7-1-1987

Physics and Modeling of Submicron Devices

Supriyo Datta
Purdue University

Mark S. Lundstrom
Purdue University

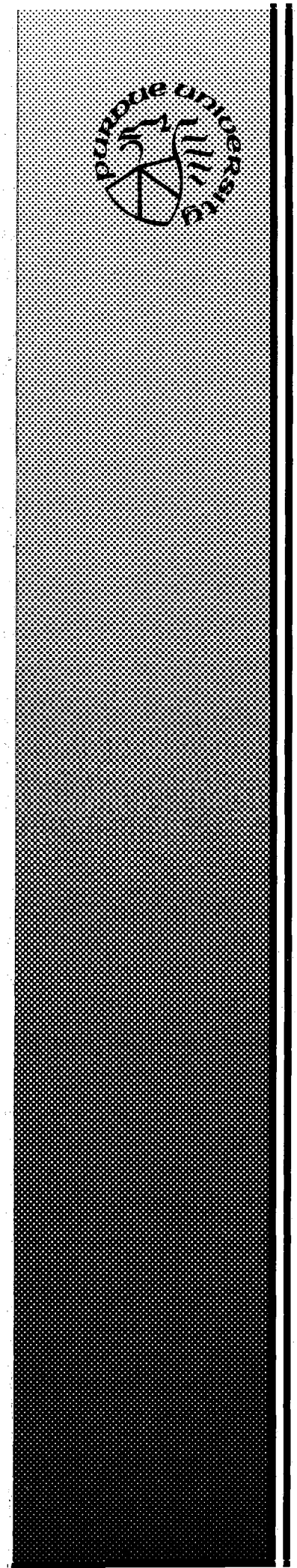
Marc Cahay
Purdue University

Amitava Das
Purdue University

H. Rob. Frohne
Purdue University

See next page for additional authors

Authors
Supriyo Datta, Mark S. Lundstrom, Marc Cahay, Amitava Das, H. Rob. Frohne, and Michael J. McLennan



Physics and Modeling of Submicron Devices

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

TR-EE 87-35
Annual Report: June 1, 1986 - July 31, 1987

School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907

Supported by the Semiconductor Research Corporation Contract Number 86-07-089

PHYSICS and MODELING

of

SUBMICRON DEVICES

Annual Report: June 1, 1986 - July 31, 1987
Supported by the Semiconductor Research Corporation
Contract Number 86-07-089

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

Purdue University
School of Electrical Engineering
Technical Report: TR-EE 87-35
West Lafayette, Indiana 47907

PREFACE

The work described in this report is directed at understanding transport physics in sub-micron heterostructure devices, at developing computational techniques for modeling such devices, and at applying these techniques to investigate new device concepts. The focus of the past year's work has been on extending our collisionless, quantum device models to treat elastic scattering processes and at applying previously-developed models to the design and study of AlGaAs/GaAs heterojunction bipolar transistors. This report describes the past year's progress in these two areas. 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

1. PROJECT OVERVIEW

1.1 Introduction

As electronic devices shrink to submicron dimensions, hot carrier effects and quantum mechanical effects are becoming increasingly important. Both effects offer the potential for improving device performance. By exploiting hot carrier effects, the speed of small bipolar and field-effect transistors may be improved significantly; quantum mechanical effects on the other hand offer the potential to realize an entirely new class of "post-shrink" devices whose performance may be orders of magnitude better than present-day devices. The objective of our work is the exploration of new, high-performance device concepts based on hot carrier and quantum transport. The work centers on the development and application of advanced device simulation programs, which will be needed to guide the design of future devices. Specific objectives of the research program are: 1) initiation of work 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 numerical device simulation programs being developed during the course of this work comprise a "tool box" that can be applied to the study and design of advanced devices.

1.2 Technical Approach

Device simulation programs are widely used in industry for the optimization of devices and for exploring new device concepts. 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 will not be 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 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 Schroedinger equation. This can be understood by noting that the relationship between Newton's law and Schroedinger'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.

During 1983-1986, our SRC-supported work focussed on the simulation of hot-electron effects in bipolar transistors using the Monte Carlo technique. This work provided insight into the nature of transport in a bipolar context and clarified some of the limitations of the drift-diffusion approach. Although the physics of hot carrier transport is now clearly 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 existing simulation tools. Work to explore new simulation strategies which accurately describe transport and recombination in a bipolar context is also underway.

Although the hot carrier work is important and continuing, the current emphasis of our work is on *quantum mechanical effects*. Compared to hot-electron effects, our understanding of quantum effects is primitive. Numerous approaches have been proposed by different workers in the field and it is not yet clear which of these

approaches will eventually provide an accurate but tractable description for quantum transport in submicron devices. Our 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.

1.3 Overview of the Report

During the past year, we have completed work on SEQUAL, our first quantum mechanical device model. SEQUAL treats the collisionless propagation of electron waves, along with their self-consistent electrostatic potential, in one-dimensional devices. Development of the program is now complete; it is available to SRC members and is described in the Appendix to this report. Application of this program to evaluate electron injection currents in heterojunction bipolar transistors is discussed in Chapter 5 of this report. We also successfully extended such calculations to simple two-dimensional structures during the past year. These calculations have been applied to the propagation of electrons from a reservoir to a well which occurs, for example, between the source and channel of a field-effect transistor. The technique and preliminary results are described in Chapter 4.

During the past year, the collisionless model has been extended to include elastic scattering mechanisms such as ionized impurity scattering. The technique is described in Chapter 2, and it is applied to simple, quantum size resistors and to a proposed quantum interference device in Chapter 3. Though much work remains to be done in extending and applying these techniques to realistic structures and devices, no major conceptual hurdles are foreseen as long as the scattering mechanisms are purely elastic. The major conceptual hurdle lies in the development of a quantum device model that includes *inelastic* scattering such as phonon scattering. Several formalisms for treating the general problem of quantum transport have been proposed; however, at this stage it is impossible to foresee which approach will eventually lead to practical device models and lend new insight into the physics of these novel structures. The accurate treatment of quantum transport, in the presence of elastic and inelastic scattering and with a self-consistent electric field is the central challenge of the research program.

1.4 Status of Purdue Device Simulation Programs

While the development of device simulation software is not the central objective of the research, when such simulation programs are developed we attempt to make them machine-independent, user-friendly, and available to SRC members. Several such programs have been developed; they are largely machine-independent and are supplied with user's manuals. Each program is based on a different simulation strategy; each has its limitations, but when intelligently used the collection of programs encompasses many of the effects important in modern devices. The development of a comprehensive, global device simulation strategy is the key objective of the project. A brief description of the available device simulation programs follows.

FISH1D: solves Poisson's equation in compositionally nonuniform semiconductors (program supplied with materials parameters for AlGaAs). The equilibrium solution and the solution under bias (assuming zero current or constant quasi-Fermi levels) is computed. Allows for partial ionization of dopants, degenerate carrier statistics, and nonparabolic, multiple conduction bands. Plots the electrostatic potential and field, carrier densities, energy band diagrams and other quantities of interest. Computes C-V characteristic and differentiates it to plot apparent carrier density versus position. The program is written in FORTRAN 77 runs and makes use of IMSL routines.

PUPHS(1D/2D): solve Poisson's equation simultaneously with the electron and hole continuity equations for compositionally nonuniform semiconductors in equilibrium or under bias. PUPHS is supplied with materials models for AlGaAs. The programs compute I-V and C-V characteristics of diodes and analyze solar cells and bipolar transistors. A separate program reads PUPHS-generated data and plots quantities of interest. Both one- and two-dimensional versions of this program are available; they are written in FORTRAN 77, and use IMSL and LINPACK routines.

DEMON: performs 1-D Monte Carlo simulation of electron transport in compositionally nonuniform devices. DEMON is supplied with materials

models for AlGaAs. User-defined field and doping profiles may be defined. A separate program plots the average carrier density, velocity, and energy versus position within the device. Histograms of the distribution function within the device may also be requested. DEMON is written in FORTRAN 77.

SEQUAL: A post-processor which, given the energy band diagram of a one-dimensional device, computes the electron current by assuming collisionless propagation of electron waves. If so directed, SEQUAL will iterate and solve Poisson's equation self-consistently. SEQUAL is written in FORTRAN 77 and makes use of IMSL libraries.

Inquiries regarding these computer simulation programs should be directed to M.S. Lundstrom at Purdue.

SRC-Supported Publications

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-86-31, 1986.

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.

Book Chapters:

[1] S. Datta, "Quantum Interference Devices," to appear in 'Physics of Quantum Electron Devices' ed. by F. Capasso, to be published by Springer-Verlag.

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," to be 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," to be 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 (also to appear in *Surface Science).*

[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; (also to appear in *J. Vac. Sci. Tech.).*

[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) with 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," to be presented at the Third International Conference on Superlattices, Microstructures and Microdevices, Chicago, August 17-20, 1987.

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] S. Datta, M. Cahay and M. McLennan, "A Scatter Matrix Approach to Quantum Transport," to appear in Phys. Rev. B.

2. A SCATTER MATRIX APPROACH TO QUANTUM TRANSPORT

2.1 Introduction

The problem of Anderson localization in two-dimensional disorder systems has been extensively studied in the past using the tight-binding Hamiltonian formalism. In this chapter, we study the problem of Anderson localization at zero temperature for a piece of *dirty* material of finite length L and width W in which the transport of electrons is modeled as regions of free propagation with occasional elastic scattering by a random array of scatterers. Each impurity is characterized by a scatter matrix which can in principle be derived for any arbitrary scattering potential, the randomness being introduced through the impurity location. The overall transmission is calculated by cascading the scatter matrices of individual scatterers and the conductance is evaluated using the Landauer formula for the case of multi-moded transport.

Numerical examples illustrating the onset of the weak and strong localization regimes as well as universal conductance fluctuations are obtained for the appropriate choice of the parameters of the model. Deviation from the universal behavior for sample length bigger than the localization length and shorter than the electron elastic mean free path are also investigated. We also show that the universal conduction result is roughly independent of the number of propagating modes throughout the sample. The technique used for cascading scatter matrices facilitates the comparison between the classical and quantum-mechanical calculation of the conductance of a given sample because the classical conductance can be deduced by considering probability scatter matrices rather than amplitude scatter matrices. Furthermore, we stress the applicability of our calculations to study the transport properties of ultra-small semiconductor devices in which the total number of propagating modes at the Fermi level can be around a few tens only. This is a complete different regime than the diffusion regime applicable to metals in which the number of propagating modes is increased by several order of magnitude.

Over the past few years, many experiments dealing with the study of electron transport in disordered submicron devices have been reported in favor of the theory of localization of Abrahams, Anderson, Licciardello and Ramakrishnan [1]. More recently, the presence of universal conductance fluctuations in disordered systems has also been established both experimentally [2-6] and theoretically [7-12]. Such fluctuations have a universal magnitude $\sim e^2/h$, independent of the sample size dimensions, provided the system behaves as a good conductor, i.e, $G > e^2/h$. Similar fluctuations with equal strength have been observed by varying the chemical potential in Si inversion layer nanostructures [5] and in ultra short channel Si MOSFET [3] and by varying the magnetic field in metallic samples [13], GaAs wires [2] various AlGaAs/GaAs systems [4,6]. Theoretically, it has also been proved that similar conductance fluctuations could be obtained in metals due to the motion of a single

carrier [8]. These fluctuations are not a time-dependent noise since they have been shown to be reproducible features in all experimental works.

The basic physics of the conductance fluctuations is due to a quantum interference effect which requires phase coherence in the wave functions over large regions of the sample. A great amount of theoretical work concerned with the problem of localization and with the size of the conductance fluctuations has been based on the usual Anderson tight-binding model [14] well appropriate to the numerical study of one and two-dimensional metals. Stone has used this model [15] to study the conductance fluctuations of metallic samples as a function of magnetic field. Extensive numerical studies of weakly disordered metals have also been performed by Giordano [16]. All simulations agree fairly well with the *universal* value predicted by the perturbative calculations. Another class of purely one-dimensional (single moded) disordered systems in which the randomness is introduced through the average spacing between impurities (spatial disorder) or through the actual shape of the potential (potential shape disorder) has also been extensively investigated in the literature [17-20]. All numerical simulations for this other category of disordered systems have also supported the scaling theory of localization.

