doping $1.0\times10^{19}/cm^{-3}$. The Aluminum fraction of the emitter is 35%. The emitter-base bias is 1.2 V. Injection energy implies average energy of the injected flux from the emitter to the base.

doped at $1.0\times10^{18}/cm^{-3}$ was injected into a 500Å wide base, doped $1.0\times10^{19}/cm^{-3}$. The flux distribution was that found by the quantum mechanical treatment described earlier. Next, we injected an unfiltered electron flux (the proportion of upper valley electrons in the flux was exactly same as it was in the emitter contact) from an energy ramp whose height was equal to the average longitudinal energy of the quantum mechanically computed flux. The results of these simulations are presented in Table 2.2, in row 1 and 2. The base transit time for the unfiltered flux was found to be twice that of filtered flux. The increase in the base transit time is mainly due to an increase in intervalley scattering rate (from 0.32% in case of filtered flux to 8.47% in case of unfiltered

flux) which randomizes the momentum and reduces the average velocity of the carriers passing through the base. It should be understood that the above example doesn't represent any physical situation, but it clearly displays the influence of effective mass filtering on the base transit time.

Next, we examined carrier transport across the same 500Å wide base but the carriers were injected from emitters with two different dopings, $1.0\times10^{18}/cm^{-3}$ and $1.0\times10^{17}/cm^{-3}$ at an emitter-base bias of 1.2V. The results are presented in Table 2.2, row 1 and row 3 respectively. The base transit time for the lightly doped emitter is shorter than for the highly doped emitter case. Enhanced tunneling in the heavily doped emitter provides better filtering but at the same time it reduces the average energy of injected carriers which increases the base transit time. These results demonstrate that the reduction in the base transit time achieved due to better effective mass filtering should be carefully weighed against the increase in base transit time due to the reduced energy of injected flux.

2.4 Conclusions

The conclusions of this study can be summarized in two respects, firstly the valley filtering across the heterojunction and second its influence on the base transit time. This study showed that due to different barrier heights for Γ, L and X-valley electrons, the filtering of injected flux at the emitter-base heterojunctions occurs even in the absence of tunneling. Tunneling additionally improves the filtering effect by allowing Γ-valley electrons to strongly tunnel to the base. Tunneling, however, reduces the average energy of the injected flux. Our study also showed that the filtering of upper valley electrons is almost independent of the mole fraction of aluminum in the emitter.

The design of the emitter-base junction to minimize the base transit time involves a compromise; a junction designed to enhance tunneling of carriers will increase the

effective mass filtering which is beneficial, but will also lower the average energy of the injected flux which reduces the effectiveness of the heterojunction launching ramp. The best way to increase the average energy of the filtered electron flux is to use a higher aluminum mole fraction in the emitter. Our study showed that a moderately doped emitter ($\sim 10^{17}$) with high aluminum mole fraction (in the range of 30% to 40%) would have the shortest base transit time. Our simulations also established that a highly doped emitter provides the best filtering effect but its base transit time is longer compared to the lightly doped case due to the lowering of the average energy of the injected flux. The reduced emitter-base junction capacitance is another advantage for the lightly-doped emitter.

Chapter 2 References

[1] H. Kroemer, *Proc. IEEE,* Vol. 71, 13 (1982).

[2] A. Marty, G. Rey, J. P. Bailbe, *Solid State Electronics,* Vol. 22, 549 (1979).

[3] N. Chand and H. Morkoc, *IEEE Tran. Electron Dev.,* Vol ED-32, 1064 (1985).

[4] D. Ankri, W. J. Schaff, C. E. C. Wood, L. F. Eastman, D. W. Woodard and L. Rathbun, *Inst. Phys. Cinf. Ser.,* Ser. 65, 431 (1983), Chap. 5.

[5] P. M. Enquist, L. P. Ramberg and L. F. Eastman, J. Appl. Phys., Vol. 18, 750 (1982).

[6] A. Das and M. S. Lundstrom, *IEEE Trans. Electron Dev.,* Vol. ED-35, 863 (1988).

[7] If γ is close to unity, β will be determined by the α since $\beta=\gamma\alpha/(1.-\gamma\alpha)$

[8] C. M. Maziar, M. E. Klausmeier-Brown, S. Bandyopadhyay, M. S. Lundstrom and S. Datta, *IEEE Trans. Electron Dev.,* Vol ED-33, 881 (1986).

[9] R. Katoh, M. Kurata and J. Yoshida, *Proceedings of IEDM,* IEDM, 11.2 (1987).

[10] H. Kroemer, *J. Vac. Sci. Technol.,* Vol 1, 126 (1083)

[11] L. P. Ramberg and T. Ishibashi, *J. Appl. Phys.,* Vol. 63, 809 (1988).

[12] There is some disagreement among researchers as to the exact percentage of Γ-valley bandgap discontinuity present in the conduction band. The different experimental figures presently range between 60% to 70%. See for example a) H. Kroemer et. al., Appl. Phys. Lett. **36**, 295 (1980), b) M. O. Watanebe et. al., Ext. Abs. 16th. Conf. on Solid State Devices and Materials, Kobe, 1984., c) J. Menedez et al., Solid State Commun. **61**, 601 (1987).

[13] M. S. Lundstrom and R.J. Schuelke, *IEEE Trans. Electron Dev.,* Vol. ED-**30**, 1151 (1983).

[14] G. C. Osbourn and D. L. Smith, *J. Vac. Sci. Technol,* Vol. 15(5), 1528 (1979).

[15] C. Mailhiot, D. L. Smith and T. C. Mcgill *J. Vac. Sci. Technol,* Vol B 1(3), 637 (1983).

[16] F. Berz, *Solid State Electronics,* Vol. 28, **1007** (1985).

[17] W. Fawcett, A. D. Boardman and S. Swain, *J. Phys. Chem. Solids,* Vol. 31, 1963 (1970).

[18] C. Jacoboni and L. Reggiani, *Rev. Modern Phys.,* Vol. 55, 645 (1983).

[19] K. Sadra, C. M. Maziar, B. G. Streetman and D. S. Tang *to appear in Appl. Phys. lett., 1988.*

[20] D. Bohm and D. Pines, *Phys. Rev.,* Vol. 92, 609 (1953).

[21] P. Lugli and D. K. Ferry, *IEEE Elec. Dev. Lett.,* Vol. EDL-6, 25 (1984)

[22] W. Walukiewicz, J. Lagowski, L. Jastrzebski and H. C. Gatos, *J. Appl. Phys.,* Vol 50, 5040 (1979).

[23] J. F. Young, P. Kelly, and N. L. Henry, *Phys. Rev. B,* Vol. 36, 4535 (1987)

Chapter 3

AN EFFICIENT METHOD FOR THE ANALYSIS OF ELECTRON WAVEGUIDES

An efficient boundary element method for obtaining the scattering parameters of any general two dimensional electron waveguide is discussed. This method has a computational cost only proportional to L^2, where L is the perimeter of the device to be modeled.

3.1 Introduction

Recently, devices have been built for the study of ballistic electron transport in which the elastic scattering from the geometry of the devices dominates that of the impurities [1,2]. These devices are made up of a network of guiding channels [1,3,4,5]. Such devices are expected to become more important in the future. This paper presents a set of powerful tools for analyzing electron waveguide networks in two dimensions. The extension to three dimensions is relatively straightforward.

The quantity of interest to device engineers and others is the conductance. Landauer [6] modeled one dimensional electron transport using a region of static potential sandwiched in between two reservoirs or contacts. The contacts are assumed to be perfectly absorbing so that any electron leaving a contact is either transmitted into the other contact or reflected back into the initial contact by the static potential in

between. The probabilities of transmission and reflection are calculated from Schrödinger's equation for the static potential in the middle. Each reservoir is assumed to be in local thermodynamic equilibrium. The dissipation of electronic energy expected in all resistors occurs in the contacts, not in between, since the potential between the contacts is static, allowing only elastic scattering in that region. Several generalizations to Landauer's original formula have been suggested for multimoded systems [7,8]. One of them, the two probe current voltage relation at finite temperature and applied bias is

$$I = \frac{2q}{h}\int dE \left[f(E) - f(E + qV_D)\right] \sum_{\alpha,\beta} |\, t_{\alpha,\beta}(E)\,|^2 \qquad (3.1)$$

where q is the electronic charge, E is the electronic energy, h is Planck's constant, f(E) is the Fermi-Dirac distribution function, and V_D is the applied bias (the difference in chemical potentials between the two contacts); the sum is over transverse modes on the left and right, β and α respectively [9,10]. The factor of two arises from the electron's spin degeneracy. The summand, $|\, t_{\alpha,\beta}(E)\,|^2$, is the probability that an electron in mode β on the left elastically scatters across the static potential into mode α on the right. The difference, $[f(E) - f(E + qV_D)]$, arises because the total current is the difference between the current entering the device from the left contact and the current entering the device form the right contact. Experiments on electron waveguide devices are usually conducted with small drain voltages, V_D, to minimize electron-electron scattering and at low temperatures to minimize electron-phonon scattering because both of these are inelastic events and cause the electron to lose its phase coherence. Under these conditions Eq. (3.1) may be simplified; the difference, $[f(E) - f(E + qV_D)] \cong -qV_D\ \partial f/\partial E$, and at zero temperature $\partial f/\partial E = -\delta(E - E_F)$ so that

$$G = \frac{I}{V_D} = \frac{2q^2}{h} \sum_{\alpha,\beta} |\, t_{\alpha,\beta}(E_F)\,|^2 \qquad (3.2)$$

Eq. (3.2) highlights the importance of the scattering parameters, $t_{\alpha,\beta}$, [11] in the calculation of the conductance.

Several methods have been employed to obtain the scattering parameters; they include the transfer matrix method [12], the recursive Green's function method [13], and abrupt junction methods [14,15]. The recursive Green's function method uses a tight binding technique that also allows for on site disorder, but has a computational cost proportional to L^4, where L is the length/width of the device to be modeled. It has been a very popular method since its introduction in 1981. The transfer matrix method provides for disorder and is very efficient, having a computational cost proportional to L^3; unfortunately, this method is numerically singular for all but small L. Another drawback is that all input and output waveguides must be in the same direction; for example, with this method it is not possible to obtain the scattering matrix of a 90 ° waveguide bend. Used in conjunction with Redheffer's rule for cascading scattering matrices this method can include disorder [16], has a cost proportional to L^3, and lacks the singularity problems of the transfer technique; unfortunately the restrictions on the direction of the input and output waveguides remain.

In Sections 3.2 and 3.3 the details of a very powerful and practical boundary element method are presented. It can be used for both eigenvalue problems and scattering problems. The method is numerically efficient; having a computational cost proportional to L^2 without disorder and proportional to L^3 with disorder (when used in conjunction with Redheffer's rule). There are no restrictions on the geometry of the junction to be solved with this method. In Section 3.4 a newly proposed quantum interference device is analyzed using the tools described in the previous sections (Fig. 3.3). This device shows a large conductance modulation.

3.2 The Boundary Element Method

3.2.1 General Method

Shown in Fig. 3.1 is a general junction. The boundary of this junction is described with a local coordinate system; $\vec{\ell}$ is directed in a counter clockwise sense along the boundary and $\vec{n}$ is directed outward, perpendicular to $\vec{\ell}$. The set of points on the boundary coincident with the ports are labeled P; the rest of the boundary points are labeled $\bar{P}$.

We match the wave function and its normal derivative across the port boundaries, P. On the remaining part of the boundary, $\bar{P}$, we apply boundary conditions to the wave function and its normal derivative; Dirichlet boundary conditions [17] are most commonly used here.

We use a single particle, single band, effective mass Hamiltonian. In the ports the potential is only a function of ℓ and the Hamiltonian is separable

$$H_{ports} = H_n + H_\ell \tag{3.3}$$

where

$$H_n \equiv \frac{p_n^2}{2m^*} = \frac{-\hbar^2}{2m^*}\frac{\partial^2}{\partial n^2} \tag{3.4}$$

and

$$H_\ell \equiv \frac{p_\ell^2}{2m^*} + V(\ell) = \frac{-\hbar^2}{2m^*}\frac{\partial^2}{\partial \ell^2} + V(\ell) \tag{3.5}$$

There the wave function may be written as

$$\Psi_\beta(\ell,n) = \frac{1}{\sqrt{k_\beta}}\chi_\beta(\ell)e^{-ik_\beta n} + \sum_\alpha S_{\alpha\beta}\frac{1}{\sqrt{k_\alpha}}\chi_\alpha(\ell)e^{ik_\alpha n} \tag{3.6}$$

where

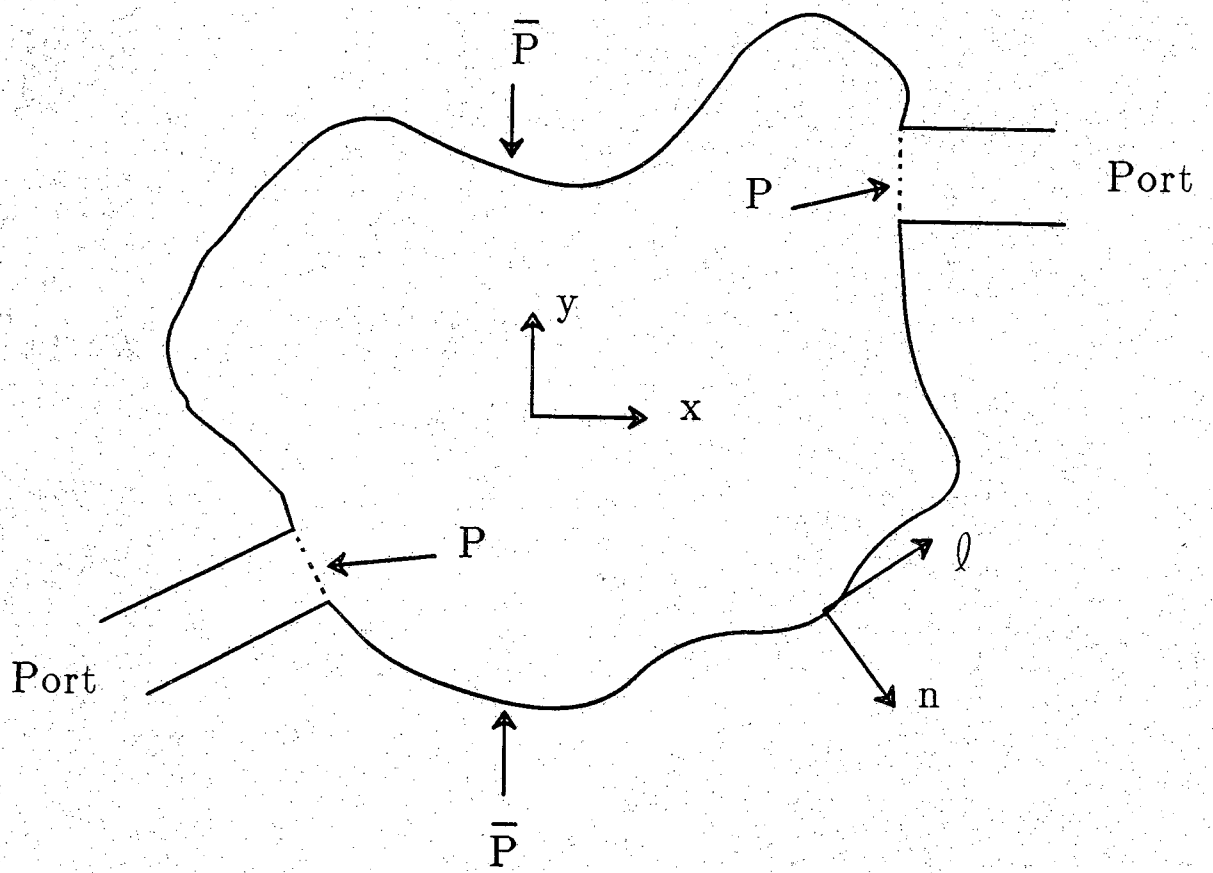


Figure 3.1 A general semiconductor waveguide junction.

$$H_\ell \chi_\alpha(\ell) = \in_\alpha \chi_\alpha(\ell) \tag{3.7}$$

and

$$E = \frac{\hbar^2 k_\alpha^2}{2m^*} + \in_\alpha \tag{3.8}$$

The $\chi_\alpha(\ell)$ are eigenmodes of the transverse Hamiltonian H_ℓ. The term $\chi_\beta(\ell)\, e^{-ik_\beta n}$ represents a wave incident on the boundary in mode β. The terms $S_{\alpha\beta}\chi_\alpha(\ell)e^{ik_\alpha n}$ represent waves scattered from the incident wave in mode β into out going waves in modes α; the factors $1/\sqrt{k_\beta}$ and $1/\sqrt{k_\alpha}$ are included to make each mode carry the same current. $\Psi_\beta(\ell,n)$ is the total wave function resulting from a wave incident on the junction in mode β.

Inside the boundary the Hamiltonian is

$$H_{in} = \frac{p^2}{2m^*} + V(x,y) = \frac{-\hbar^2 \nabla_{xy}^2}{2m^*} + V(x,y) \tag{3.9}$$

In order to apply the boundary element method, eigenfunctions, ϕ_μ, of the Hamiltonian, H_{in}, must be previously known, and a linear combination of them must be the solution, Ψ_β, inside and on the boundary, due to an incident wave from the outside in mode β

$$\Psi_\beta(x,y) = \sum_\mu c_{\mu\beta}\, \phi_\mu(x,y) \tag{3.10}$$

Only knowledge of ϕ_μ and $\partial\phi_\mu/\partial n$ on the boundary, $P \cup \overline{P}$, is required; this makes it convenient to map (x,y) to (ℓ,n). The total wave function and its normal derivative, on the boundary, due to a wave incident on P in mode β, may be expressed in terms of the eigenfunctions of H_{in}, ϕ_μ.

$$\Psi_\beta(\ell,n)\Big|_{n=0} = \sum_\mu c_{\mu\beta}\, \phi_\mu(\ell,n)\Big|_{n=0} \tag{3.11}$$

$$\left.\frac{\partial \Psi_\beta(\ell,n)}{\partial n}\right|_{n=0} = \sum_\mu c_{\mu\beta} \left.\frac{\partial \phi_\mu(\ell,n)}{\partial n}\right|_{n=0} \tag{3.12}$$

In the ports Ψ_β and $\frac{\partial \Psi_\beta}{\partial n}$ are continuous along $n = 0$ so that

$$\frac{1}{\sqrt{k_\beta}}\chi_\beta(\ell) + \sum_\alpha S_{\alpha\beta}\frac{1}{\sqrt{k_\alpha}}\chi_\alpha(\ell) = \sum_\mu c_{\mu\beta}\,\phi_\mu(\ell,0) \quad \text{for } \ell \in P \tag{3.13}$$

and

$$-i\sqrt{k_\beta}\chi_\beta(\ell) + \sum_\alpha i\sqrt{k_\alpha}S_{\alpha\beta}\chi_\alpha(\ell) = \sum_\mu c_{\mu\beta} \left.\frac{\partial \phi_\mu(\ell,n)}{\partial n}\right|_{n=0} \quad \text{for } \ell \in \bar{P} \tag{3.14}$$

Our objective is to obtain the scattering parameters, $S_{\alpha\beta}$, so we eliminate $c_{\mu\beta}$ in the next five steps [18]. Multiply Eq. (3.11) by $\frac{\partial \phi_\gamma}{\partial n}$ and then integrate along $\bar{P}$ to obtain

$$\int_{\bar{P}} \Psi_\beta \frac{\partial \phi_\gamma}{\partial n} d\ell = \sum_\mu c_{\mu\beta} \int_{\bar{P}} \phi_\mu \frac{\partial \phi_\gamma}{\partial n} d\ell \tag{3.15}$$

Multiply Eq. (3.12) by $-\phi_\gamma$ and then integrate along $\bar{P}$ to obtain

$$-\int_{\bar{P}} \phi_\gamma \frac{\partial \Psi_\beta}{\partial n} d\ell = -\sum_\mu c_{\mu\beta} \int_{\bar{P}} \phi_\gamma \frac{\partial \phi_\mu}{\partial n} d\ell \tag{3.16}$$

Multiply Eq. (3.13) by $\frac{\partial \phi_\gamma}{\partial n}$ and then integrate along P to obtain

$$\int_P \left(\frac{1}{\sqrt{k_\beta}}\chi_\beta + \sum_\alpha S_{\alpha\beta}\frac{1}{\sqrt{k_\alpha}}\chi_\alpha\right)\frac{\partial \phi_\gamma}{\partial n} d\ell = \sum_\mu c_{\mu\beta}\int_P \phi_\mu \frac{\partial \phi_\gamma}{\partial n} d\ell \tag{3.17}$$

Multiply Eq. (3.14) by $-\phi_\gamma$ and integrate along P to obtain

$$\int_P \left(i\sqrt{k_\beta}\chi_\beta - \sum_\alpha i\sqrt{k_\alpha}S_{\alpha\beta}\chi_\alpha\right)\phi_\gamma d\ell = -\sum_\mu c_{\mu\beta}\int_P \phi_\gamma \frac{\partial \phi_\mu}{\partial n} d\ell \tag{3.18}$$

Now add (3.15), (3.16), (3.17) and (3.18). The result is

$$\int_{\bar{P}}(\Psi_\beta \frac{\partial \phi_\gamma}{\partial n} - \phi_\gamma \frac{\partial \Psi_\beta}{\partial n})\, d\ell +$$

$$\int_{P}\{(\frac{1}{\sqrt{k_\beta}}\chi_\beta + \sum_\alpha S_{\alpha\beta}\frac{1}{\sqrt{k_\alpha}}\chi_\alpha)\frac{\partial \phi_\gamma}{\partial n} + \phi_\gamma(i\sqrt{k_\beta}\chi_\beta - \sum_\alpha i\sqrt{k_\alpha}S_{\alpha\beta}\chi_\alpha)\}d\ell$$

$$= \sum_\mu c_{\mu\beta} \oint(\phi_\mu \frac{\partial \phi_\gamma}{\partial n} - \phi_\gamma \frac{\partial \phi_\mu}{\partial n})d\ell = 0 \qquad (3.19)$$

The integral, $\oint(\phi_\mu \frac{\partial \phi_\gamma}{\partial n} - \phi_\gamma \frac{\partial \phi_\mu}{\partial n})d\ell$, is shown to vanish in the appendix. Eq. (3.19) is the set of integral equations we solve; it includes one equation for each value of γ. We wish to cast it in the form of a matrix equation for the scattering parameters, $S_{\alpha\beta}$, and the unknown linear combination of Ψ_β and $\partial\Psi_\beta/\partial n$ on $\bar{P}$. Ultimately we will solve the matrix equation corresponding to Eq. (3.19) on the computer. Rearranging terms leads to

$$\sum_\alpha S_{\alpha\beta}\int_P \chi_\alpha(\frac{1}{\sqrt{k_\alpha}}\frac{\partial \phi_\gamma}{\partial n} - i\sqrt{k_\alpha}\phi_\gamma)d\ell + \int_{\bar{P}}(\Psi_\beta \frac{\partial \phi_\gamma}{\partial n} - \phi_\gamma \frac{\partial \Psi_\beta}{\partial n})d\ell$$

$$= -\int_P \chi_\beta(\frac{1}{\sqrt{k_\beta}}\frac{\partial \phi_\gamma}{\partial n} + i\sqrt{k_\beta}\phi_\gamma)\, d\ell \qquad (3.20)$$

At this point we have done two steps in the ports which we have not done along the rest of the boundary. One step was applying the boundary conditions (Eqs. (13-14)). We will postpone this step in $\bar{P}$ for as long as possible to make it as easy as we can to apply any type of boundary condition desired. The other step was expanding the wave function in terms of the normal modes of the waveguides (Eq. (3.5)). Because of this, the integrals over P in Eq. (3.20) involve only known functions, which allows us to write Eq. (3.20) in the partial matrix form

$$(\mathbf{Z}_L + \mathbf{Z}_0)\mathbf{S} + \int_{\bar{P}} (\Psi_\beta \frac{\partial \phi_\gamma}{\partial n} - \phi_\gamma \frac{\partial \Psi_\beta}{\partial n}) d\ell = \mathbf{Z}_L - \mathbf{Z}_0 \tag{3.21}$$

where

$$(\mathbf{Z}_L)_{\gamma\alpha} \equiv -\int_P i\sqrt{k_\alpha} \chi_\alpha \phi_\gamma d\ell \tag{3.22}$$

$$(\mathbf{Z}_0)_{\gamma\alpha} \equiv \int_P \frac{1}{\sqrt{k_\alpha}} \chi_\alpha \frac{\partial \phi_\gamma}{\partial n} d\ell$$

$$(\mathbf{S})_{\alpha\beta} \equiv S_{\alpha\beta}$$

Really we wish to express Eq. (3.20) completely as a matrix equation. We accomplish this by expanding $\Psi_\beta(\ell,0)$ and $\left.\frac{\Psi_\beta}{\partial n}\right|_{n=0}$ in a complete set which spans $\bar{P}$, $\xi_j(\ell)$.

$$\Psi_\beta(\ell,0) = \sum_j A_{j\beta} \xi_j(\ell) \quad \text{for } \ell \in \bar{P} \tag{3.23}$$

$$\left.\frac{\partial \Psi_\beta}{\partial n}\right|_{n=0} = \sum_j B_{j\beta} \xi_j(\ell) \quad \text{for } \ell \in \bar{P} \tag{3.24}$$

Using these expansions in Eq. (3.20) we obtain the equations for $S_{\alpha\beta}$, $A_{j\beta}$ and $B_{j\beta}$,

$$\sum_\alpha S_{\alpha\beta} \int_P \chi_\alpha \left(\frac{1}{\sqrt{k_\alpha}} \frac{\partial \phi_\gamma}{\partial n} - i\sqrt{k_\alpha} \phi_\gamma\right) d\ell$$

$$+ \sum_j \left\{A_{j\beta} \int_{\bar{P}} \xi_j \frac{\partial \phi_\gamma}{\partial n} d\ell - B_{j\beta} \int_{\bar{P}} \xi_j \phi_\gamma d\ell\right\}$$

$$= -\int_P \chi_\beta \left(\frac{1}{\sqrt{k_\beta}} \frac{\partial \phi_\gamma}{\partial n} + i\sqrt{k_\beta} \phi_\gamma\right) d\ell \tag{3.25}$$

and in matrix form

$$(\mathbf{Z_L} + \mathbf{Z_0})\mathbf{S} + \mathbf{MA} + \mathbf{NB} = \mathbf{Z_L} - \mathbf{Z_0} \tag{3.26}$$

where

$$(\mathbf{M})_{\gamma j} \equiv \int_{\overline{P}} \xi_j \frac{\partial \phi_\gamma}{\partial n} d\ell \tag{3.27}$$

$$(\mathbf{N})_{\gamma j} \equiv -\int_{\overline{P}} \xi_j \phi_\gamma d\ell$$

$$(\mathbf{A})_{j\beta} \equiv A_{j\beta}$$

$$(\mathbf{B})_{j\beta} \equiv B_{j\beta}$$

The next step is to apply the boundary conditions of the specific junction of interest to Eq. (3.26); below we examine several special cases.

3.2.2 All Port Junctions

If the boundary consists entirely of ports ($\overline{P}$ has measure zero) then $\mathbf{M} = 0$ and $\mathbf{N} = 0$. In this case the solution for the scattering matrix has a form familiar from transmission line theory

$$\mathbf{S} = (\mathbf{Z_L} + \mathbf{Z_0})^{-1}(\mathbf{Z_L} - \mathbf{Z_0}) \tag{3.28}$$

3.2.3 No Port Junctions

If the boundary has no ports so that $\mathbf{Z_L} = 0$ and $\mathbf{Z_0} = 0$, then Eq. (3.26) becomes [19]

$$\mathbf{MA} + \mathbf{NB} = 0 \tag{3.29}$$

Boundary conditions must be applied to Eq. (3.29). The most commonly used one is the Dirichlet boundary condition, which corresponds to hard walls caused by an infinite

potential barrier just outside the boundary. With this boundary condition $\mathbf{A} = 0$ and the eigenvalue equation (3.29) is reduced to

$$\det[\mathbf{B}] = 0 \tag{3.30}$$

The application of the boundary element method with no ports is discussed in Ref. 20.

3.2.4 The General Junction, an Analytical Example

As it stands, Eq. (3.26) allows the solution of a more general set of scattering problems where only part of the boundary corresponds to ports. We now provide an illustrative analytical example of this, applying Eq. (3.26) to a blocked waveguide (Fig. 3.2). The results of this example, though quite obvious from a much simpler analysis, are used in Section 4. The modal functions in the input port are

$$\chi_\alpha(y) = \left[\frac{2}{W}\right]^{1/2} \sin\left(\frac{\alpha\pi y}{W}\right) \tag{3.31}$$

We choose for ϕ_γ the eigenfunctions corresponding to $V(x,y) = 0$ which already satisfy the boundary conditions, $\Psi_\beta = 0$, along the top and bottom boundaries of the waveguide,

$$\phi_\gamma(x,y) = \chi_\gamma(y)\, e^{\pm ik_\gamma x} \tag{3.32}$$

This is just one of a number of possible choices for ϕ_γ [21]. Since the boundary conditions along the top and bottom boundaries are already met, all integrals over them vanish and we label the right end as $\overline{P}$ (Fig. 3.2) where we demand that $\Psi_\beta = 0$. We expand $\partial\Psi_\beta/\partial n$ in $\overline{P}$ as

$$\frac{\partial\Psi_\beta}{\partial n} = \sum_\mu B_{\mu\beta}\chi_\mu(y) \tag{3.33}$$

Using the identity

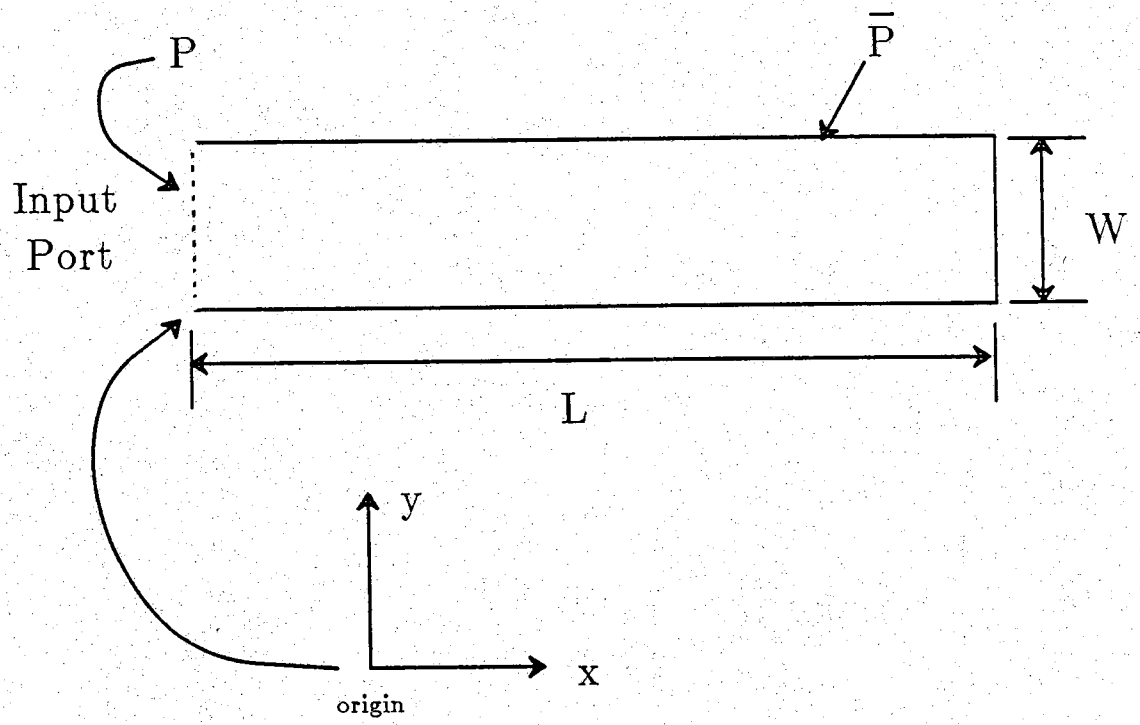


Figure 3.2 A blocked waveguide port.

$$\int_0^W \chi_\alpha(y)\chi_\beta(y)dy = \delta_{\alpha\beta} \tag{3.34}$$

along with Eqs. (3.22) and (3.27) we obtain

$$(Z_L)_{\gamma\alpha} = i\sqrt{k_\alpha}\delta_{\alpha\beta} \tag{3.35}$$

$$(Z_0)_{\gamma\alpha} = \pm i\sqrt{k_\alpha}\delta_{\alpha\beta}$$

$$(N)_{\gamma j} = -e^{\pm ik_\gamma L}\delta_{j\gamma}$$

Eq. (3.26) may be written in composite matrix form as

$$\begin{bmatrix} 2i\sqrt{k_\gamma}\delta_{\gamma\alpha} & -e^{ik_\gamma L}\delta_{\gamma\mu} \\ 0 & -e^{-ik_\gamma L}\delta_{\gamma\mu} \end{bmatrix} \begin{bmatrix} S_{\alpha\beta} \\ B_{\mu\beta} \end{bmatrix} = \begin{bmatrix} 0 \\ 2i\sqrt{k_\gamma}\delta_{\gamma\beta} \end{bmatrix} \tag{3.36}$$

The first row of Eq. (3.36) comes from the forward traveling waves of Eq. (3.28) and the second row from the backward traveling waves of Eq. (3.28). The solution for the scattering matrix is $S_{\alpha\beta} = -e^{i2k_\alpha L}\delta_{\alpha\beta}$ as could be expected. χ_α are normal modes of the waveguide, and so are uncoupled; each accumulates a phase shift of $e^{ik_\alpha L}$ on the way down, suffers a reflection of -1 (so that the wave function is zero at the end, $\overline{P}$) and then accumulates another phase shift of $e^{ik_\alpha L}$ on the way back. This simple example illustrates the basic method but not the power of the boundary element technique. In the next two sections we discuss points relating to a numerical solution of Eq. (3.26) and a non-trivial example.

3.3 Numerical Considerations

3.3.1 Truncation Approximation

The real power of Eq. (3.26) lies in its approximate numerical solution on a digital computer. With the proper choice of eigenfunctions, ϕ_γ, and expansion functions, ξ_j,

the series in Eq. (3.26) may be truncated to obtain an approximate equation for the most important scattering parameters.

$$\sum_{\alpha=1}^{\alpha_{max}}(\mathbf{Z}_L + \mathbf{Z}_0)_{\gamma\alpha}S_{\alpha\beta} + \sum_{j=1}^{j_{max}}\{(\mathbf{M})_{\gamma j}A_{j\beta} - (\mathbf{N})_{\gamma j}B_{j\beta}\}$$

$$= (\mathbf{Z}_L - \mathbf{Z}_0)_{\gamma\beta} \qquad \text{for } \gamma \in \{1,2,...,\gamma_{max}\} \tag{3.37}$$

Eq. (3.37) has the same matrix form given in Eq. (3.26). The scattering matrix, $\mathbf{S}$, is square and $\alpha_{max}/j_{max} \sim \text{length}(P)/\text{length}(\bar{P})$. The number of equations is $\gamma_{max} \geq (\alpha_{max} + j_{max})$ [22]. $\mathbf{Z}_L$ and $\mathbf{Z}_0$ have dimensions $(\gamma_{max} \times \alpha_{max})$; $\mathbf{M}$ and $\mathbf{N}$ have dimensions $(\gamma_{max} \times j_{max})$.

Care must be used in choosing the eigenfunctions, $\{\phi_\gamma \mid \gamma=1,2,...,\gamma_{max}\}$. The solution inside the device is a superposition of them and so only those eigenfunctions that are not important parts of the solution inside the device may be left out, for instance, in the previous example both positive and negative going waves must be included since the boundary, $\bar{P}$, couples these strongly, making them both important inside the device. For a general two dimensional geometry including the origin, with a constant potential, the cylindrical harmonics are a good choice.

$$\phi_\gamma = J_m(kr)\cos(m\theta + \gamma\frac{\pi}{2}) \tag{3.38}$$

where

$$k^2 = \frac{2m^*}{\hbar^2}(E - V) \tag{3.39}$$

and

$$m \equiv \text{Int}(\frac{\gamma+1}{2})$$

and J_m is the mth Bessel function of the first kind.

Care must also be exercised when choosing the set $\{\xi_j(\ell) \mid j=1,2,\ldots,j_{max}\}$ since $\Psi_\beta(\ell,0)$ and $\left.\frac{\partial \Psi_\beta}{\partial n}\right|_{n=0}$ are approximated by a linear combination of these $\xi_j(\ell)$. Ref. 20 gives two practical examples of this.

3.3.2 Network Concepts

As in any electrical network, the interconnection of elements is extremely important. Redheffer's star product rule [23], alluded to previously, yields the composite scattering matrix from the individual scattering matrices of the junctions making up the network and from the knowledge how they are connected.

Often network concepts may be used to reduce the number of computations necessary to solve a given problem. For example, if a parameter corresponding to one piece of a waveguide network is to be swept through a range of values, it is usually more efficient to break the network into two pieces, one which changes with the parameter being swept and another which is independent of the parameter. These two pieces are then connected using Redheffer's rule. The advantage stems from the fact that the junction which must be recalculated for each parameter value is smaller than the complete network. This concept was used in the next example.

It is quite practical to construct a library of scattering matrices for elemental junctions and use these to simulate very complex waveguide networks. If this is done it is important to remember that it is easy to seal a port off using Redheffer's rule, but it is difficult to open one up.

Another application of the network concept is to the problem of disorder. Since the boundary is the only place disorder may be introduced with the boundary element method, an artificial boundary is introduced which coincides with the disorder; the two scattering matrices corresponding to the two junctions resulting from this bisection are

each calculated using the boundary element method and then connected using Redheffer's rule [16].

3.4 A Numerical Example

The device of Fig. 3.3(a) is a unique kind of field effect transistor. The gate is not positioned between the source and the drain. It is a quantum interference transistor whose principle of operation is based on the wave interference between incident and reflected waves from the gate region. The conductance between the source and drain is modulated by changing the phase of the reflection coefficient at the gate.

3.4.1 Model

This device was modeled using the boundary element method to calculate the scattering matrix of a four way splitter (Fig. 3.3(b)). Redheffer's rule was used to block two of the ports by connecting them to scattering matrices calculated in the analytical example of Section 2, one with length, L, equal to zero and the other (the gate port) with a variable length, L, to obtain the two port junction of Fig. 3.3(c).

The device is three dimensional; however, it is uniform and single moded in z, so the two dimensional analysis may be applied with only a few minor modifications. Because the device is uniform in z, the z dependence of the wave function is the same throughout the device. Because we assumed hard walls and single modedness everywhere, this dependence has the form $\sin(\pi z/W_z)$. The confinement in the z dimension changes the dispersion relation, Eq. (3.8), to

$$E = \frac{\hbar^2 k_\alpha^2}{2m^*} + \in_\alpha + \frac{\hbar^2}{2m^*}\left(\frac{\pi}{W_z}\right)^2 \tag{3.40}$$

The energy, E, we used corresponds to the Fermi energy in either the source or the

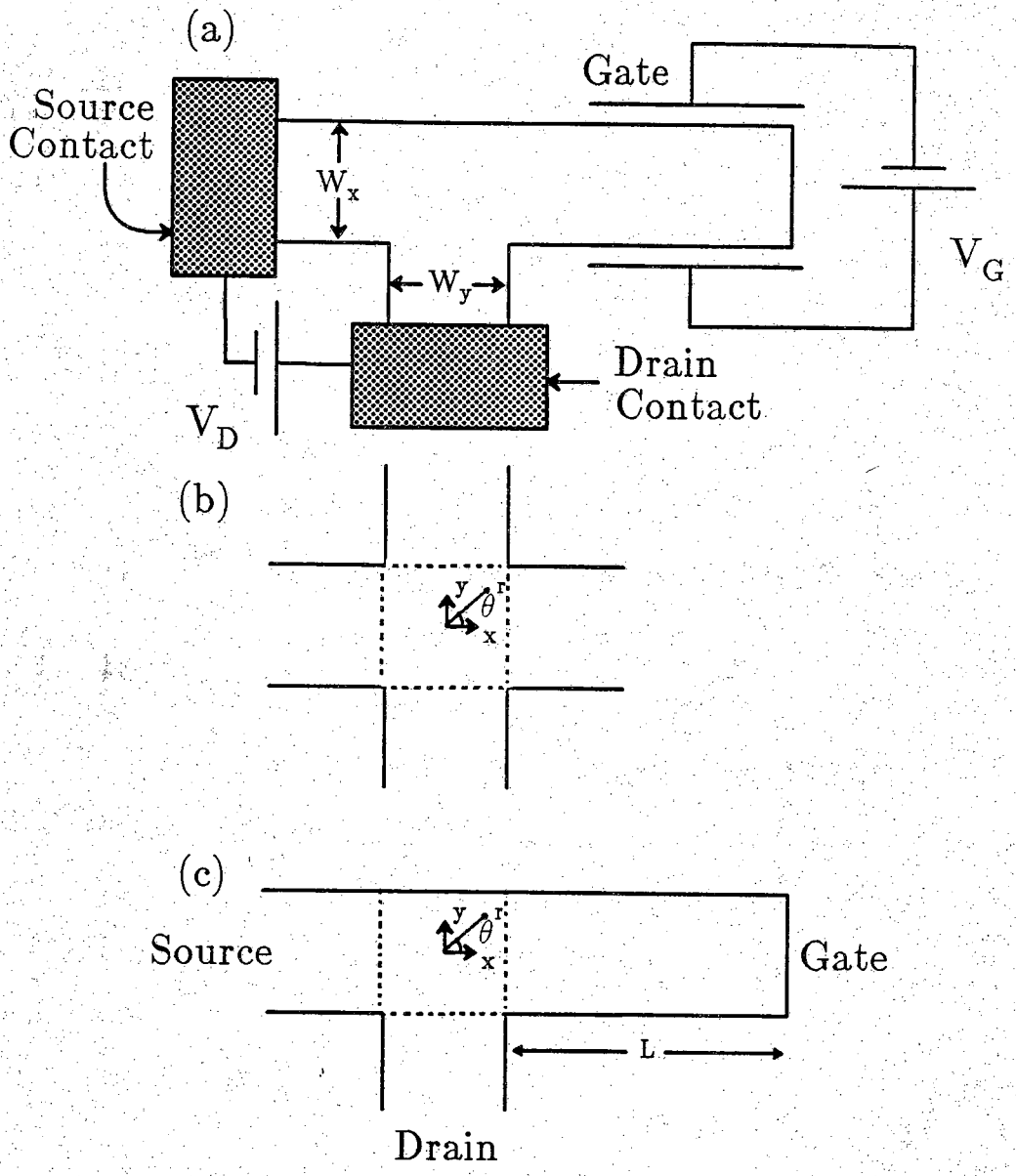


Figure 3.3 The quantum reflection field effect transistor.

drain waveguide at zero temperature with a surface carrier density of $6\times10^{11}cm^{-2}$. The width, W_z, was set at 100Å.