In the present chapter, we are mainly concerned with the second class of disordered systems which we generalize to the case of two dimensions allowing for multimode propagation throughout the disordered materials. We model the transport of electrons in a two-dimensional resistor as regions of free propagation with occasional elastic scattering by a random array of scatterers. The conductance of the sample is calculated from the overall transmission probability itself deduced by a new technique for cascading scatter matrices across subsequent sections in the device. We also stress the usefulness of that approach since the classical conductance can be deduced by considering probability scatter matrices rather than amplitude scatter matrices. We show that this is equivalent to neglecting the phases of all the Feynman paths (in real space) connecting the various modes on one side of the resistor to the other. Our model seems more appropriate to describe the transport properties of *dirty* semiconductors. In that respect, it may be very useful to analyze recent experimental data in narrow-Mosfet channels [3], GaAs [2] wires and Aharanov-Bohm structures grown by MBE [13,4,6]. In such materials, the number of propagating modes available at the Fermi level can be several order of magnitude smaller than in metallic samples. This maximum number of propagating modes considered in all our numerical examples is always less then forty.

The chapter is organized as follows. In section 2.2, we describe a technique to cascade scatter matrices across subsequent sections in the device [21]. Then, we discuss two models of impurity scattering. The first one consists of a random array of scatterers all characterized by the same scatter matrix. The randomness is introduced in the spacing between scatterers which is chosen randomly and uniformly over some range [0.,l]. The second model considers a two-dimensional piece of

semiconductor of finite length and width in which the short range impurity potential is modeled as a δ-potential. The impurities are distributed uniformly in the sample for a given impurity concentration. In section 2.3, we perform an extensive numerical study of the weak and strong localization regimes in both models. The universality of the conductance fluctuations is also studied by varying the position of a single impurity. Deviations from the universality behavior are also analyzed for samples with length longer than the localization length and shorter than the elastic mean free path. We also investigate the universality of the conductance fluctuations as a function of the number of propagating modes throughout the sample. Section 2.4 contains our conclusions.

2.2 Theory

In this section, we stress the usefulness of a new approach consisting in cascading scatter matrices through adjacent sections in a device to calculate the conductance of disordered samples according to Landauer formula [22,23]. The advantage of this new technique resides in the fact that both the classical and quantum-mechanical conductance of a specific sample can be calculated using the same technique of cascading either probability or amplitudes scatter matrices. The more difficult question on how to derive the scatter matrix for an arbitrary impurity potential is then considered. Two exactly tractable models of impurity scattering (*i.e*, for which the impurity scatter matrices can be written down explicitly) are then analyzed in some detail.

2.2.1 Conductance Formula

Consider a two-dimensional resistor with a confining potential in the y-direction; the current flows along x (Fig. 2.1 (a)). For a given confining potential we can find a set of transverse modes or subbands m=1,2,... with wavefunctions of the form

$$\Psi(x,y,t) = \phi_m(y)e^{ik_m x}e^{-iEt/\hbar} \tag{2.1}$$

We assume a parabolic dispersion relation for each mode.

$$E = \in_m + \frac{\hbar^2 k_m^2}{2m^*} \tag{2.2}$$

The conductance G is given by,

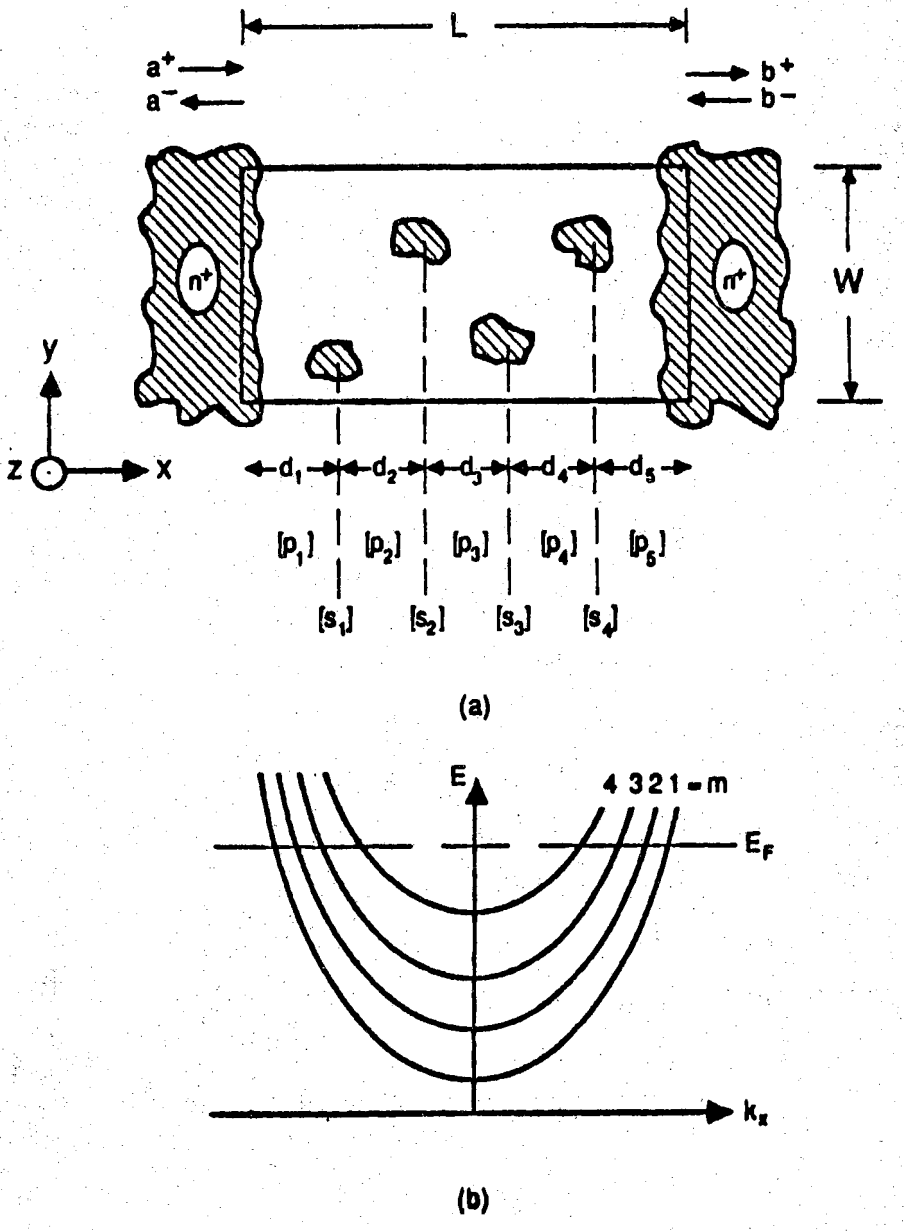


Figure 2.1 (a) A two-dimensional resistor
(b) Dispersion relations for different transverse modes

$$G = \frac{e^2}{\pi\hbar} \int dE \left(-\frac{df}{dE} \right) \sum_{m',m} |t_{m'm}(E)|^2 \tag{2.3}$$

where $t_{m',m}(E)$ is the amplitude for an electron injected from the left into subband m with energy E to be transmitted to subband m′ on the right. The energy E is conserved across the resistor since we are neglecting inelastic scattering. At low temperatures we can replace -df/dE by a delta written as

$$g \equiv \frac{G}{e^2/h} = 2 \sum_{m=1}^{M} \sum_{m'=1}^{M} T_{m',m}(E=E_f) \tag{2.4}$$

The probability $T_{m',m}$ is equal to the squared magnitude of the amplitude $t_{m',m}$. M is the total number of modes that are occupied below the Fermi level E_f. The features of the conductance described hereafter are expected to be independent of the exact analytical form of the multichannel Landauer formula [24,25]. Indeed, most of our numerical simulations will be for sample length bigger than the electron mean free path (we will however briefly comment on the use of different Landauer formulas when studying the conductance in the ballistic regime).

2.2.2 Cascading Scatter Matrices

We wish to calculate the scatter matrix [s] for the resistor connecting the incoming wave amplitudes $\{a^+\}$ and $\{b^-\}$ to the outgoing wave amplitudes $\{b^+\}$ and $\{a^-\}$.

$$\begin{pmatrix} b^+ \\ a^- \end{pmatrix} = \begin{bmatrix} t & r' \\ r & t' \end{bmatrix} \begin{pmatrix} b^- \\ a^+ \end{pmatrix} \tag{2.5}$$

$\{a^\pm\}$, $\{b^\pm\}$ are each (M×1) column vectors so that [t], [r], [t′] and [r′] are each (M×M) matrices. The matrix [t] gives us the M^2 quantities that are needed to evaluate the conductance from eq. (2.4). To calculate the scatter matrix [s] for the resistor we adopt the following model for electron propagation through it. The electrons propagate freely over varying distances $d_1, d_2, \cdots$ between impurities (Fig. 2.1 (b)). The scatter matrices $[p_1]$, $[p_2]$, $\cdots$ for these regions are given by

$$[s] = \begin{bmatrix} \phi_n & 0 \\ 0 & \phi_n \end{bmatrix} \tag{2.6}$$

$$[\phi_n] = \begin{bmatrix} e^{ik_1 d_n} & 0 & & & \\ & & 0 & & \\ 0 & e^{ik_2 d_n} & & & \\ & 0 & e^{ik_3 d_n} & & \\ & & & & e^{ik_M d_n} \end{bmatrix} \tag{2.7}$$

The only non-zero elements are the appropriate phase-shifts along the diagonal. At the impurities the electrons are scattered from one mode to another. The overall scatter matrix [s] is obtained by cascading the scatter matrices of successive sections

$$[s] = [p_1] \otimes [s_I] \otimes [p_2] \otimes [s_I] \otimes \ldots \tag{2.8}$$

Scatter matrices can be cascaded two at a time as described in appendix A. We can thus cascade $[p_1]$ and $[s_I]$ to get a composite matrix which we cascade with $[p_2]$ and so on to get the overall [s] matrix of the resistor from eq. (2.8).

2.2.3 Semiclassical Result

It will be noted that each of the transmission amplitudes $t_{m',m}$ is actually the sum of the complex amplitudes z_n of numerous Feynman paths from subband m at the left to subband m′ at the right

$$T_{m',m} = \left|\sum_n z_n\right|^2 \tag{2.9}$$

The summation index n runs over all Feynman paths originating in subband m at the left and ending in subband m′ at the right. A typical path is shown in Fig. 2.2. The number of paths is denumerably infinite but cascading the scatter matrices as described earlier automatically performs the summation for us. Now, in the semiclassical approximation we neglect the interference between scatterers, that is, between the different Feynman paths.

$$(T_{m',m})_{semi-classical} = \sum_n |z_n|^2 \tag{2.10}$$

The semiclassical T can be calculated in exactly the same way as we calculated the quantum T except that we cascade *probability* scatter matrices rather than *amplitude* scatter matrices [26]. The probability scatter matrices $[S_I]$ are obtained from the amplitude scatter matrices $[s_I]$ by replacing each element by its magnitude squared.

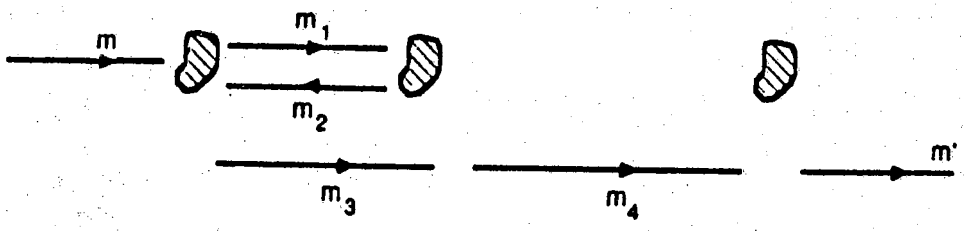


Figure 2.2 A typical Feynman path from subband m at the left to subband m' at the right

$$[S_I]_{m,n} = |[s_I]_{m,n}|^2 \tag{2.11}$$

Hence the semiclassical scatter matrix $[S_c]$ for the device is given by,

$$\begin{aligned}[S_c] &= [P_1] \otimes [S_I] \otimes [P_2] \otimes [S_I] \otimes \ldots \\ &= [S_I] \otimes [S_I] \otimes [S_I] \otimes \ldots \end{aligned} \tag{2.12}$$

since it is apparent from eqs. (2.6-2.7) that the probability scatter matrices corresponding to free propagation are identity matrices. In order to apply either eq. (2.8) or (2.12) to cascade amplitudes or probability scatter matrices, we need to calculate the scatter matrix of each individual scatterer. This is done in the next section where, by analogy with the scattering theory of nuclear physics [27], we derive an unitary scatter matrix for any arbitrary potential with finite range and confined to a two-dimensional resistor. The exact analytical form of this general scatter matrix maybe quite difficult to evaluate for an arbitrary potential. However, for the case of a δ-impurity potential, we show that the scatter matrix can be derived exactly. This model of impurity scattering will be referred later as model A. We finally consider a much simpler case of scatter matrix (model B) for which there is an even probability to be reflected or transmitted into the different modes; only the probability to be transmitted into the same mode as the incident one is different and fixed by the unitarity requirement. Even though this last model of impurity scattering doesn't correspond to any realistic physical system, it is quite interesting since the classical conductance can be derived analytically and can be shown to obey Ohm's law exactly.

2.2.4 S Matrix for an Individual Scatterer

Let us assume that for $t = -\infty$, an electron is in the state ϕ_n as given in eq. (2.1). If we wait a sufficiently long time after the scattering by the 2-D piece of semiconductor of finite width W and length L (see Fig. 2.1.(a)), the potential present in the resistor will be ineffective and the wavefunction describing the electron state for time $t \rightarrow +\infty$ can be expressed as a linear combination of the electronic states ϕ_m. This description of the scattering event is very familiar in collision theory in nuclear physics [27], the only difference being that the eigenstates ϕ_m far away from the scattering region are described as free-electron plane waves. Taking into account this analogy, we therefore can write

$$\psi_m^+ = M_{mn}\phi_n \tag{2.13}$$

where

$$M_{mn} = \langle \phi_n | \psi_m^+ \rangle \qquad (2.14)$$

is the amplitude of the eigenstate ϕ_m that is contained after the scattering has taken place, in the state that grew out of what was the electron eigenstage, ϕ_n before the scattering took place. By analogy with the treatment given in ref. [27], the matrix elements M_{mn} can be written explicitly as follows

$$M_{mn} = \langle m|n\rangle - \frac{i}{\hbar} \int dx \int dy \int dt\, \phi_m^*(\vec{\rho}) H'(\vec{\rho}) \phi_n(\vec{\rho})$$

$$- \frac{i}{\hbar^2} \int \cdots \int \phi_m^*(\vec{\rho}') H'(\vec{\rho}') G_0^+(\vec{\rho}', t'; \vec{\rho}, t) H'(\vec{\rho}) \phi_n(\vec{\rho}) d\vec{\rho}\, d\vec{\rho}'\, dt\, dt'$$

$$- \ldots \qquad (2.15)$$

where $G_0^+(\vec{\rho}', t'; \vec{\rho}, t)$ is the retarded Green's function describing the free-particle propagation in the two-dimensional resistor; $H'(\vec{\rho})$ being the general scattering potential within the 2-D piece of semiconductor.
The scatter matrix relates the current amplitudes rather than the wave amplitudes. We therefore introduce the new matrix t whose matrix elements satisfy

$$\sqrt{k_m}\, \psi_m^+ = t_{mn} \sqrt{k_n}\, \phi_n \qquad (2.16)$$

Using eqs. (2.14) and (2.15), we easily deduce the relation between the matrix M and t

$$t_{mn} = \frac{M_{mn}}{\sqrt{k_m k_n}} = \delta_{mn} - A_{mn} \qquad (2.17)$$

where

$$A_{mm} = \frac{i}{\hbar} \frac{1}{\sqrt{k_m k_n}} \int d\vec{\rho} \int dt \phi_m^*(\vec{\rho}) H'(\vec{\rho}) \phi_n(\vec{\rho})$$

$$+ \frac{i}{\hbar^2} \frac{1}{\sqrt{k_m k_n}} \int \ldots \int \phi_m^*(\vec{\rho}') H'(\vec{\rho}') G_0^+(\vec{\rho}', t'; \vec{\rho}, t) H'(\vec{\rho}) \phi_n(\vec{\rho}) d\vec{\rho} d\vec{\rho}'\, dt dt' \qquad (2.18)$$

$$+ \ldots$$

The matrix t defines the transmission part of the scatter matrix and related the ingoing (from left contact) to outgoing current amplitudes across the 2-D piece of semiconductor. The total scatter matrix S can then be written explicitly as follows

$$[S] = \begin{bmatrix} T & R' \\ R & T' \end{bmatrix} = \begin{bmatrix} I-A & -A \\ -A & I-A \end{bmatrix} \tag{2.19}$$

A being the matrix whose matrix elements are given by (2.18). Taking into account the anti-hermiticity of A, we can easily show that the scatter matrix [s] is indeed unitary. A direct calculation of the matrix elements A_{mn} in eq. (2.18) for any arbitrary potential can be quite involved. However, as will be shown next, this calculation can be performed exactly when the scattering potential is modeled as a delta-impurity potential. This specific case, even though not realistic, will be used as an illustrative example to study the problem of localization in multimoded two-dimensional resistors.

2.2.4.1 S matrix for a δ-impurity

If we model the impurity scattering in a 2-D sample (x-y plane) by a δ-impurity interaction

$$H'(x,y) = \gamma\, \delta(x-x_i)\delta(y-y_i) \tag{2.20}$$

(x_i, y_i) being the position of the ith impurity and if we assume a particle in a book confinement along the y direction, i.e., $H'(x,y) = \infty$ for $y < 0$ and $y > w$, the scatter matrix across the δ-impurity has the following form

$$[s] = \begin{bmatrix} [1+B]^{-1} & -[1+B]^{-1}B \\ -[1+B]^{-1}B & [1+B]^{-1} \end{bmatrix} \tag{2.21}$$

where B is a M×M matrix with matrix elements

$$B_{\alpha\beta} = \frac{i}{2} \frac{\tilde{\Gamma}_{\alpha\beta}}{\sqrt{k_\alpha k_\beta}} \tag{2.22}$$

where the coupling parameters $\tilde{\Gamma}_{\alpha\beta}$ are given by

$$\tilde{\Gamma}_{\alpha\beta} = \frac{4m^*\gamma}{\hbar^2} \frac{1}{w} \sin\left(\frac{m_\alpha \pi y_i}{w}\right) \sin\left(\frac{m_\beta \pi y_i}{w}\right) \tag{2.23}$$

The derivation of eq. (2.21) from the more general expression (2.19) is pretty straightforward if we use the expansion

$$[1+B]^{-1} = 1 - B + B^2 \cdots \tag{2.24}$$

and the explicit expression of the retarded Green function $G_0^+(\vec{\rho}',t';\vec{\rho},t)$ (see expression (2-25) in ref. [28]). An alternative derivation based on a direct solution of the time-independent Schrödinger equation is outlined in appendix B. As can be seen on eqs. (2.21-2.23), each individual scatter matrix is depending on the y location of the impurity in the sample introducing an additional randomization in the model different from the randomization due to the spacing between impurities along the x axis, the direction of propagation of the current. To make the distinction between these two degrees of randomization, we also studied a much simpler model of impurity scattering in which all scatterers are characterized by the same scatter matrix and the only randomization comes from the arbitrary x location of the impurities.

2.2.4.2 A Simple Model for the Impurity Scatter Matrix

More explicitly, we will assume that each impurity has the same scatter matrix $[s_I]$ given by,

$$[s_I] = e^{i[a]} \tag{2.25}$$

where [a] is a matrix with all its element equal.

$$[a]_{m,n} = \alpha \text{ for all m,n} \tag{2.26}$$

since [a] is Hermitian, the scatter matrix $[s_I]$ is unitary as it must be. The exponentiation appearing in eq. (2.25) can be performed analytically. One obtains

$$[s_I] = I + \beta[a] \tag{2.27}$$

where

$$\beta = \frac{e^{2iM\alpha}-1}{2M} \tag{2.28}$$

Thus our choice of scatter matrix implies that at each impurity, the incident mode is reflected equally into each of M modes with probability $|\beta|^2$; it is also transmitted equally into each of the other M-1 modes with probability $|\beta|^2$. For our special choice of $[s_I]$ in which an incident mode is scattered equally into every other mode we can perform the cascading in eq. (2.12) analytically. One advantage of this simple model

is that the semiclassical transmission probability after N_I impurities can be calculated exactly

$$[T_c]_{m',m} = (1 - \frac{2N_I}{\Lambda+N_I})\delta_{m',m} + \frac{N_I/M}{\Lambda+N_I} \tag{2.29}$$

where Λ is the dimensionless mean free path given by

$$\Lambda = (1 - M|\beta|^2)/M|\beta|^2 \tag{2.30}$$

The normalized semiclassical conductance g_c obtained from eq. (2.4) can also be derived analytically

$$g_c = 2M\Lambda/(\Lambda + N_I) \tag{2.31}$$

The semiclassical conductance for this model thus depends only on the number of impurities N_I and is independent of the spacing d_1, d_2,... between impurities and the wavenumbers k_1, k_2,... , k_M of the different modes. A detailed proof of the Ohmic result, eq. (2.31), is given in appendix C.

The following section is devoted to an extensive study of both models A and B including the study of the weak and strong localization regimes as well as the universal conductance fluctuations characterizing disordered systems in the mesoscopic regime. For the case of model A, we will also study the influence of the additional randomization introduced through the arbitrary y location of the impurities.

2.3 Results and Discussion

2.3.1 Simple Impurity Scatter Matrix

In the case of model B, we showed in the previous section that

$$|t_{\beta\alpha}|^2 = |r_{\beta\alpha}|^2 = |\beta|^2 \quad \text{if } \alpha \neq \beta \tag{2.32a}$$

and $$|t_{\alpha\alpha}|^2 = |1+\beta|^2 \quad \text{if } \alpha = \beta \tag{2.32b}$$

In other words, an electron incident in channel α has an equal probability to be transmitted or reflected into all modes β. Only the probability to be transmitted into the same mode is different and is given by eq. (2.29). In this model, the disorder is introduced by letting the average spacing between impurities vary arbitrarily and uniformly over some range [0,l]. For this special case, we show in appendix C that the

classical Ohm law can be derived analytically by cascading the probability scatter matrices following the prescription described at the end of section 2.2.3. The conductance is inversely proportional to the number of impurities crossed and is independent of the spacing $\{d_n\}$ between scatterers. to calculate the quantum-mechanical conductance of the sample versus the number of impurities crossed, we apply the law of composition (2.8) to the series of identical scatter matrices (2.25) interposing an arbitrarily set of propagating scatter matrices (2.6) between them.

2.3.1.1 Localization and Fluctuations

Figure 2.3 shows the resistance (g^{-1}) of a sample with M=30, Λ=33.33 as a function of the number of impurities N_I (all the device parameters are identical to those in the previous example). The distances d_n were distributed uniformly over some range such that $k_i d_n$ (i=1,M) vary between 50π and 250π. Also show for comparison is the semiclassical result (g_c^{-1}) which obeys Ohm's law. The weak and strong localization regions are evident. An estimate of the localization length is the length for which the reduced quantum conductance is about unity and can be estimated by the following relation [29]

$$\Lambda_{el} = \Lambda_{loc} \,/\, M \tag{2.33}$$

M being the number of propagating modes throughout the device. (Λ_{loc} = 1000 impurities in our specific example). When $\Lambda_{el} < L < \Lambda_{loc}$, the weak localization predicts that, when averaged over many samples, the mean quantum-mechanical conductance should be below the classical result by an amount equal to

$$\bar{g}_Q - g_c \simeq -2 \tag{2.34}$$

in reduced units e^2 / h. A simple derivation of this result for the multi-mode transport case is given in appendix D. This difference can be traced back to the fact that, due to coherent back-scattering, the averaged quantum-mechanical reflection coefficients obey the following relations

$$\langle R_{ii} \rangle \simeq \frac{2}{M} \tag{2.35}$$

for sample length such that $L >> \Lambda_{el}$. This is illustrated on Fig. 2.4 for the previously discussed example. In this figure, $\langle R_{ii} \rangle$ and $\langle R_{ij} \rangle$ stand for the average values of the diagonal and off-diagonal reflection coefficients respectively (the average being taken over the indices i and j for a given sample).