The eigenfunctions used were those of Eq. (3.38). Sinusoidal modal functions, χ_α, similar to those in the previous example were used in the waveguide ports. Eight modes were used in each of the four ports ($\alpha_{max} = 32$) and 32 eigenfunctions, ϕ_γ, were used inside ($\gamma_{max} = 32$). The matrix elements of Eq. (3.22) were computed by Rhomberg numerical integration.

The effective termination condition of the gate port was changed by varying the parameter, L. Another way of doing this would have been to change k^2 (Eq. (3.39)) in the gate region with an applied gate voltage. Two different channel widths are presented; the first, $W_x = W_y = 100Å$, allows only one propagating mode in each of the guiding channels; the second, $W_x = W_y = 400Å$, allows two propagating modes in each channel. Landauer's two probe formula, Eq. (3.2), was used to calculate the conductance from source to drain for both channel widths.

Convergence is estimated to be better than three significant figures. Current was conserved to eight decimal places. As a check the source port was also sealed off using Redheffer's rule, yielding the blocked waveguide previously discussed. The resulting scattering matrix was correct to five significant figures.

3.4.2 Results

Figure 3.4(a) shows the source to drain conductance, normalized to $2q^2/h$, in the smaller channel device. The fractional conductance modulation is 100 percent due to the single modedness of the channels. The waveform is not sinusoidal due to the interference among the many multiple reflections between the gate barrier and the splitter. If the drain port could be regarded as a small perturbation to the source gate

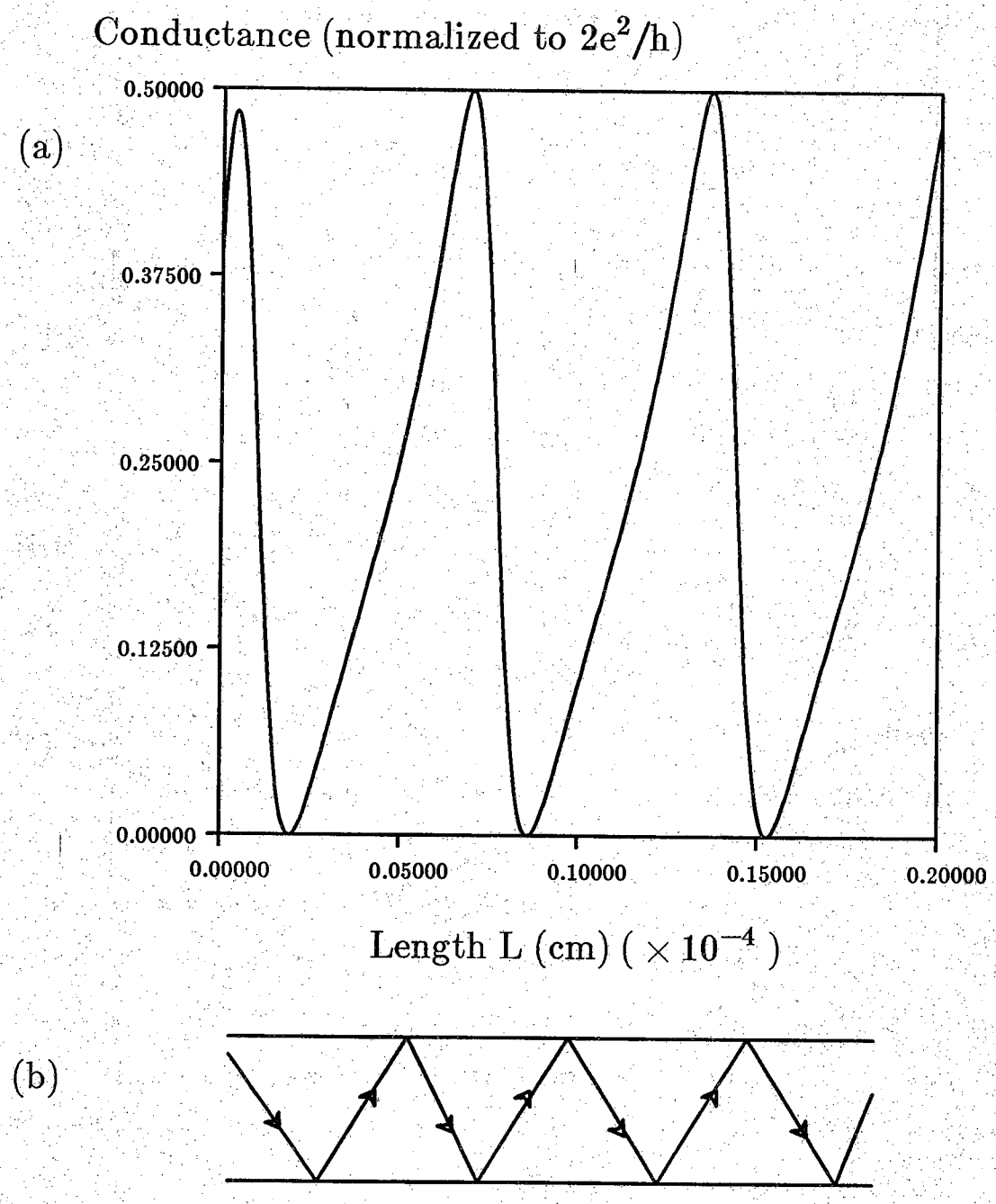


Figure 3.4 (a) Normalized conductance of the device having $W_x = W_y = 100\text{Å}$ versus the parameter L; and (b) ray diagram showing the direction of the plane wave components making up the guided wave in the channels.

waveguide, the waveform would be sinusoidal. The transconductance of the device is an oscillatory function of the gate voltage, V_G. The first peak is slightly different from the others because evanescent waves from the gate barrier are still significant for small L.

Figure 3.5(a) shows the normalized conductance along with the dominant terms making it up, $|t_{22}|^2$, $|t_{21}|^2$ and $|t_{12}|^2$. $|t_{11}|^2$ was found to be less than 10^{-3} for all L in the range of interest. The fractional conductance modulation was found to be 82 percent, which is less than that of the single moded structure as expected. An interesting and at first counter-intuitive result is the reduction in average conductance as the device channels are widened. The reason for this is that the tight confinement gives the electron a high momentum in the direction perpendicular to propagation, making it much easier to enter the drain channel. (Compare Fig. 3.4(b) and Fig. 3.5(b).) Figure 3.5 also shows that higher order modes transmit best around the corner into the drain channel. This could be expected, again because these modes have more momentum in the direction perpendicular to propagation (Fig. 3.5(b)). There is a large spike in the transmission probability $|t_{22}|^2$ at $L = 0$; this is because the angle of the k_2 vector is 44° with respect to the propagation direction so that it is almost a perfect bounce into the drain channel with $L = 0$ (Fig. 3.5(b)). The conductance as a function of L in the larger channel device is structurally more complicated than in the single moded device. As the number of propagating modes is increased there is a transition to the regime of universal conductance fluctuations which has received much attention in the literature [16,24,25,26].

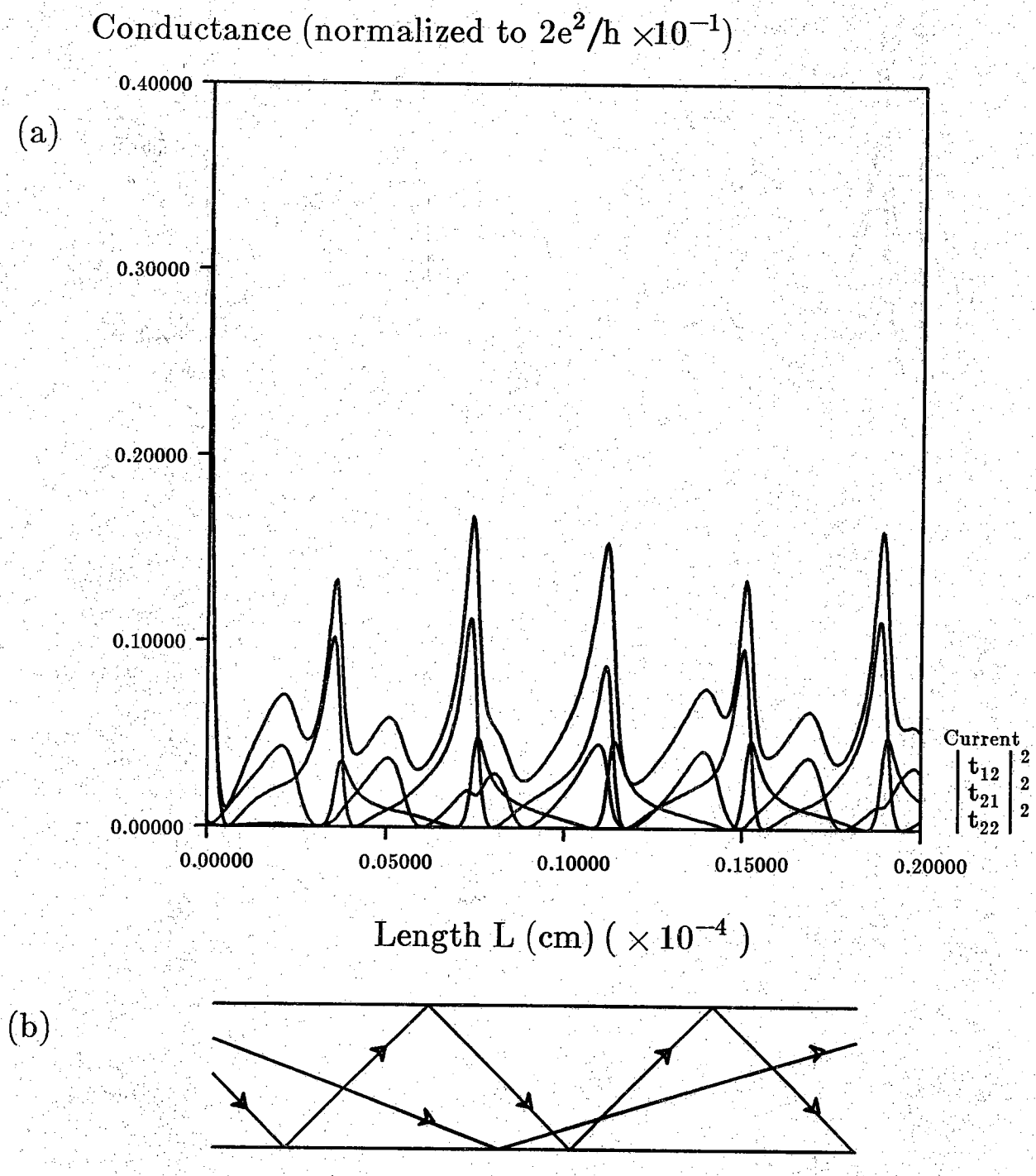


Figure 3.5 (a) Normalized conductance of the device having $W_x = W_y = 400\text{Å}$ versus the parameter L. Also shown are $|t_{22}|^2$, $|t_{21}|^2$ and $|t_{12}|^2$ versus the parameter L. (b) Ray diagram showing the direction of the plane wave components making up the guided wave in the channels for each propagating mode.

3.5 Summary

We have presented a very powerful and general set of tools for analysis of electron waveguide networks. This set of tools has been applied to a new device, and the results have been discussed.

Appendix

Given the set of solutions, ϕ_γ, to the Helmholtz equation

$$\nabla^2\phi_\gamma + k^2\phi_\gamma = 0 \tag{3.A1}$$

where

$$k^2 = \frac{2m^*}{\hbar^2}(E - V(x,y)) \tag{3.A2}$$

we show

$$\oint d\ell\, \phi_\mu \frac{\partial\phi_\gamma}{\partial n} = \oint d\ell\, \phi_\gamma \frac{\partial\phi_\mu}{\partial n} \tag{3.A3}$$

Refer to Fig. 3.1. Since ϕ_μ and ϕ_γ satisfy the Helmholtz equation

$$\begin{aligned} 0 &= \int_S (\phi_\gamma k^2\phi_\mu - \phi_\mu k^2\phi_\gamma)ds \\ &= \int_S (\phi_\mu\nabla^2\phi_\gamma - \phi_\gamma\nabla^2\phi_\mu)ds \\ &= \oint(\phi_\mu \frac{\partial\phi_\gamma}{\partial n} - \phi_\gamma \frac{\partial\phi_\mu}{\partial n})d\ell \end{aligned} \tag{3.A4}$$

The last equality is a consequence of Green's theorem [27]. Eq. (3.A4) leads directly to Eq. (3.A3).

Chapter 3 References

[1] G. Timp, H. U. Baranger, P. deVegvar, J. E. Cunningham, R. E. Howard, R. Behringer, P. M. Mankiewich, unpublished.

[2] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel and C. T. Foxon, Phys. Rev. Lett. 60, No. 9, 848 (1988).

[3] K. Ishibashi, K. Nagata, S. Ishida, K. Gamo, Y. Aoyagi, M. Kawabe, K. Murase, and S. Nambo, *Proceedings of the MRS meeting,* Dec. 4, 1986.

[4] G. Timp, A. M. Chang, P. Mankiewich, R. Behringer, J. E. Cunningham, T. Y. Chang, R. E. Howard, Phys. Rev. Lett. vol. 59, no. 6, 732 (1987).

[5] S. Washburn and R. A. Webb, Advances in Physics, 1986, Vol.35, No. 4, 375-422. (and references therein).

[6] R. Landauer, IBM J. Res. Dev. 1, 223 (1957).

[7] A. D. Stone, A. Szafer, IBM J. Res. Dev. 32, No. 3, 384 (1988); R. Landauer, Localization, Interaction, and Transport Phenomena, edited by B. Kramer, G. Bergmann and Y. Bruynseraede (Berlin, Springer) p38.

[8] M. Büttiker, Y. Imry, R. Landauer and S. Pinhas, Phys. Rev. B 31, 6207 (1985) and references therein.

[9] L. Esaki and R. Tsu, IBM J. Res. Dev. 14(61), (1970).

[10] R. Tsu and L. Esaki, Appl. Phys. Letters 22, (1973) 562-564; B. Jogai and K. L. Wang, Appl. Phys. Lett. 46(2), (1986) 167-168; J. N. Schulman, Appl. Phys. Lett. 49(12), (1986) 690-692.

[11] $t_{\alpha,\beta}$ are those scattering parameters relating an outgoing wave in mode α in the output port to an incident wave in mode β in the input port.

[12] J. N. Schulman, Y. C. Chang, Phys. Rev. B 27, No 4, 2346 (1983).

[13] P. A. Lee and D. S. Fisher, Phys. Rev. Lett. 47, No 12, 882 (1981).

[14] Alfred M. Kriman and P. Paul Ruden, Phys. Rev. B **32**, 8013 (1985).

[15] R. Frohne, S. Datta, to appear in J. Appl. Phys.

[16] M. Cahay, M. McLennan, S. Datta, Phys. Rev. B 37, No 17, 10125 (1988); S. Datta, M. Cahay, M. McLennan, Phys. Rev. B 36, No 10, 5655 (1987).

[17] Dirichlet boundary conditions correspond to equating the wave function to zero along the boundary.

[18] If the wave function inside the boundary is desired, Eqs. (9-12) may be employed to obtain the coefficients, $c_{\mu\beta}$ of Eq. (3.8).

[19] With no ports there is no incident wave and the index β should be replaced by E, the eigen-energy associated with the different eigenfunction solutions of Eq. (3.29).

[20] N. Kishi and T. Okoshi, IEEE Trans. Microwave Theory Tech., vol. MTT-35, No. 10, pp. 887 (1987).

[21] As a quick exercise the reader may like to try $\phi_\gamma(x,y) = \left(\frac{2}{W}\right)^{1/2} \sin(\frac{\gamma\pi y}{W})\sin[k_\gamma(x-L)]$.

[22] Sometimes it isn't easy to determine which eigenfunctions are most important. We have found that a nice way around this problem is to include more equations than unknowns ($\gamma_{max} > \alpha_{max} + j_{max}$). A Moore-Penrose pseudo inverse is used to obtain a least squares best fit to the equations.

[23] R. Redheffer, J. Math. Phys. 41, 1-41 (1962); T. S. Chu and T. Itoh, IEEE Trans. Microwave Theory Tech., vol. MTT 34 #2, pp. 280-284 (1986).

[24] K. Ishibashi, K. Nagata, K. Gamo, S. Namba, S. Ishia, K. Murase, M. Kawabe and Y. Aoyagi, Solid St. Comm. Vol 61, No 6, pp 385-389 (1987).

[25] W. J. Skocpol, P. M. Mankiewich, R. E. Howard, L. D. Jackel, D. M. Tennant and A. D. Stone, Phys. Rev. Lett. 56, No 26, 2865 (1986).

[26] H. U. Baranger, A. D. Stone, D. P. Divincenzo, Phys. Rev. B 37, No 11, 6521 (1988).

[27] M. R. Spiegel, "Advanced Mathematics for Engineers and Scientists", McGraw Hill Book Co. p. 154 (1971).

Chapter 4

INELASTIC SCATTERING IN QUANTUM TRANSPORT

Over the years, the semiclassical viewpoint has been the foundation for our understanding of electron transport in solids. Within this framework, electrons are treated as charged particles obeying Newton's laws of motion. For many phenomena, semiclassical descriptions based upon either drift-diffusion or the Boltzmann transport equation are sufficient. By neglecting the wave nature of matter, however, the semiclassical viewpoint fails miserably when confronted with tunneling [1], resonant tunneling [2], localization [3,4], and the quantum Hall effect [5], to name but a few examples of quantum phenomena. Since 1985, experiments on mesoscopic structures have revealed a wealth of new effects, including the Aharonov-Bohm effect [6], conductance fluctuations [7-9], and non-locality [10,11]. The term "mesoscopic" refers to structures whose overall dimensions are less than or about equal to the inelastic mean free path. In this regime, the electronic wave function remains coherent over large regions of a sample, and transport processes are strongly influenced by the electron's wave nature. As a result, these "quantum" devices bear a striking similarity to microwave networks [12-14]. In the absence of phase-breaking scattering, they can be described as such; however, some degree of scattering is present in any sample. It is the introduction of scattering into quantum transport theory that is the subject of this article. Our intent is to build upon

existing formalisms [15,16], which allow for phase-breaking processes to some degree, to provide a theory which accounts for distributed inelastic scattering.

The organization of this article is as follows. Section 4.1 provides an overview of existing transport theories, namely the Kubo formalism and the Landauer approach. In Section 4.2, we show how the Landauer formula can be extended to describe structures with distributed inelastic scattering. This is done by introducing a continuous distribution of probes, connecting each point in a structure to a conceptual reservoir of carriers in which inelastic processes can occur. A physical interpretation, relating the results of this section to the parameters of Brownian motion, is presented in Section 4.3. In Section 4.4, we demonstrate our formalism in some simple cases, including a numerical analysis of localization. Finally, in Section 4.5 we discuss possible extensions of the formalism to account for spatially correlated scattering processes, finite temperature and non-linear response.

4.1 Introduction

4.1.1 Linear Response Theory and the Kubo Formula

In the semiclassical picture, current is commonly expressed in terms of drift and diffusion [17]:

$$J_i = \sigma_{ij} E_j + eD_{ij} \nabla_j n \ , \tag{4.1.1}$$

where a summation convention is implied by the repeated indices. For small driving forces (linear response regime) the two terms can be combined, to express the current density in terms of the chemical potential μ,

$$J_i = \sigma_{ij} \nabla_j \mu \ . \tag{4.1.2}$$

Of course, Eq. (4.1.2) has been derived from a viewpoint which neglects the electronic wave nature. In the absence of strong phase-breaking scattering, the electron wave

function can remain coherent over large regions of a sample. Consequently, the current density must be inherently non-local:

$$J_i(\mathbf{r}) = \int d\mathbf{r}' \, \sigma_{ij}(\mathbf{r},\mathbf{r}')\nabla_j'\mu(\mathbf{r}') \ . \tag{4.1.3}$$

Coupled with the Kubo formula for the conductivity, this is the starting point for many linear response calculations [18-21].

The Kubo formula is a specific application of the more general fluctuation-dissipation theorem [22], which relates the equilibrium properties of a system to the nonequilibrium response. A familiar example is the Einstein relation, which expresses the diffusion coefficient D in terms of the mobility μ, for a particle experiencing Brownian motion:

$$D = \mu k_B T \ , \tag{4.1.4}$$

where k_B is the Boltzmann constant and T is the absolute temperature. In this case, diffusion is a measure of the fluctuations in a particle's velocity. Mobility is a measure of the dissipation of potential energy supplied by the driving force. Fluctuations and dissipation depend on the same physical properties of a system, and therefore the quantities are related. It is instructive to rewrite Eq. (4.1.4) as

$$\mu = \frac{1}{k_B T} \int_0^\infty dt \, \langle u(0) \, ; \, u(t) \rangle \ , \tag{4.1.5}$$

where u(t) is the velocity of the particle at time t, and the brackets $\langle \, ; \, \rangle$ represent the canonical correlation. (For a precise definition, see Ref. [22]. For our present purposes, it is sufficient to recognize this as an ensemble average.) Here, the diffusion has been written explicitly as an average of correlations in the velocity of a particle.

In the same manner, the conductivity of a sample can be expressed as an average of correlations between current operators. Assuming an excitation source with frequency ω [23],

$$\sigma(\omega) = \frac{1}{\Omega\, k_B T} \int_0^\infty dt\; e^{-i\omega t} \langle \mathbf{J}(0)\,;\, \mathbf{J}(t) \rangle \quad . \tag{4.1.6}$$

This is the Kubo conductivity for a bulk medium, where Ω is the volume of the sample. In practice, this expression is expanded in terms of the eigenmodes [15,24,25] or the Green functions [21,26] for a given Hamiltonian. We defer the details of such an expansion [15] to Appendix A. In the limit of zero temperature and dc measurements ($\omega \rightarrow 0$),

$$\sigma_{ij}(\mathbf{r},\mathbf{r}') = \frac{\hbar}{2\pi} \sum_{\alpha,\,\beta} \frac{[\mathbf{J}_{\alpha\beta}(\mathbf{r}) \otimes \mathbf{J}_{\beta\alpha}(\mathbf{r}')]_{ij}}{(E_F - E_\beta + \frac{i\hbar}{2\tau_i})(E_F - E_\alpha - \frac{i\hbar}{2\tau_i})} \quad , \tag{4.1.7}$$

where

$$J_{\alpha\beta}(\mathbf{r}) \equiv \frac{ie\hbar}{2m^*} \left[(\nabla\psi_\alpha)^* \psi_\beta - \psi_\alpha^* (\nabla\psi_\beta) \right] \tag{4.1.8}$$

τ_i is the inelastic scattering time, and E_F is the equilibrium Fermi potential.

Evaluation of the conductivity has been approached both numerically [21,27] and analytically [7,20,28] using the Green function expansion. Given a particular configuration of impurities, Green functions can be calculated numerically from a recursive technique [21], used in conjunction with the Anderson tight-binding model [29]. By generating and evaluating a large number of random samples, statistics for various moments of conductance can be compiled. This ensemble-averaging can also be performed analytically, if the conductivity is expanded in a diagrammatic series [7,20,28].

4.1.2 The Landauer Approach

In 1957, Landauer [30] approached the problem of determining conductance from a different perspective, that of a scattering problem. He proposed that current, rather than voltage, be fixed at the contacts. Near reflective obstacles, a pile-up of carriers would produce a charge imbalance, which would alter the electrostatic potential

through Poisson's equation. The resulting potential drop could be used in conjunction with the known current to define a conductance. This is the essence of his argument, although the details are somewhat different. Specifically [31], he considered a stream of electrons incident upon an obstacle in one dimension. Reflection from that obstacle produces a gradient in the carrier concentration: Electron density is increased on one side of the obstacle (i.e., the incident side), and decreased on the other. From this density gradient and the incident current, he defined a diffusion coefficient, which can be related to the conductance through the Einstein relation. The resulting conductance in one dimension is given by

$$g = \frac{2e^2}{h}\frac{T}{R} \quad , \tag{4.1.9}$$

where T and R are the transmission and reflection coefficients, respectively. Although the argument was presented with a single obstacle in mind, there is no reason why the formula should change for a string of obstacles. Any such string can always be represented by its overall transmission T and reflection R, so that Eq. (4.1.9) is valid for more complicated structures. Because of its dependence on T *and* R, this conductance has the behavior one naively expects in two limits: For highly reflective samples, $T \longrightarrow 0$ and conductance is zero; for samples with no reflection, $R \longrightarrow 0$ and conductance is infinite.

It should be noted that the arguments leading to Eq. (4.1.9) are quite general, and can be applied to a variety of similar transport problems [32]. For all subsequent discussion, however, we choose to limit ourselves to the following physical picture. In the Landauer approach, as we refer to it, a device is composed of contacts (or reservoirs) separated by regions in which transport is purely elastic (see Fig. 1). All inelastic (or phase-breaking) scattering is confined to the reservoirs, which are assumed to be "black bodies" with respect to electrons. All electrons which enter a reservoir are absorbed,

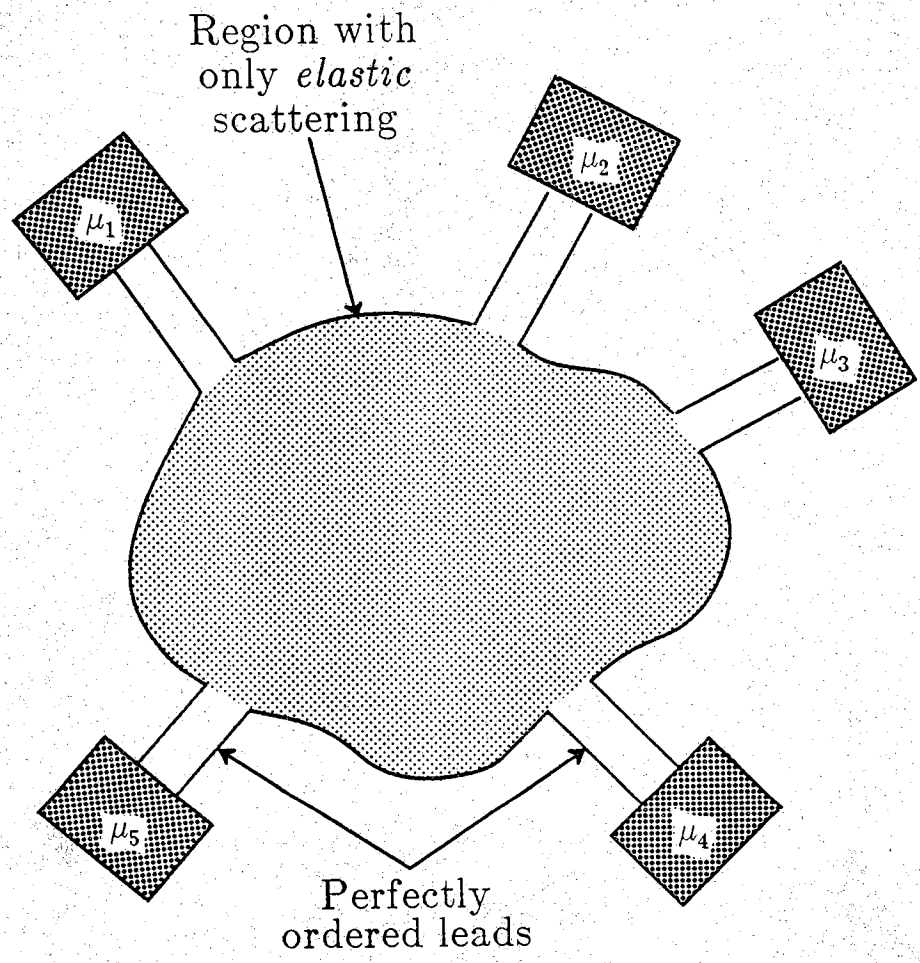


Figure 4.1 In the literature of mesoscopic structures, the Landauer approach assumes that all inelastic scattering processes are confined to contact reservoirs, between which transport is purely elastic.

and all electrons leave with completely random phases. This is the popular conception [15,24-26,33] of the Landauer approach in the literature of mesoscopic devices.

Starting from the Kubo formula, Economou and Soukoulis [24] obtained a similar expression for the conductance in one dimension,

$$g = \frac{2e^2}{h} T \quad . \tag{4.1.10}$$

Although the two conductance formulas agree in the limit $R \longrightarrow 1$, the other limit, $R \longrightarrow 0$, aroused controversy [34,35]. In the absence of reflection ($R \longrightarrow 0$, $T \longrightarrow 1$), the naive expectation is that conductance is infinite; Eq. (4.1.10), however, predicts a finite value. Because of this discrepancy, Eq. (4.1.10) was initially rejected, although recently a number of authors [15,36,37] (including Landauer himself) have explained the finite value of conductance as a contact resistance. For the physical picture described above, the conductance measured is that of the entire structure--contacts included. Even if propagation is purely ballistic in the region between contacts, there is a resistance associated with the transition of electrons from the contacts into the device. In particular, Imry [37] viewed the contacts as three-dimensional containers of a Fermi gas, connected by a one-dimensional pipe. Although there may be no reflections within the pipe, there will still be a resistance measured between containers. This is simply the spreading resistance associated with the flow of carriers from the containers into the pipe.

From the preceding discussion, it would appear that Eqs. (4.1.9) and (4.1.10) are equally plausible. It is important to clarify the circumstances in which each is valid. In fact, each can be likened to a particular experimental measurement. For the original Landauer formula, Eq. (4.1.9), the voltage was determined by a density gradient *in the immediate vicinity of the obstacle.* This corresponds to a four-probe measurement, for which a current is maintained between two contacts, and the resulting voltage is measured in a smaller region of the sample. From the previous discussion of contact

resistance, however, it is clear that Eq. (4.1.10) depends on the voltage applied *between the contacts.* This corresponds to a two-probe measurement, for which both current and voltage are measured between the same set of contacts.

It is interesting to consider whether the physical assumptions used to derive Eq. (4.1.10) correspond to a two-terminal measurement. To simulate the effect of "black body" contacts in a linear response calculation, Economou and Soukoulis connected infinitely long, perfectly conducting leads onto either side of their device. Conditions of infinite extent and perfect conduction ($T \longrightarrow 1$) guarantee that any electron entering a lead is "absorbed." Moreover, because the leads are perfect conductors, any voltage applied in the "contacts" appears only across the region of interest. In this case, the voltage used to define the conductance is the same as that applied to the entire sample, so that the measurement is indeed two-terminal.

Perhaps the clearest illustration of the four-probe assumptions buried in Landauer's original arguments has been provided by Engquist and Anderson [35]. Objecting to the measurement of voltage in terminals supplying current, the authors introduced two separate voltage probes into their model, thereby analyzing a four-probe structure explicitly. The resulting conductance was the Landauer formula of Eq. (4.1.9). A critical assumption in their model, however, was that the voltage probes were *weakly coupled* to the structure, so that the voltage could be measured without significantly disturbing the initial transport problem. In experiments on thin wires, this condition appears to be difficult to achieve [11-13]. Consequently, Eq. (4.1.10) and its generalizations are more appropriate for practical calculations [15].

In spite of the subtle differences between Eqs. (4.1.9) and (4.1.10), both are recognized in the literature as the "Landauer approach." Because Eq. (4.1.10) appears to better describe experiment, we restrict ourselves to this form, and describe its generalizations. Two important generalizations have been proposed, to account for one-

dimensional wires with finite width, and to allow for many contacts. Starting from linear response theory, and following an approach similar to that of Economou and Soukoulis, Fisher and Lee [26] derived a conductance which included the interaction between transverse scattering channels:

$$g = \frac{2e^2}{h} \mathrm{Tr}\{t^\dagger t\} \ . \tag{4.1.11}$$

where t is a matrix whose elements represent transmission from one transverse channel into another. Buttiker [33] extended the approach to include multiple contacts, by writing the current in any lead i as a function of all applied potentials:

$$I_i = \frac{e^2}{h} \sum_j \{T_{ij}\mu_j - T_{ji}\mu_i\} \ , \tag{4.1.12}$$

where μ_i is the chemical potential in contact i, and T_{ij} represents the transmission from lead j into lead i. For leads with a finite width, T_{ij} can be written as the sum of transmissions between all transverse channels,

$$T_{ij} = \sum_{\alpha,\beta} | t_{i,\alpha;j,\beta} |^2 = \mathrm{Tr}\{t_{ij}^\dagger \, t_{ij}\} \ , \tag{4.1.13}$$

where α and β index the transverse channels in leads i and j, respectively. In this manner, Eq. (4.1.11) can be viewed as a simplification of the multi-probe Landauer formula of Eq. (4.1.12), in the limit of two contacts.

The Landauer approach, embodied by Eqs. (4.1.11) and (4.1.12), represents an enormous simplification (if only conceptual) in the calculation of conductance. It relates conductance between any two contacts to the transmission coefficient, a quantity which can be obtained from random matrix theory [38,39] or transfer-matrix calculations [26,40,41]. Furthermore, it has been shown [15,26] that Eqs. (4.1.11) and (4.1.12) can be derived rigorously from linear response theory, putting the Landauer approach on equal footing with the Kubo formalism. With this foundation, we proceed in the following

sections to incorporate the effects of distributed inelastic scattering, by further generalizing the Landauer formula.

4.2 A Continuous-probe Landauer Formula

For a conventional analysis with the Landauer formula, all inelastic scattering within a device is confined to the contact regions (see Fig. 4.1). This greatly simplifies the problem, because transmission coefficients are easily calculated for the elastic propagation between contacts. For many phenomena (e.g., resonant tunneling, localization, etc.), it is important to include the effects of inelastic scattering in the region *between* contacts. To account for this, Buttiker [16,42] introduced an additional reservoir into his analysis, connected to his structure by a perfectly conducting probe. This reservoir, simply a conceptual tool, was introduced as a source of inelastic scattering. Carriers diverted into the probe would enter the reservoir and experience phase randomization; however, the probe carried no net current, so that the reservoir was not a "contact" in the physical sense. To demonstrate the model, Buttiker [42] analyzed resonant tunneling in a double-barrier structure, with his conceptual reservoir connected to the quantum well region. By varying the fraction of carriers diverted into the reservoir, he was able to demonstrate a transition from coherent resonant tunneling to sequential tunneling. In the following sections, we propose an extension of this idea--a uniform distribution of probes leading to many such reservoirs--to simulate the effects of distributed inelastic scattering.

4.2.1 Extension of the Multi-probe Landauer Formula

In the following analysis, we restrict ourselves to the linear response regime and zero temperature, to avoid unnecessary complications. These restrictions can be removed, however, as discussed in Section 4.5. In assuming zero temperature we have confined our attention to electrons with a single energy E_F, residing on the surface of a Fermi sphere. For linear response, we have neglected the effects of carrier heating, so that all electrons remain at the energy E_F. From this standpoint, we reconsider the multi-probe Landauer formula discussed in the previous section:

$$I_i = \frac{e^2}{h} \sum_j \{T_{ij}\mu_j - T_{ji}\mu_i\} \quad . \tag{4.2.1}$$

Within any structure, we introduce a uniform distribution of probes; each probe connects an infinitesimal volume to a conceptual reservoir, in which inelastic processes can occur (see Fig. 4.2). For simplicity, all reservoirs are assumed to be independent of one another, so that inelastic scattering events are *spatially uncorrelated.* This restriction is unnecessary, however, and will be removed in Section 4.5. Keeping in mind this physical picture, we extend Eq. (4.2.1) to account for the distribution of infinitesimal probes:

$$I(\mathbf{r}) = \frac{e^2}{h} \int d\mathbf{r}' \{T(\mathbf{r},\mathbf{r}')\mu(\mathbf{r}') - T(\mathbf{r}',\mathbf{r})\mu(\mathbf{r})\} \quad . \tag{4.2.2}$$

Although this generalization is straightforward, it is of little value unless we can clarify the coupling of each probe to the device. Specifically, the fraction of electrons diverted into each probe must be defined. In Buttiker's model, this was an adjustable parameter; however, we seek a definition based upon the physical properties of a system.

We assume that each probe is *weakly coupled* to the device, so that a large number of electrons introduced at any point would leak out--little by little--as they propagate through the device. Such a distribution can be described in terms of the Green function $G(\mathbf{r},\mathbf{r}')$ for the Schrödinger equation:

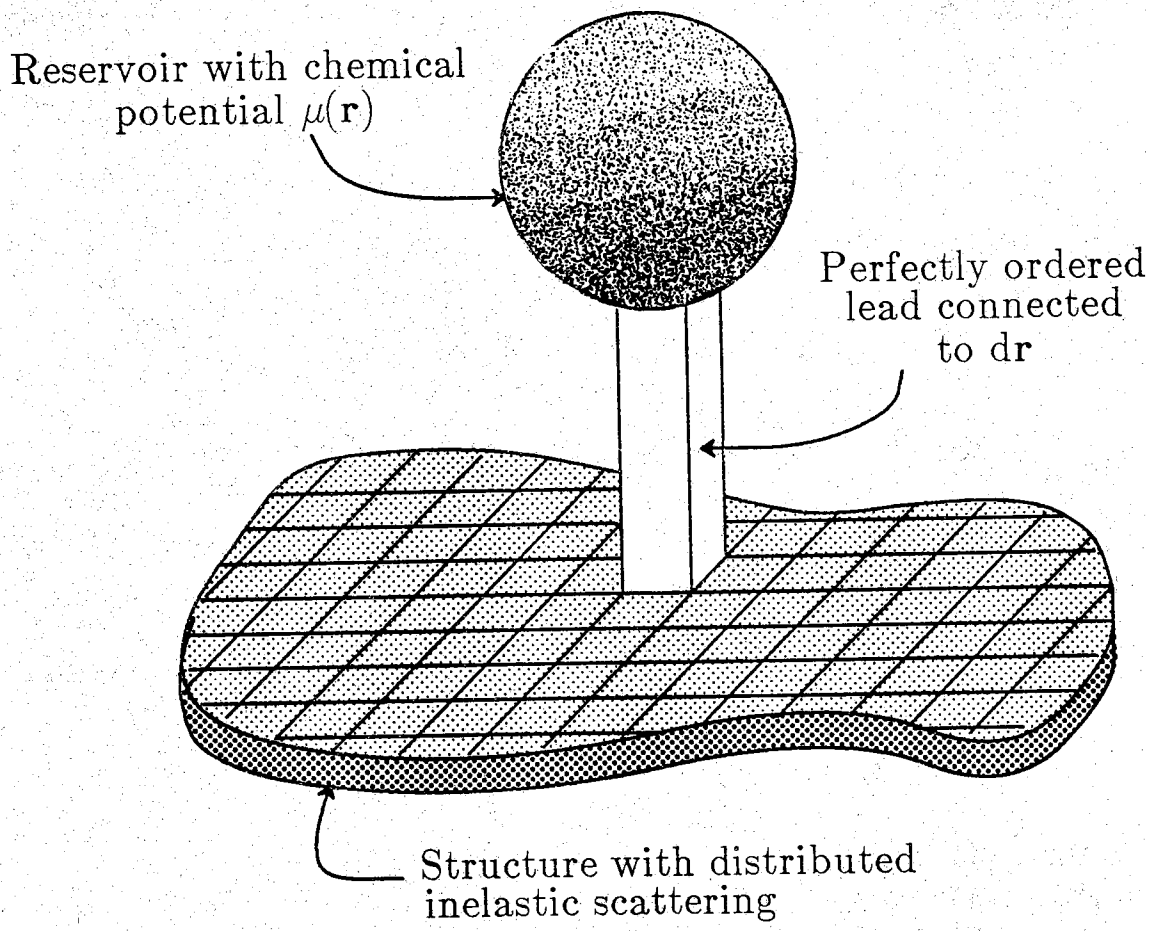


Figure 4.2 Distributed inelastic scattering is represented by a continuous distribution of probes, linking each volume element dr to a conceptual reservoir of carriers.

$$\left[E_F - H + \frac{i\hbar}{2\tau_i}\right] G(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}') \ , \tag{4.2.3a}$$

$$H = \frac{(\mathbf{p}-e\mathbf{A})^2}{2m^*} + eV(\mathbf{r}) \ . \tag{4.2.3b}$$

where τ_i is the inelastic scattering time, which is assumed to be constant. The quantity $|G(\mathbf{r},\mathbf{r}')|^2$ represents the probability that an electron, introduced at position $\mathbf{r}'$, will propagate to $\mathbf{r}$ before entering a probe and suffering an inelastic collision. In general, an imaginary potential (such as $i\hbar/2\tau_i$) causes the probability density to decay: $\nabla\cdot\mathbf{J} + \partial\rho/\partial t = -\rho/\tau_i$. In steady state, this divergence of current represents absorption of electrons by the weakly coupled reservoirs. Therefore, the current entering a probe is $e\, n(\mathbf{r})/\tau_i$, where $n(\mathbf{r})$ is the electron density at position $\mathbf{r}$. Of course, the reservoirs inject an equal current (of phase-randomized electrons) back into the device, so that overall, carriers are conserved.

Intuitively, one would expect that the transmission probability $T(\mathbf{r},\mathbf{r}')$ should be proportional to the square of the propagator $G(\mathbf{r},\mathbf{r}')$. What remains is to determine the constant of proportionality. For this, we return to the "continuous-probe" Landauer formula in Eq. (4.2.2), written in the form of an integral equation:

$$I(\mathbf{r}) + \frac{e^2}{h}\mu(\mathbf{r})\int d\mathbf{r}'\, T(\mathbf{r}',\mathbf{r}) = \frac{e^2}{h}\int d\mathbf{r}'\, T(\mathbf{r},\mathbf{r}')\mu(\mathbf{r}') \tag{4.2.4}$$

The second term on the left-hand side represents the current injected back into the device from a reservoir at $\mathbf{r}$. From the arguments presented above, this current is,

$$\frac{e^2}{h}\mu(\mathbf{r}) \int d\mathbf{r}'\, T(\mathbf{r}',\mathbf{r}) = \frac{e\, n(\mathbf{r})}{\tau_i} \ . \tag{4.2.5}$$

If the electron density varies slowly with the chemical potential, we can approximate n as $\mu(\partial n/\partial\mu)$; at low temperatures, $\partial n/\partial\mu = eN_0$, where N_0 is the density of states.

Substituting this into Eq. (4.2.5), we obtain the normalization condition for $T(\mathbf{r}',\mathbf{r})$,

$$\int d\mathbf{r}'\, T(\mathbf{r}',\mathbf{r}) = \frac{hN_0(\mathbf{r})}{\tau_i} \quad . \tag{4.2.6}$$

This, in turn, defines the constant of proportionality between $T(\mathbf{r}',\mathbf{r})$ and $|\, G(\mathbf{r}',\mathbf{r})\,|^2$,

$$T(\mathbf{r}',\mathbf{r}) \equiv \frac{hN_0(\mathbf{r})/\tau_i}{\int d\mathbf{r}'\,|\, G(\mathbf{r}',\mathbf{r})\,|^2}\,|\, G(\mathbf{r}',\mathbf{r})\,|^2 \quad . \tag{4.2.7}$$

We can simplify this expression dramatically, by recognizing that for constant τ_i, the Green function can be expanded in terms of an orthonormal set of basis functions [43],

$$G(\mathbf{r},\mathbf{r}') = \sum_\alpha \frac{\psi_\alpha(\mathbf{r})\psi_\alpha^*(\mathbf{r}')}{E_F - E_\alpha + i\hbar/2\tau_i} \quad . \tag{4.2.8}$$

From this definition, it is shown in Appendix 4.A2 that,

$$\int d\mathbf{r}\,|\, G(\mathbf{r},\mathbf{r}')\,|^2 = 2\pi\tau_i N_0(\mathbf{r}')/\hbar \quad , \tag{4.2.9}$$

where $N_0(\mathbf{r}')$ is the density of states at the Fermi level, expressed as

$$N_0(\mathbf{r}') = \sum_\alpha |\, \psi_\alpha(\mathbf{r}')\,|^2\, \delta(E_F - E_\alpha) \quad . \tag{4.2.10}$$

Substituting Eq. (4.2.9) into Eq. (4.2.7), we obtain a simplified expression for the transmission probability. We present this result below, together with Eq. (4.2.4):

$$T(\mathbf{r},\mathbf{r}') = \frac{\hbar^2}{\tau_i^2}\,|\, G(\mathbf{r},\mathbf{r}')\,|^2 \quad , \tag{4.2.11a}$$

$$I(\mathbf{r}) + \frac{e^2}{\tau_i}\mu(\mathbf{r})\, N_0(\mathbf{r}) = \frac{e^2}{h}\int d\mathbf{r}'\, T(\mathbf{r},\mathbf{r}')\, \mu(\mathbf{r}') \quad . \tag{4.2.11b}$$

In the regime of linear response and zero temperature, the above equations form the basis for our analysis. For a given Hamiltonian, the Green function defined by Eq. (4.2.3) can be calculated, and used to determine the transmission probability $T(\mathbf{r},\mathbf{r}')$

according to Eq. (4.2.11a). After specifying some combination of $I(\mathbf{r})$ and $\mu(\mathbf{r})$ as a boundary condition, $\mu(\mathbf{r})$ at all remaining points can be calculated from Eq. (4.2.11b). As an example of the boundary condition, we suppose that some external source fixes the chemical potential $\mu(\mathbf{r})$ at several points $\mathbf{r}_n$, where n indexes the boundary positions. At all other points, the current $I(\mathbf{r})$ in each probe must be zero, for conservation of carriers. Given the values $\mu(\mathbf{r}_n)$, the chemical potentials at all remaining positions can be determined by solving:

$$\mu(\mathbf{r}) \frac{h\, N_0(\mathbf{r})}{\tau_i} = \int d\mathbf{r}'\, T(\mathbf{r},\mathbf{r}')\, \mu(\mathbf{r}') \quad \text{for } \mathbf{r} \neq \mathbf{r}_n \ . \tag{4.2.12}$$

The current at each position $\mathbf{r}_n$ is non-zero, and can be calculated from the "continuous-probe" Landauer formula of Eq. (4.2.2). Alternatively, we might suppose that the current is fixed at positions $\mathbf{r}_n$, and that the chemical potentials $\mu(\mathbf{r}_n)$ must be determined. Both approaches are equivalent. For each boundary position $\mathbf{r}_n$, either $I(\mathbf{r}_n)$ or $\mu(\mathbf{r}_n)$ is specified, and the remaining quantity must be determined. At all other positions, $I(\mathbf{r})$ is zero, and $\mu(\mathbf{r})$ is calculated from Eq. (4.2.12).

4.2.2 Connection to the Kubo Formalism

To supplement the arguments presented above, a rigorous justification of Eq. (4.2.11a) can be derived from the Kubo formula for the conductivity (see Ref. [15] and Appendix 4.A1):

$$\sigma_{mn}(\mathbf{r},\mathbf{r}') = \frac{\hbar}{2\pi} \sum_{\alpha,\beta} \frac{[\mathbf{J}_{\alpha\beta}(\mathbf{r}) \otimes \mathbf{J}_{\beta\alpha}(\mathbf{r}')]_{mn}}{(E_F - E_\beta + i\hbar/2\tau_i)(E_F - E_\alpha - i\hbar/2\tau_i)} \ , \tag{4.2.13a}$$

where

$$\mathbf{J}_{\alpha\beta} = \frac{ie\hbar}{2m^*}[(\nabla\psi_\alpha)^*\psi_\beta - \psi_\alpha^*(\nabla\psi_\beta)] \ , \tag{4.2.13b}$$

and $\otimes$ represents the direct product of vectors $\mathbf{J}_{\alpha\beta}$ and $\mathbf{J}_{\beta\alpha}$, to form a tensor indexed by m and n. This conductivity can be related to the conductance between leads i and j by integrating over the cross sections of each lead [20],

$$g_{ij} = \int d(\mathbf{S}_i)_m \int d(\mathbf{S}'_j)_n \ \sigma_{mn}(\mathbf{r},\mathbf{r}') \ , \tag{4.2.14}$$

where the vector $\mathbf{S}_i$ is normal to the cross section. We can identify the conductance g_{ij} with the transmission probability T_{ij}, by rewriting the multi-probe Landauer formula (Eq. (4.2.1)), and by taking $\mu_i = 0$ as a potential reference,

$$I_i \equiv \sum_j g_{ij}\,\mu_j = \frac{e^2}{h}\sum_j \{T_{ij}\mu_j - T_{ji}\mu_i\} \tag{4.2.15a}$$

$$g_{ij} \equiv \frac{e^2}{h}T_{ij} \quad \text{for } i\neq j \ . \tag{4.2.15b}$$

Returning to the limit of a continuous distribution of probes, we combine Eqs. (4.2.13a), (4.2.14) and (4.2.15b), to obtain

$$\frac{e^2}{h}T(\mathbf{r},\mathbf{r}') = \int dS \int dS' \ \frac{\hbar}{2\pi}\sum_{\alpha,\beta}\frac{[\mathbf{J}_{\alpha\beta}(\mathbf{r})\cdot\hat{\mathbf{n}}(\mathbf{r})][\mathbf{J}_{\beta\alpha}(\mathbf{r}')\cdot\hat{\mathbf{n}}(\mathbf{r}')]}{(E_F - E_\beta + i\hbar/2\tau_i)(E_F - E_\alpha - i\hbar/2\tau_i)} \ , \tag{4.2.16}$$

where the integrals with respect to S and S′ are performed over the infinitesimal cross sections of probes at $\mathbf{r}$ and $\mathbf{r}'$, respectively. To show the vector nature of dS explicitly, we have written $d\mathbf{S} = dS\,\hat{n}$, where $\hat{\mathbf{n}}(\mathbf{r})$ is the unit normal for the cross section of the probe at $\mathbf{r}$. Now, $\mathbf{J}\cdot\hat{\mathbf{n}}$ integrated over the cross section of a probe defines the current entering that probe; this current was previously assumed to be $e\,\mathbf{n}(\mathbf{r})/\tau_i$. Hence, we can write Eq. (4.2.16) as

$$T(\mathbf{r},\mathbf{r}') = \frac{\hbar^2}{\tau_i^2} \sum_{\alpha,\beta} \frac{\psi_\alpha^*(\mathbf{r})\psi_\beta(\mathbf{r})\psi_\beta^*(\mathbf{r}')\psi_\alpha(\mathbf{r}')}{(E_F - E_\beta + i\hbar/2\tau_i)(E_F - E_\alpha - i\hbar/2\tau_i)} \quad . \tag{4.2.17}$$

Recalling the expansion of the Green function in a complete set of basis functions (Eq. (4.2.8)), we arrive at our proof:

$$T(\mathbf{r},\mathbf{r}') = \frac{\hbar^2}{\tau_i^2} \left| G(\mathbf{r},\mathbf{r}') \right|^2 \quad . \tag{4.2.18}$$

4.2.3 Dissipation of Power

Irreversibility has been introduced into our model through the phase-randomizing action of the reservoirs. Because of this, our system is open, and capable of dissipating power. In the following discussion, we will obtain an expression for the conductance of a sample, based upon a calculation of dissipated power. As a prelude to this, we reconsider our expression for current:

$$I(\mathbf{r}) = \frac{e^2}{h} \int d\mathbf{r}' \left\{ T(\mathbf{r},\mathbf{r}')\mu(\mathbf{r}') - T(\mathbf{r}',\mathbf{r})\mu(\mathbf{r}) \right\} \quad . \tag{4.2.19}$$

We have changed nothing by writing this as

$$I(\mathbf{r}) = \frac{e^2}{h} \int d\mathbf{r}' \left\{ g_S(\mathbf{r},\mathbf{r}')[\mu(\mathbf{r}') - \mu(\mathbf{r})] + g_A(\mathbf{r},\mathbf{r}')[\mu(\mathbf{r}') + \mu(\mathbf{r})] \right\} \quad , \tag{4.2.20a}$$

where

$$g_S(\mathbf{r},\mathbf{r}') = \frac{e^2}{h} \left(\frac{T(\mathbf{r},\mathbf{r}') + T(\mathbf{r}',\mathbf{r})}{2} \right) \quad , \tag{4.2.20b}$$

$$g_A(\mathbf{r},\mathbf{r}') = \frac{e^2}{h} \left(\frac{T(\mathbf{r},\mathbf{r}') - T(\mathbf{r}',\mathbf{r})}{2} \right) \quad . \tag{4.2.20c}$$

From these equations we can show that the current $I(\mathbf{r})$ is identically zero in equilibrium, as it must be. In equilibrium, the chemical potential μ is a constant, so that the

first term in Eq. (4.2.20a) vanishes. The second term, however, requires more careful consideration. In the absence of a magnetic field, we have a detailed balance between points $\mathbf{r}$ and $\mathbf{r}'$,

$$T(\mathbf{r},\mathbf{r}') = T(\mathbf{r}',\mathbf{r}) \quad \rightarrow \quad g_A(\mathbf{r},\mathbf{r}') = 0 \quad \text{for } B = 0 \ .$$

This can be understood by expanding $|\, G(\mathbf{r},\mathbf{r}')\,|^2$ in a complete set of basis functions:

$$|\, G(\mathbf{r},\mathbf{r}')\,|^2 = \sum_{\alpha,\beta} \left(\frac{\psi_\alpha(\mathbf{r})\psi_\alpha^*(\mathbf{r}')}{E_F - E_\alpha + i\hbar/2\tau_i} \right) \left(\frac{\psi_\beta^*(\mathbf{r})\psi_\beta(\mathbf{r}')}{E_F - E_\beta - i\hbar/2\tau_i} \right) \ . \tag{4.2.21}$$

In the absence of a magnetic field, the eigenstates can be chosen to be real, since ψ_α and ψ_α^* are degenerate [44]. Because of this, Eq. (4.2.21) is unaltered by an interchange of $\mathbf{r}$ and $\mathbf{r}'$, proving that $g_A(\mathbf{r},\mathbf{r}') = 0$ in this case.

In the presence of a magnetic field, however, we do not have detailed balance, although any outflow from $\mathbf{r}'$ in one direction must be balanced by an inflow in another. To show this, we must evaluate the integral of $g_A(\mathbf{r},\mathbf{r}')$,

$$\int d\mathbf{r}\ g_A(\mathbf{r},\mathbf{r}') = \frac{e^2\hbar}{4\pi\tau_i^2} \int d\mathbf{r} \left(|\, G(\mathbf{r},\mathbf{r}')\,|^2 - |\, G(\mathbf{r}',\mathbf{r})\,|^2 \right) \ . \tag{4.2.22}$$

We consider the first term on the right-hand side, written in terms of its expansion from Eq. (4.2.21). As a consequence of completeness,

$$\int d\mathbf{r}\ \psi_\alpha^*(\mathbf{r})\ \psi_\beta(\mathbf{r}) = \delta_{\alpha\beta} \ , \tag{4.2.23}$$

so that the sum over β can be performed, and the first term simplifies to

$$\int d\mathbf{r}\, |\, G(\mathbf{r},\mathbf{r}')\,|^2 = \int d\mathbf{r} \sum_\alpha \frac{\psi_\alpha^*(\mathbf{r}')\psi_\alpha^*(\mathbf{r}')}{(E_F - E_\alpha)^2 + \hbar^2/4\tau_i^2} \ . \tag{4.2.24}$$

Following a similar procedure to simplify the second term in Eq. (4.2.22), we write the total,

$$\int d\mathbf{r}\, g_A(\mathbf{r},\mathbf{r}') = \frac{e^2\hbar}{4\pi\tau_i^2} \int d\mathbf{r} \sum_\alpha \left(\frac{|\psi_\alpha(\mathbf{r}')|^2 - |\psi_\alpha(\mathbf{r}')|^2}{(E_F - E_\alpha)^2 + \hbar^2/4\tau_i^2} \right) = 0 \qquad (4.2.25)$$

This result is unchanged if we integrate with respect to $\mathbf{r}'$. In general, therefore,

$$\int d\mathbf{r}\, g_A(\mathbf{r},\mathbf{r}') = \int d\mathbf{r}'\, g_A(\mathbf{r},\mathbf{r}') = 0 \ , \qquad (4.2.26)$$

so that the total current flowing out of any point is zero, in equilibrium.

To calculate the power dissipated in a sample, we sum the losses between pairs of "terminals" $\mathbf{r}$ and $\mathbf{r}'$. We confine our attention to the case of no magnetic field, so that $g_A(\mathbf{r},\mathbf{r}') = 0$. The total power P_0 dissipated in a sample can then be expressed as

$$P_0 = \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}'\, g_S(\mathbf{r},\mathbf{r}')[\mu(\mathbf{r}') - \mu(\mathbf{r})]^2 \ . \qquad (4.2.27)$$

It is evident that in the absence of magnetic field, we can represent any medium as a continuous network of resistors. As shown in Fig. 4.3, any two volume elements $d\mathbf{r}'$ and $d\mathbf{r}$ are connected by a conductance $g_S(\mathbf{r},\mathbf{r}')d\mathbf{r}'d\mathbf{r}$. If a small potential difference $\Delta\mu$ is impressed across a pair of terminals, then the conductance g_0 seen from the terminals is obtained by equating $g_0(\Delta\mu)^2$ to the total power P_0 dissipated in the network,

$$g_0 = \frac{P_0}{(\Delta\mu)^2} = \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}'\, g_S(\mathbf{r},\mathbf{r}') \left[\frac{\mu(\mathbf{r}') - \mu(\mathbf{r})}{\Delta\mu} \right]^2 \ . \qquad (4.2.28)$$

For two-terminal measurements, this equation can be used to compute the conductance of any arbitrary structure with distributed inelastic scattering. Of course, it requires a knowledge of the chemical potential $\mu(\mathbf{r})$, everywhere within the device. Unless we assume a form for $\mu(\mathbf{r})$, this requires a solution of Eqs. (4.2.11a) and (4.2.11b), as described previously. Nevertheless, this is a convenient starting point for conductance calculations, as we shall see in the following sections.

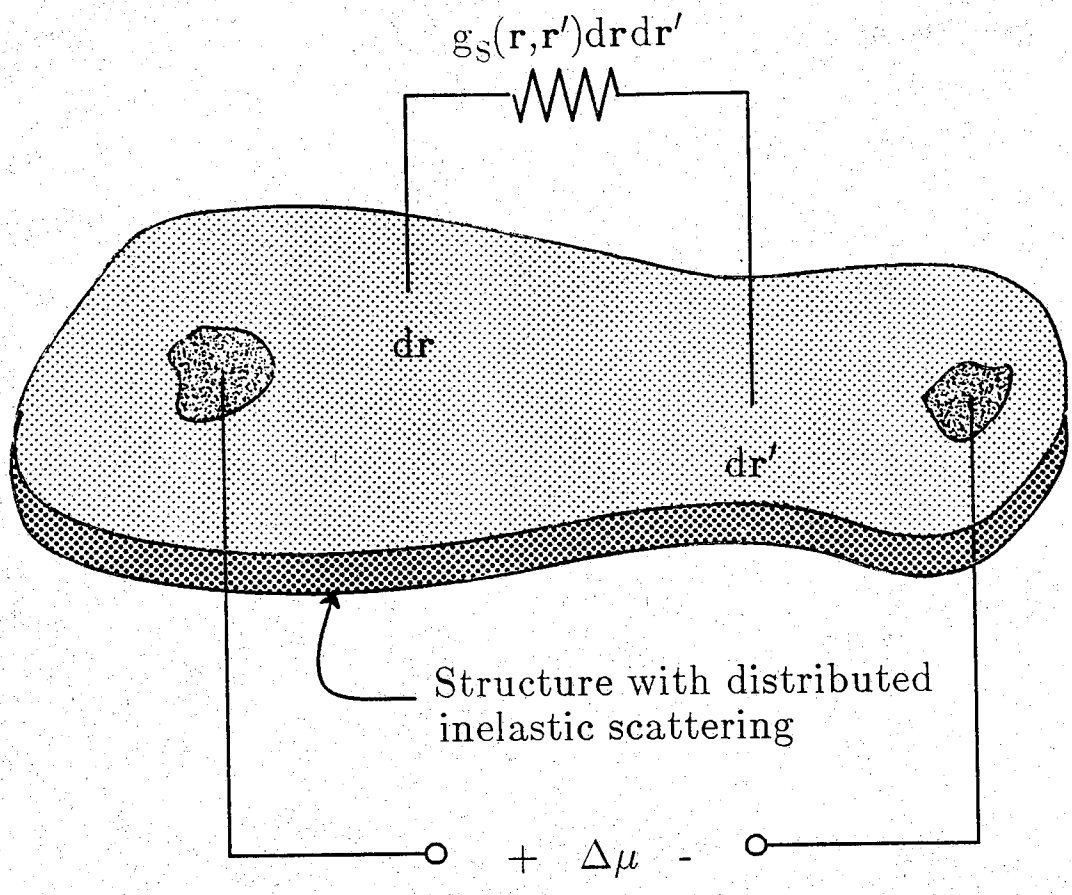


Figure 4.3 In the absence of magnetic fields, any medium can be represented as a continuous network of resistors.

4.3 A Physical Picture of the Transport Process

4.3.1 Derivation

We will now reproduce the results of the previous section, based upon a new conceptual approach to inelastic scattering. If we assume that inelastic scattering processes are delta-correlated spatially, we can view every inelastic scattering event as a *quantum measurement of the position of the electron.* We can then view the transport of electrons as a series of "hops." A "hop" from $\mathbf{r}'$ to $\mathbf{r}$ starts with one inelastic scattering event at $\mathbf{r}'$, and ends with another at $\mathbf{r}$; between $\mathbf{r}'$ and $\mathbf{r}$ the electron propagates elastically with a fixed energy. We can write

$$\nabla\cdot\mathbf{J} = I(\mathbf{r}) = e\int d\mathbf{r}' \left\{\nu(\mathbf{r},\mathbf{r}')\, n(\mathbf{r}') - \nu(\mathbf{r}',\mathbf{r})\, n(\mathbf{r})\right\} \quad , \tag{4.3.1}$$

where $n(\mathbf{r})$ is the electron density, and $\nu(\mathbf{r},\mathbf{r}')$ is the hopping frequency between $\mathbf{r}'$ and $\mathbf{r}$. The function $\nu(\mathbf{r},\mathbf{r}')$ is determined from the Schrödinger equation, and contains all quantum interference effects.

From this point of view, the problem of quantum transport can be conceptually divided into two separate problems on two different length scales. On a scale greater than the inelastic diffusion length (L_i), we have a problem of Brownian motion. In this limit, Eq. (4.3.1) can be viewed as a substitute for the classical drift-diffusion equation. But this macroscopic motion is influenced by the microscopic dynamics through the function $\nu(\mathbf{r},\mathbf{r}')$, which is determined quantum mechanically. On a length scale shorter than L_i, the function $\nu(\mathbf{r},\mathbf{r}')$ accounts for the electronic wave nature, so that quantum effects can be observed.

To obtain the hopping frequency $\nu(\mathbf{r},\mathbf{r}')$, we assume (as we did in Section 4.2.1) that the number of electrons inelastically scattered in a volume element $d\mathbf{r}'$ per unit time is given by $d\mathbf{r}'\, n(\mathbf{r}')/\tau_i$. If we assume that a fraction $P(\mathbf{r},\mathbf{r}')d\mathbf{r}$ will suffer their next inelastic event in the volume element $d\mathbf{r}$, then

$$\nu(\mathbf{r},\mathbf{r}') = P(\mathbf{r},\mathbf{r}')/\tau_i \quad . \tag{4.3.2}$$

$P(\mathbf{r},\mathbf{r}')$ is the probability that an electron injected at $\mathbf{r}'$ will propagate to $\mathbf{r}$ *without suffering another inelastic scattering event.* From this definition, we recognize that $P(\mathbf{r},\mathbf{r}')$ must be proportional to $|G(\mathbf{r},\mathbf{r}')|^2$, where $G(\mathbf{r},\mathbf{r}')$ is the Green function of the Schrödinger equation, defined by Eq. (4.2.3). Noting that $\int d\mathbf{r}\, P(\mathbf{r},\mathbf{r}')$ must equal one, we can write

$$P(\mathbf{r},\mathbf{r}') = \frac{|G(\mathbf{r},\mathbf{r}')|^2}{\int d\mathbf{r}\ |G(\mathbf{r},\mathbf{r}')|^2} \quad . \tag{4.3.3}$$

It is shown in Appendix 4.A2 that

$$\int d\mathbf{r}\ |G(\mathbf{r},\mathbf{r}')|^2 = 2\pi\tau_i N_0(\mathbf{r}')/\hbar \quad , \tag{4.3.4}$$

where $N_0(\mathbf{r}')$ is the density of states at the Fermi level, given by

$$N_0(\mathbf{r}') = \sum_\alpha |\psi_\alpha(\mathbf{r}')|^2\, \delta(E_F - E_\alpha) \quad . \tag{4.3.5}$$

Combining Eqs. (4.3.3) and (4.3.4), we obtain the desired expression for the hopping frequency $\nu(\mathbf{r},\mathbf{r}')$,

$$\nu(\mathbf{r},\mathbf{r}') = \frac{\hbar\,|G(\mathbf{r},\mathbf{r}')|^2}{2\pi N_0(\mathbf{r}')\tau_i^2} \quad . \tag{4.3.6}$$

Finally, we will show that Eq. (4.3.1) is equivalent our previous result (Eq. (4.2.2)). If the electron density varies slowly with the chemical potential, we can approximate n by $\mu(\partial n/\partial\mu)$; at low temperatures, $\partial n/\partial\mu = eN_0$, so that Eq. (4.3.1) can be written as

$$I(\mathbf{r}) = e^2 \int d\mathbf{r}' \{\nu(\mathbf{r},\mathbf{r}')\, N_0(\mathbf{r}')\, \mu(\mathbf{r}') - \nu(\mathbf{r}',\mathbf{r})\, N_0(\mathbf{r})\, \mu(\mathbf{r})\} \quad . \tag{4.3.7}$$

From Eq. (4.3.6) and the definition of $T(\mathbf{r},\mathbf{r}')$,

$$T(\mathbf{r},\mathbf{r}') = hN_0(\mathbf{r}')\nu(\mathbf{r},\mathbf{r}') = \frac{\hbar^2}{\tau_i^2} |G(\mathbf{r},\mathbf{r}')|^2 , \tag{4.3.8}$$

which leads to our previous result:

$$I(\mathbf{r}) = \frac{e^2}{h}\int d\mathbf{r}'\{T(\mathbf{r},\mathbf{r}')\mu(\mathbf{r}') - T(\mathbf{r}',\mathbf{r})\mu(\mathbf{r})\} . \tag{4.3.9}$$

Although somewhat arm-waving, this derivation presents an alternative viewpoint which is helpful in clarifying our physical assumptions. Moreover, a connection to semiclassical transport theory can be established by studying the moments of the distribution of hopping frequencies, as we shall see in the following section.

4.3.2 Moments of the Hopping Frequency Distribution Function

In the physical picture described above, transport was viewed as a series of "hops" between inelastic scattering events. Buried in the distribution of hopping frequencies $\nu(\mathbf{r},\mathbf{r}')$ is the quantum mechanical behavior of the electron. By considering moments of this distribution function, we will obtain obtain expressions for drift and diffusion which also account for the wave nature of electrons. As a starting point, we consider a bulk homogeneous medium with a uniform potential drop along x. We recall the expression for the dissipated power from Section 4.2.3,

$$P_0 = \frac{1}{2}\int d\mathbf{r} \int d\mathbf{r}' \, g_S(\mathbf{r},\mathbf{r}') \left[(x-x')\frac{\partial\mu}{\partial x}\right]^2 . \tag{4.3.10}$$

Because $g_S(\mathbf{r},\mathbf{r}')$ depends only on the difference $\mathbf{r}-\mathbf{r}'$ (the Green function has this property in the absence of a magnetic field), it is convenient to make a change of variables,

$$\frac{P_0}{\Omega} = \frac{1}{2}\left(\frac{\partial\mu}{\partial x}\right)^2 \int d(\mathbf{r}-\mathbf{r}') \, (x-x')^2 \, g_S(|\mathbf{r}-\mathbf{r}'|) , \tag{4.3.11}$$

since the integration over $(\mathbf{r}+\mathbf{r}')$ simply yields the volume Ω. For convenience, we

choose the origin of our coordinate system such that $\mathbf{r}' = 0$. Recalling Eq. (4.3.8) and the definition of $g_S(\mathbf{r},\mathbf{r}')$, we can express the dissipated power in terms of the distribution of hopping frequencies,

$$\frac{P_0}{\Omega} = \frac{e^2}{2}\left(\frac{\partial\mu}{\partial x}\right)^2 \int d\mathbf{r}\, x^2\, N_0 \left(\frac{\nu(\mathbf{r},\mathbf{0}) + \nu(\mathbf{0},\mathbf{r})}{2}\right) , \tag{4.3.12}$$

where we have assumed that $N_0(\mathbf{r}) = N_0$, a constant independent of position. This is strictly correct only in an ensemble-averaged sense, which is the spirit of our present analysis. Equating (P_0/Ω) to $\sigma(\partial\mu/\partial x)^2$ we obtain the conductivity σ.

$$\sigma = e^2 N_0 D = \frac{e^2}{2} N_0 \int d\mathbf{r}\, x^2 \left(\frac{\nu(\mathbf{r},\mathbf{0}) + \nu(\mathbf{0},\mathbf{r})}{2}\right) . \tag{4.3.13}$$

From this, we can identify the diffusion coefficient D as the second moment of the hopping frequency distribution,

$$D \equiv \frac{1}{2}\int d\mathbf{r}\, x^2 \left(\frac{\nu(\mathbf{r},\mathbf{0}) + \nu(\mathbf{0},\mathbf{r})}{2}\right) . \tag{4.3.14}$$

This relationship is well-known in the study of ordinary Brownian motion [45], and thereby gives support to our physical interpretation of the transport process. We define the x-component of the drift velocity as the first moment of the hopping frequency distribution,

$$(v_d)_x \equiv \int d\mathbf{r}\, x \left(\frac{\nu(\mathbf{r},\mathbf{0}) + \nu(\mathbf{0},\mathbf{r})}{2}\right) . \tag{4.3.15}$$

In the present analysis, we have not only made contact with semiclassical drift-diffusion, we have indicated how it can be extended. By computing $\nu(\mathbf{r},\mathbf{r}')$ from the Schrödinger equation, we have incorporated quantum mechanical effects into the calculation of D and v_d.

4.4 Simple Examples

4.4.1 Analytical Examples Using Classical Dynamics

Before considering a more complicated numerical example, we present some simple results obtained by using classical dynamics. Assuming a constant inelastic scattering time τ_i, we can describe the hopping distribution by

$$\nu(\mathbf{r},\mathbf{0}) = \nu(\mathbf{0},\mathbf{r}) = \int_0^\infty \frac{dt}{\tau_i^2}\, \delta(\mathbf{r}-\mathbf{r}_0(t))\, e^{-t/\tau_i} \quad , \tag{4.4.1}$$

where $\mathbf{r}_0(t)$ describes the classical trajectory of the electron, as obtained from Newton's laws. The damping exponential represents the effect of the imaginary potential $i\hbar/2\tau_i$ in the Schrödinger equation. As discussed in Section 4.2, electrons propagating through a device are continuously lost to perfectly absorbing reservoirs, thereby suffering an inelastic collision. For convenience, we have assumed that $\mathbf{r}_0(t=0) = \mathbf{0}$. In general, we will be concerned with moments of the hopping frequency distribution,

$$E_\nu\{x^n\} \equiv \int d\mathbf{r}\, \frac{x^n}{n!}\nu(\mathbf{r}) = \int d\mathbf{r}\, \frac{x^n}{n!} \int_0^\infty \frac{dt}{\tau_i^2}\, \delta(\mathbf{r}-\mathbf{r}_0(t))\, e^{-t/\tau_i} \quad . \tag{4.4.2}$$

If we interchange the order of integration and use the delta function to our advantage, this simplifies to

$$E_\nu\{x^n\} = \int_0^\infty \frac{dt}{\tau_i^2}\, \frac{[x_0(t)]^n}{n!}\, e^{-t/\tau_i} \quad , \tag{4.4.3}$$

where $x_0(t)$ is the x-component of the trajectory $\mathbf{r}_0(t)$. In the following discussion, we will evaluate this expression for a few simple classical trajectories, to determine the drift velocity $\mathbf{v}_d$ and diffusion coefficient D, according to the definitions justified in Section 4.3.2:

$$(v_d)_x \equiv E_\nu\{x\} \ , \tag{4.4.4a}$$

$$D \equiv E_\nu\{x^2\} \ . \tag{4.4.4b}$$

To start with, we consider a classical electron moving without any influence from electric and magnetic fields. Assuming the particle has an initial velocity v_x in the x-direction, the x-component of the trajectory is $x_0(t) = v_x t$. Because we are interested in the collective behavior for an ensemble of electrons, we must average over all possible initial states. We assume that the initial velocity of an electron is uniformly distributed over all directions in a sample, due to the action of scattering processes. With these assumptions, we evaluate Eqs. (4.4.4a) and (4.4.4b):

$$<(v_d)_x> = \left\langle \int_0^\infty \frac{dt}{\tau_i^2} \, v_x t \, e^{-t/\tau_i} \right\rangle = <v_x> = 0 \ , \tag{4.4.5a}$$

$$<D> = \left\langle \int_0^\infty \frac{dt}{\tau_i^2} \, v_x^2 t^2 \, e^{-t/\tau_i} \right\rangle = <v_x^2 \tau_i> \neq 0 \ , \tag{4.4.5b}$$

where we have used the angle brackets $<\cdot>$ to denote ensemble averaging. Without the influence of fields, the drift velocity of an ensemble of electrons is zero; diffusion, however, is not.

As a second example, we consider the effect of an electric field in the x-direction, causing an electron to accelerate: $x_0(t) = v_x t + a_x t^2/2$, where $a_x = eE/m^*$ is the acceleration due to the field. We evaluate the drift velocity, and obtain,

$$<(v_d)_x> = \left\langle \int_0^\infty \frac{dt}{\tau_i^2} (v_x t + \frac{a_x t^2}{2}) \, e^{-t/\tau_i} \right\rangle = <v_x> + <a_x \tau_i> \ . \tag{4.4.6}$$

After averaging over all initial velocities, we again find that $<v_x> = 0$. Because of the acceleration, however, the drift velocity is not zero,

$$\langle (v_d)_x \rangle = \frac{e\tau_i}{m^*} E_x \ . \tag{4.4.7}$$

Finally, we assume a magnetic field along the z-axis, and an electric field in the x-y plane. In Appendix 4.A3, the x-component of the classical trajectory is shown to be:

$$x_0(t) = \left(\frac{a_x}{\omega_c^2} - \frac{v_y}{\omega_c} \right) (1 - \cos\omega_c t) + \left(\frac{a_y}{\omega_c^2} + \frac{v_x}{\omega_c} \right) \sin\omega_c t - \frac{a_y}{\omega_c} t \ , \tag{4.4.8}$$

where we have introduced the cyclotron frequency for an electron $\omega_c = |e|B/m^*$. Substituting Eq. (4.4.8) into Eq. (4.4.3) and performing the necessary integrals, we obtain the drift velocity,

$$(v_d)_x = \left(\frac{a_x}{\omega_c} - v_y \right) \frac{\omega_c \tau_i}{1 + \omega_c^2 \tau_i^2} + \left(\frac{a_y}{\omega_c} + v_x \right) \frac{1}{1 + \omega_c^2 \tau_i^2} - \frac{a_y}{\omega_c} \ . \tag{4.4.9}$$

If we average over all initial velocities, $\langle v_x \rangle$ and $\langle v_y \rangle$ vanish, leaving only the terms involving acceleration. By substituting in the acceleration due to the electric field,

$$\langle (v_d)_x \rangle = \frac{e\tau_i}{m^*} \left[\frac{E_x}{1 + \omega_c^2 \tau_i^2} - \frac{E_y \, \omega_c \tau_i}{1 + \omega_c^2 \tau_i^2} \right] \ . \tag{4.4.10}$$

This velocity defines the x-component of the drift current, which we can use to determine conductivity:

$$\sigma_{xx} = \left. \frac{e \, n \, (v_d)_x}{E_x} \right|_{E_y=0} = \frac{\sigma_0}{1 + \omega_c^2 \tau_i^2} \ , \tag{4.4.11a}$$

$$\sigma_{xy} = \left. \frac{e \, n \, (v_d)_x}{E_y} \right|_{E_x=0} = -\sigma_0 \frac{\omega_c \tau_i}{1 + \omega_c^2 \tau_i^2} \ . \tag{4.4.11b}$$

Following a similar derivation for the y-component of the drift velocity, we can define σ_{yx} and σ_{yy}, and produce the usual magnetoconductivity tensor [46],

$$\sigma = \frac{\sigma_0}{1 + \omega_c^2 \tau_i^2} \begin{bmatrix} 1 & -\omega_c \tau_i \\ \omega_c \tau_i & 1 \end{bmatrix} . \tag{4.4.12}$$

In summary, we have shown that some familiar results can be obtained by assuming a classical motion of electrons. Having developed some confidence, we proceed to demonstrate quantum mechanical effects in a numerical example.

4.4.2 Numerical Example of Localization

In recent years, considerable attention has been given to the effects of disorder on electron transport [3,4]. In particular, Anderson has shown [47] that for a sufficiently high degree of disorder, and in the absence of inelastic scattering, conductance can *decrease exponentially* with length. Electron wave functions become spatially localized, having envelope functions that decay exponentially with distance. As the overlap between localized states decreases, the conductance vanishes. In the presence of inelastic scattering, however, electrons interacting with phonons can "hop" between localized states. For this rare circumstance, inelastic scattering actually *improves* the conductance.

A illustration of this effect is well suited to our model. In principle, we can calculate the conductivity for any chain of randomly spaced impurities. We consider the average conductivity of many such chains, as a function of the inelastic scattering time τ_i. We will show that, as τ_i increases, conductivity falls off exponentially. Furthermore, if we neglect interference effects, the localization behavior is destroyed, and conductance is constant for large τ_i. In the following sections we describe our numerical solution in detail. We derive an expression for the conductivity of a quasi one-dimensional wire, which must be evaluated numerically. A scattering matrix technique is presented for the calculation of the Green function, and we summarize our results.

4.4.2.1 Conductivity For a Quasi One-dimensional Wire

In previous discussion, we developed an expression for the conductivity by considering the power dissipated within a sample:

$$P_0 = \frac{1}{2}\int d\mathbf{r} \int d\mathbf{r}' \, (x-x')^2 \left(\frac{\partial \mu}{\partial x}\right)^2 g_S(\mathbf{r},\mathbf{r}') \ , \tag{4.4.13a}$$

where

$$g_S(\mathbf{r},\mathbf{r}') = \frac{e^2}{2h}\{T(\mathbf{r},\mathbf{r}') + T(\mathbf{r}',\mathbf{r})\} \ . \tag{4.4.13b}$$

We return to this derivation, restricting our attention to a quasi one-dimensional wire: a two-dimensional electron gas whose width is smaller than an inelastic length L_i. As a result of confinement, the allowed values of the transverse momentum are quantized. Transport can be described in terms of a finite number of transverse scattering channels or modes. In effect, we are describing an electron "wave guide." From this standpoint, we choose a mode-based representation for $T(\mathbf{r},\mathbf{r}')$,

$$T(\mathbf{r},\mathbf{r}') \leftrightarrow T_{mn}(x,x') \ , \tag{4.4.14}$$

which represents the probability that an electron introduced into mode n at position x' will arrive in mode m at x before being inelastically scattered. In the absence of magnetic field, we have detailed balance, so that $T_{mn}(x,x')$ is the same under an interchange of x and x', and m and n. In this mode-based representation, the dissipated power can be expressed as,

$$P_0 = \frac{e^2}{4h}\left(\frac{\partial \mu}{\partial x}\right)^2 \sum_{mn} \int dx \int dx' \, (x-x')^2 \, T_{mn}(x,x') \ , \tag{4.4.15}$$

where we have included an extra factor of 1/2 to avoid double-counting between pairs of modes. From a property of the Green function, $T_{mn}(x,x')$ is dependent only on the

difference $(x-x')$. It is convenient to change our variables of integration to $(x-x')$ and $(x+x')$,

$$\frac{P_0}{L} = \frac{e^2}{4h}\left(\frac{\partial\mu}{\partial x}\right)^2 \sum_{mn} \int d(x-x')\,(x-x')^2\, T_{ij}(x-x') \quad , \tag{4.4.16}$$

where L is the length of the wire. For convenience, we choose $x' = 0$ (equivalent to another change of variables), and equate P_0/L to $\sigma(\partial\mu/\partial x)^2$. By averaging over all possible impurity configurations, we write the ensemble-average conductivity as,

$$\langle\sigma\rangle = \frac{e^2}{4h} \int dx\, x^2 \left\langle \sum_{mn} T_{mn}(x) \right\rangle \quad . \tag{4.4.17}$$

We assume that in an ensemble-average sense, the transmission function can be characterized by an exponential decay length L_D. For classical transport, this decay length is simply the inelastic length L_i; it represents the decreasing probability that an electron will travel farther than L_i before being inelastically scattered. For quantum transport, however, L_D must be determined by explicit calculation. We intend to show that for strong disorder, L_D becomes the localization length. In any event, we define

$$\left\langle \sum_{mn} T_{mn}(x) \right\rangle \equiv T_0\, e^{-|x|/L_D} \quad , \tag{4.4.18}$$

where T_0 is a constant chosen to satisfy the normalization of the ensemble-averaged function as follows. In Section 4.2.1, we defined the normalization of $T(\mathbf{r},\mathbf{r}')$ as,

$$\int d\mathbf{r}'\, T(\mathbf{r}',\mathbf{r}) = \frac{hN_0(\mathbf{r})}{\tau_i} \quad , \tag{4.4.19a}$$

For our mode-based representation, this becomes,

$$\sum_{m} \int dx\, T_{mn}(x) = \frac{hN_{0,n}}{\tau_i} \quad , \tag{4.4.19b}$$

where $N_{0,n}$ is the density of states for mode n, evaluated at the injection point $x'=0$. By

summing over all injected modes n, we determine our normalization condition,

$$\int dx \left\langle \sum_{mn} T_{mn}(x) \right\rangle = \frac{h}{\tau_i} \sum_{n} N_{0,n} \ , \tag{4.4.20}$$

which we equate to the integral of the function defined by Eq. (4.4.18), to determine the constant T_0,

$$T_0 = \frac{h}{L_D \tau_i} \sum_{n} N_{0,n} \ . \tag{4.4.21}$$

We can now evaluate the ensemble-averaged conductivity from Eqs. (4.4.17) and (4.4.18). If we assume that the wire is sufficiently long,

$$\langle \sigma \rangle = \frac{e^2}{4h} T_0 \int_{-\infty}^{\infty} dx \, x^2 \, e^{-|x|/L_D} = \frac{e^2}{h} T_0 L_D^3 \ , \tag{4.4.22}$$

and substituting in the definition of T_0 from Eq. (4.4.21), we obtain our final result:

$$\langle \sigma \rangle = e^2 \left(\sum_{n} N_{0,n} \right) \frac{L_D^2}{\tau_i} \equiv e^2 N_0 D \ , \tag{4.4.23}$$

which is the usual expression for conductivity; however, the details of a quantum mechanical solution have been buried in the calculation of L_D. Our numerical solution, therefore, will proceed as follows. Given a chain of randomly spaced impurities, we calculate the Green function of the Schrödinger equation, and sum the function $T_{mn}(x)$ over all modes. This composite function is then fit to a decaying exponential on either side of the injection point, and the decay length L_D is determined. By averaging over many random samples, we determine an average decay length $\langle L_D \rangle$. A final concern is the averaging process used to determine this quantity. Because of the random impurity spacing, values of L_D will fluctuate from sample to sample. In the localized regime, this fluctuation can be quite large. Anderson et. al. [48] have shown that the inverse of the localization length has a well behaved distribution. For this reason, we compute the

mean of $1/L_D$ and invert it, to determine our average $\langle L_D \rangle$. With this result, we compute the average conductivity from Eq. (4.4.23).