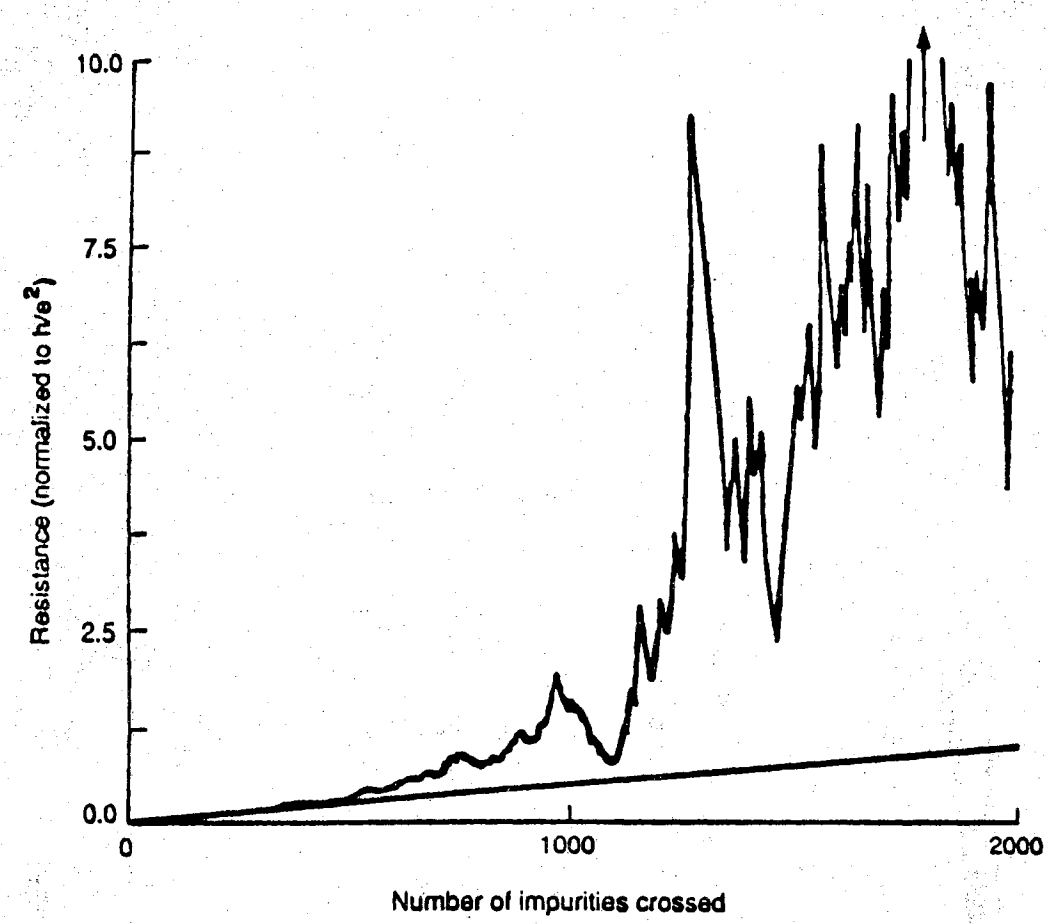


Figure 2.3 Normalized resistance vs. number of impurities N_I for a particular sample (M=30, Λ=33.33)

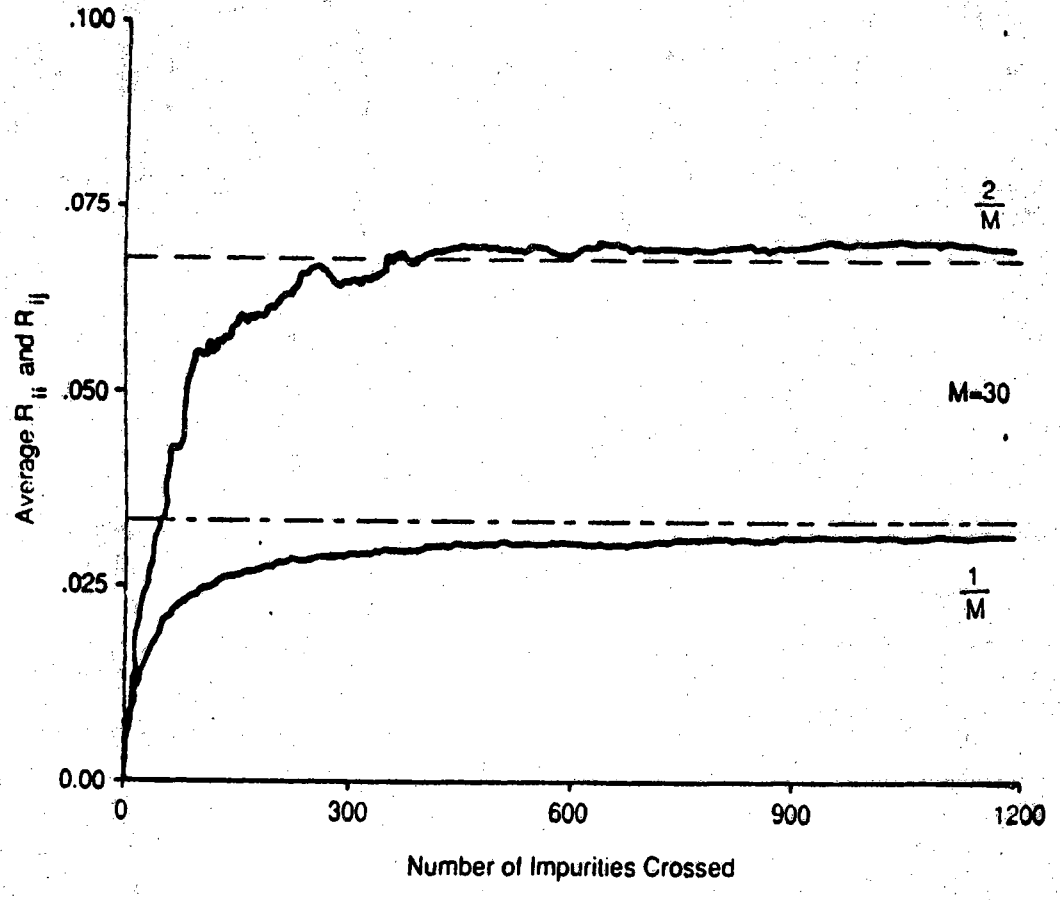


Figure 2.4 Average diagonal and off diagonal reflection coefficients for example in Fig. 2.3

According to the ergodic hypothesis of Lee and co-workers [9], in the metallic regime, i.e, when $\Lambda_{el} < L < \Lambda_{loc}$, the sample specific fluctuations of the conductance over an ensemble of samples which differ only in their microscopic impurity configuration can be viewed as similar to the statistical fluctuations in conductance obtained by moving a single impurity in a given sample. For the 30 modes case, the average quantum- mechanical conductance g_q was calculated by moving one impurity only. More precisely, the middle impurity of a sample with fixed number N_I of impurities was moved by varying distances up to one μm. For this case, as illustrated on Fig. 2.5, we obtain $g_Q - g_{cl} \sim -1.5$ over the entire weak localization regime in quite good agreement with (2.34). Fig. 2.6 illustrates the fluctuations in conductance while moving the middle impurity of a sample with 600 impurities (as before, $\Lambda_{el} = 33.33$ and $M = 30$). The horizontal axis in this picture represents the distance by which the middle impurity was moved from its original position. We notice the large conductance fluctuations whose variance $0.52\ e^2/h$ is in close agreement with the universal value $0.53\ e^2/h$ [9]. This variance was calculated over the entire weak localization length and extended into the strong localization region. As seen in Fig. 2.7, the size of the conductance fluctuations is approximately constant over the weak localization regime and gradually decreases below its universal value while extending the length of the sample into the strong localization regime. This is in agreement with the theoretical prediction and the numerical simulations [16,8] performed on an Anderson tight- binding hamiltonian.

In the ballistic regime, i.e, for $L < \Lambda_{el}$, the conductance fluctuations were also calculated using the two probe Landauer formula (2.4). However, recent experiments performed on metallic rings have employed a four probe measuring configuration to measure the size of the conductance fluctuations in the presence of an external magnetic field. It has been suggested that a more appropriate Landauer formula to describe those type of measurements is given by

$$g = 2 \sum_{m=1}^{M} \frac{T_m}{R_m} \tag{2.36}$$

where T_m and R_m are the total transmission and reflection probabilities for electron impinging from the contacts into the m^{th} mode, *i.e*,

$$T_m = \sum_{m'} T_{m',m} \tag{2.37a}$$

and

$$R_m = \sum_{m'} R_{m'm} = 1 - T_m. \tag{2.37b}$$

Since it is precisely in the ballistic regime that eq. (2.4) and eq. (2.36) give essentially different results, we have calculated the size of the conductance fluctuations in this

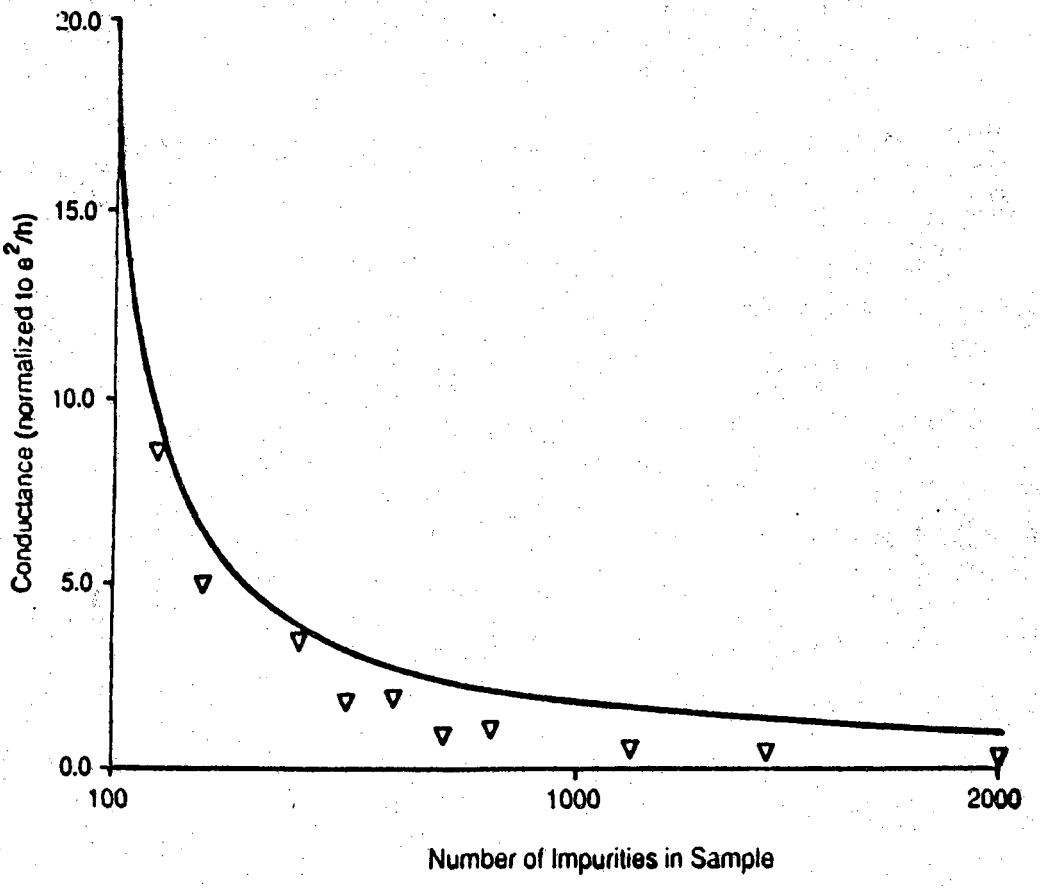


Figure 2.5 Average quantum-mechanical conductance for increasing sample length compared to the classical value (full curve). The average is taken after motion of the middle impurity of the sample.

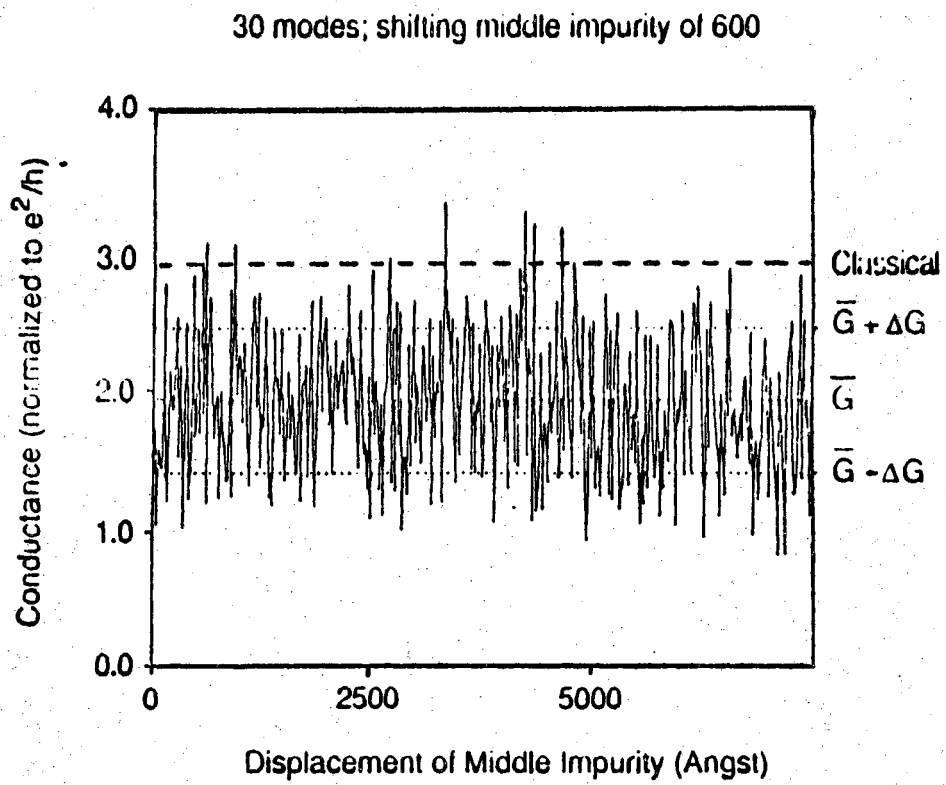


Figure 2.6 Universal conductance fluctuations of a sample containing 600 impurities due to the motion of the middle impurity a distance d to the right from its original position. The classical conductance is also shown for comparison.

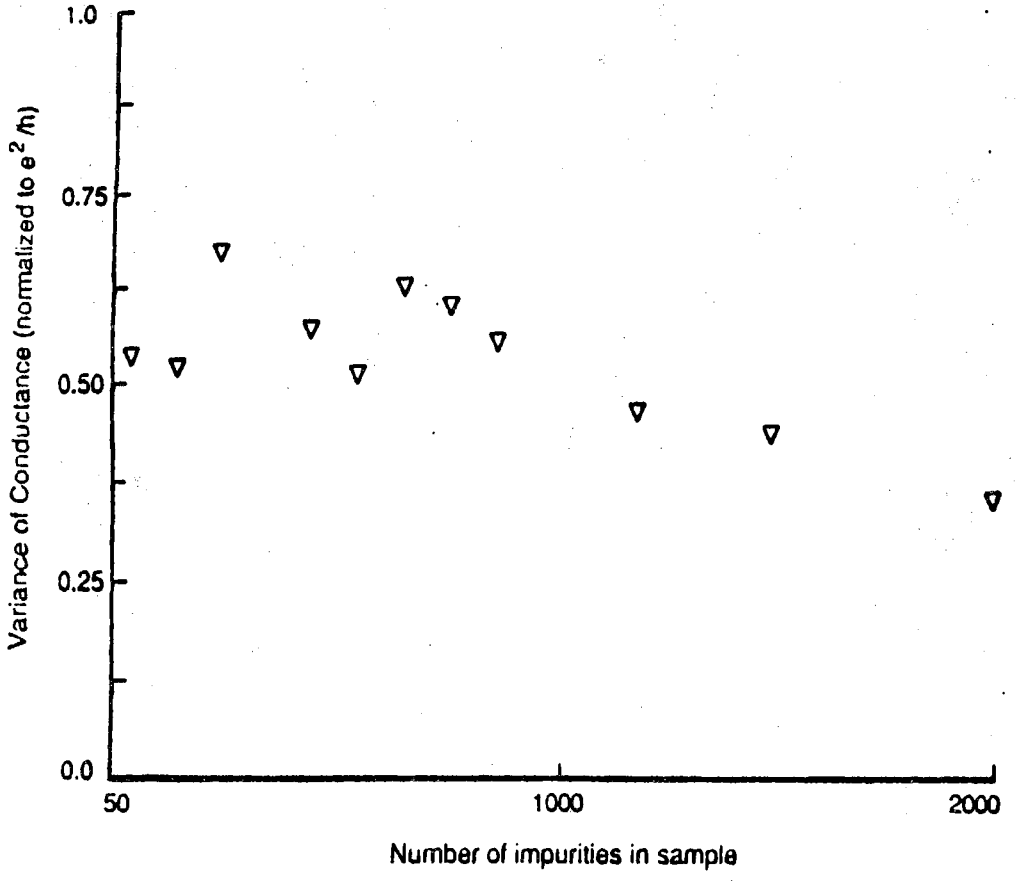


Figure 2.7 Variance of the universal conductance fluctuations (due to the motion of the middle impurity) while increasing the sample length from the weak to the strong localization region. The parameters of the sample are the same as in Fig. 2.3

regime using both eqs. (2.4) and (2.36) (As seen on Fig. 2.8). The conductance fluctuations obtained using eq. (2.4) decrease slightly below the universal result in the ballistic regime (we expect $\Delta g = 0.$ for samples containing a single impurity). This agrees with the results obtained using a tight-binding Hamiltonian [16] for which the universality of the conductance fluctuations was found to be quite robust and valid even for very short samples. On the contrary, the value of ΔG using eq. (2.36) are found to be several times e^2/h. Therefore, while performing four-probe conductance measurements, we should expect the size of the conductance fluctuations to be larger then their universal value if the transport in the device is nearly ballistic (The possibility of ballistic transport in Aharanov-Bohm structures grown by MBE has been analyzed in ref. [30]) .

Finally, in Fig. 2.9 , we show that the size of the conductance fluctuations due to the motion of a single impurity stays approximately constant while increasing the number of propagating modes through the sample. All the sample considered in this simulation had a fixed number of impurities (N_I=100) which puts all of them in the weak localization regime. They were also all characterized by the same value of the parameter α entering the impurity scatter matrix (2.25). (This means that, while characterized by the same impurity configuration, those samples have slightly different elastic mean free path. In Fig. 2.9, each curve is labelled by its elastic mean free path Λ calculated using eq. (2.30)). Following an argument due to Lee [12], this tends to prove that the ratio $\Delta G/G$ doesn't behave as $1/M$ as one would expect by assuming that all the transmission coefficients T_{ij} for the disordered sample are uncorrelated complex random variables.

The theoretical predictions concerning the universality of the conductance fluctuations have been deduced from perturbation theory [8,9] to lowest order in $(k_f\Lambda)^{-1}$, where k_f is the Fermi wavevector and Λ is the elastic mean free path. This condition was met in all our numerical calculations since the average spacing $\{ d_n \}$ between impurities was chosen such that $k_f d_n \simeq 50\pi$ for the mode with the highest transverse energy. While considering model B for impurity scattering, the requirement $k_f\Lambda \gg 1$ needed to observe the weak, strong localization effects as well as the universal conductance fluctuations should always be kept in mind. Indeed, failure to meet this requirement can introduce spurious results in the conductance calculations. In fact, if the average spacing between scatterers is so small ($d_n \longrightarrow 0$) then each scatter matrix for the free propagation between impurities reduces to

$$[s] = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \qquad (2.38)$$

Physically, this means that all the phases of the different modes are small and not fully randomized over 2π between scatterers. In that case, the propagating scatter

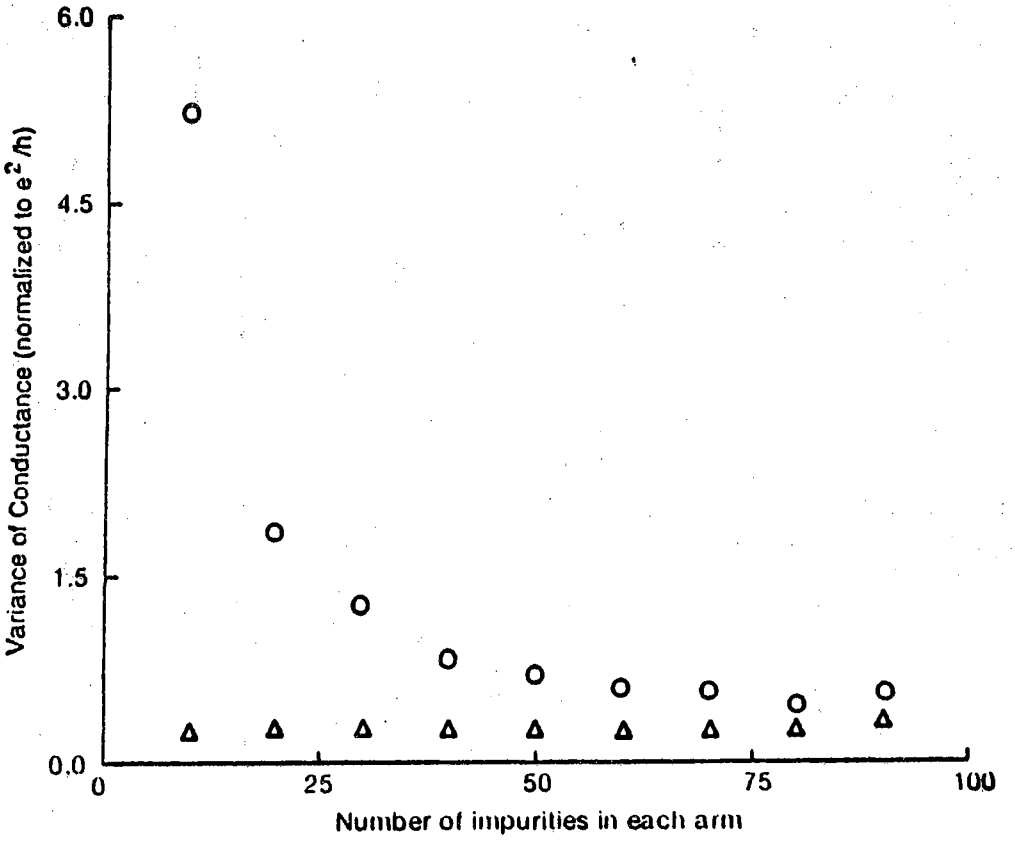


Figure 2.8 Size of the universal conductance fluctuations (due to the motion of a single impurity) calculated using a two-probe (eq. (2.4) in the text) and four-probe (see eq. (2.36)) conductance formula respectively. The parameters characterizing the sample are the same as in Fig. 2.3

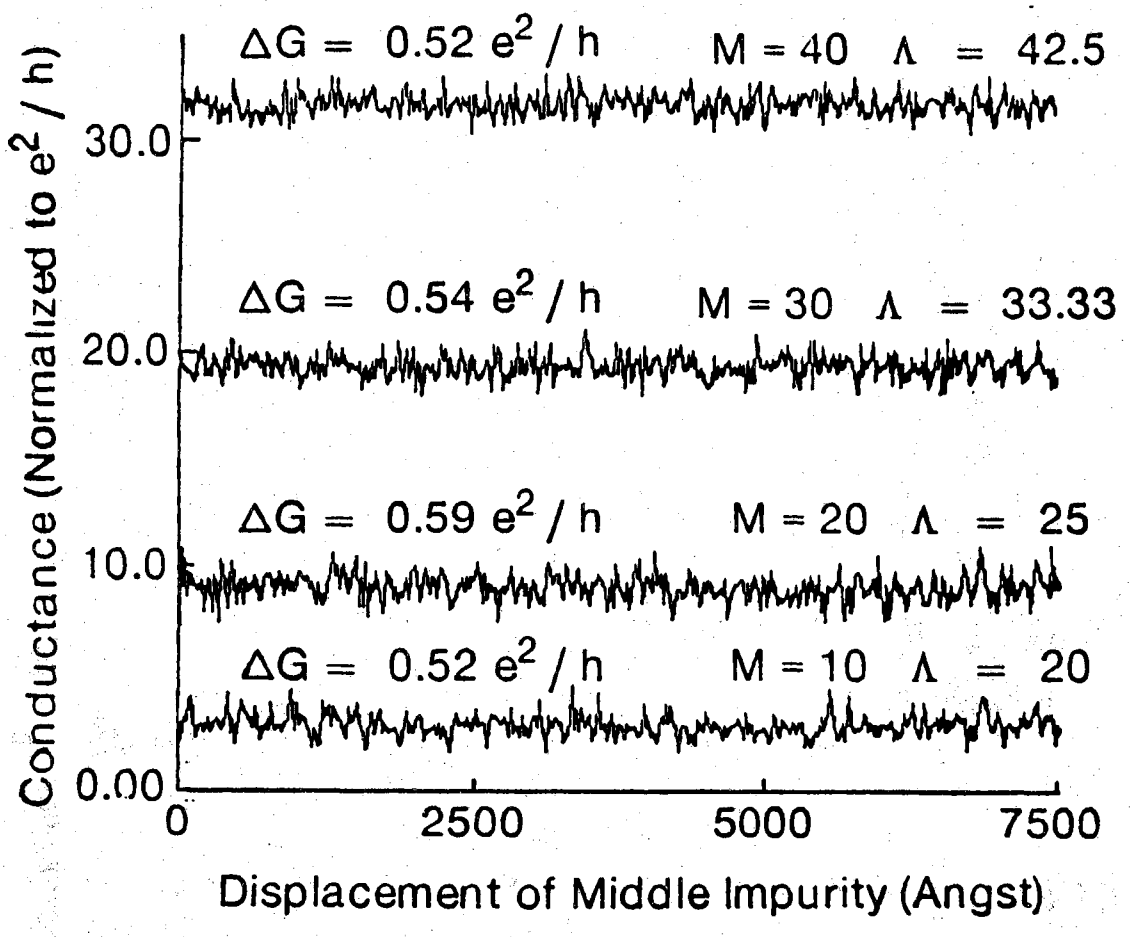


Figure 2.9 Universal conductance fluctuations after displacement of the middle impurity of the sample while increasing the number of propagating modes. The impurity is moved along the x direction by a distance d to the right of its original position. The number of propagating modes is indicated as well as the value of the elastic mean free path (eq. (2.30) in the text)

matrices can be neglected when cascading all the scatter matrices across the device. We must then cascade the scatter matrix (2.25) with itself a number of times equal to the number of impurities crossed. This can be done exactly analytically. In the limit where the number of impurities in the sample goes to infinity, the conductance can then be shown to converge to the following limit

$$g \longrightarrow M-1 \tag{2.39}$$

a limit independent of the constant α in eq. (2.25). If the average spacing between impurities is gradually increased, we then obtain a quantum-mechanical conductance which doesn't show the saturation just discussed.