4.4.2.2 Numerical Evaluation of the Green Function

We now seek a numerical solution for the Green function of the Schrödinger equation, defined by

$$\left(E_F - H + \frac{i\hbar}{2\tau_i} \right) G(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}') \ , \tag{4.4.24a}$$

$$H = \frac{-\hbar^2}{2m^*}\nabla^2 + eV(\mathbf{r}) \ . \tag{4.4.24b}$$

Because our quasi one-dimensional wire bears a striking resemblance to a wave guide, we will borrow from the solution techniques developed for the analysis of microwave networks. In particular, we recognize that the confining potential in the transverse direction gives rise to a set of modes m, so that the wave function in the regions between scatterers can be represented as [49],

$$\psi^{\pm}(x,y;k_m) = \phi_m(y)\, e^{\pm ik_m x} \ , \tag{4.4.25}$$

where the $\pm$ signs represent propagation in the positive and negative x-directions, respectively (see Fig. 4.4a). For simplicity, we assume that each mode has a parabolic dispersion relation,

$$E = \epsilon_m + \frac{\hbar^2 k_m^2}{2m^*} \ , \tag{4.4.26}$$

where ϵ_m is the minimum energy of mode m. Because ϵ_m increases as m increases, there will only be a finite number of modes which are allowed to propagate at a given Fermi energy E_F. A natural representation of the wave function, therefore, is a column vector at each point x whose M elements specify the (complex) amplitude of each mode. The

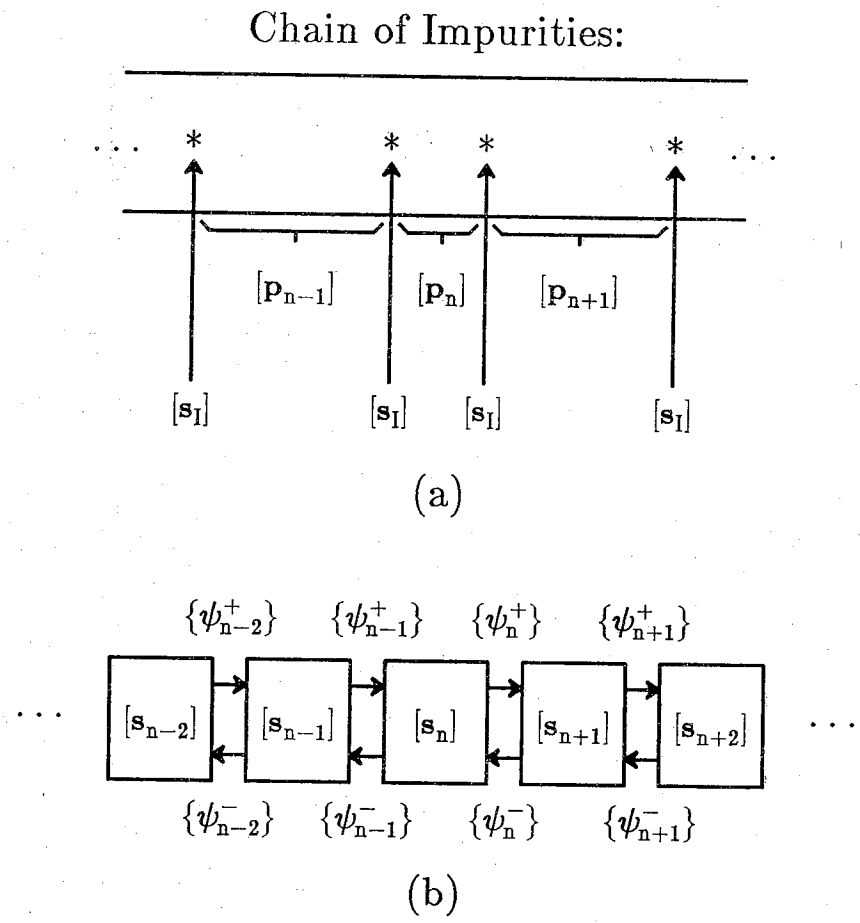


Figure 4.4 (a) A quasi one-dimensional chain of impurities is represented by a series of scattering matrices, which couple the amplitudes of transverse modes. (b) By combining scattering matrices for each impurity and an adjacent region of free propagation, the problem is simplified.

effect of an impurity is to couple the amplitudes of different modes, a process which can be described by a scattering matrix:

$$\begin{pmatrix} \psi_n^- \\ \psi_{n+1}^+ \end{pmatrix} = \begin{bmatrix} \mathbf{r} & \mathbf{t}' \\ \mathbf{t} & \mathbf{r}' \end{bmatrix} \begin{pmatrix} \psi_n^+ \\ \psi_{n+1}^- \end{pmatrix} , \tag{4.4.27}$$

where $\{\psi_n^{+-}\}$ is an $M\times 1$ column vector of amplitudes for waves traveling in the positive (negative) direction, at a position x_n. The elements of the scattering matrix ([r], [t], [r′], [t′]) are themselves $M\times M$ matrices.

Any sample can be modeled with a series of scattering matrices, representing collisions with elastic scatterers and free propagation in the regions between. In general, the scattering matrix for each impurity should be different, depending on the impurity strength, position in the transverse direction, etc. We avoid this complication, however, and define the scattering matrix for each impurity as [49],

$$[\mathbf{s}_I]_{ij} = \delta_{ij} + \frac{\beta}{\alpha} , \tag{4.4.28a}$$

where

$$\beta = \frac{e^{2iM\alpha} - 1}{2M} , \tag{4.4.28b}$$

and M is the number of modes. The parameter α is related to the scattering strength, so that the elastic mean free path can be determined from

$$\Lambda_{el} = \frac{1 - M|\beta|^2}{M|\beta|^2} . \tag{4.4.29}$$

We also require a scattering matrix to represent the regions of free propagation between impurities. Although there is no coupling between modes in these regions, each mode acquires a phase shift proportional to the length of the region, so that the scattering matrix is

$$[\mathbf{p}_n]_{ij} = \delta_{ij}\, e^{ik_j d_n} \quad , \tag{4.4.30}$$

where d_n is the length of propagation region n. The wave vector k_j for mode j is defined by

$$k_j = \left[\frac{2m^*}{\hbar^2}\left(E_F - \epsilon_j + \frac{i\hbar}{2\tau_i}\right)\right]^{1/2} \tag{4.4.31}$$

Because of the complex potential $i\hbar/2\tau_i$ in Eq. (4.4.24a), k_j will be complex. This accounts for the loss of electrons into the continuous distribution of reservoirs, as discussed in Section 4.2.1.

We can combine the effects of any two scattering matrices with the following law of composition: $[\mathbf{s}_{12}] \equiv [\mathbf{s}_1] \otimes [\mathbf{s}_2]$ is defined by the sub-matrices

$$\mathbf{r}_{12} = \mathbf{r}_1 + \mathbf{t}_1'\mathbf{r}_2\,[\mathbf{I} - \mathbf{r}_1'\mathbf{r}_2]^{-1}\,\mathbf{t}_1 \quad , \tag{4.4.32a}$$

$$\mathbf{t}_{12} = \mathbf{t}_2\,[\mathbf{I} - \mathbf{r}_1'\mathbf{r}_2]^{-1}\,\mathbf{t}_1 \quad , \tag{4.4.32b}$$

$$\mathbf{r}_{12}' = \mathbf{r}_2' + \mathbf{t}_2\,[\mathbf{I} - \mathbf{r}_1'\mathbf{r}_2]^{-1}\,\mathbf{r}_1'\mathbf{t}_2' \quad , \tag{4.4.32c}$$

$$\mathbf{t}_{12}' = \mathbf{t}_1'\,[\mathbf{I} + \mathbf{r}_2[\mathbf{I} - \mathbf{r}_1'\mathbf{r}_2]^{-1}\mathbf{r}_1']\,\mathbf{t}_2' \quad . \tag{4.4.32d}$$

Therefore, the scattering matrix for a chain of impurities can be calculated by combining the individual scattering matrices:

$$[\mathbf{s}_c] = [\mathbf{s}_I] \otimes [\mathbf{p}_1] \otimes [\mathbf{s}_I] \otimes [\mathbf{p}_2] \otimes \cdots \quad . \tag{4.4.33}$$

For subsequent analysis, we define the column vector $\{\psi_n^{\pm}\}$ to represent the wave amplitudes at the end of interval n and to the immediate left of impurity $n+1$, as shown in Fig. 4.4b. We can combine the scattering matrices for each impurity and the adjacent region of propagation into a single matrix,

$$[\mathbf{s}_n] = [\mathbf{s}_I] \otimes [\mathbf{p}_n] \ , \tag{4.4.34}$$

which connects column vectors $\{\psi_{n-1}^{\pm}\}$ and $\{\psi_n^{\pm}\}$.

With an understanding of scattering matrices, we proceed to solve for the Green function of the Schrödinger equation, which we present in a simplified form,

$$\left[\frac{d^2}{dx^2} + k_j^2\right] G_j(x) = \delta(x)\, \delta_{ij} \ . \tag{4.4.35}$$

where i is the injected mode. Near $x = 0$, we have two boundary conditions. The first is that the wave function must be continuous,

$$G_j(x{=}0^-) \ = \ G_j(x{=}0^+) \ . \tag{4.4.36a}$$

The second, obtained by integrating Eq. (4.4.35), describes the effect of injection into mode i,

$$\frac{d}{dx}G_j(x)\bigg|_{x=0^+} - \frac{d}{dx}G_j(x)\bigg|_{x=0^-} = \delta_{ij} \ . \tag{4.4.36b}$$

On either side of $x = 0$, we represent the solution $G_j(x)$ as:

$$G_j(x{=}0^-) \ = \ \{\mathbf{a}^+\}_j\, e^{ik_jx} \ + \ \{\mathbf{a}^-\}_j\, e^{-ik_jx} \ , \tag{4.4.37a}$$

$$G_j(x{=}0^+) \ = \ \{\mathbf{b}^+\}_j\, e^{ik_jx} \ + \ \{\mathbf{b}^-\}_j\, e^{-ik_jx} \ , \tag{4.4.37b}$$

where $\{\mathbf{a}^{\pm}\}$ and $\{\mathbf{b}^{\pm}\}$ are $M{\times}1$ column vectors describing the wave function amplitude in each mode. Expressing the boundary conditions in terms of these column vectors, we obtain,

$$\{\mathbf{a}^+\}_j + \{\mathbf{a}^-\}_j \ = \ \{\mathbf{b}^+\}_j + \{\mathbf{b}^-\}_j \tag{4.4.38a}$$

$$ik_j\, (\{\mathbf{b}^+\}_j - \{\mathbf{b}^-\}_j - \{\mathbf{a}^+\}_j + \{\mathbf{a}^-\}_j) \ = \ \delta_{ij} \tag{4.4.38b}$$

For an arbitrary chain with N impurities, we can combine the scattering matrices to the

left and right of the injection point into two composite matrices, $[\mathbf{S}_L]$ and $[\mathbf{S}_R]$, as shown in Fig. 4.5. This greatly simplifies an expression of the remaining two boundary conditions. If we assume that the contacts at each end of the chain are perfectly absorbing, we can define $\{\mathbf{a}^+\}$ and $\{\mathbf{b}^-\}$ by what is reflected from each half of the chain:

$$\{\mathbf{a}^+\} = [\mathbf{r}_L'] \{\mathbf{a}^-\} \ , \tag{4.4.39a}$$

$$\{\mathbf{b}^-\} = [\mathbf{r}_R] \{\mathbf{b}^+\} \ . \tag{4.4.39b}$$

We now have four equations to solve for our four unknown vectors $\{\mathbf{a}^\pm\}$ and $\{\mathbf{b}^\pm\}$. Of course, these same equations must solved for injection into each mode i. To organize all of the column-vector solutions, we define the following matrices,

$$[\mathbf{b}_s^+] = \left\{ i[\mathbf{k}] \left([\mathrm{I} - \mathbf{r}_R] + [\mathrm{I} - \mathbf{r}_L'][\mathrm{I} + \mathbf{r}_L']^{-1}[\mathrm{I} + \mathbf{r}_R] \right) \right\}^{-1} \ , \tag{4.4.40a}$$

$$[\mathbf{a}_s^-] = [\mathrm{I} + \mathbf{r}_L']^{-1}[\mathrm{I} + \mathbf{r}_R][\mathbf{b}_s^+] \ , \tag{4.4.40b}$$

where we have introduced the matrix $[\mathbf{k}]_{ij} \equiv k_j \delta_{ij}$. Each column i of the matrices $[\mathbf{b}_s^+]$ and $[\mathbf{a}_s^-]$ is the solution vector for injection into mode i. In the same way, we can organize all of the column vectors $\{\psi_n^\pm\}$ for each node n into a matrix $[\psi_{s,n}^\pm]$. All subsequent discussion will deal directly with these matrices.

What remains is to determine the amplitudes $[\psi_{s,n}^\pm]$ at all other points in the chain. We restrict our attention to the left half of the chain, although the same analysis can be applied to the right half as well. In general, we consider waves incident from the right on two scatterers, as shown in Fig. 4.6. Assuming that the output on the left-hand side is completely absorbed by the contact, we can write the wave amplitudes between the two scatterers by summing the contributions from multiply reflected waves:

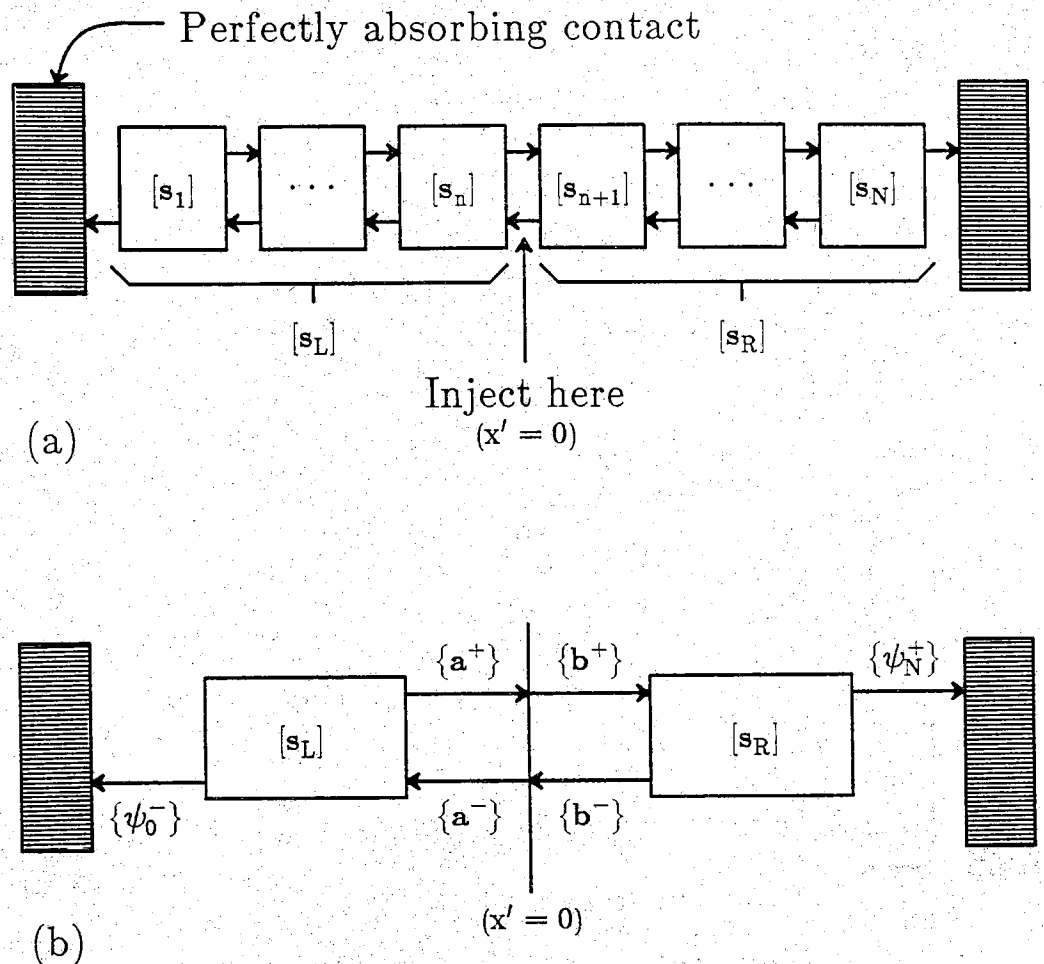


Figure 4.5 (a) Scattering matrices on either side of the injection point can be combined into two composite scattering matrices, $[s_L]$ and $[s_R]$. (b) Simplified problem for Green function solution.

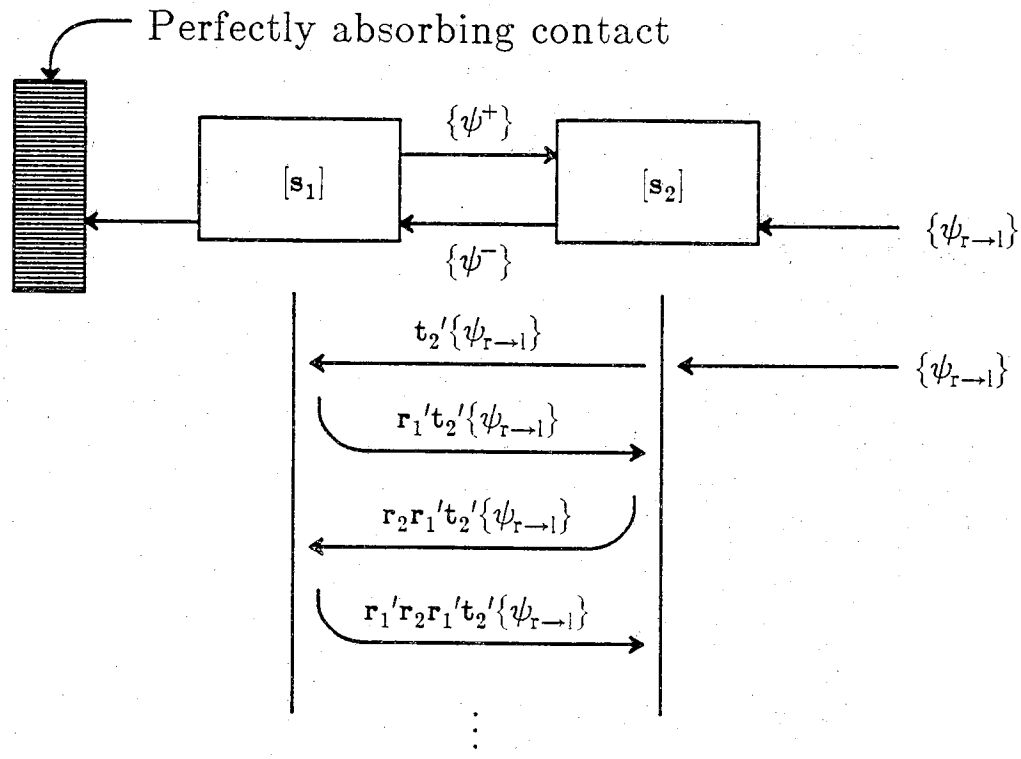


Figure 4.6 Given an incident wave amplitude $\{\psi_{r\to l}\}$, we can calculate the amplitudes $\{\psi^+\}$ and $\{\psi^-\}$ between two scatters by summing multiply reflected paths.

$$[\psi^+] = \mathbf{r}_1'\mathbf{t}_2'\psi_{r\to l} + \mathbf{r}_1'\mathbf{r}_2\mathbf{r}_1'\mathbf{t}_2'\psi_{r\to l} + \mathbf{r}_1'\mathbf{r}_2\mathbf{r}_1'\mathbf{r}_2\mathbf{r}_1'\mathbf{t}_2\psi_{r\to l} + \cdots$$

$$= \mathbf{r}_1'[\mathrm{I} - \mathbf{r}_2\mathbf{r}_1']^{-1}\mathbf{t}_2'\psi_{r\to l} \ , \tag{4.4.41a}$$

$$[\psi^-] = [\mathrm{I} - \mathbf{r}_2\mathbf{r}_1']^{-1}\mathbf{t}_2'\psi_{r\to l} \ . \tag{4.4.41b}$$

Similar expressions can be obtained (for analysis of the right-hand side), if we consider a wave incident from the left and absorbed on the right,

$$[\psi^+] = [\mathrm{I} - \mathbf{r}_1'\mathbf{r}_2]^{-1}\mathbf{t}_1\psi_{l\to r} \ , \tag{4.4.42a}$$

$$[\psi^-] = \mathbf{r}_2[\mathrm{I} - \mathbf{r}_1'\mathbf{r}_2]^{-1}\mathbf{t}_1\psi_{l\to r} \ . \tag{4.4.42b}$$

Using these expressions, we can determine $[\psi^{\pm}_{s,n}]$ at all remaining points with an iterative technique, which is illustrated in Fig. 4.7. We define the scattering matrix $[\mathbf{c}_n]$ as the composite of all matrices up to the point where we will determine $[\psi^{\pm}_{s,n}]$,

$$[\mathbf{c}_n] \equiv [\mathbf{s}_1] \otimes [\mathbf{s}_2] \otimes \cdots \otimes [\mathbf{s}_n] \ . \tag{4.4.43}$$

For the amplitude immediately left of the injection point, we have the solution $[\mathbf{a}_s^-]$, determined from Eq. (4.4.40b); this is our initial group of incident vectors $[\psi_{r\to l}]$. We can calculate $[\psi^{\pm}_{s,n}]$ between matrices $[\mathbf{c}_n]$ and $[\mathbf{s}_{n+1}]$ from Eqs. (4.4.41). For the amplitudes at the next point, we use the group of incident vectors $[\psi^-_{s,n}]$ which we have just determined. We evaluate $[\psi^{\pm}_{s,n-1}]$ from Eqs. (4.4.41), between matrices $[\mathbf{c}_{n-1}]$ and $[\mathbf{s}_n]$. This solution, in turn, defines our next group of incident vectors, and the process repeats until all amplitudes have been determined.

Having calculated the amplitudes $[\psi^{\pm}_{s,n}]$ at each node n, we can calculate the form of the transmission function,

$$S(x) \equiv \sum_{ij} \left| [\psi^+_{s,n}]_{ij} + [\psi^-_{s,n}]_{ij} \right|^2 \sim \sum_{ij} T_{ij}(x_n) \ . \tag{4.4.44}$$

Proper normalization of this function is immaterial. On either side of the injection

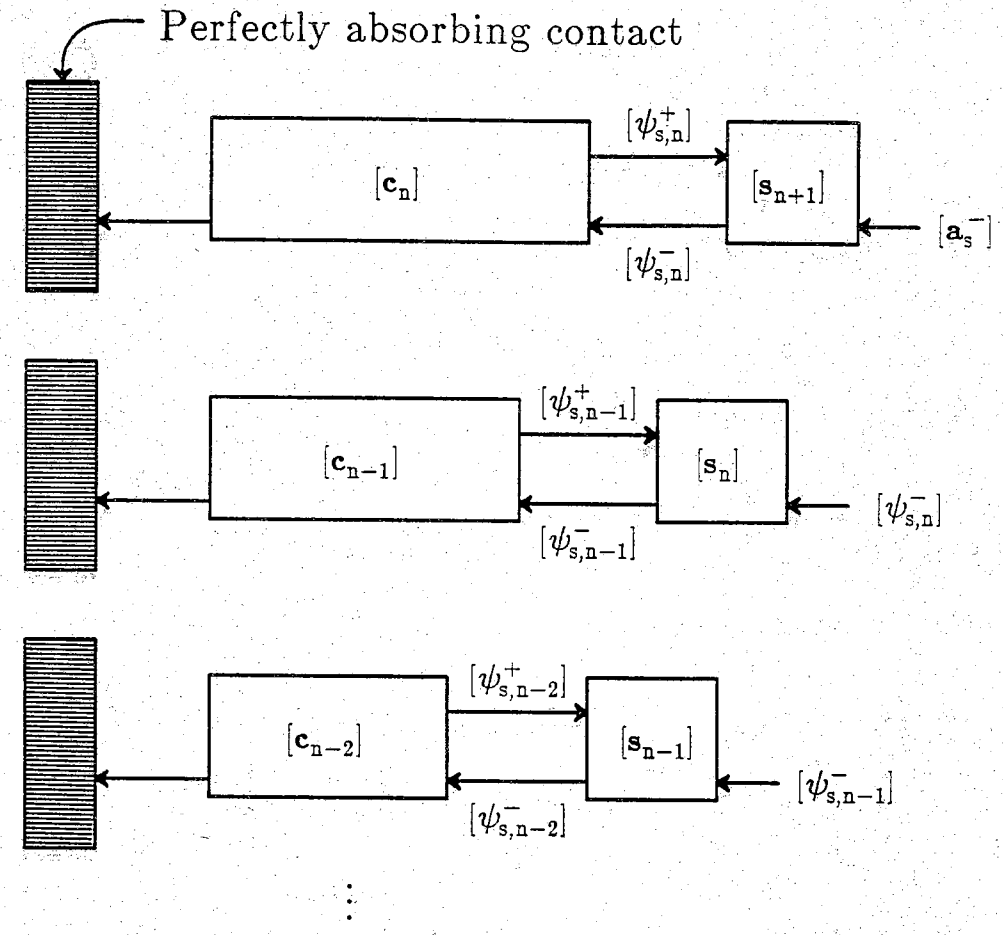


Figure 4.7 The wave amplitudes $[\psi^{\pm}_{s,n}]$ at each node n can be determined from an iterative technique. The initial incident wave $[\mathbf{a}_s^-]$ is used to determine $[\psi^{\pm}_{s,n}]$, which in turn provides the next incident wave $[\psi^-_{s,n}]$ for the solution at node *n-1*. The process is repeated until the wave function has been determined at all points.

point, S(x) is fit to a straight line on a logarithmic scale, and it is the resulting slope (i.e., $1/L_D$) which is physically meaningful.

Finally, we recognize that a semiclassical analysis can be performed by using squared-magnitudes in place of (complex) amplitudes,

$$[\mathbf{s}_I] \rightarrow [\mathbf{S}_I]_{ij} \equiv |[\mathbf{s}_I]_{ij}|^2 , \tag{4.4.45a}$$

$$[\mathbf{p}_n] \rightarrow [\mathbf{P}_n]_{ij} \equiv |[\mathbf{p}_n]_{ij}|^2 , \tag{4.4.45b}$$

$$[\psi^{+-}_{s,n}] \rightarrow [\Psi^{+-}_{s,n}] \tag{4.4.45c}$$

where we have used capital letters to distinguish the probabilistic solution. In addition to the changes shown above, we must replace $i[\mathbf{k}] \rightarrow \mathrm{Re}\{[\mathbf{k}]\}$ in Eq. (4.4.40a), and the transmission function is calculated as:

$$S(x) \rightarrow S_{scl}(x) \equiv \sum_{ij} \{[\Psi^{+}_{s,n}]_{ij} + [\Psi^{-}_{s,n}]_{ij}\} . \tag{4.4.46}$$

All other aspects of the solution, however, are identical. The ability to neglect phase information allows us to properly demonstrate localization: For large inelastic scattering times τ_i, we will show that quantum mechanical solution exhibits localization, while the semiclassical (probabilistic) solution does not.

4.4.2.3 Results and Conclusions

Using the model developed in the previous sections, we have numerically computed the Green function for many random samples, with a range of scattering times τ_i. Each sample had five transverse modes and 400 impurities, and the injection point was immediately left of the 200^{th} impurity. Impurities for each sample were randomly placed with a uniform distribution of spacings, and the impurity strength was chosen to be $\alpha = 0.3$, so that the elastic length Λ_{el} was 4 impurities. An estimate of the localization length [50] is $\Lambda_{loc} = M\Lambda_{el}$, where M is the number of modes. In this case $M = 5$, so

that samples should exhibit localization when the inelastic length L_i exceeds 20 impurities. Inelastic scattering times τ_i were chosen between 10^{-12} s and 10^{-8} s, corresponding to inelastic lengths between 0.2 and 95 impurities. All results were verified against longer chains, to ensure that absorption at the contacts had little effect on the analysis.

To provide some insight into the transmission function S(x) defined by Eq. (4.4.44), we have plotted the results for two arbitrary samples, with different inelastic scattering times τ_i. In Fig 4.8, we present the function S(x) for $\tau_i = 10^{-10}$ s, which corresponds to an inelastic length L_i of about 9 impurities; hence, this sample is in the so-called weak localization regime. Although the function is predominantly characterized by its exponential decay, small fluctuations are clearly visible. These fluctuations are direct evidence of quantum mechanical interference effects. If the same sample were analyzed using the semiclassical procedure (Eq. (4.4.46)), the fluctuations would disappear, and only the exponential decay would remain. As the inelastic scattering time increases, the fluctuations become larger. In Fig 4.9, we present the function S(x) for $\tau_i = 10^{-8}$ s, which corresponds to an inelastic length of about 95 impurities. Although the general character of exponential decay remains, the fluctuations have added considerable scatter to the decay length L_D.

Statistical averages of the decay length L_D were obtained by by averaging $1/L_D$ for 100 independent random samples. Both semiclassical and quantum results, obtained for a range of scattering times τ_i, are shown in Fig. 4.10. For small τ_i, the exponential decay of the function S(x) is dominated by inelastic scattering; consequently, semiclassical and quantum results are in close agreement. As τ_i increases, however, the semiclassical decay length continues to increase, while the quantum result approaches a constant value. This limiting decay length is the localization length Λ_{loc}. We emphasize that the only difference between the two calculations is that the semiclassical analysis deals with *probabilities*, rather than probability *amplitudes*, so that all phase information is lost.

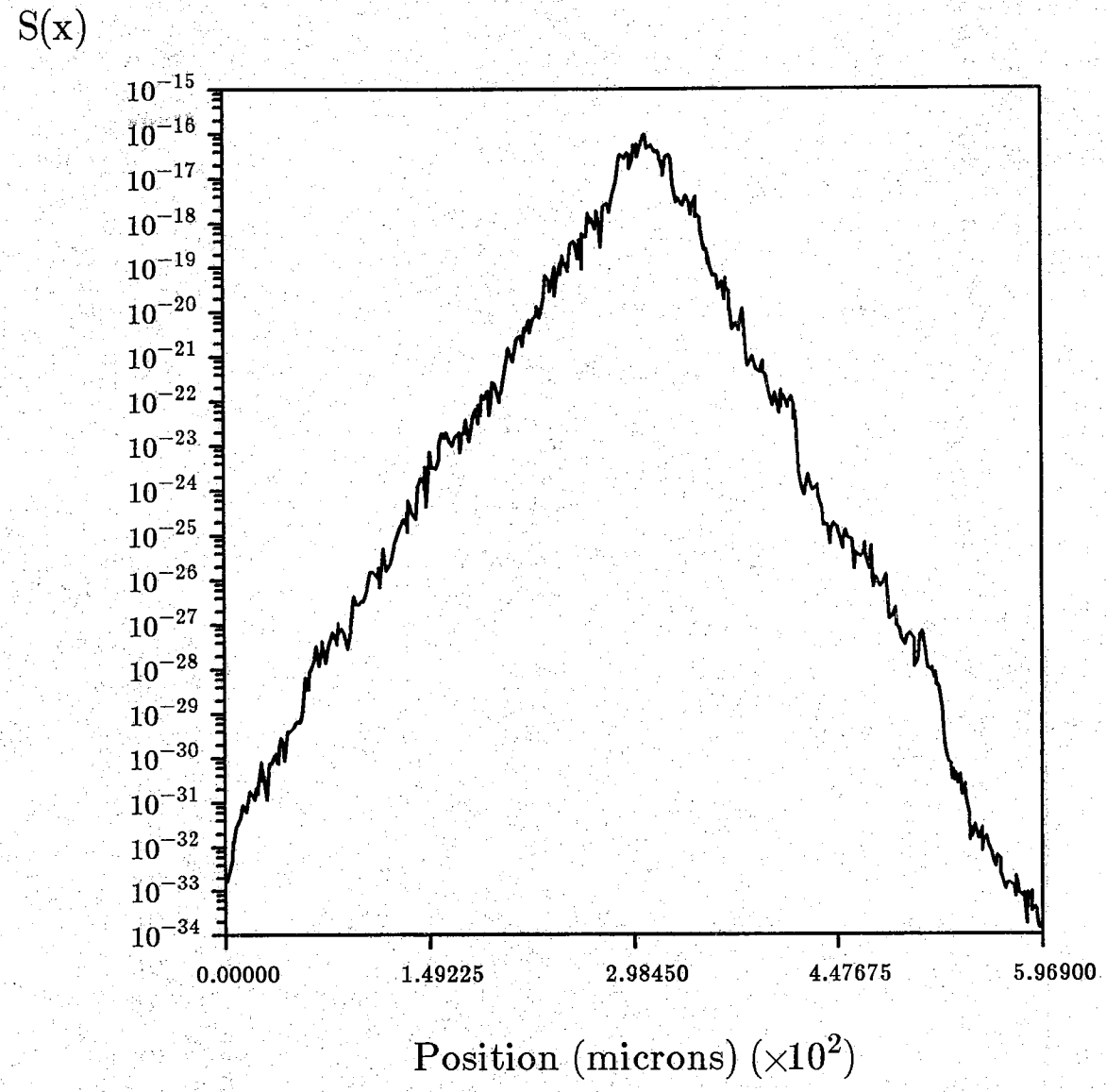


Figure 4.8 The function S(x) defined in Eq. (4.4.44), for an arbitrary sample with $\tau_i = 10^{-10}$ s. The injection point is in the middle of the sample.

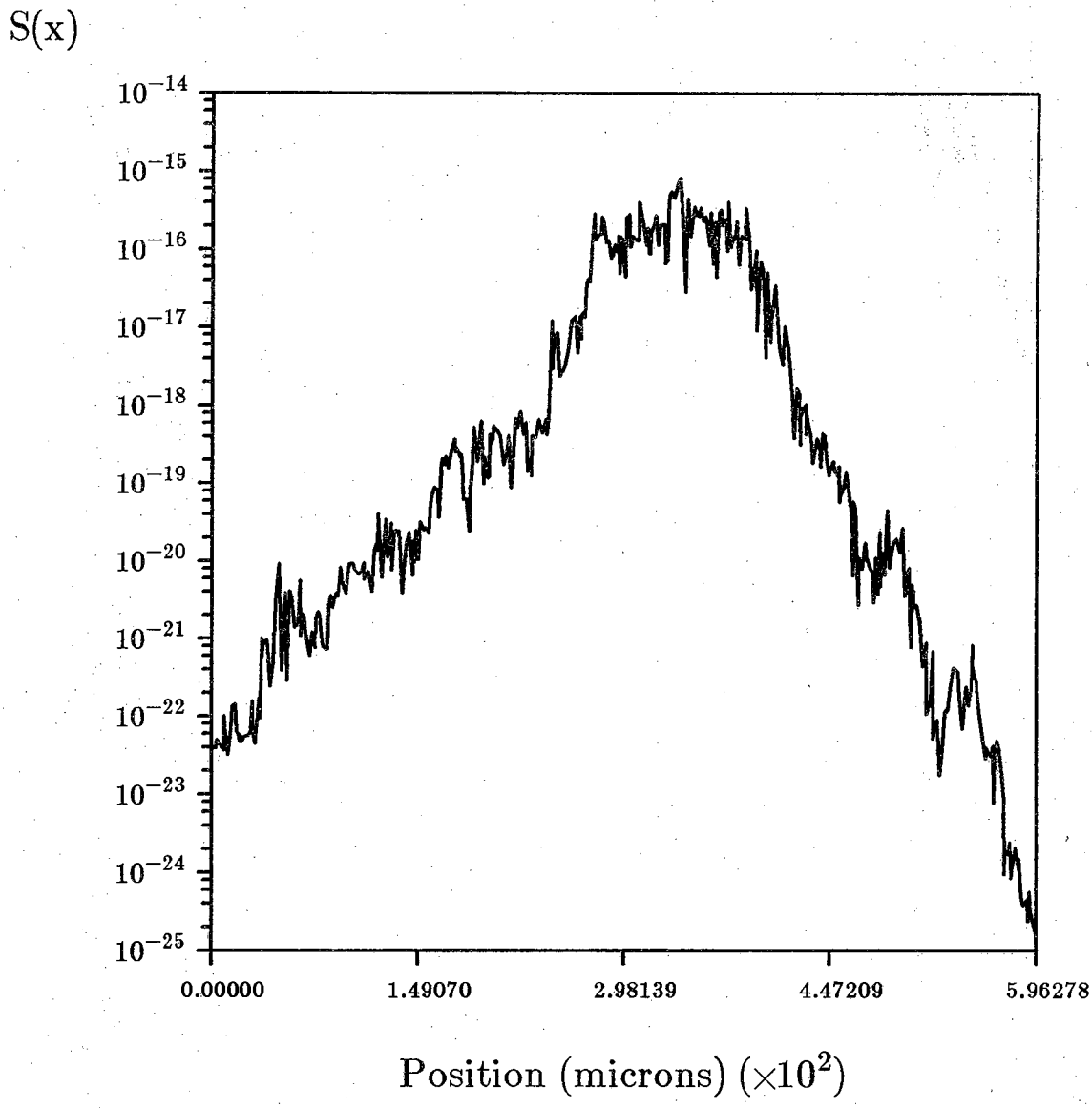


Figure 4.9 The function S(x) for an arbitrary sample with $\tau_i = 10^{-8}$ s.

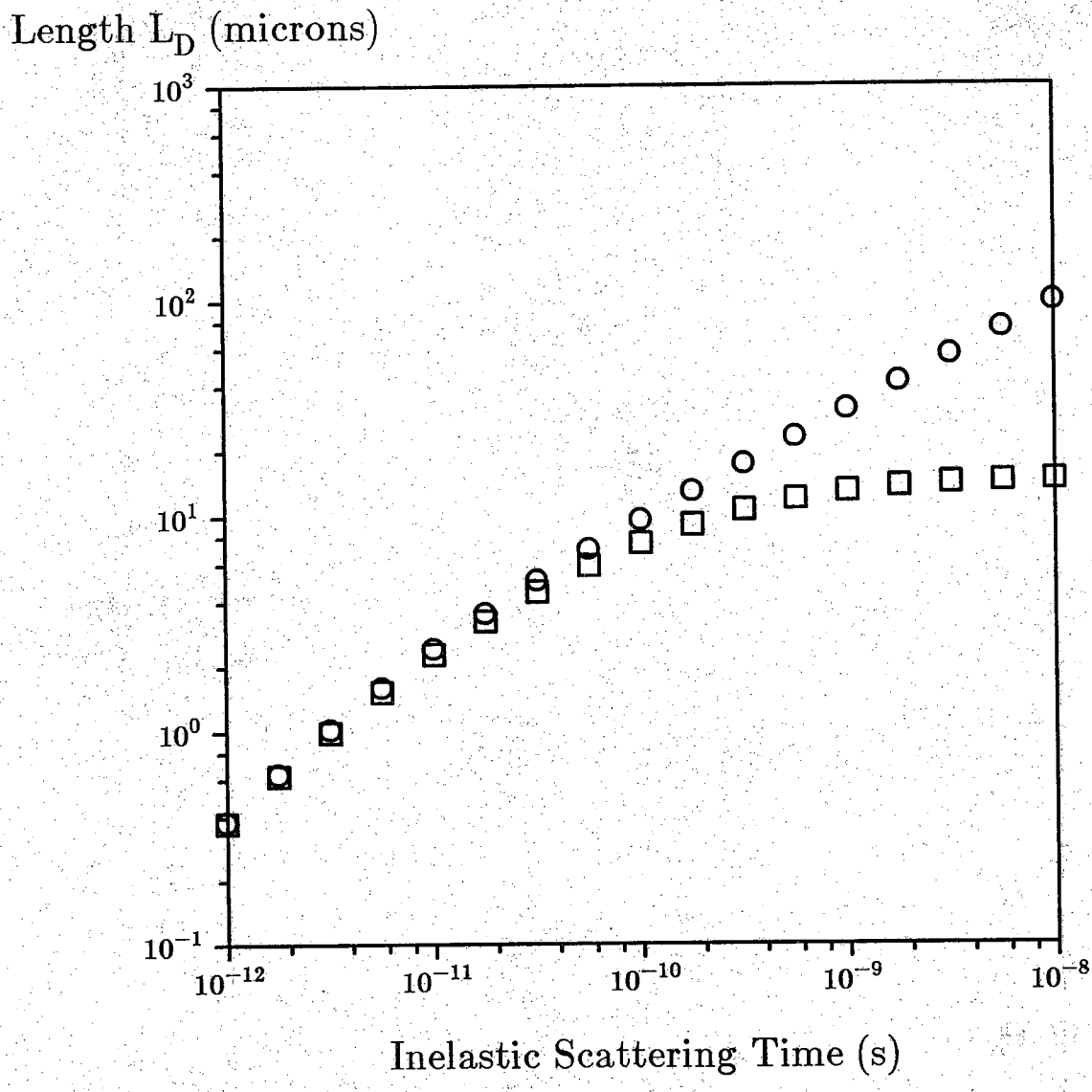


Figure 4.10 Average decay length L_D for semiclassical (○) and quantum (□) analyses. As the inelastic scattering time increases, the quantum result approaches the localization length Λ_{loc}.