2.3.2 δ-impurity Scatterers

The resistance versus length of a 2-D resistor (GaAs) doped $10^{11}cm^{-2}$ was calculated for samples of width 2000 Å and characterized by different impurity configurations (Fig. 2.10 shows a typical sample 2000 Å wide and 20000 Å long in which the positions of the δ-scatterers are chosen randomly with a uniform distribution in both x and y directions for a given impurity concentration). For comparison, the *classical* resistance was also calculated for the same samples using the technique of cascading probability scatter matrices as described in Section 2.2.3). Several remarks have to be made.

As can be seen on Fig. 2.11, we first notice that, even in the weak localization regime (i.e when $R << h/2e^2$), a full quantum-mechanical gives a variation of resistance versus length far less simple than the classical Ohm law. Deviation from this result can even be noticed quite wide samples. Furthermore, different impurity locations give rise to drastically oscillations patterns in the variation of the resistance of the sample versus its length. The classical resistances are also slightly different for different samples and, even though close from being linear, the resistance variation versus length doesn't obey strictly Ohm law. In fact, due to the *particle in a box* confinement potential in the y direction, the conductance of the two-dimensional resistors is strongly depending on the y positions of the different impurities both classically and quantum-mechanically. However, while considering wider and wider samples, the classical variation from sample to sample (with a fixed length) is supposed to dissapear: this corresponds to the thermodynamic limit which stays that the classical resistance should only depend on macroscopic parameters such as the impurity concentration. Quantum-mechanically, however, the *universal conductance result* predicts that the conductance fluctuations while considering different samples should be of the order of e^2/h no matter how wide the samples are, *i.e*, no matter how large is the number of propagating modes throughout the sample. Obtaining the variance of the

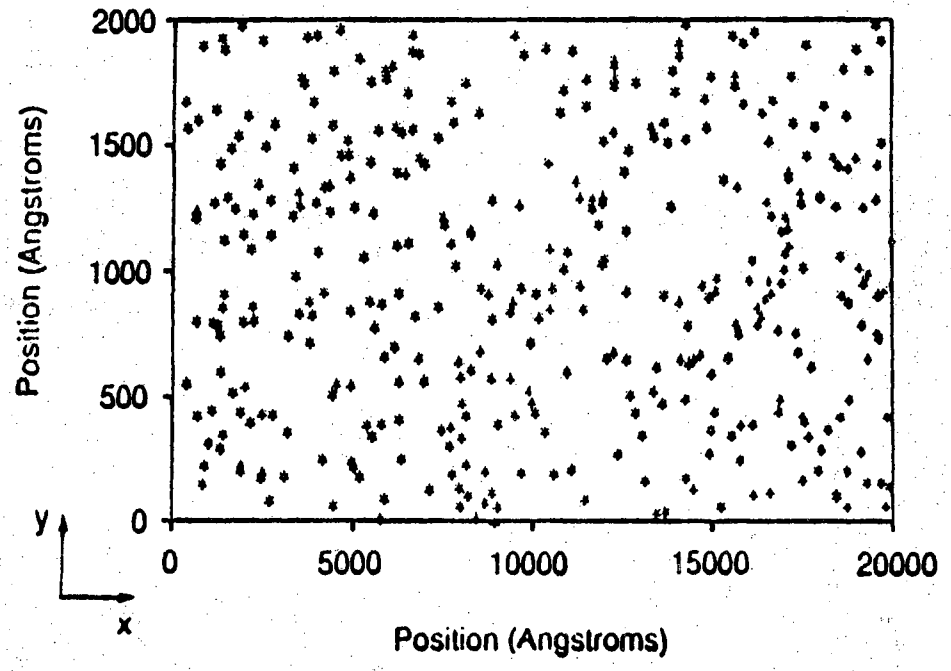


Figure 2.10 Model B: A specific sample 2×10^5Å long and 2×10^3Å wide sample.

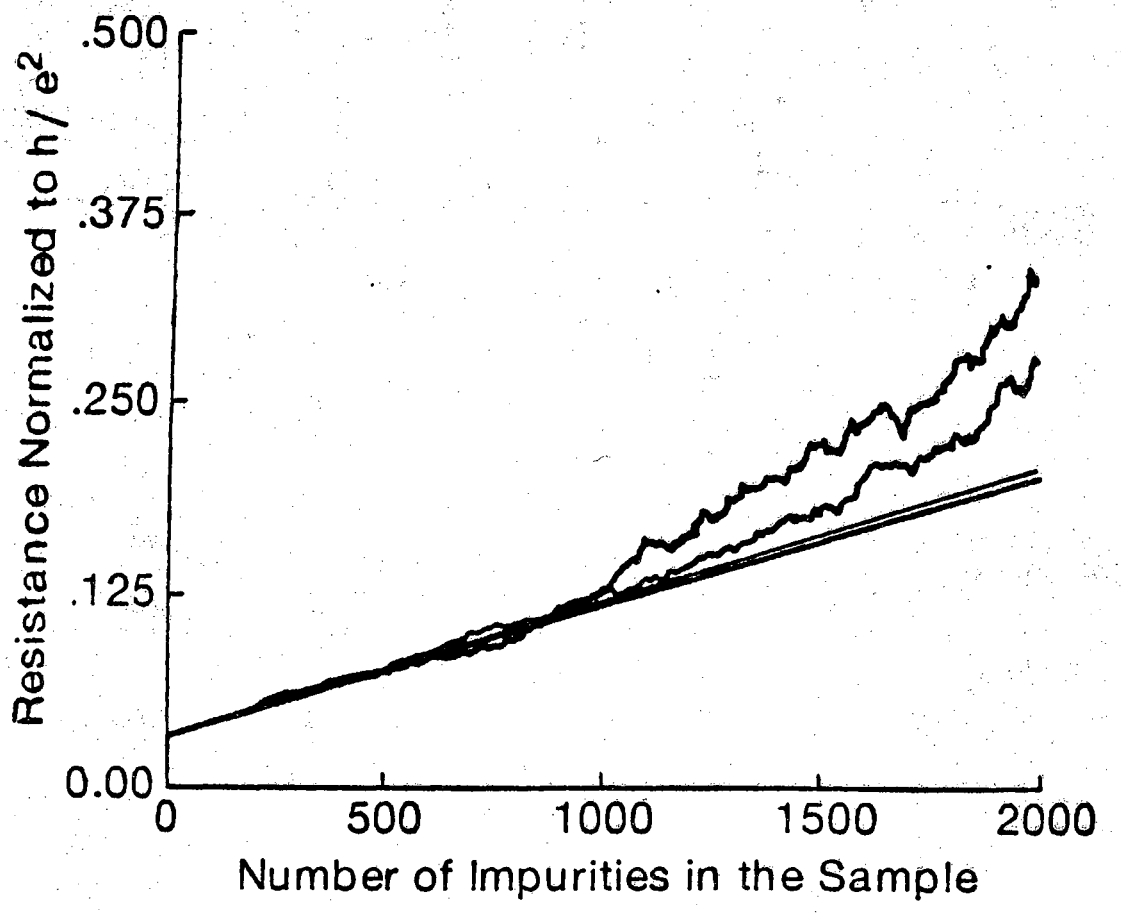


Figure 2.11 Resistance versus length for samples (A,B) such as the one shown in Fig. 2.10 differing by their impurity configuration. (The samples are 2000Å wide)

conductance fluctuations over a statistically meaningful number of completely different samples is quite time consuming. However, since conductance fluctuations of the same magnitude can be obtained by the motion of a single impurity [8], we calculated (both classically and quantum-mechanically) the variance of the conductance fluctuations by changing the y-position of the middle impurity only from one side of the resistor to the other by step of 10 Å. Fig. 2.12 the classical and quantum-mechanical conductance variation of a sample (2×10^3Å wide and 10^5Å long) while moving the middle impurity as described above. The quantum-mechanical fluctuations are seen to be quite large ($\Delta G \simeq 0.2e^2/h$) whereas no sensible fluctuation is perceptible in the classical case.

2.4 Conclusions

In this chapter, we have studied the problem of Anderson localization for the case of multi-moded transport using a technique for cascading scatter matrices associated to subsequent sections in a device. This study is complementary to the numerical simulations performed on tight-binding Hamiltonians [16,15] and generalizes to the case of multi-moded transport earlier numerical calculations performed on purely 1-D random array of scatterers [17,20]. The technique can in principle be applied for any arbitrary impurity potential with finite range. The conductance of disordered samples was calculated using a scatter matrix approach which furnishes directly the reflection and transmission coefficients in the different modes. One other advantage of cascading scatter matrices rather than transfer matrices is that the latter blow up exponentially while increasing the length of the device. In our case, elastic scattering requires all scatter matrices to be unitary, an easy check to perform numerically while cascading the different scatter matrices across the device. Another advantage of cascading scatter matrices is that the classical conductance can also be derived by replacing all amplitude scatter matrix by the probability scatter matrices. All our numerical simulations agree with the scaling theory of localization both in the weak and strong localization regimes. Furthermore, we investigated the universality of the conductance fluctuations in the mesoscopic range (i.e for sample length such that $\Lambda < L < \Lambda_{loc}$) by varying the position of one impurity only. All numerical simulations agree well with the size of the fluctuations predicted theoretically for quasi-one dimensional systems, i.e., 0.53 e^2/h [8]. We also calculated the size of the fluctuations for sample length bigger than the localization length as expected theoretically. When varying the microscopic configuration of the sample, the size of the fluctuations is predicted to be below the universal result and is continuously decreasing as a function of length (as expected theoretically) in the strong localization regime. We have also shown that the universal conductance result is independent of the number of propagating modes throughout the device.

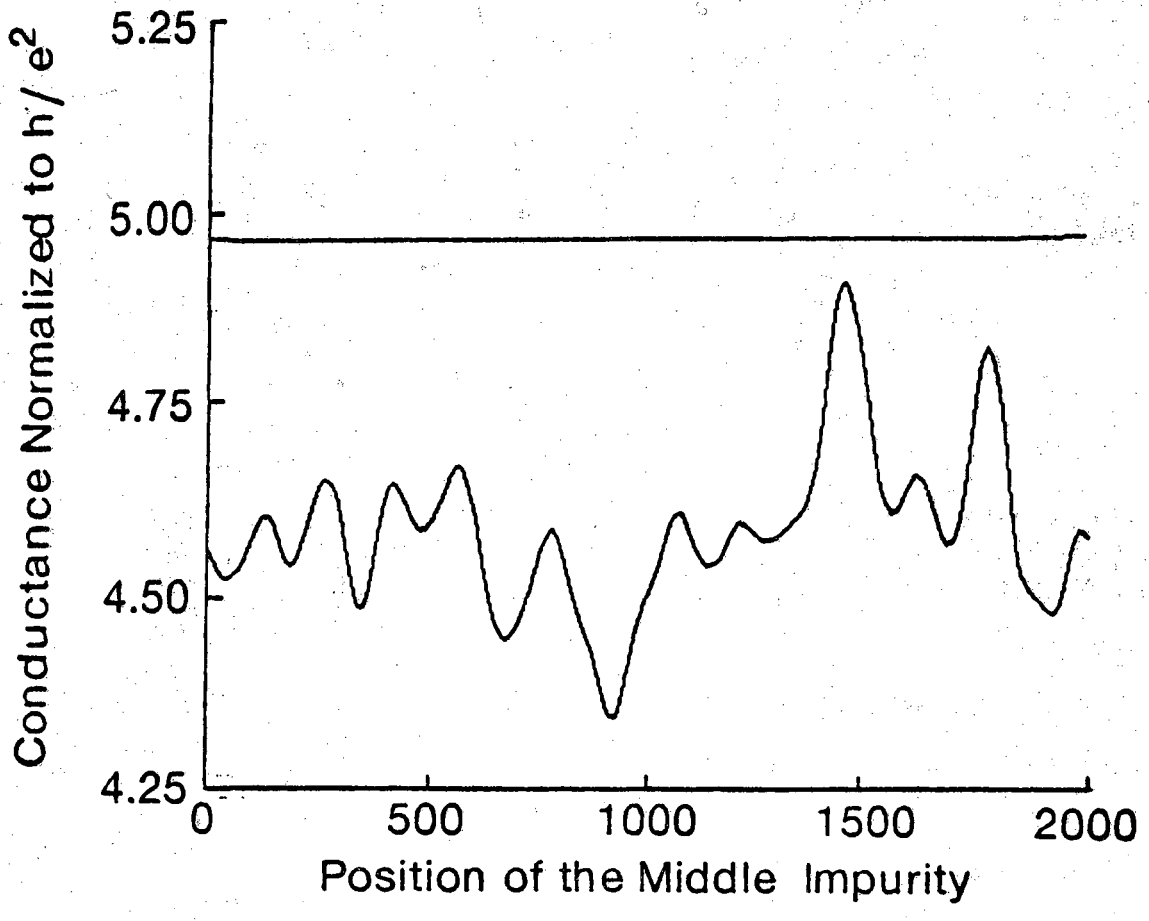


Figure 2.12 Classical versus quantum-mechanical conductance of a sample such as the one shown in Fig. 2.10 after moving the middle impurity from one side of the resistor to the other (the x position of the impurity being unchanged). The sample is 2×10^3 Å wide and 10^5 Å long

Recently, various groups have reported the observation of universal conductance fluctuations in ultrasmall GaAs wires and in Aharanov-Bohm heterostructures. In those ultrasmall structures, the spatial quantization reduces the number of propagating channels available at the Fermi level to be of the order of a few tens only. This number is by several orders of magnitude smaller than the number of propagating modes available at the Fermi energy in metallic samples and makes our model particularly useful to investigate the influence of impurity scattering in ultrasmall semiconductor heterostructures. Preliminary results concerning the size of the conductance modulation in the recently proposed Electrostatic Aharanov- Bohm effect will be discussed in the next chapter [31].

Appendix A: Cascading Scatter Matrices

In this appendix, we will derive the composite scatter matrix for a region consisting of two subsequent sections of a device described by their individual scatter matrices.

Let us consider a specific interval in a device. As discussed in section 2.2.2) , a scatter matrix relates incoming wave amplitudes $\{a^+\}$ and $\{b^-\}$ to the outgoing wave amplitudes $\{b^+\}$ and $\{a^-\}$ as follows

$$\begin{pmatrix} b^+ \\ a^- \end{pmatrix} = \begin{bmatrix} t & r' \\ r & t' \end{bmatrix} \begin{pmatrix} b^- \\ a^+ \end{pmatrix}. \tag{A1}$$

The elements of the scatter matrix (r,t',t,r') represent the amplitudes for electrons to be transmitted (t,t') or reflected (r,r') from that interval. It follows that, in a region of two intervals, an electron could be multiply reflected between two interfaces. By summing the amplitudes of electrons following an infinite number of multiply reflected paths, the overall transmitted and reflected amplitudes can be determined. This analysis is pictured in Figures A1 and A2. In Fig. A1, electrons incident from the left of the region could be transmitted straight through the structure, or could experience an infinite number of multiple reflections before being transmitted. The total transmitted amplitude is the sum of the contributions from each possible path:

$$t = t_2 \left[1 + r_1' r_2 + (r_1' r_2)^2 + \dots \right] t_1 \tag{A2}$$

where the subscripts indicate the interval in which transmission or reflection occurs. The phase acquired by the wavefunction in the course of multiple reflections is contained in the definition of the scatter matrix elements and need not be added in this analysis. Since the infinite series of multiply reflected terms is a geometric series, the composite transmitted amplitude can be simplified:

$$t = t_2 \left| 1 - r_1' r_2 \right|^{-1} t_1 \tag{A3}$$

Following a similar derivation (see Figs. A1 and A2), the remaining elements of the composite scatter matrix can be determined, completing the analysis:

$$r = r_1 + t_1' r_2 [1 - r_1' r_2]^{-1} t_1 \tag{A4}$$

$$t' = t_1' [1 + r_2 [1 - r_1' r_2]^{-1} r_1'] t_2' \tag{A5}$$

$$r' = r_2' + t_2 [1 - r_1' r_2]^{-1} r_1' t_2' \tag{A6}$$

Even if the product $r_1' r_2$ were to approach unity, making the inverse of $[1 - r_1' r_2]$

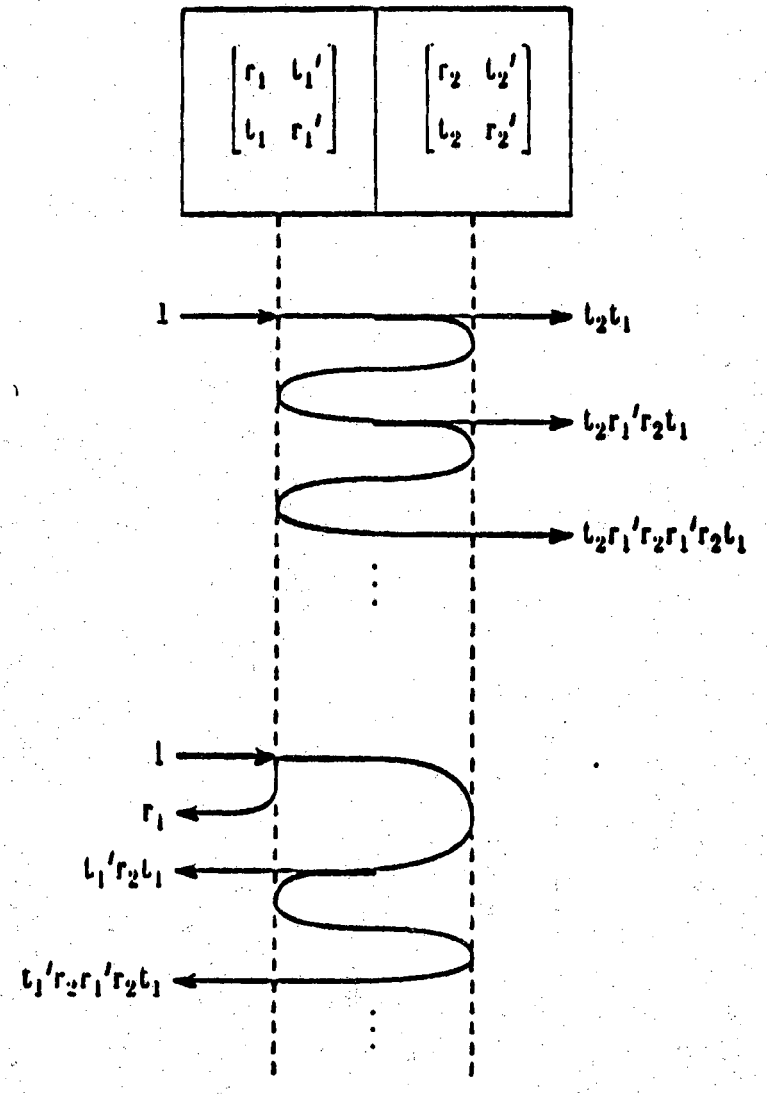


Figure A1. Derivation of t and r for a composite scatter matrix, by considering multiple reflections of the wavefunction.

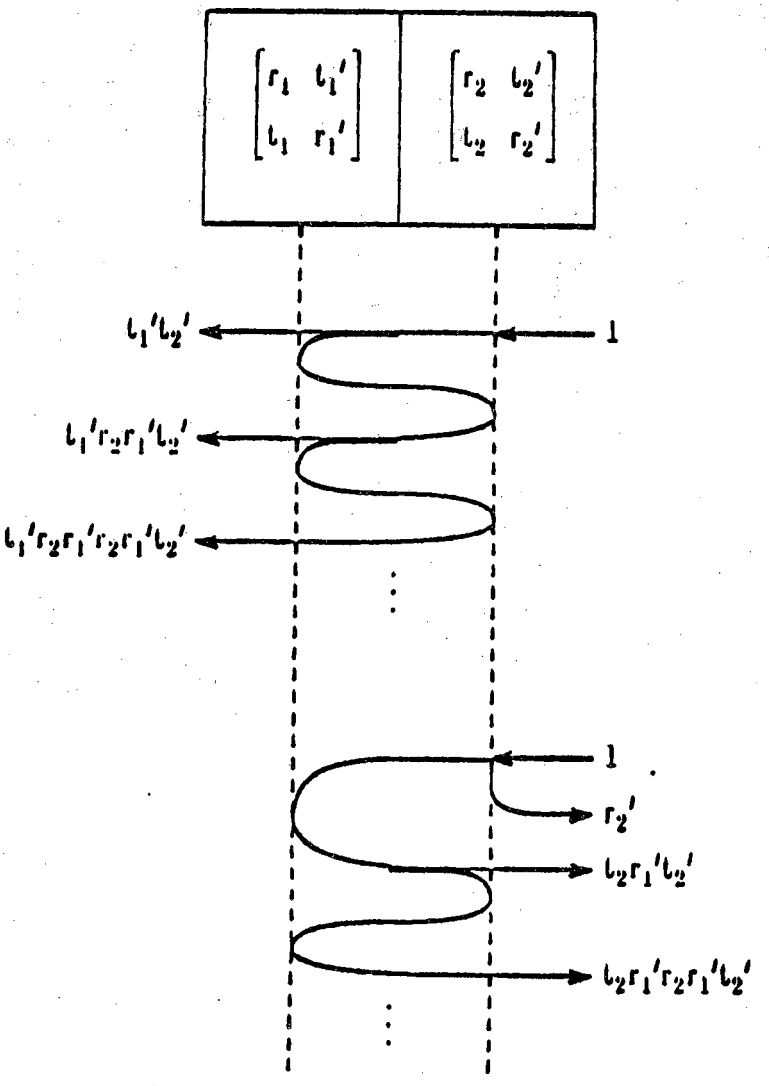


Figure A2. Derivation of t' and r' for a composite scatter matrix, by considering multiple reflections of the wavefunction.

large, transmitted amplitudes t_1' and t_2 would approach zero, by particle conservation laws. Since many computers allow small numbers to underflow benignly to zero, a scatter matrix solution can be implemented with relatively little error checking.

An alternative way to deduce the set of relations (A3-A6) is the following. For each section in the device, we make the transformation from the scatter matrix to the associate transfer matrix. The latter relates the wave amplitudes on one side of the section ($a^{\pm}$) to the wave amplitudes on the other side ($b^{\pm}$), *i.e*,

$$\begin{pmatrix} b^+ \\ b^- \end{pmatrix} = [W] \begin{pmatrix} a^+ \\ a^- \end{pmatrix} = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{pmatrix} a^+ \\ a^- \end{pmatrix}. \tag{A7}$$

The different blocks W_{ij} (i,j=1,2) can be expressed in terms of the various matrices (r, r', t, t') composing the scatter matrix of the same section,

$$W_{11} = t - r' t'^{-1} r \tag{A8}$$

$$W_{12} = r' t'^{-1} \tag{A9}$$

$$W_{21} = t'^{-1} r \tag{A10}$$

$$W_{22} = t^{-1} . \tag{A11}$$

Reciprocally, the various blocks (r, t', t, t') composing the scatter matrix can be expressed in terms of the various block W_{ij} (i,j=i,2),

$$r = -W_{22}^{-1} W_{21} \tag{A12}$$

$$r' = W_{12} W_{22}^{-1} \tag{A13}$$

$$t = W_{11} - W_{12} W_{22}^{-1} W_{21} \tag{A14}$$

$$t' = W_{22}^{-1} \tag{A15}$$

Since the law of composition for transfer matrices is a simple matrix multiplication, we can easily calculate the transfer matrix associated to two subsequent sections in a device. Then, using the set of equations (A12) to (A15) in which the W_{ij} are the blocks of the overall transfer matrix, we get equations (A3) to (A6), defining the blocks of the composite scatter matrix.