Therefore, the appearance of the localization length is a direct result of the interference effects included in a quantum analysis.

Evidence of localization is even more striking in a plot of the diffusion coefficient $D \equiv L_D^2/\tau_i$, presented in Fig. 4.11. Because the conductivity is directly proportional to D (Eq. (4.4.23)), we will consider the two quantities interchangeably. For large τ_i, the quantum diffusion coefficient vanishes exponentially. The semiclassical result, however, remains constant at the value dictated by the elastic length Λ_{el}. As τ_i decreases, inelastic scattering destroys the wave function coherence which is vital to localization. Although initially this improves conduction, the point is reached where inelastic scattering dominates both semiclassical and quantum results. Beyond this point, conductivity decreases, and the two viewpoints converge to the same result.

In a quantum analysis with moderate inelastic scattering times, the electron wave function can remain coherent over large regions of a sample. Effectively, the electron "sees" a larger area of the device than what is normally presumed from classical current paths. Evidence of this behavior can be seen in our model, if we consider the statistical fluctuations in L_D. Specifically, we consider the variance $\mathrm{var}(L_D) \equiv \langle L_D^2 \rangle - \langle L_D \rangle^2$ for two cases: injection into the same sample at different points, and injection into different samples at the same point. Results of this calculation are presented in Fig. 4.12. For τ_i small, inelastic scattering dominates, and fluctuations in either case are small. For τ_i large, however, the fluctuations obtained for *different samples* are clearly larger. This reflects a greater correlation amongst values of L_D obtained within the same sample, demonstrating the non-local nature of mesoscopic structures. We note that, as a consequence of averaging $1/L_D$, fluctuations become constant in the strong localization regime. This indicates that, as shown by Anderson and co-workers [48], the distribution of $1/L_D$ is well behaved, even in strong localization.

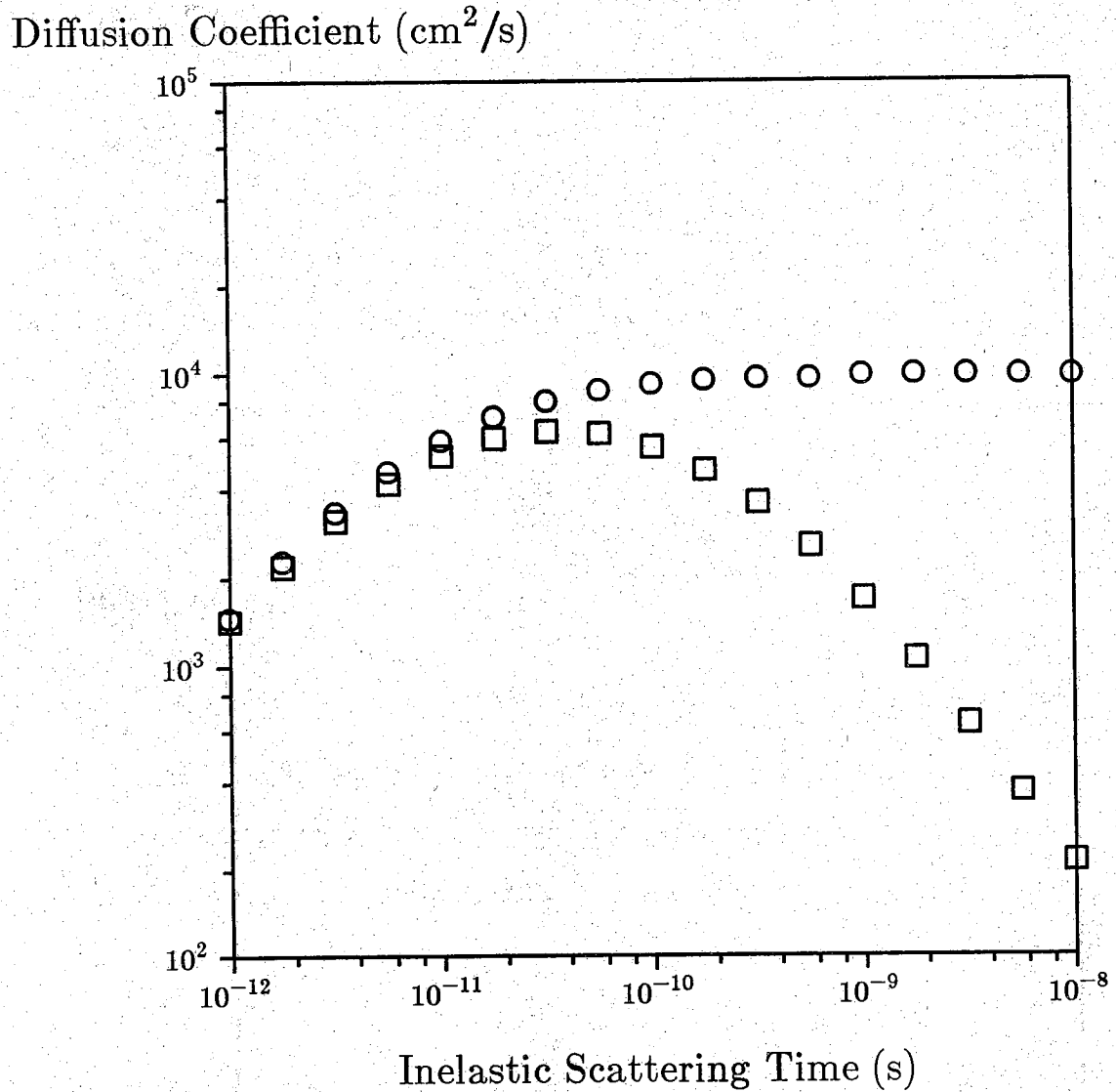


Figure 4.11 Average diffusion coefficient $D \equiv L_D^2/\tau_i$ for semiclassical (○) and quantum (□) analyses. As the inelastic scattering time increases, the quantum result vanishes exponentially. This is evidence of strong localization.

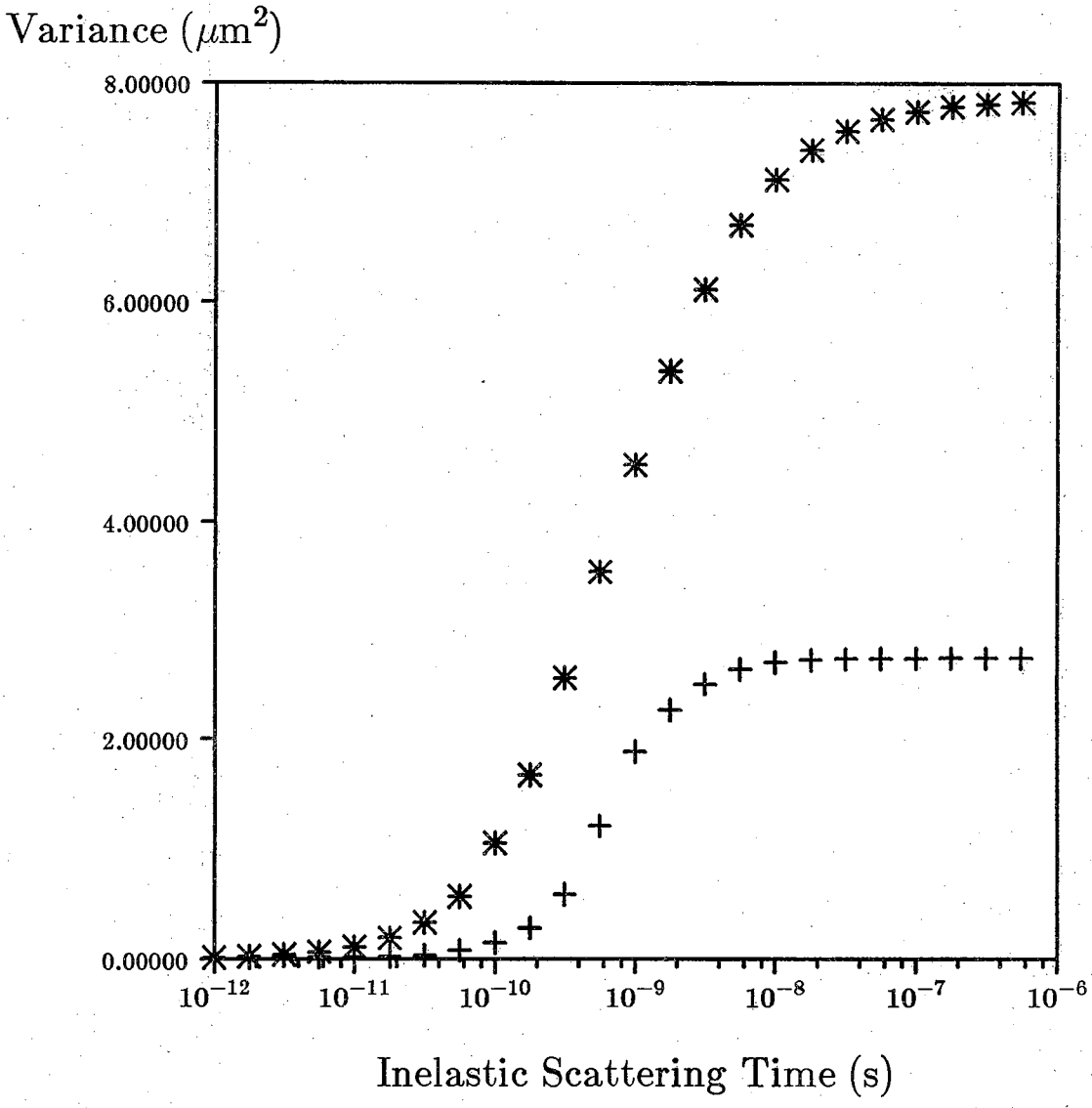


Figure 4.12 Fluctuations in the decay length L_D obtained by injecting different samples at the same point (*), and by injecting the same sample at different points (+). Fluctuations are smaller for injection within the same sample, because of the non-local character of electrons in mesoscopic devices.

4.5 Extensions of the Formalism

In the preceding sections, we made some assumptions that greatly simplified our analysis. In particular, we assumed that inelastic scattering events were spatially uncorrelated processes, that our system was at zero-temperature, and that we were eliciting a linear response. These restrictions can be removed, however, as indicated in this section.

In our extension of the multi-probe Landauer formula, we assumed a continuous distribution of probes, connecting each point in a sample to a conceptual reservoir of carriers (see Fig. 4.2). Inelastic scattering events occurred only within the reservoirs, and all reservoirs acted *independently* of one another. Because of the infinitesimal probe size and independent action of the reservoirs, all inelastic scattering events were spatially uncorrelated. This model is not entirely realistic, since phonon scattering processes are indeed correlated spatially. However, we can always find some basis in which scattering processes are uncorrelated. One possibility is a superlattice representation, for which states are indexed by a combination of position and momentum variables. In this case, the volume elements connected to each reservoir have a finite size, and states within each volume element are indexed by the modes (corresponding to periodic boundary conditions) of the probes. Scattering events between different modes are assumed to be uncorrelated, although they have a spatial correlation which corresponds to the size of the volume elements. Such a change of basis has little effect on our analysis. Although in Sections 4.2 and 4.3 we used a position representation, the Green function can be defined for any complete basis, so that the framework of our model is more flexible than we have previously indicated.

By restricting our analysis to zero-temperature, we confined our attention to electrons at a single energy, the Fermi energy E_F. At finite temperatures, however, electrons propagate within a spread of energies around the Fermi level. These energies

represent parallel channels of conduction, whose contributions must be weighted (according to occupation) and summed. In the limit of linear response, the occupation at each energy E is described the the Fermi-Dirac factor $f_i(E)$. The subscript i reminds us that for each state i, the distribution function is characterized by a quasi Fermi potential μ_i. We have intentionally used the discrete index i to avoid specifying a particular basis. In light of previous discussion, i could well index both the modes and the finite-size blocks in the superlattice representation. In any event, the current which was previously expressed as

$$I_i = \frac{e^2}{h} \sum_j \{T_{ij}\, \mu_j - T_{ji}\, \mu_i\} \ , \tag{4.5.1}$$

can be generalized to finite temperatures (and linear response) as:

$$I_i = \frac{e}{h} \sum_j \int dE' \, \{T_{ij}(E')\, f_j(E') - T_{ji}(E')\, f_i(E')\} \ , \tag{4.5.2a}$$

where

$$f_i(E) = \frac{1}{e^{(E-e\mu_i)/k_BT} + 1} \ . \tag{4.5.2b}$$

We can easily show that this extension reduces to Eq. (4.5.1) in the limit of zero-temperature. To do this, we consider the Taylor series expansion,

$$f_i = f_i \Big|_{\mu_i=0} + \mu_i \frac{df_i}{d\mu_i} \Big|_{\mu_i=0} + \frac{\mu_i^2}{2} \frac{d^2 f_i}{d\mu_i^2} \Big|_{\mu_i=0} + \cdots \ . \tag{4.5.3}$$

As temperature approaches zero, we can approximate $df/d\mu_i$ by a delta function,

$$\frac{df_i}{d\mu_i} = -e\frac{df_i}{dE} \rightarrow e\delta(E-E_F) \qquad \text{for } T \rightarrow 0 \ ,$$

so that Eq. (4.5.2a) becomes,

$$I_i = \frac{e}{h} \sum_j \Big[\int dE'\, f_0(E')\{T_{ij}(E') - T_{ji}(E')\}$$

$$+ \; e\int dE'\, \{T_{ij}(E')\mu_j - T_{ji}(E')\mu_i\}\delta(E-E_F)$$

$$+ \; e^2\int dE'\, \{T_{ij}(E')\mu_j^2 - T_{ji}(E')\mu_i^2\}\delta'(E-E_F) + \cdots \Big] \qquad (4.5.4)$$

where $\delta'(E)$ represents the derivative of a delta function, which can be evaluated from integration by parts,

$$\int dE\, F(E)\delta^{(n)}(E-E_F) \;=\; (-1)^n \left.\frac{d^nF}{dE^n}\right|_{E=E_F} . \qquad (4.5.5)$$

If $T_{ij}(E)$ varies slowly with energy, we can neglect the second-order (and all higher order) terms of μ in Eq. (4.5.4). In addition, we can neglect the zero-order term, since

$$\sum_j \{T_{ij}(E) - T_{ij}(E)\} \;=\; 0 \;, \qquad (4.5.6)$$

as shown in Section 4.2.3. This leaves only the first-order term, which reproduces Eq. (4.5.1).

Note that, aside from requiring an extra integration over energy, we have added little to the complexity of the problem. By assuming linear response, we have assumed the form of the distribution function, so that the number of unknowns (i.e., μ_i for each state i) remains the same.

For non-linear response, we cannot assume that the distribution function is characterized by a single variable; on the contrary, the value of $f_i(E)$ must be determined at all energies E, in addition to all states i. To determine $f_i(E)$, we consider the current flowing in a (conceptual) probe. We will assume that the distribution function is known for all reservoirs connected to external sources. Therefore, we are concerned only with the *unconnected* reservoirs, so that the current flowing in the associated probe is zero.

Current injected from the reservoir at a particular energy E will be proportional to the distribution function $f_i(E)$,

$$I_i^{out}(E) = \frac{e}{h} \sum_j T_{ji}(E)\, f_i(E) \quad . \tag{4.5.7}$$

All incoming current is absorbed by the reservoir, and redistributed in energy according to the detailed nature of the scattering processes. In this respect, scattering processes can be characterized by a function $\Gamma_{ij}(E,E')$, which describes the probability that an electron entering the reservoir in a state i at energy E' will be scattered to the state j and energy E. Of course, the states i and j are assumed to belong to the same reservoir (i.e., modes of the probe, in the superlattice representation). If states i and j belong to different reservoirs, the value of $\Gamma_{ij}(E,E')$ must be zero. Because carriers are conserved, the output current $I_i^{out}(E)$ depends on all of the input currents which have been redistributed in energy,

$$I_i^{out} = \frac{e}{h} \sum_j \sum_k \int dE'\, \Gamma_{ij}(E,E')\, T_{jk}(E')\, f_k(E') \quad . \tag{4.5.8}$$

Together with Eq. (4.5.7), this yields a relationship for the distribution function,

$$f_i(E) \sum_j T_{ji}(E) = \sum_j \sum_k \int dE'\, \Gamma_{ij}(E,E')\, T_{jk}(E')\, f_k(E') \quad , \tag{4.5.9}$$

which embodies the solution for non-linear response.

One other source of non-linearity involves the electrostatic potential. To this point, we have treated electrons as non-interacting *neutral* particles; in general, however, fluctuations in the electron density will induce changes in the electrostatic potential, as dictated by the Poisson equation. Changes in the potential feed back into the solution for the Green function, thereby altering the electron density. Strictly speaking, a proper analysis requires a self-consistent solution for both electron density and electrostatic potential. For linear response, however, corrections due self-consistency may

be unimportant. To understand this, we consider the current in the absence of a magnetic field,

$$I_i = \frac{e^2}{h} \sum_j T_{ij} (\mu_j - \mu_i) \ . \tag{4.5.10}$$

As the structure is driven away from equilibrium, variation in the current can be expressed as,

$$\delta I_i = \frac{e^2}{h} \sum_j \{\delta T_{ij} (\mu_j - \mu_i) + T_{ij} (\delta\mu_j - \delta\mu_i)\} \ , \tag{4.5.11}$$

where δT_{ij} represents the influence of self-consistency. If the system is sufficiently close to equilibrium (as for linear response), the difference $(\mu_j - \mu_i)$ will be small, so that self-consistent effects may be neglected. Although this is not a rigorous proof, it is an indication that self-consistency is important only for non-linear response.

Appendix 4.A1 The Kubo Formula from Linear Response Theory

In Sections 4.1.1 and 4.2.2 we presented the Kubo formula for the conductivity,

$$\sigma_{ij}(\mathbf{r},\mathbf{r}') = \frac{\hbar}{2\pi} \sum_{\alpha,\beta} \frac{[\mathbf{J}_{\alpha\beta}(\mathbf{r}) \otimes \mathbf{J}_{\beta\alpha}(\mathbf{r}')]_{ij}}{(E_F - E_\beta + i\hbar/2\tau_i)(E_F - E_\alpha - i\hbar/2\tau_i)} \ , \tag{4.A1.1}$$

which we will now derive from linear response theory. We assume that a magnetic vector potential $\mathbf{A}(\mathbf{r},t)$ of a single frequency ω is the driving force, eliciting a linear response from our system. The effect of this response is to introduce an additional term into the Hamiltonian,

$$H' = \int d\mathbf{r}\, \mathbf{J}(\mathbf{r},t)\cdot\mathbf{A}(\mathbf{r},t) \ . \tag{4.A1.2}$$

A result of linear response theory is that the response of the system measured by $\mathbf{J}(\mathbf{r},t)$ can be calculated from [22],

$$J_i(\mathbf{r},t) = \int d\mathbf{r}'\, dt'\, R_{ij}(\mathbf{r},t\ ;\ \mathbf{r}',t')\, A_j(\mathbf{r}',t') \ , \tag{4.A1.3}$$

where $R_{ij}(\mathbf{r},t\ ;\ \mathbf{r}',t')$ is the response function defined by,

$$R_{ij}(\mathbf{r},t\ ;\ \mathbf{r}',t') \equiv \frac{\theta(t-t')}{i\hbar} < [\mathbf{J}(\mathbf{r},t),\ \mathbf{J}(\mathbf{r}',t')] > \ . \tag{4.A1.4}$$

Here, $\theta(t)$ is the unit step function, and the brackets $<\cdot>$ denote an averaging of the expectation values for all many-body states $|\Psi>$. In general, we can choose a gauge for which the scalar potential is identically zero,

$$\mathbf{E}(\mathbf{r},t) = -\frac{\partial}{\partial t}\mathbf{A}(\mathbf{r},t) = -i\omega\mathbf{A}(\mathbf{r},t) \ , \tag{4.A1.5}$$

so that Eq. (4.A1.3) becomes,

$$J_i(\mathbf{r},t) = \int d\mathbf{r}'\, dt'\, \sigma_{ij}(\mathbf{r},t\ ;\ \mathbf{r}',t')\, E_j(\mathbf{r}',t') \ , \tag{4.A1.6}$$

where we have identified the conductivity tensor $\sigma(\mathbf{r},t\ ;\ \mathbf{r}',t')$ with the response function $\mathbf{R}(\mathbf{r},t\ ;\ \mathbf{r}',t')$ as follows:

$$-i\omega\sigma(\mathbf{r},t\,;\,\mathbf{r}',t') \equiv \mathbf{R}(\mathbf{r},t\,;\,\mathbf{r}',t') \ . \tag{4.A1.7}$$

Our objective, therefore, is to evaluate the response function of Eq. (4.A1.4), to obtain the Kubo conductivity. If we assume a complete set of eigenstates ψ_α (with energy E_α) for the unperturbed Hamiltonian, we can expand the current density as,

$$\mathbf{J}(\mathbf{r},t) = \sum_{\alpha\ \alpha'} \mathbf{J}_{\alpha\alpha'}(\mathbf{r})\ a_\alpha^\dagger(t)\ a_{\alpha'}(t) \ , \tag{4.A1.8a}$$

where

$$\mathbf{J}_{\alpha\alpha'} \equiv \frac{ie\hbar}{2m^*}\left[(\nabla\psi_\alpha^*)\psi_{\alpha'} - \psi_\alpha^*(\nabla\psi_{\alpha'})\right] \ , \tag{4.A1.8b}$$

and $a_\alpha^\dagger(t)$ and $a_\alpha(t)$ are the creation and annihilation operators acting on state α. Substituting Eq. (4.A1.8a) into Eq. (4.A1.4), we obtain

$$\mathbf{R}(\mathbf{r},\mathbf{r}'\,;\,t) = \sum_{\alpha\ \alpha'\ \beta\ \beta'} \mathbf{J}_{\alpha\alpha'}(\mathbf{r}) \otimes \mathbf{J}_{\beta\beta'}(\mathbf{r}')\ \frac{\theta(t)}{i\hbar}$$

$$< [a_\alpha^\dagger(t)\ a_{\alpha'}(t),\ a_\beta^\dagger(0)\ a_{\beta'}(0)] > \ , \tag{4.A1.9}$$

where we used $\otimes$ to denote a direct product of vectors, yielding a tensor. For convenience, we have taken $t' = 0$. The commutator above can be simplified [51],

$$< [a_\alpha^\dagger(t)\ a_{\alpha'}(t),\ a_\beta^\dagger(0)\ a_{\beta'}(0)] > \ = \ <a_\alpha^\dagger(t)\ a_{\beta'}(0)><a_{\alpha'}(t)\ a_\beta^\dagger(0)> \ -$$

$$<a_\beta^\dagger(0)\ a_{\alpha'}(t)><a_{\beta'}(0)\ a_\alpha^\dagger(t)> \ . \tag{4.A1.10}$$

We will consider only one of the ensemble-averaged terms above, although the same arguments can be applied to all four terms. As we pointed out earlier, the brackets $<\cdot>$ denote averaging of the expectation values for all possible states $|\Psi>$. We write this explicitly as,

$$<a_\alpha^\dagger(t)\, a_{\beta'}(0)> \equiv \Big\langle <\Psi \,|\, a_\alpha^\dagger(t)\, a_{\beta'}(0) \,|\, \Psi> \Big\rangle_\Psi \ , \tag{4.A1.11}$$

and the time evolution of the creation and annihilation operators is understood to be,

$$a_\alpha(t) \ = \ a_\alpha(0)\, e^{-iE_\alpha t/\hbar}\, e^{-|t|/2\tau_i} \ \equiv \ a_\alpha(0)\, G_\alpha(t) \ , \tag{4.A1.12a}$$

$$a_\alpha^\dagger(t) \ = \ a_\alpha^\dagger(0)\, G_\alpha^*(t) \ = \ a_\alpha^\dagger(0)\, G_\alpha(-t) \ , \tag{4.A1.12b}$$

so that Eq. (4.A1.11) becomes,

$$<a_\alpha^\dagger(t)\, a_{\beta'}(0)> \ = \ G_\alpha^*(t) \Big\langle <\Psi \,|\, a_\alpha^\dagger(0)\, a_{\beta'}(0) \,|\, \Psi> \Big\rangle_\Psi \ . \tag{4.A1.13}$$

We recognize that the expectation value is zero, for $\alpha \neq \beta'$. This is because the state vectors resulting from annihilation in two different states are orthogonal:

$$<\Psi \,|\, a_\alpha^\dagger\, a_{\beta'} \,|\, \Psi> \ = \ <a_\alpha\, \Psi \,|\, a_{\beta'}\, \Psi> \ = \ 0 \quad \text{for} \quad \alpha \neq \beta' \ .$$

We also recognize that, for operators acting on the same state $\alpha = \beta'$, the operator $a_\alpha^\dagger\, a_\alpha$ is the number operator, yielding the number of electrons (0 or 1) in state α. When we average over all possible state vectors $|\Psi>$, this yields the probability of finding an electron in state α, which is the Fermi-Dirac factor f_α:

$$<a_\alpha^\dagger(t)\, a_{\beta'}(0)> \ = \ G_\alpha^*(t) \Big\langle <\Psi \,|\, a_\alpha^\dagger\, a_\alpha \,|\, \Psi> \Big\rangle_\Psi \delta_{\alpha\beta'} \ = \ G_\alpha^*(t)\, f_\alpha\, \delta_{\alpha\beta'} \tag{4.A1.14}$$

In a similar manner, we can simplify the remaining terms in Eq. (4.A1.10). We note that, as a property of the creation and annihilation operators, $a_\alpha\, a_\alpha^\dagger \ = \ 1 - a_\alpha^\dagger\, a_\alpha$, giving rise to factors $(1 - f_\alpha)$ after ensemble-averaging. Eq. (4.A1.10) simplifies to

$$< [a_\alpha^\dagger(t)\, a_{\alpha'}(t),\, a_\beta^\dagger(0)\, a_{\beta'}(0)] >$$

$$= \ \delta_{\alpha\beta'}\, \delta_{\beta\alpha'}\ G_\beta(t)\, G_\alpha(-t)\, [f_\alpha(1 - f_\beta) - f_\beta(1 - f_\alpha)] \ . \tag{4.A1.15}$$

Substituting this result into Eq. (4.A1.9), and performing the sums over α' and β', we obtain the response function,

$$\mathbf{R}(\mathbf{r},\mathbf{r}'\,;\,t) \;=\; \sum_{\alpha\beta} \frac{\theta(t)}{i\hbar}\, G_\beta(t)\, G_\alpha(-t)\, (f_\alpha - f_\beta)\, [\mathbf{J}_{\alpha\beta}(\mathbf{r}) \otimes \mathbf{J}_{\beta\alpha}(\mathbf{r}')] \quad , \qquad (4.A1.16)$$

where we have neglected terms $(1 - f_\alpha)$, which are unimportant for the electronic states that contribute to conduction. The response function is related to the conductivity through the frequency of our source (Eq. (4.A1.7)). Because of this, it is convenient to work in the frequency domain. The Green functions $G_\beta(t)$ and $G_\alpha(-t)$ have Fourier transform counterparts,

$$G_\beta(t) \;\leftrightarrow\; G_\beta^R(\omega) \;\equiv\; \frac{i\hbar}{\hbar\omega - E_\beta + i\eta} \quad , \qquad (4.A1.17a)$$

$$G_\alpha(-t) \;\leftrightarrow\; G_\alpha^A(\omega) \;\equiv\; \frac{-i\hbar}{\hbar\omega - E_\alpha - i\eta} \quad , \qquad (4.A1.17b)$$

Typically, the energy $i\eta$ is a true infinitesimal [15], introduced only to define the Green functions for $\omega = E_\alpha/\hbar$. In our analysis, however, we can identify the imaginary energy $i\eta$ with the loss term in the Schrödinger equation, $i\hbar/2\tau_i$. This identification has been made previously [21], with the observation that such a term simulates the effect of inelastic scattering, if scattering events are spatially uncorrelated; this is precisely the physical picture presented in Sections 4.2 and 4.3. In the frequency domain, the product $G_\beta(t)\, G_\alpha(-t)$ can be represented by a sort of convolution integral,

$$\mathbf{R}(\mathbf{r},\mathbf{r}';\omega) \;=\; \sum_{\alpha\beta} \frac{1}{i\hbar}\, [\mathbf{J}_{\alpha\beta}(\mathbf{r}) \otimes \mathbf{J}_{\beta\alpha}(\mathbf{r}')] \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi}\, G_\beta^R(\epsilon+\omega)\, G_\alpha^A(\epsilon)\, [f_\alpha - f_\beta] \quad . \qquad (4.A1.18)$$

It is apparent from Eqs. (4.A1.17) that, for τ_i large ($\eta \rightarrow 0$), the functions $G_\alpha^A(\omega)$ and $G_\alpha^R(\omega)$ are quite peaked near frequencies $\omega = E_\alpha/\hbar$. Because these frequencies will dominate the integral over ϵ, we can represent the Fermi-Dirac factors over the entire range of frequencies without significantly affecting the result:

$$\mathbf{R}(\mathbf{r},\mathbf{r}';\omega) = \sum_{\alpha\beta} \frac{1}{i\hbar} [\mathbf{J}_{\alpha\beta}(\mathbf{r}) \otimes \mathbf{J}_{\beta\alpha}(\mathbf{r}')]$$

$$\int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi} [f(\epsilon) - f(\epsilon+\omega)]\, G^{R}_{\beta}(\epsilon+\omega)\, G^{A}_{\alpha}(\epsilon) \ . \qquad (4.A1.19)$$

Finally, we can relate the response function to the conductivity, by using Eq. (4.A1.7),

$$\sigma(\mathbf{r},\mathbf{r}';\omega) = \frac{1}{2\pi\hbar} \sum_{\alpha\beta} [\mathbf{J}_{\alpha\beta}(\mathbf{r}) \otimes \mathbf{J}_{\beta\alpha}(\mathbf{r}')]$$

$$\int_{-\infty}^{\infty} d\epsilon \left(\frac{f(\epsilon) - f(\epsilon+\omega)}{\omega} \right) G^{R}_{\beta}(\epsilon+\omega)\, G^{A}_{\alpha}(\epsilon) \ . \qquad (4.A1.20)$$

In the limit of dc conductivity ($\omega \longrightarrow 0$),

$$\lim_{\omega\to 0} \frac{f(\epsilon+\omega) - f(\epsilon)}{\omega} = \frac{df}{d\omega} = \hbar \frac{df}{dE} \ ,$$

and in the limit of zero temperature, $-df/dE \longrightarrow \delta(E-E_F)$, so that the integral in Eq. (4.A1.20) can be performed, to yield our final result,

$$\sigma(\mathbf{r},\mathbf{r}';\omega) = \frac{\hbar}{2\pi} \sum_{\alpha\beta} \frac{[\mathbf{J}_{\alpha\beta}(\mathbf{r}) \otimes \mathbf{J}_{\beta\alpha}(\mathbf{r}')]}{(E_F - E_\beta + i\hbar/2\tau_i)(E_F - E_\alpha - i\hbar/2\tau_i)} \ , \qquad (4.A1.21)$$

as was to be shown.

Appendix 4.A2 Normalization of the Green Function

In Section 4.3.1, we stated the following normalization,

$$\int d\mathbf{r} \ |\, G(\mathbf{r},\mathbf{r}')\,|^2 = \frac{2\pi\tau_i}{\hbar} N_0(\mathbf{r}') \ , \qquad (4.A2.1)$$

which we intend to prove in this appendix. We can expand the Green function in a complete set of basis functions ψ_α, which are the eigenfunctions of the Hamiltonian [43],

$$G(\mathbf{r},\mathbf{r}') = \sum_{\alpha} \frac{\psi_\alpha(\mathbf{r})\psi_\alpha^*(\mathbf{r}')}{E_F - E_\alpha + i\hbar/2\tau_i} \quad , \tag{4.A2.2}$$

so that the normalization integral becomes,

$$\int d\mathbf{r} \; |\, G(\mathbf{r},\mathbf{r}') \,|^2 = \int d\mathbf{r} \sum_{\alpha\beta} \frac{\psi_\alpha(\mathbf{r})\psi_\alpha^*(\mathbf{r}')\psi_\beta^*(\mathbf{r})\psi_\beta(\mathbf{r}')}{(E_F - E_\alpha + i\hbar/2\tau_i)(E_F - E_\beta - i\hbar/2\tau_i)} \quad . \tag{4.A2.3}$$

Because the functions ψ_α form a complete set,

$$\int d\mathbf{r} \; \psi_\beta^*(\mathbf{r})\psi_\alpha(\mathbf{r}) = \delta_{\alpha\beta} \quad , \tag{4.A2.4}$$

which is simply a statement of completeness. This is precisely the integral required by Eq. (4.A2.3), so that after summing over β, our result simplifies to

$$\int d\mathbf{r} \; |\, G(\mathbf{r},\mathbf{r}') \,|^2 = \frac{2\tau_i}{\hbar} \sum_{\alpha} |\, \psi_\alpha(\mathbf{r}') \,|^2 \frac{\hbar/2\tau_i}{(E_F - E_\alpha)^2 + (\hbar/2\tau_i)^2} \quad . \tag{4.A2.5}$$

If we define $x \equiv \hbar/2\tau_i$ and $\epsilon \equiv E_F - E_\alpha$, we can better appreciate the form of this function,

$$\int d\mathbf{r} \; |\, G(\mathbf{r},\mathbf{r}') \,|^2 = \frac{2\tau_i}{\hbar} \sum_{\alpha} |\, \psi_\alpha(\mathbf{r}') \,|^2 \frac{x}{\epsilon^2 + x^2} \quad . \tag{4.A2.6}$$

We recognize that one definition of the Dirac delta function is

$$\lim_{x \to 0} \frac{x}{\epsilon^2 + x^2} = \pi \, \delta(\epsilon) \quad , \tag{4.A2.7}$$

so that if the inelastic scattering time τ_i is large,

$$\int d\mathbf{r} \; |\, G(\mathbf{r},\mathbf{r}') \,|^2 = \frac{2\tau_i}{\hbar} \sum_{\alpha} |\, \psi_\alpha(\mathbf{r}') \,|^2 \delta(E_F - E_\alpha) \quad . \tag{4.A2.8}$$

In the limit of zero-temperature, the density of states is defined as

$$N_0(\mathbf{r}) \equiv \sum_{\alpha} |\psi_\alpha(\mathbf{r})|^2 \delta(E_F - E_\alpha) \ , \tag{4.A2.9}$$

and we obtain our normalization condition Eq. (4.A2.1), as was to be shown.

Appendix 4.A3 Classical Trajectory in Electric and Magnetic Fields

In Section 4.4.1, we required the classical trajectory of an electron in the presence of uniform, constant electric and magnetic fields. This appendix contains a derivation of the trajectory, starting with an expression for the velocity from classical mechanics,

$$\mathbf{F} = m^* \frac{d\mathbf{v}}{dt} = e(\mathbf{E} + \mathbf{v}\times\mathbf{B}) \ , \tag{4.A3.1}$$

If we assume a magnetic field oriented along the z-axis, and an electric field in the x-y plane, this simplifies to two coupled differential equations,

$$m^* \frac{d(v)_x}{dt} = eE_x + eB(v)_y \ , \tag{4.A3.2a}$$

$$m^* \frac{d(v)_y}{dt} = eE_y - eB(v)_x \ , \tag{4.A3.2b}$$

from which we eliminate $(v)_y$, to obtain

$$\frac{d^2(v)_x}{dt^2} + \omega_c^2 (v)_x = \omega_c^2 \frac{E_y}{B} \ . \tag{4.A3.3}$$

We have introduced the cyclotron frequency for an electron $\omega_c \equiv |e|B/m^*$ to simplify the constants. Solutions to Eq. (4.A3.3) are of the form,

$$(v)_x = \alpha \sin\omega_c t + \beta \cos\omega_c t + \frac{E_y}{B} \ . \tag{4.A3.4a}$$

Substituting the solution for $(v)_x$ into Eq. (4.A3.2a), we obtain a corresponding solution for $(v)_y$,

$$(v)_y = -\alpha \cos\omega_c t + \beta \sin\omega_c t - \frac{E_x}{B} \quad . \tag{4.A3.4b}$$

We assume that the components v_x and v_y of the initial velocity are known, so that the constants α and β are determined from Eqs. (4.A3.4) at t=0,

$$v_x = \beta + \frac{E_y}{B} \quad , \tag{4.A3.5a}$$

$$v_y = -\alpha - \frac{E_x}{B} \quad . \tag{4.A3.5b}$$

Solving for α and β and substituting the results into Eq. (4.A3.4a), we obtain the x-component of velocity

$$(v)_x = \left(\frac{a_x}{\omega_c} - v_y\right)\sin\omega_c t + \left(\frac{a_y}{\omega_c} + v_x\right)\cos\omega_c t - \frac{a_y}{\omega_c} \quad , \tag{4.A3.6}$$

where we have introduced a_i as the acceleration due to the electric field, $a_i \equiv eE_i/m^*$. The x-component of the trajectory can be determined from Eq. (4.A3.6) by integrating over time,

$$x_0(t) = \int_0^t (v)_x \, dt'$$

$$= \left(\frac{a_x}{\omega_c^2} - \frac{v_y}{\omega_c}\right)(1 - \cos\omega_c t) + \left(\frac{a_y}{\omega_c^2} + \frac{v_x}{\omega_c}\right)\sin\omega_c t - \frac{a_y}{\omega_c} t \quad , \tag{4.A3.7}$$

which is our final result.

Chapter 4 References

[1] Nobel lectures by L. Esaki, I. Giaever and B. Josephson, *Rev. Mod. Phys.*, **46**, 237 (1974).

[2] Esaki, L., *IEEE J. Quantum Electron.*, **QE-22**(9), 1611 (1986).

[3] Lee, P. A. and Ramakrishnan, T. V., *Rev. Mod. Phys.*, **57**(2), 287 (1985).

[4] Fowler, A. B., Wainer, J. J. and Webb, R. A., *IBM J. Res. Develop.*, **32**(3), 372 (1988).

[5] Azbel, M. Ya., *IBM J. Res. Develop.*, **32**(1), 52 (1988).

[6] Washburn, S. and Webb, R. A., *Adv. Phys.*, **35**(4), 375 (1986).

[7] Lee, P. A. and Stone, A. D., *Phys. Rev. Lett.*, **55**(15), 1622 (1985); Lee, P. A., Stone, A. D., and Fukuyama, H., *Phys. Rev. B* **35**(3), 1039 (1987).

[8] Washburn, S., *IBM J. Res. Develop.*, **32**(3), 335 (1988).

[9] Kaplan, S. B. and Hartstein, A., *IBM J. Res. Develop.* **32**(3), 347 (1988).

[10] Umbach, C. P., Santhanam, P., van Haesendonck, C. and Webb, R. A., *Appl. Phys. Lett.*, **50**(18), 1289 (1987).

[11] Baranger, H. U., Stone, A. D. and DiVincenzo, D. P., *Phys. Rev. B* **37**(11), 6521 (1988).

[12] Skocpol, W. J., Mankiewich, P. M., Howard, R. E., Jackel, L. D. and Tennant, D. M., *Phys. Rev. Lett.*, **58**(22), 2347 (1987).

[13] Timp, G., Chang, A. M., Mankiewich, P., Beheringer, R., Cunningham, J. E., Chang, T. Y. and Howard, R. E., *Phys. Rev. Lett.*, **59**(6), 732 (1987).

[14] Datta, S. and Lundstrom, M. S., to appear in *Superlattices and Microstructures*, Academic Press).

[15] Stone, A. D. and Szafer, A., *IBM J. Res. Develop.*, **32**(3), 384 (1988).

[16] Buttiker, M., *Phys. Rev. B* **33**(5), 3020 (1986).

[17] Pierret, R. F., *Modular Series on Solid State Devices: Semiconductor Fundamentals,* Vol. I, Chapter 3, Addison-Wesley, Reading, Massachusetts, 1983, First Edition.

[18] Maekawa, S., Isana, Y. and Ebisawa, H., *J. Phys. Soc. Jpn.*, **56**(1), 25 (1987).

[19] Kane, C. L., Serota, R. A. and Lee, P. A., *Phys. Rev. B* **37**, 6701 (1987).

[20] Kane, C. L., Lee, P. A. and DiVincenzo, D. P., *Phys. Rev. B* **38**, 2995 (1988).

[21] Thouless, D. J. and Kirkpatrick, S., *J. Phys. C: Solid State Phys.*, **14**, 235 (1981).

[22] Kubo, R., *Rep. Prog. Phys.*, **29**, 255 (1968).

[23] Kubo, R., *Science,* **233**, 330 (1986).

[24] Economou, E. N. and Soukoulis, C. M., *Phys. Rev. Lett.*, **46**(9), 618 (1981).

[25] Langreth, D. C. and Abrahams, E., *Phys. Rev. B* **24**(6), 2978 (1981).

[26] Fisher, D. S. and Lee, P. A., *Phys. Rev. B* **23**(12), 6851 (1981)

[27] Xie, X. C. and DasSarma, S., *Phys. Rev. B* **38**(5), 3529 (1988).

[28] DiVincenzo, D. P. and Kane, C. L., *Phys. Rev. B* **38**(5), 3006 (1988).

[29] Anderson, P. W., *Phys. Rev.*, **109**, 1492 (1958).

[30] Landauer, R., *IBM J. Res. Develop.*, **1**, 223 (1957).

[31] Landauer, R., *Philos. Mag.*, **21**, 863 (1970).

[32] Landauer, R., *IBM J. Res. Develop.*, **32**(3), 306 (1988).

[33] Buttiker, M., *Phys. Rev. Lett.*, **57**(14), 1761 (1986).

[34] Thouless, D. J., *Phys. Rev. Lett.*, **47**(13), 972 (1981); Economou, E. N. and Soukoulis, C. M., *Phys. Rev. Lett.*, **47**(13), 973 (1981).

[35] Engquist, H.-L. and Anderson, P. W., *Phys. Rev. B* **24**(2), 1151 (1981).

[36] Landauer, R., *Z. Phys. B,* **68**, 217 (1987).

[37] Imry, Y., "Physics of Mesoscopic Systems," *Directions in Condensed Matter Physics,* G. Grinstein and G. Mazenko, Eds., World Scientific Press, Singapore, 1986, p. 101.

[38] Imry, Y., *Europhys. Lett.,* **1**(5), 249 (1986).

[39] Muttalib, K. A., Pichard, J.-L. and Stone, A. D., *Phys. Rev. Lett.,* **59**(21), 2475 (1987).

[40] Giordano, N., *Phys. Rev. B* **36**(8), 4190 (1987).

[41] Pendry, J., *IBM J. Res. Develop.,* **32**(1), 137 (1988).

[42] Buttiker, M., *IBM J. Res. Develop.,* **32**(1), 63 (1988).

[43] Schiff, L. I., *Quantum Mechanics, Chapter 9, Section 37, McGraw-Hill, New York, 1968, Third Edition.*

[44] Landau, L. D. and Lifshitz, E. M., *Quantum Mechanics (Non-relativistic Theory),* p. 55, Addison-Wesley, Reading, Massachusetts, 1965, Second Edition.

[45] Einstein, A., *Investigations on the Theory of the Brownian Movement,* Dover Publications, 1956.

[46] Kittel, C., *Introduction to Solid State Physics,* p. 155, John Wiley and Sons, Inc., New York, 1986, Sixth Edition.

[47] Anderson, P. W., *Phys. Rev.,* **109**(5), 1492 (1958).

[48] Anderson, P. W., Thouless, D. J., Abrahams, E. and Fisher, D. S., *Phys. Rev. B* **22**(8), 3519 (1980).

[49] Cahay, M., McLennan, M. and Datta, S., *Phys. Rev. B* **37**(17), 10125 (1988).

[50] Thouless, D. J., *Phys. Rev. Lett.*, **39**, 1167 (1977).

[51] Doniach, S. and Sondheimer, E. H., *Green's Functions for Solid State Physicists*, p. 114, Benjamin, Reading, Massachusetts, 1974.

Appendix A
QUANTUM DEVICES (A REVIEW)

The following is a text of the keynote lecture delivered at the session on Transport and Spectroscopy in 1-Dimensional Systems at the 4th International Conference on Superlattices, Microstructures and Microdevices, Trieste, Italy, August 8-12, 1988 (to appear in *Superlattices and Microstructures*, Academic Press).

1. What is a quantum device?

Device analysis has traditionally been based on semiclassical transport theory (Boltzmann equation, drift-diffusion equation etc.) which views electrons as particles that obey Newton's law in an external electric field and are scattered occasionally by phonons and impurities. Simply stated, a quantum device is one that can only be understood taking the wave nature of electrons into account. A variety of quantum effects in electronic transport have been observed and studied over the years such as tunneling, resonant tunneling, weak and strong localization and the quantum Hall effect. In this talk, however, we will concentrate on three interrelated quantum effects that have been observed since 1985: the Aharonov-Bohm effect, conductance fluctuations and non-local effects. We will review these exciting developments emphasizing their implications for electronic devices of the future.

2. Why aren't quantum effects more common?

It is well-known that electrons possess wave-like properties and show interference much like electromagnetic waves. Why is it that in dealing with microwave networks or integrated optics we routinely worry about interference and yet we ignore it completely in the analysis of conventional electronic devices? There are three distinct reasons. Firstly, the wave nature manifests itself only when the device size is comparable to a wavelength which is ~1cm. for microwaves, ~1μm. for light and ~100Å for electrons in semiconductors. As we know, geometrical optics is usually adequate for describing ordinary optical systems using lenses and prisms while wave optics is necessary only when the medium varies on the scale of microns as in a diffraction grating or in integrated optics. Similarly, Newton's laws provide an adequate description of electronic motion in vacuum tubes or large semiconductor devices. But a wave description based on the Schrödinger equation can become necessary in describing electron transport in submicron structures. Secondly, inelastic processes due to phonons and other electrons within the device destroy phase coherence and interference phenomena. (Strictly speaking, phase-breaking is caused by any scattering process in which the scatterer changes its state. This includes scattering by a magnetic impurity if it changes its spin state, though the process may be elastic). Assuming an inelastic scattering time ~40ps. at T~1K and an average electron velocity ~2.5×10^7 cm/sec. we obtain an inelastic mean free path ~10μm. (This number will be significantly smaller at higher temperatures and for hot electrons. We will concentrate on 'cold' electrons near the Fermi energy.) Advances in microfabrication have now made it possible to build devices much smaller than a micron so that there is hardly any scattering within the device. In fact, the usual description of 'quantum devices' is based on the assumption that all inelastic

scattering takes place in the contacts and not in the device. Under these conditions, we can expect electronic devices to behave more like microwave networks than like ordinary circuit elements.

Finally, it should be noted that while in electromagnetics monochromatic waves and single-moded waveguides are quite common, electron waves in solids commonly have a large spread in their energy (analogous to the frequency of classical waves) and electron waveguides are commonly multimoded (the 'modes' of an electron waveguide are commonly referred to as subbands.) This gives rise to a large spread in the wavelengths of the electrons which tends to wash out interference effects. It is for this reason that electron microscopists, who use the wave properties of electrons to determine crystal structures, try to obtain a monoenergetic and well-collimated beam of electrons. At low temperatures and low voltages only electrons near the Fermi level contribute to the conductance so that the energy spread is small. The number of modes M is approximately equal to $n_s^{1/2}W$ where n_s is the areal density of electrons and W is the width of the device. A simple estimate shows that with $n_s = 3.6\times10^{11}\text{cm}^{-2}$, a single-moded quantum wire needs to be $\lesssim$200Å. This may seem somewhat discouraging, but recent experiments have shown that semiconductor wires have a fairly wide depletion layer (3000-4000 Å) surrounding the actual conducting channel, much like the cladding layer in an optical fiber. This means that with adequate process control it may be possible to fabricate single-moded quantum wires with a physical width $\sim$·4μm. which is well within the reach of present-day technology [1-4]. Recently by using a gate to reduce gradually the width of a point contact, the number of transverse modes (and hence the conductance) was observed to decrease in discrete steps to one [4]. Such single-moded quantum wires raise the possibility of duplicating with electron waveguides many of the device concepts that are well-known in integrated optics.

There is another point that we would like to emphasize. We usually regard multimodedness as an undesirable element that severely reduces interference effects. It is generally believed that if there are M different modes, then the magnitude of interference effects is $\sim 1/M$. However, this is only true if the different modes have completely random phases. Experiments on metallic wires and rings indicate that in many physical configurations the different modes are not totally uncorrelated, leading to quantum interference effects far in excess of the 1/M estimate. For example, conductance oscillations $\sim .1\%$ due to the Aharonov-Bohm (A-B) effect were observed in metallic rings with $M \sim 10^5 - 10^6$ [5]. Conductance fluctuations $\sim e^2/h$ have been observed in metallic and semiconductor samples with widely varying values of M [5-8]. This shows that multiple modes are often correlated despite the apparent randomness. More interestingly, with the present advances in microtechnology, it may be feasible to engineer structures where multiple modes are correlated in such a way as to produce large interference effects. For example, the device described in Section 5 is ideally expected to show nearly 100% interference effects even as $M \rightarrow \infty$. Of course, much experimental work remains to be done before we know the degree of correlation amongst the multiple modes that can actually be achieved with real semiconductor structures. This is an important consideration since a major concern about single-moded quantum devices is their low current capability.

3. Theoretical Background

There are two types of device concepts that we will discuss. The first type is based on the Aharonov-Bohm effect where the current in a *two-port quantum device* (Fig. 1) is controlled by changing the potential (vector or scalar) distribution within it. The second type is based on non-local effects where the current in a *multi-port quantum device* (Fig. 2) is controlled by changing the boundary conditions at one port.

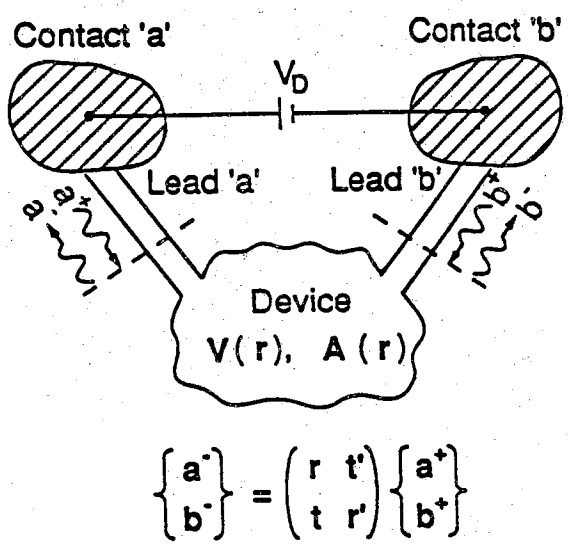
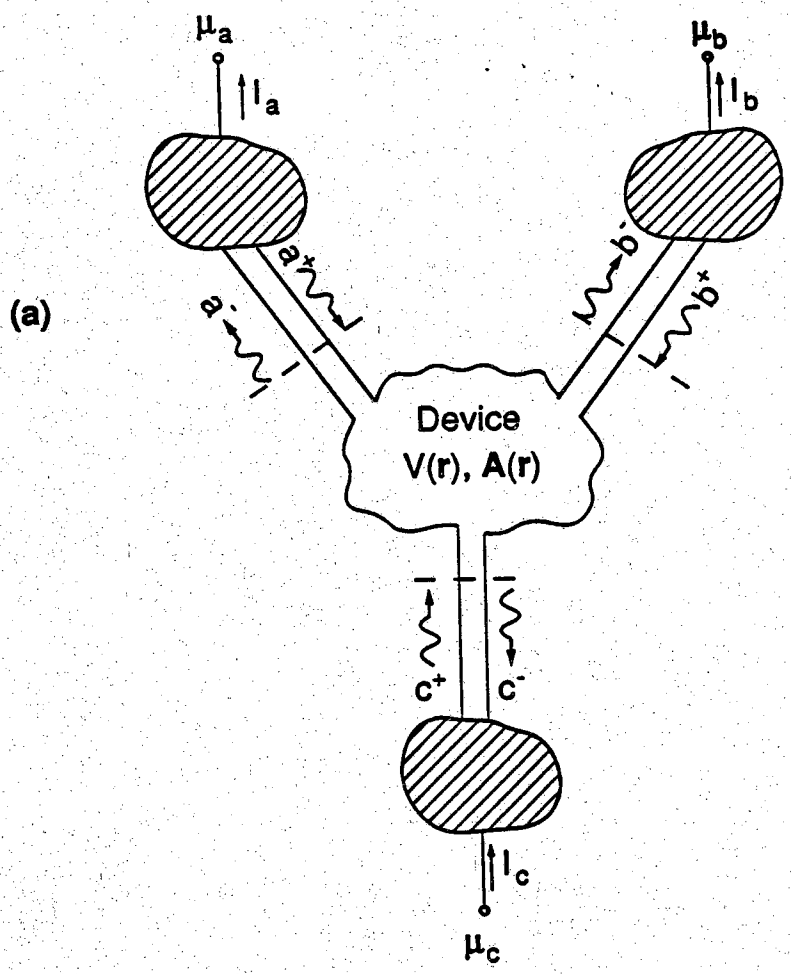


Fig. 1: Schematic diagram of an arbitrary device connected to the contacts through two leads; all inelastic scattering and dissipation is assumed to occur only in the contacts and not within the device. In general, the problem is to calculate the scattering matrix $S = \begin{pmatrix} r & t' \\ t & r' \end{pmatrix}$ for the device, given the potentials V(r), A(r).



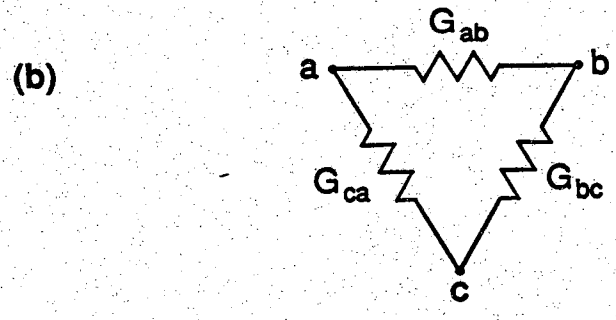


Fig. 2: (a) Schematic diagram of a multiport (3-port) quantum device. As in Fig. 1, it is assumed that inelastic scattering and dissipation takes place only in the contacts.

(b) Equivalent resistor network in zero magnetic field.

(a) Two-probe Landauer Formula

Before getting into specific structures, let us briefly state the basic approach that is commonly used to describe the behavior of quantum devices [9-13]. The device is assumed to be connected to two contacts by perfect leads that allow electron waves to propagate freely without any scattering. Inelastic scattering processes (and hence dissipation) are assumed to take place only in the contacts and *not within the device.* Under these conditions we can describe the propagation of electrons with a fixed energy E from lead 'a' to lead 'b' using a scattering matrix (Fig. 1).

$$\begin{Bmatrix} \mathbf{a}^- \\ \mathbf{b}^- \end{Bmatrix} = \begin{bmatrix} \mathbf{r}(E) & \mathbf{t}'(E) \\ \mathbf{t}(E) & \mathbf{r}'(E) \end{bmatrix} \begin{Bmatrix} \mathbf{a}^+ \\ \mathbf{b}^+ \end{Bmatrix} \tag{1}$$

Here $\mathbf{a}^+,\mathbf{a}^-(\mathbf{b}^+,\mathbf{b}^-)$ are the amplitudes of the incoming and outgoing waves respectively at lead 'a' ('b'). The lead is viewed as a waveguide having a discrete set of modes or subbands. The amplitudes $\mathbf{a}^+,\mathbf{a}^-$ are both $(M_a\times 1)$ column vectors where M_a is the number of modes in lead 'a'. Similarly $\mathbf{r}$ is an $(M_a\times M_a)$ matrix, $\mathbf{t}$ is an $(M_b\times M_a)$ matrix and so on. If we know the scattering matrix for a device, we can calculate the current I for a voltage V_D from the following expression. (Actually $|t_{nm}|^2$ in eq. (2) should be replaced by $(|t_{nm}|^2 + |t_{mn}|^2)/2$. However, even though in a magnetic field $|t_{nm}|^2$ may not equal $|t_{mn}|^2$ the two terms are equal after summing over n and m [14]).

$$I = \frac{2e}{h}\int dE[f(E) - f(E+eV_D)] \sum_{m=1}^{M_a}\sum_{n=1}^{M_b} |t_{nm}(E)|^2 \tag{2}$$

Here f(E) is the Fermi-Dirac factor. For small voltages we can simplify eq. (2) to

$$G = \frac{I}{V_D} = \frac{2e^2}{h}\int dE\left(-\frac{\partial f}{\partial E}\right) \sum_{m=1}^{M_a}\sum_{n=1}^{M_b} |t_{nm}(E)|^2 \tag{3a}$$

At low temperatures $-\partial f/\partial E \simeq \delta(E-E_F)$, so that we can further simplify eq. (3a).

$$G = \frac{2e^2}{h} \sum_{m=1}^{M_a} \sum_{n=1}^{M_b} |t_{nm}(E_F)|^2 \tag{3b}$$

This formula is known as the two-probe Landauer formula; it reduces the problem of calculating the conductance to a scattering problem not unlike those encountered in microwave circuits or integrated optics. We will not go into the details but the scattering matrix $\mathbf{S} = \begin{pmatrix} \mathbf{r} & \mathbf{t}' \\ \mathbf{t} & \mathbf{r}' \end{pmatrix}$ can, in principle, be calculated from the Schrödinger equation if we are given the scalar and vector potentials $V(\mathbf{r})$, $\mathbf{A}(\mathbf{r})$ in the device.

$$\left[\frac{(\mathbf{p}-e\mathbf{A})^2}{2m^*} + eV\right]\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \tag{4}$$

$\mathbf{p}$ is the momentum operator defined by $\mathbf{p}=-i\hbar\nabla$. If there is no magnetic field present then the vector potential $\mathbf{A}$ can be taken to be zero and eq. (4) simplifies to

$$\nabla^2\Psi = -\frac{2m^*}{\hbar^2}(E - eV)\Psi \tag{5}$$

The scalar potential $eV(\mathbf{r})$ includes any discontinuities in the band-edge due to heterojunctions, band-bending due to space charge as well as any microscopic scattering potential due to defects, impurities etc. It will be noted that eq. (5) is very similar to Maxwell's equation used in integrated optics.

$$\nabla^2\mathcal{E} = -\omega^2\mu\epsilon(\mathbf{r})\mathcal{E} \tag{6}$$

Here $\mathcal{E}$ is the electric field, ω is the radian frequency, μ is the permeability and ϵ is the spatially varying dielectric constant; we have assumed $\mathcal{E} \cdot \nabla\epsilon = 0$ for simplicity. Comparing eq. (5) to eq. (6) it is evident that electron waves moving through a medium with a varying potential $V(\mathbf{r})$ is analogous to light moving through a medium with a varying dielectric constant (or refractive index). As we might expect, the quantum interference devices we will discuss have well-known optical or microwave analogs.

It is easy to see that eq. (2) when applied to a resonant tunneling diode [15], reduces to the familiar expression [10]. In a resonant tunneling diode, neglecting any impurities or defects, the potential $V(\mathbf{r})$ varies only in the direction of propagation so that electrons are not scattered between different transverse modes.

$$t_{nm}(E) = T_m(E)\delta_{nm} \tag{7a}$$

Substituting eq. (7a) into eq. (1),

$$I = \frac{2e}{h}\int dE[f(E) - f(E+eV_D)]\sum_m |\, T_m(E)\,|^2 \tag{7b}$$

Usually the cross-sectional area, A, is fairly large so that one can replace the sum over modes m by an integral over transverse wavevectors $\mathbf{k}_t$ to obtain the familiar expression

$$\frac{I}{A} = \frac{2e}{h}\int dE[f(E) - f(E+eV_D)]\int \frac{d\mathbf{k}_t}{4\pi^2} |\, T(E,\mathbf{k}_t)\,|^2 \tag{7c}$$

We will not discuss this device further since it is very well-known and is the subject of a separate session. We mention it only to reassure ourselves that the conceptual framework we are discussing is only a generalization of concepts that we are quite familiar with.

(b) Multi-probe Landauer Formula

So far we have discussed structures having only two leads (Fig. 1). However, the same basic formalism can be extended to describe structures with multiple leads such as the one shown in Fig. 2a [13]. Such multiport planar networks have found extensive use in millimeter-wave integrated circuits and integrated optics [16]. We could calculate a (N×N) scattering matrix describing an N-port network, for electrons with energy $E=E_F$. At low temperatures and low voltages only these electrons contribute to the conductance (eq. (3b)). With N=3, we have,

$$\begin{Bmatrix} a^- \\ b^- \\ c^- \end{Bmatrix} = \begin{bmatrix} r_{aa} & t_{ab} & t_{ac} \\ t_{ba} & r_{bb} & t_{bc} \\ t_{ca} & t_{cb} & r_{cc} \end{bmatrix} \begin{Bmatrix} a^+ \\ b^+ \\ c^+ \end{Bmatrix} \tag{8a}$$

Each element of the scattering matrix is itself an (M×M) matrix if there are M modes in each lead. Büttiker [13] showed that the currents I_i(i=a,b,c) are related to the chemical potentials μ_i(i=a,b,c) by the relation (Fig. 2a)

$$I_i = \frac{2e^2}{h} \sum_{j=a,b,c} T_{ij}\mu_j - T_{ji}\mu_i \tag{8b}$$

where

$$T_{ij} = \sum_{m,n=1}^{M} |(t_{ij})_{m,n}|^2 \tag{8c}$$

In the absence of any magnetic fields, $T_{ij} = T_{ji}$ so that eq. (8b) can be written as

$$I_i = \frac{2e^2}{h} \sum_{j=a,b,c} T_{ij}(\mu_j - \mu_i) \tag{9a}$$

Eq. (9a) is basically the equation one obtains by applying Kirchhoff's laws to the three-node resistor network shown in Fig. 2b. The conductance G_{ab} connecting nodes 'a' and 'b' is given by

$$G_{ab} = \frac{2e^2}{h} T_{ab} \tag{9b}$$

Similarly G_{ac} is given by

$$G_{ac} = \frac{2e^2}{h} T_{ac} \tag{9c}$$

and so on. In general this concept can be extended to model any N-port structure as a network with N nodes connected pairwise through N(N-1)/2 resistors, as long as there is

no magnetic field. A quantum device is thus non-local in character. A change in potential at one node changes the currents at all other nodes.

(c) Continuous-probe Landauer Formula

In both Figs. 1 and 2 it is assumed that there is no inelastic scattering within the device; all the dissipation takes place in the contacts. Büttiker has applied the multiprobe formula to simulate the effect of inelastic scattering within a structure by connecting a side branch leading to a phase-randomizing reservoir. We have shown [17] that distributed inelastic scattering processes (provided the scattering rate depends only on the local electron density $n(\mathbf{r})$ and is equal to $n(\mathbf{r})/\tau_i$ where τ_i is the inelastic scattering time) can be modeled with a continuous distribution of side branches, linking the main structure to reservoirs. Generalizing eq. (8b) to an infinite number of probes distributed uniformly throughout the structure (not just at the boundaries) we can write

$$I(\mathbf{r})d\mathbf{r} = \frac{2e^2}{h} d\mathbf{r} \int d\mathbf{r}' \{T(\mathbf{r},\mathbf{r}')\mu(\mathbf{r}') - T(\mathbf{r}',\mathbf{r})\mu(\mathbf{r})\} \tag{10a}$$

The problem is to determine $T(\mathbf{r},\mathbf{r}')$. We have shown that

$$T(\mathbf{r},\mathbf{r}') = \frac{\hbar^2}{\tau_i^2} \left| G_F(\mathbf{r},\mathbf{r}') \right|^2 \tag{10b}$$

$G_F(\mathbf{r},\mathbf{r}')$ is the Green function of the Schrödinger equation including an imaginary potential $i\hbar/2\tau_i$.

$$(E_F - H + \frac{i\hbar}{2\tau_i}) G_F(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}') \tag{10c}$$

$$H = (\mathbf{p} - e\mathbf{A})^2/2m^* + eV \tag{10d}$$

In the absence of any magnetic fields, $T(\mathbf{r},\mathbf{r}') = T(\mathbf{r}',\mathbf{r})$, so that we can view the structure as a continuous network of resistors. Every volume element $d\mathbf{r}'$ is connected to

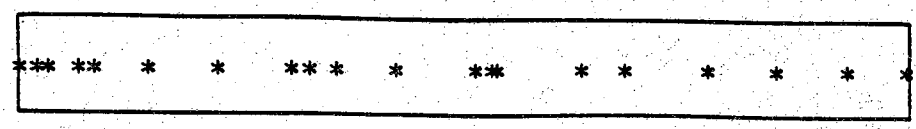
every other volume element dr by a conductance $G(\mathbf{r},\mathbf{r}')d\mathbf{r}d\mathbf{r}' = (2e^2/h)\ T(\mathbf{r},\mathbf{r}')d\mathbf{r}d\mathbf{r}'$. It can be shown that because of the imaginary potential $i\hbar/2\tau_i$, the conductances $G(\mathbf{r},\mathbf{r}')$ only extend over a finite spatial range $|\mathbf{r}-\mathbf{r}'| \leq L_i$ where L_i is the inelastic mean free path. In general we have a continuous network where every point is connected to every other point within a radius equal to the inelastic mean free path; the individual conductances are calculated from eqs. (10). The non-locality of quantum devices is thus restricted to a range equal to an inelastic mean free path.

4. Conductance Fluctuations

Consider a rather mundane device: a simple resistor with a random array of static scatterers (Fig. 3a). We can calculate its conductance G from eq. (3), once we have determined its scattering matrix. We will not go into the details of how we calculate the scattering matrix but suppose we move a *single* scatterer out of an array of 100 scatterers and calculate the conductance G as a function of the location of the scatterer. As we might expect, the conductance fluctuates (Fig. 3b) because of interference amongst the many multiply reflected waves [18]. However, one would normally expect the size of the fluctuations to go down as $\sim 1/M$. The surprising fact is that the size of the fluctuation ΔG is $\sim e^2/h$ *independent of* M [19,20]. Such 'universal conductance fluctuations' have been observed experimentally in metallic wires and narrow MOSFET's as a function of the magnetic field and the Fermi level (this is more convenient but is believed to be equivalent to moving one scatterer within the sample) [1,6-8].

Another interesting fact is that if we neglect interference effects altogether and calculate the transmission probabilities semiclassically (such as by using Monte Carlo simulation) we obtain a larger value than we get quantum mechanically. The semiclassical conductance G_C is larger than the average quantum conductance $\langle G \rangle$ by $\sim 2e^2/h$. A

(a)



(b)

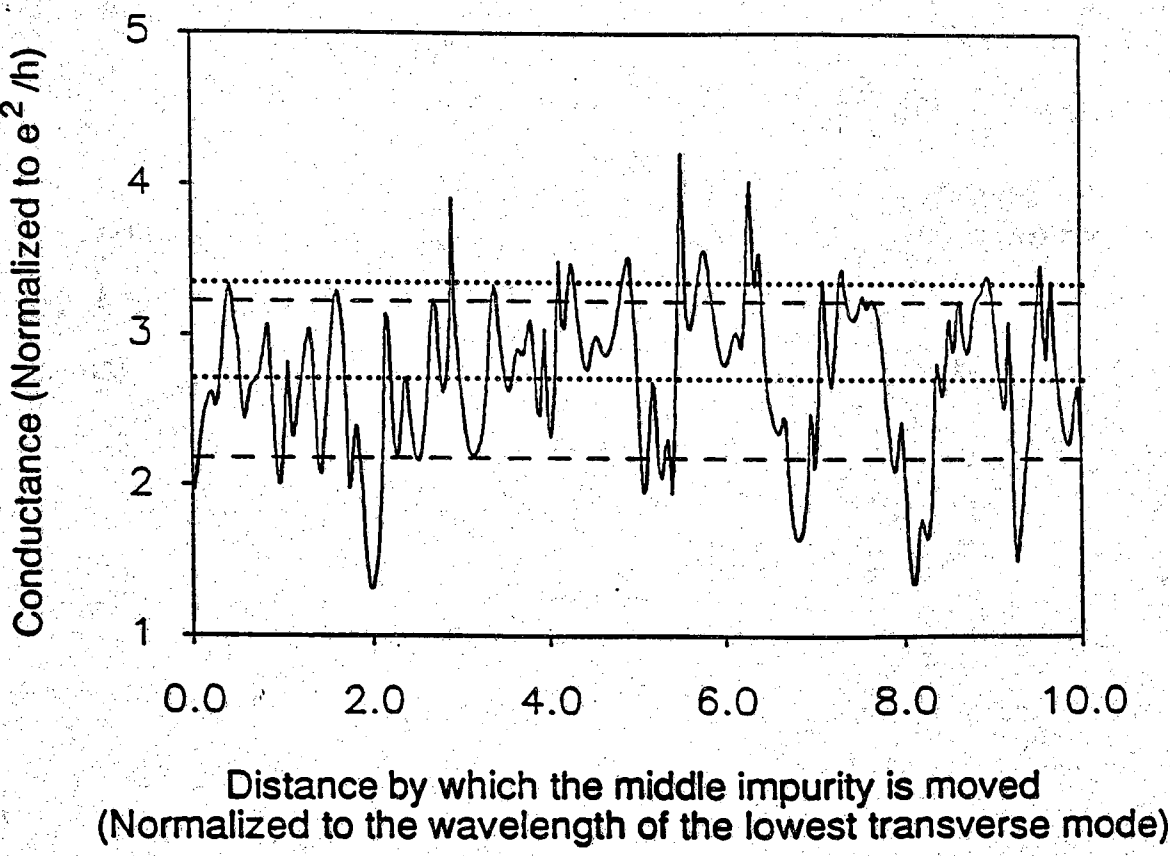


Fig. 3: (a) A resistor with an array of static scatters

(b) Conductance fluctuations as a function of the location of the middle scatterer. The remaining scatterers in the array of 100 are not moved.

small applied magnetic field destroys interference effects and causes the conductance to increase from $<G>$ to G_C. This is known as the weak localization effect and has been observed experimentally in thin metal films and MOSFET's [21].

The two effects mentioned in this section - weak localization and conductance fluctuations - do not appear to have any device applications and were discussed mainly for completeness. However, both these phenomena illustrate an important point: even under apparently random conditions quantum interference effects may not decrease as 1/M due to hidden correlations.

5. Aharonov-Bohm Effect

Most of the work done on the Aharonov-Bohm effect in solids [1,5] has utilized lithographically defined ring structures of the form shown in Fig. 4a. This structure provides two alternative paths between the two leads and seems like an obvious structure for observing quantum interference. In fact an optical analog of this structure known as the Mach-Zender interferometer (Fig. 4) is used as an optical modulator in integrated optics [22]. The transmission of light through the structure is modulated by changing the phase difference between the two arms through their refractive indices; this makes the wavelengths in the two channels different for the same frequency. In view of the similarity between eqs. (5) and (6) 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. This requires an electric field $\mathcal{E}$ *in the plane of the ring* and is known as the electrostatic Aharonov-Bohm effect which has not yet been clearly demonstrated in solids, though there is some preliminary experimental evidence [23]. What has been observed unequivocally in solids is the (magnetic) Aharonov-Bohm effect where the vector potential **A** of one arm is changed with respect to the other by applying a magnetic field **B** *perpendicular to the plane of the ring.*

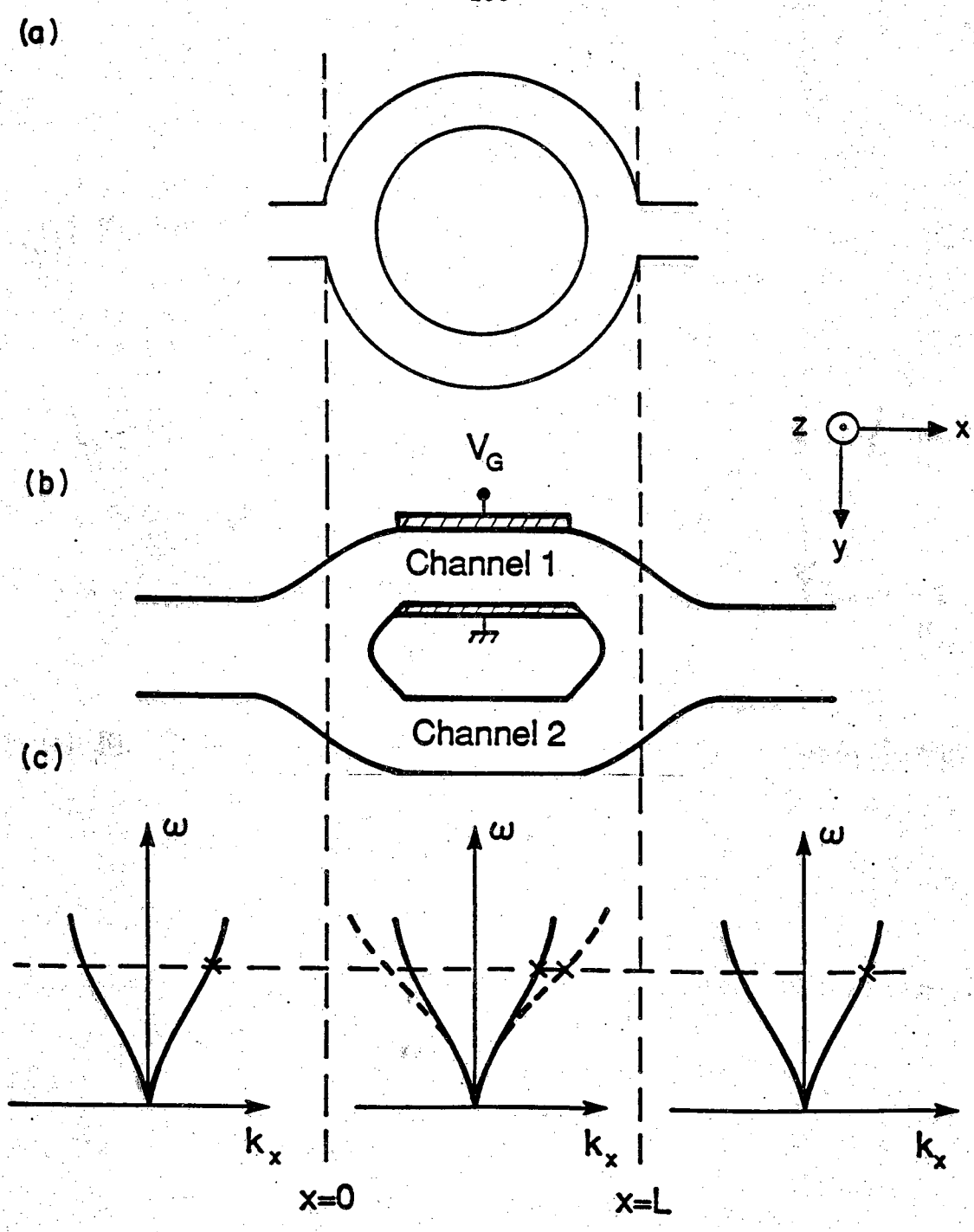


Fig. 4: (a) Ring structure commonly used in Aharonov-Bohm experiments in solids.

(b) An optical analogy: the waveguide Mach-Zender interferometer. The refractive index in Channel 1 is controlled with the gate voltage V_G through the electro-optic effect.

(c) Dispersion curves $\omega(k_x,k_y = 0)$ for the lowest guided mode in the three regions.

Neglecting multiple reflections, we can write

$$t_{nm} \simeq t_{nm}^{(1)} + t_{nm}^{(2)} \tag{11a}$$

where the superscripts 1 and 2 represent the transmission amplitudes through the two arms of the ring in the absence of any external fields. Assuming that $t_{nm}^{(1)}$ and $t_{nm}^{(2)}$ have approximately the same magnitude but differ in phase by θ_{nm} we can write from eq. (11a),

$$|t_{nm}|^2 = 2|t_{nm}^{(1)}|^2(1 + \cos\theta_{nm}) \tag{11b}$$

A magnetic field changes the phase-shift θ_{nm} and thereby modulates the transmission probabilities $|t_{nm}|^2$. Assuming that the different transmission probabilities $|t_{nm}^{(1)}|^2$ are nearly equal and that $\theta_{nm}(B) = \theta_{nm}(0) + \alpha_{nm}B$ we can write from eqs. (11b) and (3b)

$$G \simeq G_0 \sum_{m,n} 1 + \cos(\theta_{nm}(0) + \alpha_{nm}B) \tag{12}$$

If we assume that the coefficients α_{nm} are nearly equal then it is apparent from eq. (12) that the conductance will oscillate as a function of the magnetic field B with a period equal to $2\pi/\alpha$. If the zero field phase $\theta_{nm}(0)$ corresponding to different t_{nm} are completely random, it can be shown that the percentage conductance modulation $\Delta G/G \sim 1/M$, M being the number of modes [12]. However, as we have mentioned earlier, experiments on metallic rings have shown conductance modulations far in excess of $1/M$ indicating that the zero field phases are partially correlated. Experiments have also revealed the importance of using rings whose thickness is small compared to the diameter. This can be understood as follows. It can be shown that the coefficients α_{nm} are approximately equal to $eA/\hbar$ where A is the area enclosed by the two arms of the ring. This means that there is a spread in the values of α_{nm} proportional to the difference in the areas enclosed by the inner and outer diameters. If this spread is large,

there is no unique period to the Aharonov-Bohm oscillations and the conductance variation appears more like random fluctuations [5].

It seems that for device applications the electrostatic Aharonov-Bohm effect would be more suitable than the magnetic one. There is an important distinction between the magnetic and the electrostatic Aharonov-Bohm effects. It can be shown that for the electrostatic effect the phase-shift is proportional to the potential difference V_{12} between the two arms of the ring: $\theta_{nm}(V_{12}) = \theta_{nm}(0) + \alpha'_{nm}V_{12}$. The coefficients α'_{nm} are approximately equal to $e\tau_t/\hbar$ where τ_t is the transit time for an electron through the ring. This makes the electrostatic effect difficult to observe since under the usual conditions of diffusive transport since there is a large spread in the transit times [24]. The situation is analogous to the magnetic effect in rings with poor aspect ratios (= inner diameter/outer diameter). For this reason it seems that ballistic structures with minimal multiple reflections which minimize the spread in transit times are more suitable for observing the electrostatic effect. The structure that we have proposed as a possible design for a 'Quantum Interference Transistor' (QUIT) is shown in Fig. 5a [25-30]. It is basically an ordinary Field Effect Transistor (FET) with a barrier in the middle of the channel. The length L is small enough that electrons travel *ballistically* across it. The channels are narrow enough to be *single-moded* in the z-direction. One way to fabricate this structure may be to grow a barrier layer over the entire film, interrupt the growth to etch away the barrier where it is not needed and then continue the growth process. The major challenge lies in ensuring the quality of the regrown interfaces. At this stage it is not clear whether this structure or the lithographically defined ring structure will eventually prove more suitable for device applications. An obvious concern about the structure in Fig. 5a is its poor aspect ratio. However, we believe that if the structure is single-moded in the z-direction, the aspect ratio is not very important. This is because every electron has a unique wavefunction in the z-direction (shown in Fig. 5a)

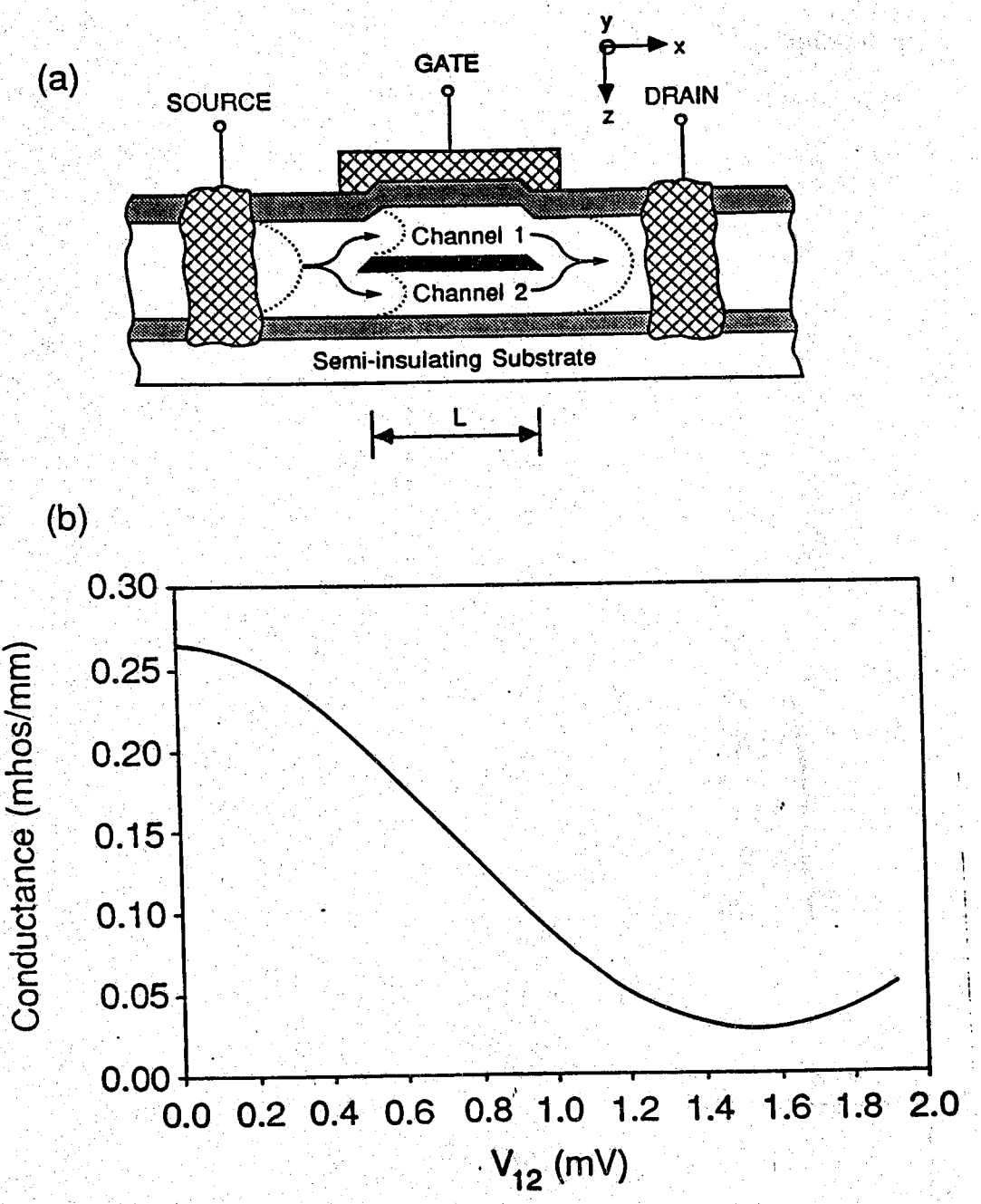


Fig. 5: A proposed Quantum Interference Transistor. The structure consists of a conducting channel with a barrier in the middle. Also shown is the calculated conductance versus potential difference between the channels.

and hence a unique phase-shift in an electric or magnetic field; hence there is, theoretically, no spread in θ.

Experiments on rings in high-mobility GaAs/Al GaAs heterostructures with $M < 10$ have shown conductance modulations $\sim$10% in a magnetic field. The obvious way to increase the percentage conductance modulation ($\Delta G/G$) is to reduce the number of modes; however, this does not increase the absolute conductance modulation (ΔG). If we could ensure perfect symmetry between the two channels in the structure shown in Fig. 5a then it is theoretically possible to attain a 100% modulation in the conductance even if the structure is very wide in the y-direction so that $M \rightarrow \infty$ [25-30]. This is because the conditions of perfect symmetry, ballistic transport and single-modedness in the z-direction imply a perfect correlation among the different modes so that the zero field phases $\theta_{mn}(0)$ are all zero, leading to a $\Delta G/G$ far in excess of the $1/M$ estimate. To what extent this perfect correlation can be implemented in real semiconductor structures, only future experiments can tell.

Since we have assumed only a single mode in the z-direction the subscripts n,m in eq. (11c) refer to modes in the y-direction. In a ballistic structure there is no scattering among these modes so that with the zero-field phases equal to zero we can write

$$G \simeq G_0 \sum_m (1 + \cos \alpha'_m V_{12}) \tag{13}$$

To compute the phase-shift $\alpha'_m V_{12}$ induced by a gate voltage we note that the dispersion relations for the two channels can be written as,

$$E = \epsilon_{1,m} + \frac{\hbar^2 k_1^2}{2m^*} \tag{14a}$$

$$E = \epsilon_{2,m} + \frac{\hbar^2 k_2^2}{2m^*} \tag{14b}$$

where k_1 and k_2 are the wave numbers in the x-direction and $\epsilon_{1,m}$ and $\epsilon_{2,m}$ are the

energies at the bottom of subband 'm' in channels 1 and 2 respectively. Without a gate voltage, the two channels have been assumed perfectly symmetric so that $\epsilon_{1,m} = \epsilon_{2,m}$. But an applied gate voltage changes the average potential in channel 1 with respect to that in channel 2. It can be shown from lowest order perturbation theory that

$$\epsilon_{1,m} - \epsilon_{2,m} = eV_{12} \tag{15a}$$

where

$$V_{12} = \langle 1 | V(z) | 1 \rangle - \langle 2 | V(z) | 2 \rangle \tag{15b}$$

$|1\rangle$ and $|2\rangle$ are the wavefunctions in the z-direction in channels 1 and 2 respectively as shown in Fig. 4a. Using eqs. (14a), (14b) and (15a) we have

$$\alpha'_m V_{12} = (k_1 - k_2)L = eV_{12}L/\hbar v \tag{16a}$$

where

$$v = \hbar(k_1 + k_2)/2m^* \tag{16b}$$

The velocity v for a given energy E is different for different modes, m, since $\epsilon_{1,m}$ and $\epsilon_{2,m}$ are different (eqs. (14a,b)). Consequently, the coefficients α'_m varies slowly from one mode to another. Fig. 5b shows the normalized conductance as a function of V_{12} calculated for a structure that is assumed very wide in the y-direction so that the summation over m in eq. (13) can be replaced by an integral over k_y. A conductance modulation $\Delta G/G \sim 90\%$ is predicted even though $M \rightarrow \infty$! In this structure $\Delta G/G$ is far in excess of 1/M because we have deliberately engineered it to ensure that with $V_{12} = 0$, every mode m interferes constructively leading to a maximum in the conductance. Future experiments will show to what extent this perfect correlation amongst the modes can be implemented and how deleterious the inevitable asymmetries in real semiconductor structures will be. One way to get a large percentage modulation $\Delta G/G$ in the conductance despite the asymmetries is to restrict the number of modes by making the

structure narrower in y; the absolute magnitude of the conductance modulation ΔG, however, would decrease.

The attractive feature of this proposed Quantum Interference Transistor is the very small potential ($\sim$mV) that is required to operate it. This may result in large transconductances and low power-delay products [28]. It seems that the speed of operation of this device should ultimately be limited by the transit-time across the two-channel region of length L. However, as L is decreased the potential difference V_{12} required to switch the device is increased and becomes comparable to the potential required to deplete the channel if L becomes comparable to the DeBroglie wavelength of electrons ($kL \sim 2\pi$).

>From an applied point of view an important question is the temperature range over which the device can operate. Temperature affects the operation of the device in two distinct ways. Firstly, it increases the energy spread of the electrons so that the variation of $\delta\theta$ over E (as well as k_y) is important as we discussed earlier. Secondly, it increases the inelastic scattering in the device. It seems that even at 77K, L can be made small enough that inelastic scattering should not be excessive. However, only further experiments can answer these questions conclusively.

6. Non-local Effects

Consider the four-probe Hall bridge shown in Fig. 6a. Ordinarily we would expect $(V_b - V_a)$ to be some fraction of the voltage V applied across the structure. However, if inelastic scattering is negligible within the structure, then even this simple everyday structure holds some surprises. According to our discussion in Section 2 we could view this multiport quantum device as an interconnected network of resistors as shown in Fig. 6b. To calculate $(V_b - V_a)$ we can calculate a Thevenin equivalent as shown in Fig. 6c with

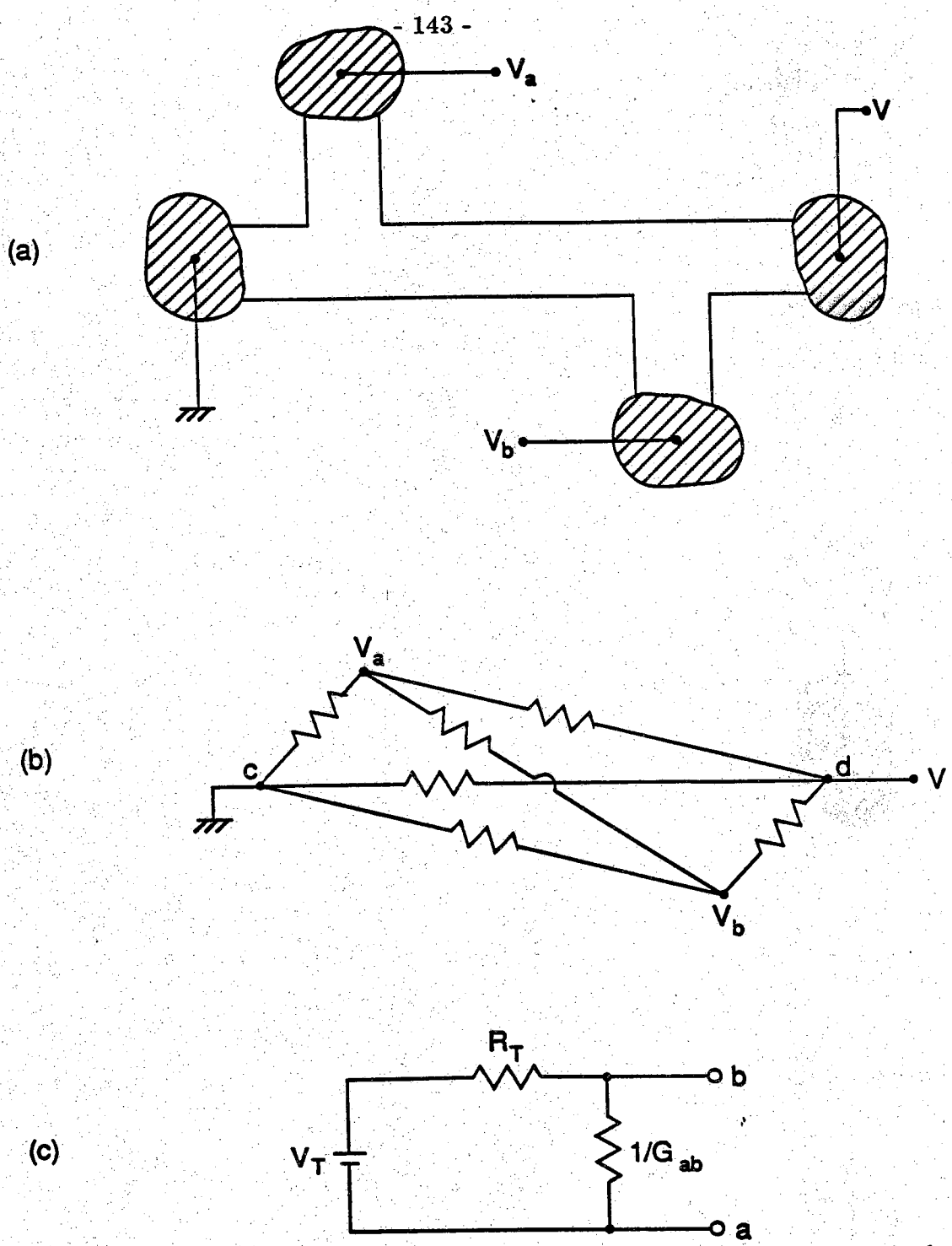


Fig. 6: A four-probe Hall bridge whose dimensions are much smaller than an inelastic mean free path.

(a) Configuration.

(b) Equivalent Resistor Network.

(c) Thevenin's equivalent for the network in (b).

$$V_T = V\left(\frac{G_{bc}}{G_{bc}+G_{bd}} - \frac{G_{ac}}{G_{ac}+G_{ad}}\right) \tag{17a}$$

$$R_T = \frac{1}{G_{bc}+G_{bd}} + \frac{1}{G_{ac}+G_{ad}} \tag{17b}$$

It is apparent from eq. (17a) that depending on the relative values of the different conductances V_T could even be negative as pointed out by Büttiker [30]. It is hard to rationalize a voltage drop that is opposite to the direction of current flow, though it follows quite simply from our resistor network model. One way to 'explain' it is to say that the voltage probes are affected by everything within an inelastic mean free path and do not just measure the local potential. A variety of such non-local effects have recently been observed [2,3,7,8,31,32].

Multiport structures of the type shown in Fig. 2a can be used to implement transistors of a somewhat different type than that discussed in Section 8.2 [33,34]. Consider the 3-port network shown in Fig. 7a; 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. As every microwave engineer knows, the transmission between two ports is influenced by the load conditions at the third port. Thus, if we change the phase of the reflection coefficient at port 3 by changing the gate potential, it should affect the current that flows between ports 1 and 2 labeled source and drain. The surprising feature of this device is that the gate is not positioned between the source and the drain as we are accustomed to expect in electronic devices. It can be located anywhere within an inelastic mean free path. However, this 'remote control' is well-known in microwave networks, and may find useful applications in quantum devices of the future.

For a quantitative description of the operation of the device in Fig. 7a we need the scattering matrix describing the 3-way splitter.

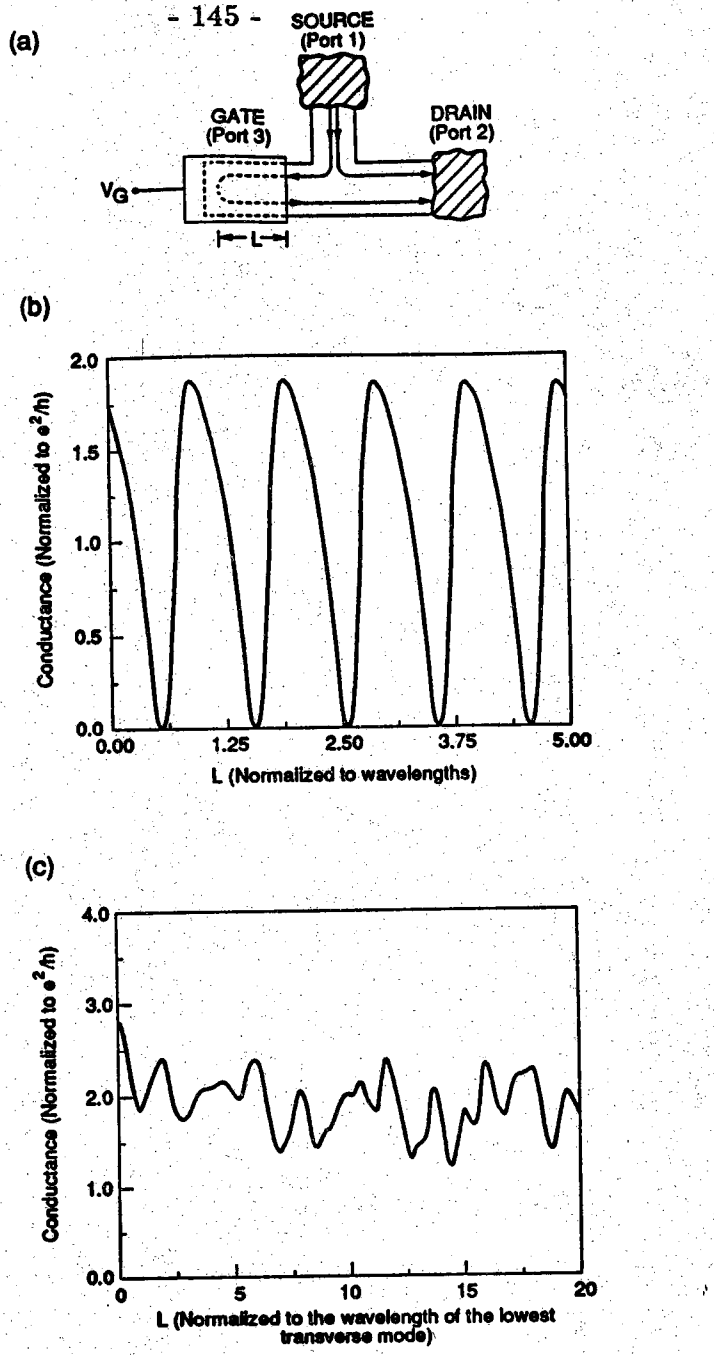


Fig. 7: (a) A proposed quantum transistor with a remote gate. The gate potential changes the phase difference between the two primary paths between the source and the drain.

(b) Normalized conductance as a function of the length L in wavelengths for a device with one propagating mode (M=1).

(c) Normalized conductance as a function of the length L in wavelengths (of the lowest mode having the shortest wavelength) for a device with eight propagating modes (M=8).

$$\begin{Bmatrix} \mathbf{S}^- \\ \mathbf{D}^- \\ \mathbf{G}^- \end{Bmatrix} = \begin{bmatrix} \mathbf{r}_{SS} & \mathbf{t}_{SD} & \mathbf{t}_{SG} \\ \mathbf{t}_{DS} & \mathbf{r}_{DD} & \mathbf{t}_{DG} \\ \mathbf{t}_{GS} & \mathbf{t}_{GD} & \mathbf{r}_{GG} \end{bmatrix} \begin{Bmatrix} \mathbf{S}^+ \\ \mathbf{D}^+ \\ \mathbf{G}^+ \end{Bmatrix} \tag{18a}$$

We have used $\mathbf{S}^{\pm}$, $\mathbf{D}^{\pm}$, $\mathbf{G}^{\pm}$ to denote the incoming and outgoing wave amplitudes at the source, drain and gate respectively; if the leads are multimoded then the amplitudes are column vectors as we discussed earlier. If the wave amplitudes $\mathbf{G}^+$ and $\mathbf{G}^-$ at the gate are related by the reflection matrix $\mathbf{R}$

$$\mathbf{G}^+ = \mathbf{R}\,\mathbf{G}^- \tag{18b}$$

we can show from eqs. (18a) and (18b) that

$$\begin{Bmatrix} \mathbf{S}^- \\ \mathbf{D}^- \end{Bmatrix} = \begin{bmatrix} \mathbf{r} & \mathbf{t}' \\ \mathbf{t} & \mathbf{r}' \end{bmatrix} \begin{Bmatrix} \mathbf{S}^+ \\ \mathbf{D}^+ \end{Bmatrix} \tag{19}$$

where

$$\mathbf{r} = \mathbf{r}_{SS} + \mathbf{t}_{SD}(\mathbf{I} - \mathbf{R}\mathbf{r}_{GG})^{-1}\mathbf{R}\mathbf{t}_{DS} \tag{20a}$$

$$\mathbf{t}' = \mathbf{t}_{SD} + \mathbf{t}_{SG}(\mathbf{I} - \mathbf{R}\mathbf{r}_{GG})^{-1}\mathbf{R}\mathbf{t}_{GD} \tag{20b}$$

$$\mathbf{t} = \mathbf{t}_{DS} + \mathbf{t}_{DG}(\mathbf{I} - \mathbf{R}\mathbf{r}_{GG})^{-1}\mathbf{R}\mathbf{t}_{GS} \tag{20c}$$

$$\mathbf{r}' = \mathbf{r}_{DD} + \mathbf{t}_{DG}(\mathbf{I} - \mathbf{R}\mathbf{r}_{GG})^{-1}\mathbf{R}\mathbf{t}_{GD} \tag{20d}$$

It is apparent from eq. (20c) that the transmission $\mathbf{t}$ from the source to the drain (which determines the conductance) can be modulated by changing the reflection coefficient at the gate ($\mathbf{R}$). We can get some physical insight by expanding eq. (20c) in a geometric series as follows.

$$\mathbf{t} = \mathbf{t}_{DS} + \mathbf{t}_{DG}\,\mathbf{R}\,\mathbf{t}_{GS} + \mathbf{t}_{DG}\,\mathbf{R}\,\mathbf{r}_{GG}\,\mathbf{R}\,\mathbf{t}_{GS} + \mathbf{t}_{DG}(\mathbf{R}\,\mathbf{r}_{GG})^2\,\mathbf{R}\,\mathbf{t}_{GS} + \cdots \qquad (21)$$

The first term on the right in eq. (31) is the amplitude for direct transmission from port 1 to port 2 while the succeeding terms are the amplitudes for transmission after one, two, three, ... reflections at port 3. If $\mathbf{r}_{GG}$ is small, then we can write approximately

$$\mathbf{t} \simeq \mathbf{t}_{DS} + \mathbf{t}_{DG}\,\mathbf{R}\,\mathbf{t}_{GS} \qquad (22)$$

With this approximation, we could view the device in Fig. 7a as providing two primary paths from the source to the drain with the gate controlling their interference, much like the Aharonov-Bohm device. Figs. 7b, 7c show the calculated conductance for two specific devices having one mode and eight modes respectively. We assume that the reflection coefficient $\mathbf{R}$ is given by

$$R_{nm} = \delta_{nm}\, e^{ik_m L} \qquad (23)$$

where k_m is the wavenumber of the mth mode at the Fermi level and L is the effective distance of the gate from the junction. Fig. 7b shows that the conductance is modulated by 100% as the length L is changed. On the other hand for a device with eight modes, the variation in the conductance with L looks more like conductance fluctuations $\sim e^2/h$. It seems that devices of the type shown in Fig. 7a will only be useful if the number of modes is small. Using single-moded quantum wires at low temperatures we can conceive of a variety of devices analogous to well-known microwave and integrated optical devices. Even the polarization of electromagnetic waves has its analog in the spin of electrons. It may be possible to exploit this degree of freedom in narrow gap semiconductors having a large spin-orbit coupling that can be controlled with a gate potential [35].

7. Quantum Networks

The device in Fig. 7a utilized one port in a 3-port structure to modulate the conductance connecting the other two ports. We could go a step further and utilize one port in an (N+1)-port structure to control the network of N(N-1)/2 resistors formed by the remaining N ports. Consider for example the structure in Fig. 8a with four ports connected to the side branches of a main waveguide. The gate potential shifts the standing wave pattern in the main waveguide and thus alters the couplings between different ports. A strong coupling occurs when two ports are both located on the peaks of the standing wave pattern while a port located near a trough of the standing wave pattern is only weakly coupled to the other ports. The coupling between two ports determines the magnitude of the conductance connecting the corresponding nodes in the equivalent resistor network (eqs. (9b,c)). We thus have a programmable resistor network (Fig. 8b) rather than a single resistor as in a conventional transistor. By connecting each node to a non-linear device such as a resonant tunneling diode (this can be done vertically without external connections), we could implement a highly interconnected pattern of non-linear elements. If we model this device as a capacitance C in parallel with a non-linear conductance described by $V = f(I)$ then Kirchoff's law applied to the circuit in Fig. 8b yields

$$C\frac{dV_i}{dt} + I_i + V_i(\sum_j G_{ij}) = \sum_j G_{ij}V_j \qquad (25)$$

where $V_j = f(I_j)$. Eq. (25) is very similar to the dynamical equations used to describe neural networks [36]. Clearly major hurdles remain to be overcome before such exotic devices become practicable. But it should be noted that the true power and utility of quantum devices may eventually lie not in the implementation of conventional transistors with a source, a drain and a gate (where the low current capability is a major hur-

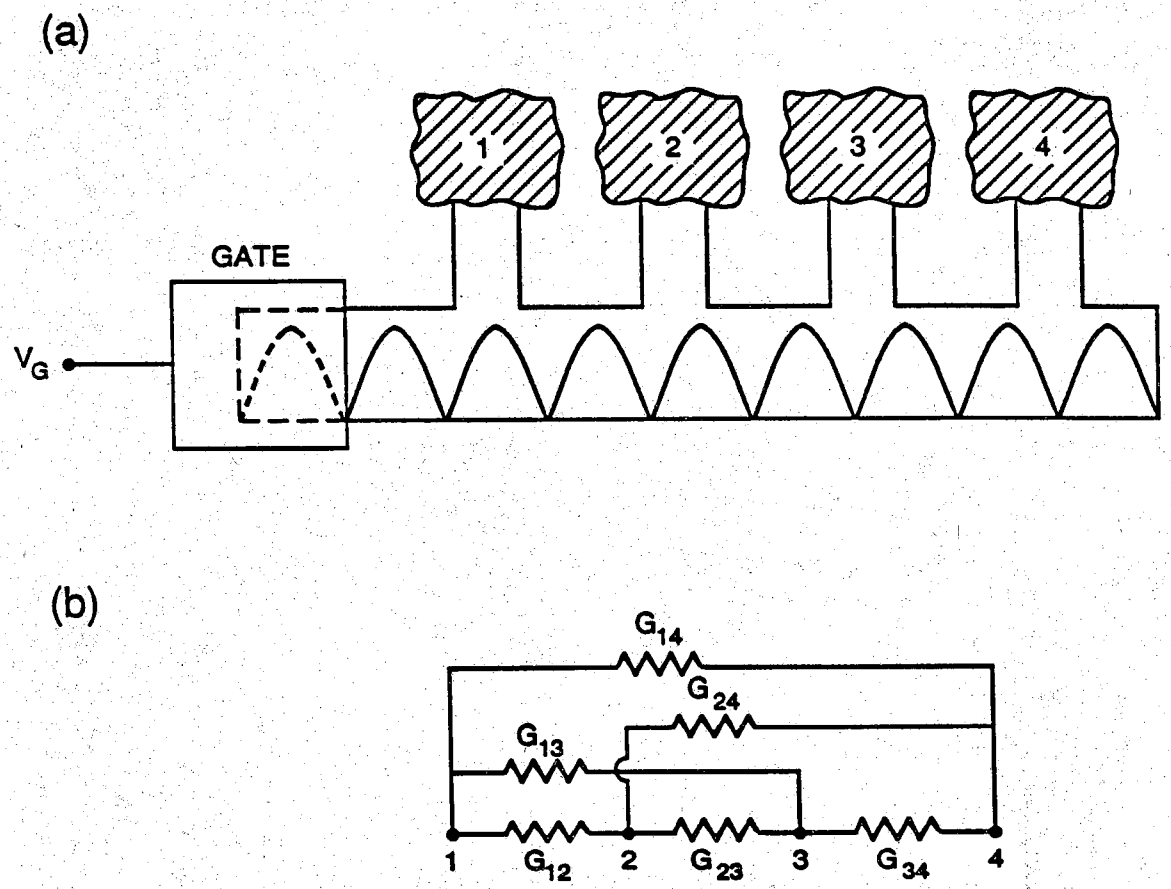


Fig. 8: A programmable multiterminal quantum network.

(a) Configuration.

(b) Equivalent circuit. The dotted lines also show a non-linear conductance connected to each node.

dle), but in the implementation of programmable multiterminal quantum networks that may lead to radically new concepts for electronic devices.

8. Summary

In summary, we have discussed theoretically the possibility of two types of quantum devices. The first is a two-port quantum device with two alternative paths connecting the ports. The current is modulated by controlling the interference between the paths through the potential distribution in the device. A very small potential is needed to turn off the current so that large transconductances and low power-delay products are expected.

The second type of device uses one port in an (N+1)-port structure to control the network of N(N-1)/2 resistors formed by the remaining N-ports. If N=2 we have a transistor with a remote gate. If N>2 we have a programmable resistor network that could be used to implement neural networks. Alternatively we could implement complex logical functions with a source and a drain controlled by multiple gates.

Acknowledgements

This work was supported by the National Science Foundation under Grant no. ECS-83-51-036, by the Semiconductor Research Corporation under Contract no. 87-SJ-089, and by the Office of Naval Research under Contract no. N00014-87-K-0693. The author would like to thank R. Frohne for Figs. 5 and 7 and M. McLennan for Fig. 3 and for carefully reading the manuscript. The author is deeply indebted to his colleague Mark S. Lundstrom for delivering this talk on his behalf at the conference, for many helpful suggestions, and for his continued collaboration and friendship.

Appendix A References

[1] G. Timp, A. M. Chang, P. Mankiewich, R. Behringer, J. E. Cunningham, T. Y. Chang and R. E. Howard, Phys. Rev. Lett. **59**, 732 (1987).

[2] C. J. B. Ford, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, C. T. Foxon, J. J. Harris and C. Roberts, J. Phys. C.: Solid State Phys. **21**, L325 (1988).

[3] G. Timp, A. M. Chang, J. E. Cunningham, T. Y. Chang, P. Mankiewich, R. Behringer and R. E. Howard, Phys. Rev. Lett. **58**, 2814 (1987); G. Timp, H. U. Baranger, P. DeVegvar, J. E. Cunningham, R. E. Howard, R. Behringer and P. M. Mankiewich, Phys. Rev. Lett. (to appear); P. G. N. DeVegvar, G. Timp, P. M. Mankiewich, J. E. Cunningham, R. Behringer and R. E. Howard, preprint.

[4] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel and C. T. Foxon, Phys. Rev. Lett. **60**, 848 (1988); D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie and G. A. C. Jones, J. Phys. C.: Solid State Phys. **21**, L209 (1988).

[5] The work on Aharonov-Bohm effect is reviewed in S. Washburn and R. A. Webb, Adv. Phys. **35**, 375 (1986).

[6] S. Washburn, C. P. Umbach, R. B. Laibowitz and R. A. Webb, Phys. Rev. **B32**, 4789 (1985).

[7] W. J. Skocpol, P. M. Mankiewich, R. E. Howard, L. D. Jackel, D. M. Tennant and A. D. Stone, Phys. Rev. Lett. **58**, 2347 (1987); A. Benoit, C. P. Umbach, R. B. Laibowitz and R. A. Webb, Phys. Rev. Lett. **58**, 2343 (1987).

[8] T. J. Thornton, M. Pepper, H. Ahmed, D. Andrews and G. J. Davies, Physical Review Letters **56**, 1198 (1986).

[9] R. Landauer, IBM J. Res. Dev. **1**, 223 (1957).

[10] R. Tsu and L. Esaki, Appl. Phys. Lett. **22**, 562 (1973).

[11] R. Landauer, Philos. Mag. **21**, 863 (1970).

[12] M. Büttiker, Y. Imry, R. Landauer and S. Pinhas, Phys. Rev. **B31**, 6207 (1985).

[13] M. Büttiker, Phys. Rev. Lett. **57**, 1761 (1986); M. Büttiker, IBM J. Res. Dev. **32**, 63 (1988).

[14] D. S. Fisher and P. A. Lee, Phys. Rev. **B23**, 6851 (1981).

[15] E. R. Brown, T. C. L. G. Sollner, W. D. Goodhue and C. D. Parker, Appl. Phys. Lett. **50**, 83 (1987); F. Capasso, S. Sen, A. C. Gossard, A. L. Hutchinson and J. H. English, IEEE Electron Device Letters **EDL-7**, 573 (1986).

[16] T. Okoshi, *Planar Circuits for Microwaves and Lightwaves,* Springer Series in Electrophysics 18, Springer-Verlag, 1985.

[17] S. Datta and M. McLennan, Technical Report TR-EE 88-42, Purdue University.

[18] M. Cahay, M. McLennan and S. Datta, Phys. Rev. **B**, 10125 (1988).

[19] A.D. Stone, Phys. Rev. Lett. **54**, 2692 (1985).

[20] P.A. Lee, A.D. Stone and H. Fukuyama, Phys. Rev. **B35**, 1039 (1987) and references therein.

[21] G. Bergmann, Phys. Reports, **107**, 1 (1984) and references therein.

[22] F.J. Leonberger, C.E. Woodward and D.L. Spears, IEEE Trans. on Circuits and Systems, **CAS–26**, 1125 (1979); W. E. Martin, Appl. Phys. Lett. **26**, 560 (1975).

[23] S. Washburn, H. Schmid, D. Kern and R. A. Webb, Phys. Rev. Lett. **59**, 1791 (1987).

[24] S. Datta, M. Cahay and M. McLennan, Phys. Rev. **B36**, 5655 (1987).

[25] S. Datta et. al.: Second International Conference on Modulated Semiconductor Structures, Kyoto, Sept. 1985 (Surf. Sci. **174**, 439 (1986)). A lithographic ring structure operating on a similar principle has been described by A.B. Fowler, US Patent No. 45503320 (1985).

[26] S. Datta et. al., Phys. Rev. Lett. **55**, 2344 (1985).

[27] S. Datta, M.R. Melloch, S. Bandyopadhyay and M.S. Lundstrom, Appl. Phys. Lett. **48**, 487 (1986).

[28] S. Bandyopadhyay, S. Datta, and M.R. Melloch, Superlattices and Microstructures **2**, 539 (1986); S. Bandyopadhyay et. al., Proceedings of the International Electron Devices Meeting, IEEE Catalog No. 86CH2381-2.

[29] S. Datta and S. Bandyopadhyay, Phys. Rev. Lett. **58**, 717 (1987).

[30] M. Büttiker, IBM Journal of Research and Development **32**, 317 (1988).

[31] H.U. Baranger, A.D. Stone and D.P. Di Vencenzo, Phys. Rev. **B37**, 6521 (1988).

[32] C. P. Umbach, P. Santhanam, C. van Haesendonck and R. A. Webb, Appl. Phys. Lett. **50**, 1289 (1987).

[33] A more detailed version of this paper will appear in S. Datta, Quantum Interference Devices, chapter in Physics of Quantum Electron Devices, edited by F. Capasso, to be published in 1989 by Springer-Verlag.

[34] S. Datta, Proceedings of the International Conference on Solid State Devices and Materials, Tokyo, Aug. 1988.

[35] Y.A. Bychkov and E.I. Rashba, J. Phys. **C17**, 6039 (1984).

[36] H. Sompolinsky, A Crisanti and H. J. Sommers, Phys. Rev. Lett. **61**, 259 (1988) and references therein.

Appendix B

SRC-SUPPORTED PUBLICATIONS AND TALKS

Reports:

[1] M.S. Lundstrom, S. Datta, R.J. Schuelke, S. Bandyopadhyay, and P.H. Sorlie, "Physics and Modeling of Heterostructure Semiconductor Devices," Purdue University Technical Report, TR-EE 84-35, August 1984 (also available as an SRC annual report).

[2] M.S. Lundstrom, S. Datta, S. Bandyopadhyay, P. Sorlie, M. Klausmeier-Brown, C. Maziar, and M. Cahay, "Physics and Modeling of Heterostructure Semiconductor Devices," Purdue University Technical Report, TR-EE 85-14, 1985, (also available as an SRC annual report).

[3] M. S. Lundstrom, S. Datta, S. Bandyopadhyay, M. Cahay, A. Das, T. E. Dungan, M. E. Klausmeier-Brown, C. M. Maziar, M. J. McLennan, "Physics and Modeling of Heterostructure Semiconductor Device," TR-EE-86-311, 1986.

[4] S. Datta, M.S. Lundstrom, M. Cahay, A. Das, H.R. Frohne, and M.J. McLennan, "Physics and Modeling of Sub-Micron Devices," Purdue University Technical Report, TR-EE-87-35, 1987.

Theses:

[1] Robert J. Schuelke, "Numerical Simulation of Semiconductor Heterostructures," Ph.D. Thesis, Purdue Univ., Dec. 1984.

[2] Paul H. Sorlie, "Hall Effect Characterization of GaAs Compounds," M.S.E.E. Thesis, Purdue Univ., Aug. 1985.

[3] Supriyo Bandyopadhyay, "Electron Transport in Sub-Micron Devices," Ph.D. Thesis, Purdue Univ., Dec. 1985.

[4] Martin E. Klausmeier-Brown, "Monte Carlo Studies of Electron Transport in III-V Heterostructures," M.S.E.E. Thesis, Purdue Univ., May 1986.

[5] Christine M. Maziar, "A Critical Analysis of Electron Transport in AlGaAs/GaAs Heterojunction Bipolar Transistors," Ph.D. Thesis, Aug. 1986.

[6] Michael J. McLennan, "Quantum Ballistic Transport in Semiconductor Heterostructure," M.S.E.E. Thesis, Purdue Univ., May 1987.

[7] Marc Cahay, "Quantum Mechanical Analysis of Ultrasmall Devices," Ph.D. Thesis, Purdue Univ., Dec. 1987.

[8] R. Frohne, "Analysis of Electron Waveguides," Ph.D. Thesis, Purdue Univ., Dec. 1988.

Conference Presentations:

[1] M.S. Lundstrom and J.L. Gray, "Solution of Poisson's Equation in III-V Heterostructures," presented at the Second SIAM-IEEE Conference on Numerical Simulation of VLSI Devices, Boston, Mass., Nov. 12, 1984.

[2] S. Bandyopadhyay, S. Datta, M.S. Lundstrom, and C.M. Maziar, "A Two-Temperature Distribution Function for High-Field Transport in GaAs," presented at the March Meeting of the American Physical Society, Baltimore MA., March 25-29, 1985.

[3] S. Bandyopadhyay, M. E. Klausmeier-Brown, C. M. Maziar, S. Datta, and M. S. Lundstrom, "Transport Parameters in Sub-Micron Devices," presented at the 2nd International Conf. on Simulation of Semiconductor Devices and Processes, Swansea, U.K., July 1987.

[4] M. Cahay, S. Bandyopadhyay, M. McLennan, S. Datta, and M. S. Lundstrom, "Quantum Transport in Ultrasmall Structures", Am. Phys. Soc. Meeting, Las Vegas, April 1987.

[5] M. Cahay, S. Bandyopadhyay, S. Datta, and M. S. Lundstrom, "Quantum Mechanical Device Modeling," presented at the 2nd International Conf. on Simulation of Semiconductor Devices and Processes, Swansea, U.K., July 1987.

[6] S. Datta, S. Bandyopadhyay, M. R. Melloch, R. Reifenberger, M. Miller, M. Vaziri, T. Dungan, and R. Noren, "Aharonov-Bohm Oscillations Due to Quantum Interference Between Parallel Quantum Wells," presented at the 2nd International Conf. Modulated Semiconductors, Kyoto, Japan, Sept., 1985 (*Surface Science* , 174, 439 (1986)).

[7] S. Datta, S. Bandyopadhyay, M. R. Melloch, R. Reifenberger, and M. S. Lundstrom, "Aharonov-Bohm Effect in Semiconductor Microstructures", Am. Phys. Soc. Meeting, Las Vegas, April 1987.

[8] S. Bandyopadhyay, S. Datta, M. R. Melloch, R. Reifenberger, and M. S. Lundstrom, "Transmission Matrix Analysis of the Aharonov-Bohm Effect", Am. Phys. Soc. Meeting, Las Vegas, April 1987.

[9] M. R. Melloch. S. Bandyopadhyay, S. Datta, R. Noren, M. S. Lundstrom, K. Tan, and T. Dungan, "Aharonov-Bohm Effect in an MBE Grown Double Quantum Well", Sixth Molecular Beam Epitaxy Workshop, Aug. 14-16, 1985, Minneapolis, MN.

[10] M. S. Lundstrom, "Transport Limits to Bipolar Speed," presented at the SRC Topical Research Conference on Bipolar Device Technology, Tempe, AZ, April 24-25, 1986.

[11] S. Bandyopadhyay, M. R. Melloch, S. Datta, J. A. Cooper, Jr., and M. S. Lundstrom, "A Novel Quantum Interference Transistor (QUIT) wth extremely Low Power-Delay Product and Very High Transconductance," presented at IEEE Electron Devices Meeting, Los Angeles, CA, Dec. 1986.

[12] M. Cahay, M. McLennan, S. Datta, and M. S. Lundstrom, "Self-Consistent Calculation of I-V Characteristics of Ultra-Small Devices," presented at Intern. Conf. on Numerical Modeling of Semiconductor," Los Angeles, CA, Dec. 1986.

[13] M. Cahay, M. McLennan and S. Datta, "Analysis of electron propagation through at two-dimensional random array of scatters," presented at the March meeting of the American Physical Society, New York, March 16-20, 1987.

[14] S. Bandyopadhyay, B. Das, R. Reifenberger, S. Datta, W.P. Hong and P.K. Bhattacharya, "Aharonov-Bohm effect in a double quantum well structure with (InAs) (GaAs) monolayer superlattices," presented at the March meeting of the American Physical Society, New York, March 16-20, 1987.

[15] S. Datta and S. Bandyopadhyay, "Analysis of the Aharonov-Bohm effect in the ballistic regime," presented at the March meeting of the American Physical Society, New York, March 16-20, 1987.

[16] S. Datta, "A new concept for a quantum interference transistor," presented at the workshop on ballistic electrons, Santa Barbara, March 22-26, 1987. (INVITED).

[17] M. Cahay, M. McLennan, S. Datta, "Analysis of electron propagation through narrow n^+ GaAs wires," presented at the Third International Conference on Superlattices, Microstructures and Microdevices, Chicago, August 17-20, 1987.

[18] S. Datta, "Quantum Networks," International Conference of Electronic Materials, Tokyo, Japan, June 1988 (INVITED).

[19] S. Datta, "Quantum Devices," Fourth International Conference on Superlattices, Microstructures and Microdevices, Trieste, Italy, August 1988 (KEYNOTE LECTURE).

[20] S. Datta, "Quantum Interference Transistors," International Conference on Solid State Devices and Materials, Tokyo, Japan, August 1988 (INVITED).

[20] A. Das and M. S. Lundstrom, "Modeling Electron Injection Across Semiconductor Barriers," presented at the Numerical Analysis of Semiconductor Processes and Devices, NUPAD2, San Diego, CA, 1988.

Journal Publications:

[1] M.S. Lundstrom, "Boundary Conditions for pn Heterojunctions," *Solid-State Electron.,* Vol. 27, pp. 491-496, 1984.

[2] R.J. Schuelke and M.S. Lundstrom, "Thermionic Emission-Diffusion Theory of Isotype Heterojunctions," *Solid-State Electron.,* 1985.

[3] J.L. Gray and M.S. Lundstrom, "Numerical Solution of Poisson's Equation with Application to C-V Analysis of Heterostructure Capacitors," *IEEE Trans. Electron Dev.,* Vol. ED-32, pp. 2102-2109, 1985.

[4] R.J. Schuelke, C.M. Maziar, and M.S. Lundstrom, "Open-Circuit Voltage Enhancement in Graded-Bandgap Solar Cells," *Solar Cells,* Vol. 15, pp. 73-78, 1985.

[5] C.M. Maziar, M.E. Klausmeier-Brown, S. Bandyopadhyay, M.S. Lundstrom, and S. Datta, "Monte Carlo Evaluation of Electron Transport in Heterojunction Bipolar Transistor Base Structures," *IEEE Trans. Electron Dev.,* Vol. ED-33, pp. 881-886, 1986.

[6] S. Datta, M. R. Melloch, S. Bandyopadhyay, R. Noren, M. Vaziri, M. Miller, and R. Reifenberger, "Novel Interference Effects Between Parallel Quantum Wells", *Phys. Rev. Lett.,* 55 (21), 2344, (1985)

[7] S. Datta, M. R. Melloch, S. Bandyopadhyay, M. S. Lundstrom, "Proposed Structure for Large Quantum Interference Effects", *Appl. Phys. Lett.,* Feb. 17, 1986.

[8] M. S. Lundstrom, "An Ebers-Moll Model for the Heterostructure Bipolar Transistor," *Solid-State Electron,* Vol. 29, pp. 1173-1179, 1986.

[9] S. Bandyopadhyay, M. E. Klausmeier-Brown, C. M. Maziar, S. Datta, M. S. Lundstrom, "Rigorous Technique to Couple Monte Carlo and Drift-Diffusion Models for Computationally Efficient Device Simulation," *IEEE Trans. Electron Dev.,* Vol. ED-34, pp. 392-399, 1987.

[10] M. Cahay, M. McLennan, S. Datta, and M. S. Lundstrom, "Importance of Space-Charge Effects in Resonant Tunneling Devices," *Applied Physics Letters,* Vol. 50, p. 612, 1987.

[11] C. M. Maziar and M. S. Lundstrom, "Caughey-Thomas Parameters for Electron Mobility Calculations in GaAs," *Electronics Letters,* Vol. 22, pp. 565-566, 1986.

[12] C. M. Maziar, M. E. Klausmeier-Brown and M. S. Lundstrom, "Proposed Structure for Collector Transit Time Reduction in AlGaAs/GaAs Bipolar Transistor," *Electron Device Lett.,* Vol. EDL-8, pp. 483-486, 1986.

[13] C. M. Maziar and M. S. Lundstrom, "On the Estimation of Base Transit Time in ALGaAs/GaAs Bipolar Transistors," *Electron Device Letters,* Vol. EDL-8, pp. 90-92, 1987.

[14] C. M. Maziar and M. S. Lundstrom, "Monte Carlo Simulation of GaAs Schottky Barrier Behavior," *Electronics Letters,* Vol. 23, pp. 61,62, 1987.

[15] S. Bandyopadhyay, C.M. Maziar, S. Datta and M.S. Lundstrom, "An Analytical Technique for Calculating High Field Transport Parameters in Semiconductors," *J. Appl. Phys.,* 60, 278-284 (1986).

[16] S. Bandyopadhyay, S. Datta and M.R. Melloch, "Aharonov-Bohm Effect in Semiconductor Microstructures - Novel Device Possibilities," *Superlattices and Microstructures,* 2, 539-542 (1986), also presented at the Second International Conference on Superlattices, Microstructures and Microdevices held at Göteborg, Sweden, August 17-20, 1986.

[17] S. Datta and S. Bandyopadhyay, "Aharonov-Bohm effect in semiconductor microstructures," *Phys. Rev. Lett.* 58, 717-720 (1987).

[18] A. Das and M.S. Lundstrom, "Numerical Study of Emitter-Base Junction Design for AlGaAs/GaAs Heterojunction Bipolar Transistors," *IEEE Trans. Electron Dev.,* Vol. ED-35, 1988.

[19] S. Datta, M. Cahay, and M. McLennan, "A Scatter Matrix Approach to Quantum Transport," *Phys. Rev. B.,* Vol. 36, pp. 5655-5658, 1987.

[20] M. Cahay, M. McLennan, and S. Datta, "Conductance of an Array of Elastic Scatterers: A Scattering Matrix Approach," *Phys. Rev. B.*, vol. 37, 10125 (1988).

[21] R. Frohne and S. Datta, "Electron Transfer between Regions with Different Confining Potentials," to appear in J. Appl. Phys.

[21] S. Datta, "Quantum Devices," to appear in *Superlattices and Microstructures.*

Invited Talks at SRC Companies:

[1] S. Datta, "Electron Transport in Small Devices," Kodak Research Laboratories, May 1985.

[2] M.S. Lundstrom, "Numerical Simulation of Bipolar Heterostructures," Kodak Research Laboratories," August 20, 1985.

[3] S. Datta, "A New Concept for a Quantum Interference Device," Texas Instruments, November 13, 1985.

[4] M.S. Lundstrom, "Bipolar Heterostructures: Device Physics and Modeling," Honeywell Solid State Electronics Division," December 17, 1985.

[5] M.S. Lundstrom, "Bipolar Heterostructures: Device Physics and Modeling," Sperry Corporation, Semiconductor Division, December 18, 1985.

[6] S. Datta, "A New Concept for a Quantum Interference Transistor," Bell Laboratories, Murray-Hill, April 10, 1987.

[7] S. Datta, "Quantum transport in ultrasmall devices," IBM T. J. Watson Research Center, Yorktown Heights, May 8, 1987.

Appendix C

SEQUAL RELEASE 2.1

The following pages summarize recent enhancements to the program SEQUAL (Semiconductor Electrostatics by QUantum AnaLysis). At the request of various users, some commands have been added to allow better control over the analysis. These additions are explained, and example calculations are presented, which illustrate the improvements.

1

SEQUAL Release 2.1

At the request of various users, several enhancements have been added to SEQUAL. The purpose of this document is to explain the additions, and to illustrate the use of new input keys in example calculations. Important changes are as follows:

- A card *kzgrid* was added to the input deck, allowing the user to specify the mesh in k_z-space (see SEQUAL User's Manual, p. 8). Although the automatic (default) mesh is recommended for most purposes, it is sometimes desirable to "zoom-in" on a particular region of the transmission coefficient in energy space; an example of this use is presented near the end of this document.

- A key *itvar* was added to the *solve* card, allowing the user more control over self-consistent calculations. Previously, the current density was used to determine convergence: When the required number of significant figures had been attained (specified by *prec* on the *solve* card), iteration was terminated. For equilibrium calculations, an exact cancellation of oppositely flowing currents is difficult to achieve. In this case, the current density should not be used to control convergence. Furthermore, the current density is identically zero for bound state calculations, and therefore provides no information. The new key *itvar* allows the user to specify which program variable, current density or electrostatic potential (or both), should determine convergence.

- For SEQUAL 2.0, results of a self-consistent calculation were written only upon convergence of the final solution. Because the execution time for a self-consistent calculation can be quite long, the partial results from each iteration are now written (and over-written), according to the output requests of the user (see card *output* in the user's manual).

- Previously, it was assumed that starting position for any device (i.e., the first value specified for "z" in an input file) was zero; this restriction has been removed. For many practical devices, an "interesting" region of the

potential is surrounded by large regions over which the potential is constant. Because the length of a device can substantially affect the execution time of SEQUAL, it is prudent to trim the output from classical analysis programs. It is therefore convenient that the position-space grid be allowed to start from any value.

- A (small) bug in the solution of Poisson's equation (subroutine POISS) was corrected. This particular problem occurred **only** when the electron density was zero at some node, since a necessary logarithm became undefined. This problem prevented a self-consistent calculation of bound states, because the bound state wave function vanishes at the ends of the device. An example of a self-consistent calculation for the bound states of a AlGaAs/GaAs interface is presented near the end of this document.

The remainder of this update contains the following. Changes necessary to upgrade SEQUAL for use with IMSL **10.0** are presented below. A summary of the syntax for the *kzgrid* and *solve* cards is presented, followed by two example problems illustrating their use. Example output is presented for analyses of a double-barrier resonant tunneling device, and an AlGaAs/GaAs interface.

Conversion to IMSL 10.0

Solution of the bound states in SEQUAL 2.1 requires a solution of the eigensystem for a real general matrix. A subroutine in the IMSL library is used for this purpose.[†] SEQUAL 2.1 is released to be compatible with IMSL 9.2, although the most current version of IMSL is 10.0. To upgrade SEQUAL for use with this version, the following changes must be made:

[†] If IMSL is unavailable, any comparable eigensystem solver could be substituted in the subroutine BSTATE. If no such alternative is available, the bound state solution must be removed from the program (see the Sequal User's Manual, p. 59).

- Immediately before the first executable line in the main program, insert the following:

```
c
c zzz zzz zzz  MACHINE DEPENDENT CODE  zzz zzz zzz
c
c If IMSL version 10.0 is not available, comment out
c the following lines:
c
      common /worksp/  rwksp
      real rwksp(126022)
c
c ---------------------------------------------------------------
c Create workspace for IMSL eigenvalue solver...
c ---------------------------------------------------------------
c
      if (imsl) then
        call iwkin(126022)
      endif
```

- At the start of subroutine BSTATE, add:

```
      external evcrg, epirg
      real epirg
```

- In subroutine BSTATE, the call to IMSL subroutine "eigrf" must be changed to the following **two** IMSL calls:

```
c
c  zzz zzz zzz  MACHINE DEPENDENT CODE  zzz zzz zzz
c
c  If IMSL routines are unavailable, comment out the
c    following 2 lines, so that SEQUAL can be compiled:
c
      call evcrg(nodes,matrix,zimax,val,vec,zimax)
      bsperf = epirg(nodes,nodes,matrix,zimax,val,vec,zimax)
```

kzgrid specify k_z-grid for propagating states

Default State:

kzgrid auto=true from=0.0 to=0.0 steps=1

auto The key *auto* is a switch controlling the k_z-grid specification. If true, the automatic (default) grid is selected, and all other *kzgrid* keys are ignored. If false, the remaining keys specify the k_z-grid. Points in the user-specified grid are uniformly distributed in k_z-space, and form the basis for isolating wavefunction maxima (see Sequal User's Manual, p. 8). Because the user-specified grid may not always be sufficiently dense, the automatic grid is highly recommended for typical calculations.

from

to The energy range for the user-defined k_z-space grid is specified by the values assigned to *from* and *to*, in units of eV. For a given computation, only a single energy range is allowed; however, results from several energy ranges can be obtained by multiple assignments to these two keys.

steps The interval of k_z-space corresponding to the energy range defined by *from* and *to* is broken into an equal number of intervals, according to the value assigned to *steps*. As for the automatic grid, the step size is uniform with respect to k_z, **not** energy. Both the user-defined and automatic grids form a skeletal grid, to which k_z-nodes corresponding to wave function maxima are added. It is this composite grid which is used for output of the transmission coefficient. Hence, the number of points on this grid may be larger than the value assigned to *steps*, and the spacing (even in k_z-space) may not always be uniform.

solve — specify parameters controlling the analysis

Default State:

solve prec=3 itmax=9 itvar=jv states=prop inject=both

prec The number of significant figures desired for important quantities can be assigned to the *prec* key of the *solve* card. "Important" quantities include the electron density, the current density, and the Fermi-level in each contact. Since SEQUAL will struggle admirably (at the expense of CPU time) to achieve whatever precision is specified, the value should be kept within reasonable limits.

itmax For a self-consistent analysis, SEQUAL solves iteratively for the electrostatic potential and the electron density. Convergence is achieved when the number of significant figures in some quantity (between iterations) settles down to the number requested. To avoid excessive use of CPU time in obtaining convergence, SEQUAL terminates iteration when the number of iterations exceeds the value assigned to *itmax*. Therefore, iteration can be suppressed by specifying zero as the maximum number of iterations. If results are written to output files (see description of the *output* card), iteration can be continued at any point, by using the output of a previous run as the input device description file (see Figure 4.2 in the Sequal User's Manual).

itvar Self-consistent calculations are terminated when the quantity specified by *itvar* achieves the requested precision. This quantity can be current density, electrostatic potential or both:

Value	Quantity which determines convergence
j	Current density
v	Electrostatic potential
jv, vj	Current density **and** electrostatic potential

For bound state and equilibrium calculations, current density is zero, and the electrostatic potential is the recommended quantity for determining convergence.

states Both propagating and bound electronic states can be considered in the solution of the Schrödinger equation, according to the value assigned to *states*:

Value	Implication
prop	Consider propagating electronic states
bound	Consider bound electronic states
all	Consider propagating and bound states

The default is to consider only the propagating states, since bound states do not contribute to current density. Furthermore, the population of bound states is correct only for devices in equilibrium. For small deviations from equilibrium, however, the bound-state result obtained is a reasonable approximation to the correct solution. The consideration of bound states is particularly important for a self-consistent solution. Because the electrostatic potential is determined from the electron density, the bound-state contribution (even if it is only approximate) can significantly alter the final result.

inject In the solution of Schrödinger's equation for propagating states, electrons can be injected into the device from two contacts. Each contact provides a separate contribution to both electron density and current density. The value assigned to *inject* determines which of the contributions will be calculated:

Value	Implication
l-to-r	Consider electrons propagating from left to right (inject from left contact)
r-to-l	Consider electrons propagating from right to left (inject from right contact)
both	Inject from both contacts

Because the vast majority of devices require an analysis with injection from both contacts, the default value is "both." In special cases,

however, the contribution from one contact may be insignificant. Consider, for example, the calculation of electron current for a p-n junction. Injection from the p-type contact is unnecessary, since the current component would be negligible.

2

Example 1

The following represents a typical example use of the *kzgrid* card. A double-barrier resonant tunneling structure is analyzed with the default k_z-space grid, and the resulting transmission coefficient is shown in Fig. 1 below. Squares mark the actual points output from SEQUAL. These points are a combination of the uniform (automatic) grid and extra k_z-points corresponding to wave function maxima. Near the first peak in the transmission coefficient, the k_z-grid appears to be a bit sparse. SEQUAL uses considerably higher resolution for integration of the wavefunction, although these additional points are not printed out for the transmission coefficient. Therefore, to increase the output resolution of the transmission coefficient, we are left with two alternatives. The first is to reduce the automatic step size by assigning a fraction to the *kscale* parameter (see detailed discussion of the *maxima* card in the Sequal User's Manual). Although this solves the problem, it is a tremendous waste of CPU time, since we are only interested in increasing the resolution in a narrow band of energy. Instead, we add a *kzgrid* card to original input deck, and obtain the output attached on the following pages. A plot of the transmission coefficient with enhanced resolution is shown in Fig 2.

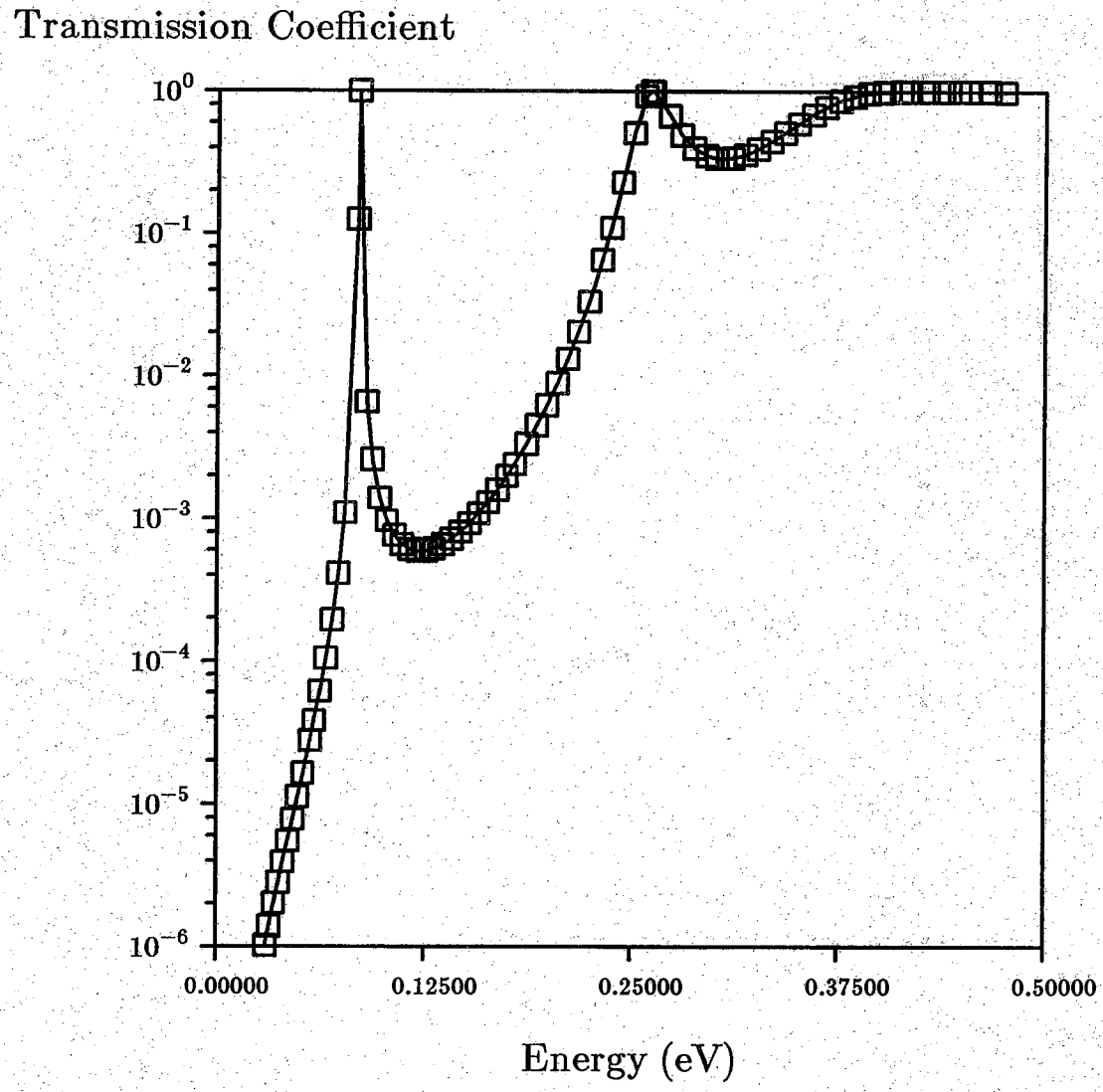


Fig. 1 A plot of transmission coefficient for a resonant tunneling structure. Squares (□) mark the grid points obtained from an analysis using the default (automatic) grid.

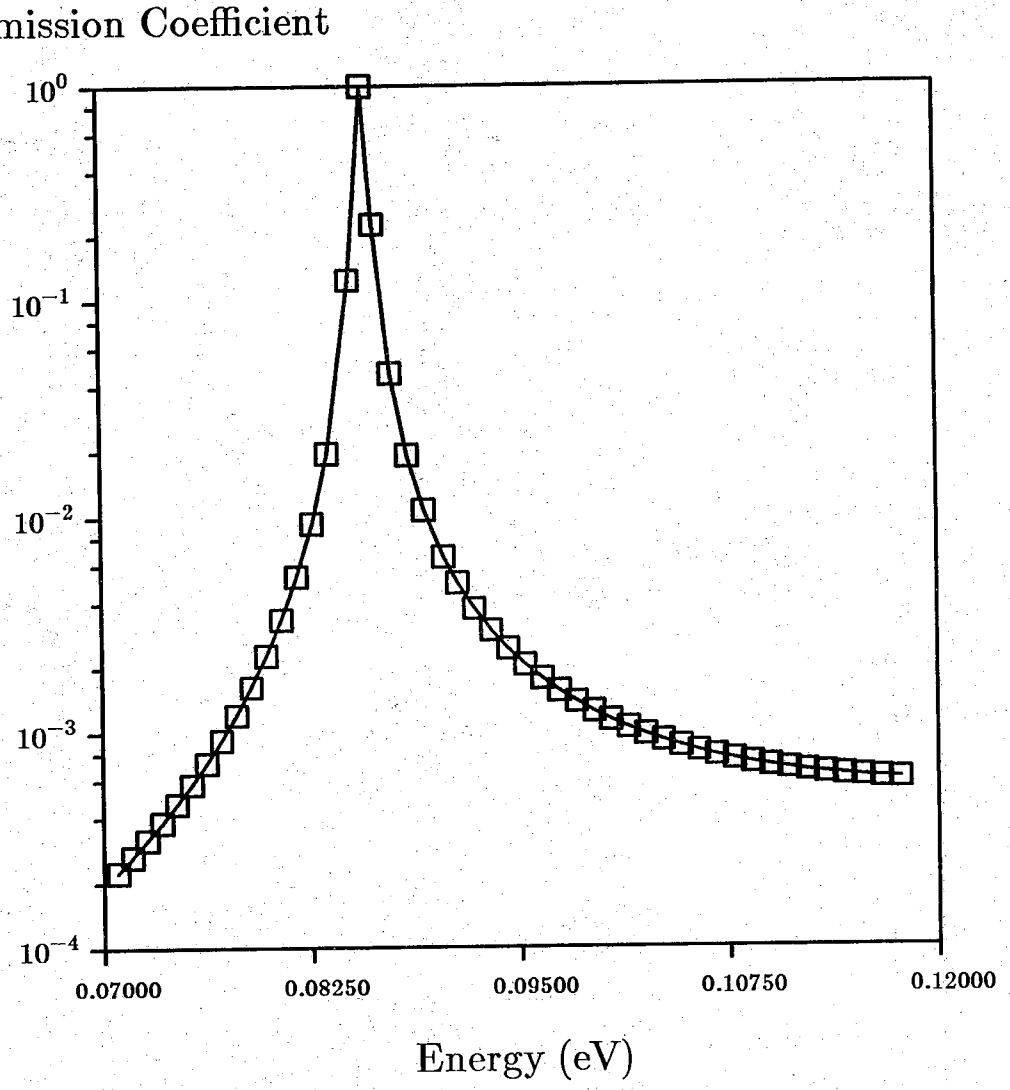
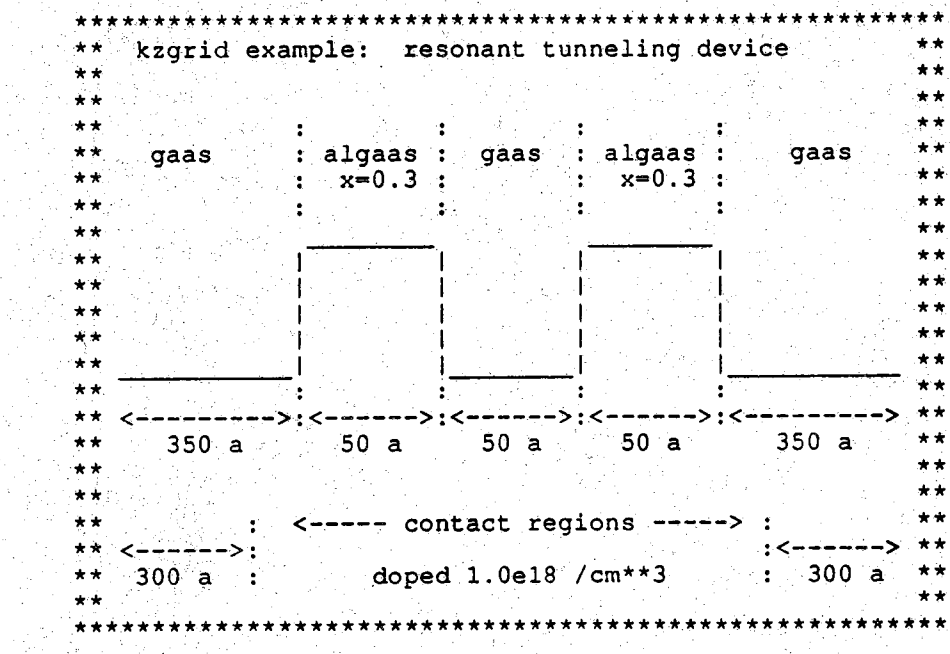


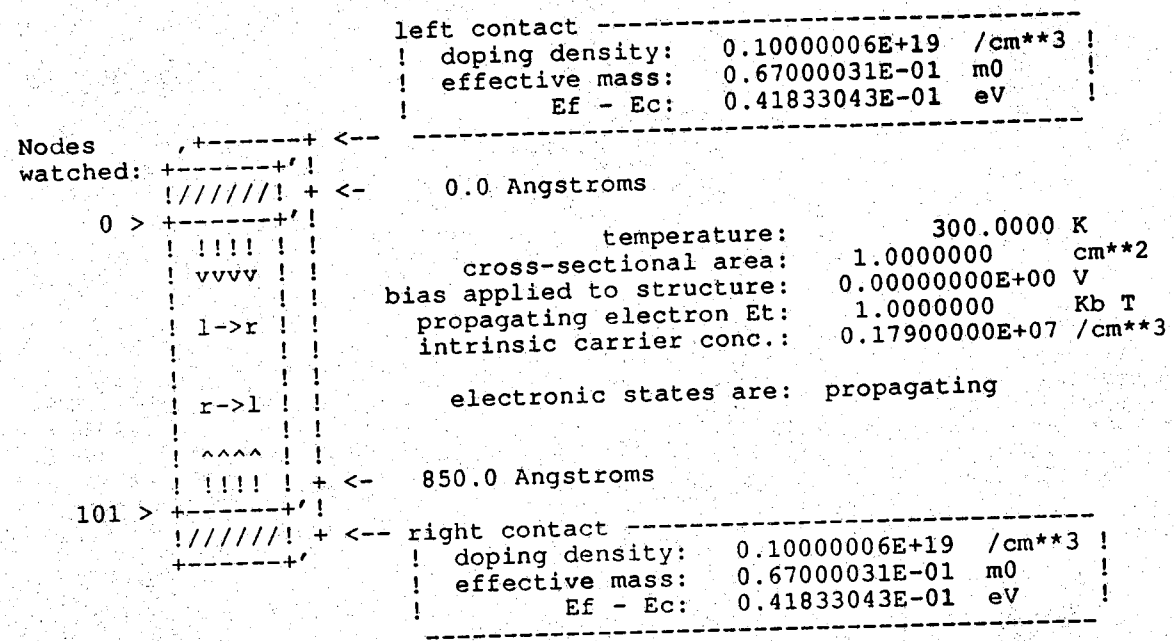
Fig. 2 A plot of transmission coefficient in a small energy range near the first resonance. Squares (□) mark the grid points obtained from an analysis with a user-specified grid.

```
------- XXXX  XXXXX   XXXX   XX  XX   XXXX   XX -----------------------------
      XX     XX     XX  XX  XX  XX  XX  XX  XX                      SEQUAL 2.1
----- XXX   XXXX   XX  XX  XX  XX  XXXXXX  XX ------------- Purdue University
       XX  XX     XX  X   XX  XX  XX  XX  XX                      August 1988
-- XXXX   XXXXX   XXX X   XXXXX  XX  XX  XXXXXX -----------------------------

SEQUAL:  input deck
-------------------------------------------------------------------------------
```



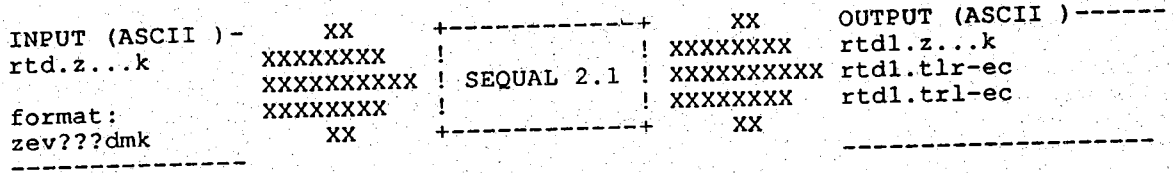

```
title   example of manual kzgrid selection

    >>>>  use output of previous self-consistent analysis  <<<<
    >>>>   for input in present calculation.               <<<<

input   file=rtd.z...k  format=zev???dmk

device  temp=300.0  bias=0.0
solve   itmax=0  prec=3  states=prop

    >>>>  select manual kz-space grid to zoom in on first   <<<<
    >>>>   quasi-bound state.                               <<<<

kzgrid  auto=false  from=0.07  to=0.12  steps=50

print   format1=* format2=* verbose=true
output  file=rtd1  data=dt
-------------------------------------------------------------------------------

Executing a total of  1 calculation(s).
```

example of manual kzgrid selection



PROPAGATING Kz-GRID: user-defined energy window:

From energy:	0.70000052E-01	eV
To energy:	0.12000006	eV
Number of steps:	50	



```
SEQUAL 2.1                                                        page   3
calculation  1 of  1                          Statistics and Iteration Data
-------------------------------------------------------------------------------
example of manual kzgrid selection
===============================================================================
```

WAVEFUNCTION FOR PROPAGATING ELECTRONS: Use of kz-space nodes

Iteration:	# 0
kz-space nodes ...	
used in isolating maxima:	73
used in integration:	288
miscellaneous:	108
TOTAL:	469

WAVEFUNCTION FOR PROPAGATING ELECTRONS: Integration concerns

Iteration:	# 0
kz-space maxima found:	6
Average maxima separation / kz-step	
... left-to-right:	3.6942
... right-to-left:	3.6958
Number of kz-space intervals integrated:	12
Gauss-Legendre integration	
highest order:	10
lowest order:	6
average order:	7

example of manual kzgrid selection

TYPE	REMARK	SUGGESTION / EXPLANATION
Warning	Using a manual kz-grid for injection of electrons in propagating states.	Electron/current density obtained may be a fraction of the proper value.

```
SEQUAL 2.1                                                        page   5
calculation  1 of  1                                        Final Results
-------------------------------------------------------------------------------
example of manual kzgrid selection
===============================================================================
```

Current:		Voltage:	
LEFT-to-RIGHT:	-0.490938E+04 A	In Input File:	0.000000E+00 V
RIGHT-to-LEFT:	-0.490959E+04 A	Applied Bias:	0.000000E+00 V
Total Current:	0.210938E+00 A	Total Voltage:	0.000000E+00 V

3

Example 2

As a demonstration of the *itvar* key, we perform a self-consistent analysis of bound states at an AlGaAs/GaAs interface. Because we are considering only bound states, current density is zero. In this case, the electrostatic potential alone should determine convergence. To specify this, we include "*itvar*=v" on the *solve* card. Output for this example is presented on pages that follow. (For a schematic view of the device, see output page 1.)

The calculated sheet density for electrons confined near the interface is $n_s = 1.74 \times 10^{12}$ cm^{-2}. It is important to recognize that this density includes electrons in *both* AlGaAs and GaAs layers, as illustrated in Figs. 3 and 4. Fig. 3 presents the self-consistent conduction band profile, and Fig. 4, the corresponding electron density, for a region of the structure near the interface. Although typically the AlGaAs layer is assumed to be depleted, in this case a substantial amount of the sheet density n_s can be attributed to electrons in the AlGaAs layer.

Note that, according to SEQUAL, the device appears to be biased (see output pages 5 and 8 on the following pages). This is because the input structure was taken from FISH1D, a semiclassical heterojunction analysis program. In FISH1D, all valleys (including Γ and X) are taken into account to determine the contact Fermi potentials, whereas only the Γ-valley is considered in SEQUAL. Therefore, a "bias" appears across the structure in SEQUAL, although the same structure is in equilibrium for FISH1D. The "bias" can be removed in SEQUAL by applying additional bias to counteract it. However, in this example that is unnecessary. Bound states are populated according to the Fermi level of the **back** contact, so this false "bias" can be ignored.

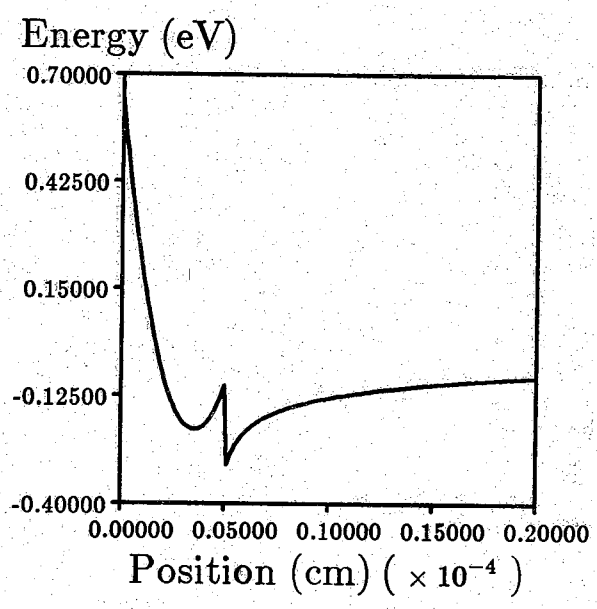


Fig. 3 The self-consistent conduction band diagram for an AlGaAs/GaAs interface, obtained by considering only bound states.

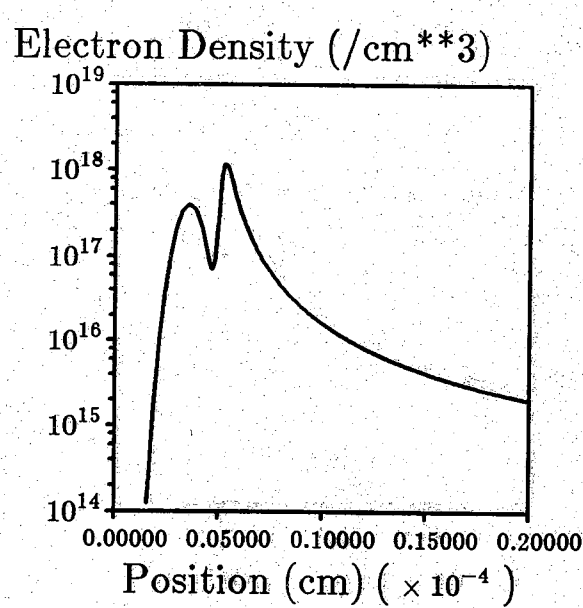
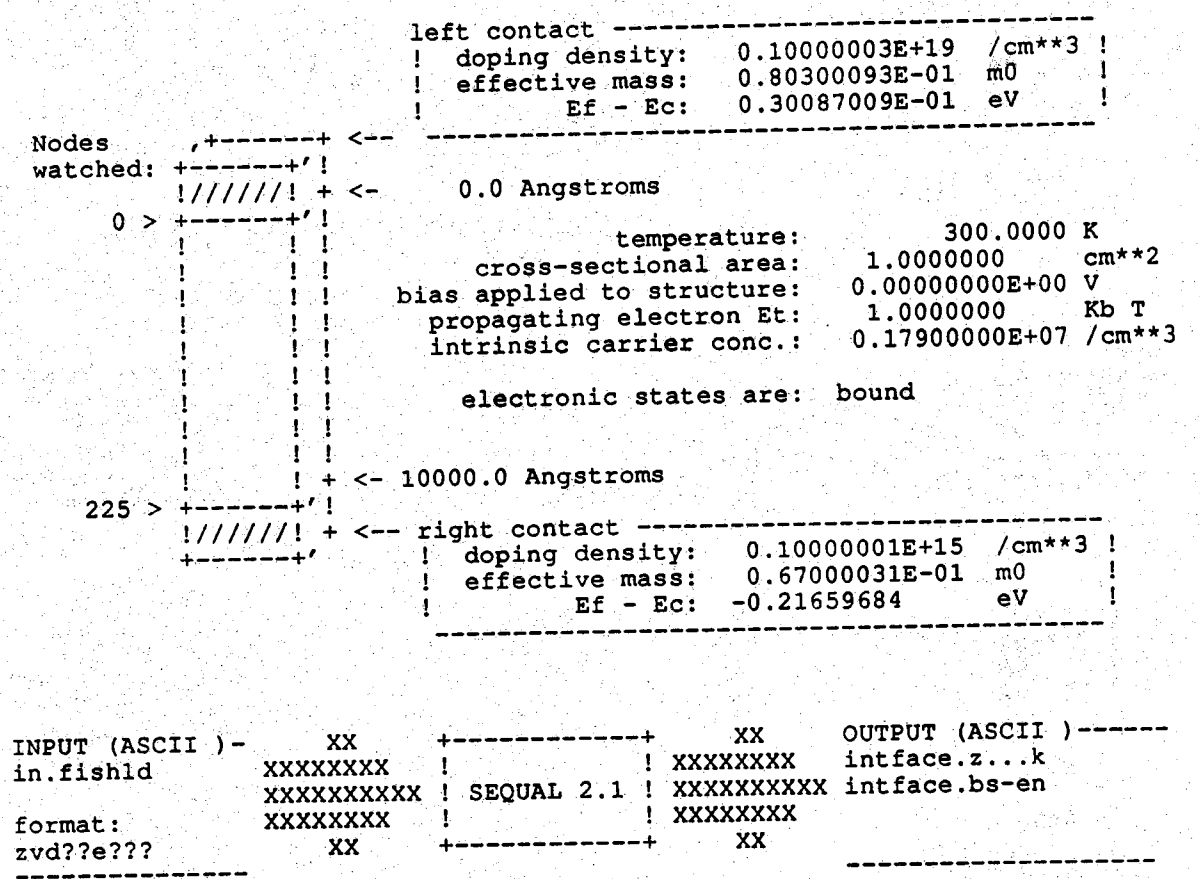


Fig. 4 The total electron density in all bound states, for the region of the conduction band shown above. Notice that the electron density in the AlGaAs layer (left hump) is substantial.

```
------- XXXX  XXXXX   XXXX   XX  XX   XXXX   XX ------------------------------
      XX     XX     XX  XX  XX  XX  XX  XX  XX                    SEQUAL 2.1
----- XXX   XXXX   XX  XX  XX  XX  XXXXXX  XX ------------- Purdue University
       XX  XX     XX  X   XX  XX  XX  XX  XX                    August 1988
-- XXXX   XXXXX   XXX X   XXXXX  XX  XX  XXXXXX ------------------------------

SEQUAL:  input deck
-------------------------------------------------------------------------------


   ********************************************************
   **  bound state example:  algaas / gaas interface     **
   **                                                    **
   **   -------------------------                        **
   **  |    schottky barrier     | <-- phi = 0.86        **
   **  |-------------------------| - - -                 **
   **  |     algaas (x=0.3)      |   ^                   **
   **  |                         |   |  500 angst        **
   **  |   nd = 1.0e18 /cm**3    |   v                   **
   **  |-------------------------| - - -                 **
   **  |          gaas           |   ^                   **
   **  |                         |   |  9500 angst       **
   **  |   nd = 1.e14 /cm**3     |   v                   **
   **   -------------------------  - - -                 **
   **                                                    **
   ********************************************************

title   bound-state example:  algaas/gaas interface

    >>>>  output from fish1d has position in microns.   <<<<
    >>>>  specify scale factor for proper length scale. <<<<

input   file=in.fish1d  format=zvd??e???
scale   cm=1.e4

matter  nodes=50/225  emass=0.0803/0.067  krel=11.9/12.847
doping  nodes=50/225  nd+=1.e18/1.e14

device  temp=300.0  bias=0.0

    >>>>  for bound states, there is no current.  set itvar=v  <<<<
    >>>>   to use convergence of the electrostatic potential   <<<<
    >>>>   as the criterion for a self-consistent solution.    <<<<

solve   itmax=10  prec=3  states=bound  itvar=v

print   format1=* format2=* verbose=true
output  file=intface  data=db
-------------------------------------------------------------------------------

Executing a total of  1 calculation(s).
```

```
SEQUAL 2.1                                                          page   2
preliminary information                          Summary of Input Information
-------------------------------------------------------------------------------
bound-state example:  algaas/gaas interface
===============================================================================

For all calculations, the following assumptions apply,
  regardless of data read from input file(s).

     ,+-------+ node   0
    +-------+'!                          effective mass =  0.803001E-01 m0
    !       ! !                 dielectric constant =   11.9000      e0
    !       ! + node  50
    +-------+'+ node  51
    +-------+'!                          effective mass =  0.670000E-01 m0
    !       ! !                 dielectric constant =   12.8470      e0
    !       ! + node 225
    +-------+'


     ,+-------+ node   0
    +-------+'!                donor doping density =  0.100000E+19 /cm**3
    !       ! + node  50
    +-------+'+ node  51
    +-------+'!                donor doping density =  0.100000E+15 /cm**3
    !       ! + node 225
    +-------+'
```

bound-state example: algaas/gaas interface



SEQUAL 2.1 page 4
calculation 1 of 1 Statistics and Iteration Data

bound-state example: algaas/gaas interface

Iteration Number	Current (A)	Significant Figs in Current	Max Change in V(z)	Significant Figs in V(z)
1	0.0000000E+00	3	0.946E-02	1
2	0.0000000E+00	3	0.197E-02	2
3	0.0000000E+00	3	0.509E-03	3

bound-state example: algaas/gaas interface

TYPE	REMARK	SUGGESTION / EXPLANATION
Warning	Obtaining convergence in a self-consistent calculation may be difficult.	Contacts should be more heavily doped than the internal device structure.
Caution	Population of bound-states may be incorrect, since device is under bias.	Device must be in (or near) equilibrium for Fermi-Dirac statistics to apply.

SEQUAL 2.1 page 6
calculation 1 of 1 Bound-State Calculation

bound-state example: algaas/gaas interface

NOTE: Confidence in the bound-state solution is high

Bound-State Energy (eV)	Electron Dens. (/cm**2)	Occupation (% of total)
-0.23346710	0.78350470E+12	45.018539
-0.19914645	0.35672346E+12	20.496582
-0.18572903	0.19334090E+12	11.108964
-0.16791409	0.12185246E+12	7.0013876
-0.15970147	0.77001327E+11	4.4243374
-0.14297843	0.41610449E+11	2.3908491
-0.13474482	0.34421989E+11	1.9778156
-0.12861872	0.24305697E+11	1.3965540
-0.11890638	0.16508072E+11	0.94851923
-0.11065400	0.12072239E+11	0.69364566
-0.10395455	0.95209677E+10	0.54705477
-0.99294603E-01	0.83339674E+10	0.47885251
-0.95357955E-01	0.69041439E+10	0.39669758
-0.90595961E-01	0.55891395E+10	0.32114017
-0.85954428E-01	0.46443397E+10	0.26685411
-0.81701398E-01	0.39352571E+10	0.22611165
-0.77840924E-01	0.33894725E+10	0.19475198
-0.74338734E-01	0.29628815E+10	0.17024100
-0.71162641E-01	0.26251244E+10	0.15083414
-0.68283856E-01	0.23555379E+10	0.13534427
-0.65848231E-01	0.21486211E+10	0.12345529
-0.63814104E-01	0.20005722E+10	0.11494869
-0.61324317E-01	0.18459904E+10	0.10606676
-0.59660703E-01	0.16749463E+10	0.96238911E-01
-0.58285769E-01	0.16415347E+10	0.94319165E-01
-0.55165745E-01	0.14038211E+10	0.80660641E-01
-0.54835804E-01	0.14185697E+10	0.81508040E-01
-0.51128648E-01	0.12003448E+10	0.68969309E-01
-0.50750017E-01	0.12002565E+10	0.68964243E-01
-0.47476452E-01	0.10425728E+10	0.59904065E-01
-0.46183787E-01	0.10021391E+10	0.57580810E-01
-0.44137236E-01	0.91646848E+09	0.52658349E-01
-0.41250527E-01	0.82743885E+09	0.47542900E-01
-0.41059058E-01	0.81377587E+09	0.46757862E-01
-0.38195815E-01	0.72830413E+09	0.41846838E-01
-0.36037434E-01	0.67760666E+09	0.38933869E-01
-0.35514742E-01	0.65661005E+09	0.37727434E-01
-0.32983769E-01	0.59524966E+09	0.34201797E-01
-0.30652590E-01	0.55390106E+09	0.31825993E-01
-0.30578151E-01	0.54215910E+09	0.31151343E-01
-0.28274480E-01	0.49596058E+09	0.28496861E-01
-0.26054014E-01	0.45527603E+09	0.26159197E-01
-0.25317345E-01	0.45730099E+09	0.26275564E-01
-0.23898955E-01	0.41872691E+09	0.24059173E-01
-0.21793514E-01	0.38631347E+09	0.22196755E-01

bound-state example: algaas/gaas interface

Bound-State Energy (eV)	Electron Dens. (/cm**2)	Occupation (% of total)
-0.20363860E-01	0.38428851E+09	0.22080414E-01
-0.19726913E-01	0.35665690E+09	0.20492751E-01
-0.17682794E-01	0.32975872E+09	0.18947251E-01
-0.15651178E-01	0.30423834E+09	0.17480899E-01
-0.15441049E-01	0.31453030E+09	0.18072255E-01
-0.13621282E-01	0.28147712E+09	0.16173095E-01
-0.11583120E-01	0.26009414E+09	0.14944464E-01
-0.97666979E-02	0.24839613E+09	0.14272321E-01
-0.95288977E-02	0.24009014E+09	0.13795078E-01
-0.74497312E-02	0.22146555E+09	0.12724947E-01
-0.53382628E-02	0.20421982E+09	0.11734042E-01
-0.34537381E-02	0.19271066E+09	0.11072751E-01
-0.31791739E-02	0.18766365E+09	0.10782760E-01
-0.81478781E-03	0.17179670E+09	0.98710805E-02
TOTAL:	0.17404044E+13 /cm**2 in 59 bound-states	

```
SEQUAL 2.1                                                        page   8
calculation  1 of  1                                        Final Results
-------------------------------------------------------------------------
bound-state example:  algaas/gaas interface
=========================================================================

Current:                                  Voltage:
-----------------------------------       -----------------------------------
   LEFT-to-RIGHT:   0.000000E+00 A           In Input File:  -0.888461E+00 V
   RIGHT-to-LEFT:   0.000000E+00 A            Applied Bias:   0.000000E+00 V
                    --------------                            --------------
   Total Current:   0.000000E+00 A          Total Voltage:   -0.888461E+00 V
```