Appendix B: Scatter Matrix for a δ-impurity

In this appendix, we derive the exact analytical form of the scatter matrix across a δ-impurity. Our start point is the one-electron effective mass Schrödinger equation

$$\left[-\frac{\hbar^2}{2m^*} \nabla^2 + E_c(x) + H'(\vec{r}) \right] \psi(\vec{r},t) = i\hbar \frac{\partial \psi(\vec{r},t)}{\partial t} \tag{B1}$$

written for an electron propagating through a device with arbitrary conduction band energy profile $E_c(x)$ (x being the direction of propagation of the current). In eq. (B1), $H'(\vec{r})$ is the scattering part of the Hamiltonian which we assume to be time-independent since our main concern is to study the influence of impurity scattering on the conductance of dirty samples at absolute zero. For elastic scattering, the electron total energy is conserved while traversing the device and therefore the time-dependent part of the wave function can be eliminated using the ansatz

$$\psi(\vec{r},t) = \phi(\vec{r}) e^{-\frac{iEt}{\hbar}} \tag{B2}$$

If H' is identically equal to zero, the band energy profile (or external potential energy) is assumed to be simple enough so we can write the eigenstates of the Schrödinger eq. (B1) explicitly. For instance, for devices grown by MBE, the potential energy profile varies only in the direction of the axis of growth of the structure (which we choose as the x-direction). The eigenstates of the Schrödinger eq. (B1) can then be characterized by the two *good* quantum numbers $\vec{k}_t$ and E, $\vec{k}_t$ being the component of the wavevector in the (y,z) plane and E the electron total energy. Both these quantities are conserved if no scattering is present. The electron eigenstates describing the transverse motion of the electrons can then be written as

$$\frac{e^{i(\vec{k}\cdot\vec{\rho})}}{\sqrt{A}} \tag{B3}$$

for which we use hereafter the shorthand notation $|\alpha>$; $\vec{\rho}$ is the radius vector in the y-z plane. The normalizing factor $\sqrt{A}$ is introduced because we use periodic boundary conditions in y,z to determine the allowed values of $\vec{k}_t$ and E. Those different states are then orthogonal, *i.e.*,

$$\int dy\, dz\, |\alpha>^* |\beta> = \delta_{\alpha\beta} \tag{B4}$$

and form a complete set of eigenstates on which the general solution of eq. (A1) for a fixed energy can be expanded when the scattering part of the Hamiltonian $H'(\vec{r},t)$ is non-zero,

$$\phi(\vec{r}) = \sum_{\alpha} C_\alpha(x)\, |\, \alpha> \tag{B5}$$

For different type of confining potential in the y-z directions such as a simple box of width W and height H, the good quantum numbers would then be the total energy E and a set of two integers n and m characterizing the particle in a box confinement in the transverse direction. The electron energy dispersion relation would in this case be given by

$$E = \epsilon_n^y + \epsilon_m^z + \frac{\hbar^2 k_x^2}{2m^2} \tag{B6a}$$

where

$$\epsilon_n^y = \frac{n^2 \pi^2 \hbar^2}{2m^* W^2} \tag{B6b}$$

and

$$\epsilon_m^z = \frac{m^2 \pi^2 \hbar^2}{2m^* H^2} \tag{B6c}$$

For this special case, the transverse part of the electron eigenstates can be explicitly written as follows

$$|\alpha> = \sqrt{\frac{2}{W}} \sin(\frac{m\pi y}{W}) \sqrt{\frac{2}{H}} \sin(\frac{n\pi z}{H}) \tag{B7}$$

It is then straightforward to calculate the charge density $\rho(\vec{r})$ and current density J(x) associated to the wavefunction (B5). We obtain respectively

$$\rho(x) = e\, \psi^* \psi = e \sum_{\alpha} C_\alpha^*(x)\, C_\alpha(x) \tag{B8}$$

and

$$J(x) = e \sum_{\alpha} \frac{i\hbar}{2m^*} \left| \left(\frac{d\Psi^*}{dx} \right) \Psi - \Psi^* \left(\frac{d\Psi}{dx} \right) \right| \tag{B9}$$

Here $\rho(x)$ and $J(x)$ are respectively the charge and current densities integrated over y,z and t so that the cross-terms do not appear. Using eq. (B2) and making use of the orthogonality condition (B4), the Schrödinger eq. (B1) can be written

$$\frac{d^2 C_\alpha}{dx^2} + k_\alpha^2 C_\alpha = \sum_{\beta} \Gamma_{\alpha\beta}(x)\, C_\beta \tag{B10}$$

where

$$k_\alpha^2 = \frac{2m^*}{\hbar^2} \left| E - \frac{\hbar^2 k_t^2}{2m^*} - E_c(x) \right| \tag{B11}$$

and

$$\Gamma_{\alpha\beta}(x) = \frac{2m^*}{\hbar^2} \int dydz \; |\alpha\rangle^* H' \, |\beta\rangle \tag{B12}$$

Equation (B10) is our main result. The problem of solving the scattering problem is then reduced to the calculation of the coefficients $\Gamma_{\alpha\beta}$ and to the solution of the system of coupled differential equations (B10). Equation (B10) has indeed to be written for all modes α's considered in any particular problem. Not that the sum over β in the second member of eq. (B10) do include the term $\alpha=\beta$. In practice, the number of modes has to be reduced to some finite number M.

In the set of couple differential equations (B10), both quantities k_α^2's and $\Gamma_{\alpha\beta}$'s are function of the variable x and depend respectively on the exact shape of the conduction band energy profile and the interacting potential. A scatter matrix relates the modes indicent on an obstacle from either direction to the modes leaving the obstacle in either direction. In order to actually calculate the scatter matrix of each individual section, we first introduce the following new set of variables

$$C_\alpha^{\pm}(x) = \frac{1}{2} \sqrt{\frac{\hbar k_\alpha}{m^*}} \left(C_\alpha \pm \frac{1}{ik_\alpha} \frac{dC_\alpha}{dx} \right) \tag{B13}$$

which can easily be inverted to give

$$C_\alpha = \frac{1}{2\beta_\alpha} \left(C_\alpha^+ + C_\alpha^- \right) \tag{B14a}$$

$$\dot{C}_\alpha = \frac{ik_\alpha}{2\beta_\alpha}\,(C_\alpha^+ - C_\alpha^-) \tag{B14b}$$

where, by definition,

$$\beta_\alpha = \frac{1}{2}\sqrt{\frac{\hbar\, k_\alpha}{m^*}} \tag{B15}$$

The $C_\alpha^{+,-}$ represent the amplitude of the current density in mode α traveling along the negative and positive x axis respectively. Indeed, one can easily show using eq. (B13) and the definitions (B14a,b) that the current density can be written as follows

$$J(x) = \sum_\alpha J_\alpha^+ - J_\alpha^- \tag{B16}$$

where

$$J_\alpha^+ = (C_\alpha^+)^* C_\alpha^+ \text{ and } J_\alpha^- = (C_\alpha^-)^* C_\alpha^- \tag{B17}$$

By definition, the scatter matrix across a finite section located between x_0 and $x_0+\ell$ is the matrix which satisfies

$$\begin{bmatrix} C_B^+(x_0+\ell) \\ \cdot \\ \cdot \\ \hline C_\alpha^-(x_0) \\ \cdot \\ \cdot \end{bmatrix} = S \begin{bmatrix} C_B^-(x_0+\ell) \\ \cdot \\ \cdot \\ \hline C_\alpha^+(x_0) \\ \cdot \\ \cdot \end{bmatrix} \tag{B18}$$

the index α and β referring to the different modes on the left and right side of the section $[x_0, x_0+\ell\,]$ respectively. This scatter matrix cannot be derived exactly analytically for arbitrary conduction band energy profile $E_c(x)$ and coupling parameters $\Gamma_{\alpha\beta}(x)$. However, the scatter matrix has a quite simple analytical form in a specific case which we now consider in detail.

Scatter matrix across a δ - impurity

If we model the impurity scattering in a 2-D sample (x-y plane) by a δ - impurity interaction

$$H'(x,y) = \gamma\delta(x-x_i)\ \delta(y-y_i) \tag{B19}$$

the set of differential equations (B10) can be written as follows

$$\frac{d^2C_\alpha}{dx^2} + k_\alpha^2 C_\alpha = \sum_\beta \tilde{\Gamma}_{\alpha\beta}\ \delta(x-x_i)C_\beta \tag{B20}$$

where $\tilde{\Gamma}_{\alpha\beta}$ is given by eq. (2.23) in the text. Integrating both sides of eq. (B1) from $x_i-\epsilon$ to $x_i+\epsilon$ (ϵ being a small positive quantity) and taking into account the assumed continuity of the C_α's, we obtain

$$\dot{C}_\alpha(x_i+\epsilon) - \dot{C}_\alpha(x_i-\epsilon) = \sum_\beta \tilde{\Gamma}_{\alpha\beta}C_\beta(x_i+\epsilon) \tag{B21}$$

which we rewrite

$$\dot{C}_\alpha(x_i+\epsilon) = \dot{C}_\alpha(x_i-\epsilon) + \sum_\beta \tilde{\Gamma}_{\alpha\beta}C_\beta(x_i+\epsilon) \tag{B22}$$

Since all C_α's are continuous, we have

$$C_\alpha(x_i+\epsilon) = C_\alpha(x_i-\epsilon) \tag{B23}$$

Dividing eq. (B22) by ik_α and adding the obtained result to eq. (B23), we get

$$\begin{aligned} C_\alpha(x_i+\epsilon)+\frac{1}{ik_\alpha}\dot{C}_\alpha(x_i+\epsilon) &= \frac{1}{\beta_\alpha}\ C_\alpha^+(x_i+\epsilon) \\ &= \frac{1}{\beta_\alpha}C_\alpha^+(x_i-\epsilon)-\sum_\beta \frac{\tilde{\Gamma}_{\alpha\beta}}{ik_\alpha}C_\beta(x_i) \end{aligned} \tag{B24}$$

Now, using eqs. (B14a-b), we have

$$C_\beta(x_i+\epsilon) = \frac{1}{2\beta_\beta}\ [C_\beta^+(x_i+\epsilon)+C_\beta^-(x_i-\epsilon)] \tag{B25}$$

Plugging this last result in eq. (B6), we finally derive

$$C_\alpha^+(x_i+\epsilon) = C_\alpha^+(x_i-\epsilon)+\frac{\beta_\alpha}{2ik_\alpha}\sum_\beta \tilde{\Gamma}_{\alpha\beta}\frac{1}{\beta_\beta}(C_\beta^+(x_i+\epsilon)+C_\beta^-(x_i-\epsilon)) \tag{B26}$$

or equivalently

$$C_\alpha^+(x_i+\epsilon)-\sum_\beta \frac{1}{2i k_\alpha}\tilde{\Gamma}_{\alpha\beta}\left(\frac{\beta_\alpha}{\beta_\beta}\right)C_\beta^+(x_i+\epsilon) =$$

$$C_\alpha^+(x_i-\epsilon)+\sum_\beta \frac{\tilde{\Gamma}_{\alpha\beta}}{2ik_\alpha}\left(\frac{\beta_\alpha}{\beta_\beta}\right)C_\beta^-(x_i+\epsilon) \tag{B27}$$

valid for all modes α. The equations for the different modes can be written in a matrix form (which is done here for the case of two modes only for simplicity)

$$\begin{bmatrix} 1-\dfrac{\tilde{\Gamma}_{11}}{2ik_1} & \dfrac{-1}{2ik_1}\tilde{\Gamma}_{12}\dfrac{\beta_1}{\beta_2} \\ \dfrac{-1}{2ik_2}\tilde{\Gamma}_{21}\dfrac{\beta_2}{\beta_1} & 1-\dfrac{1}{2ik_2}\tilde{\Gamma}_{22} \end{bmatrix} \begin{bmatrix} c_1^+(x_i+\epsilon) \\ c_2^+(x_i+\epsilon) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & \dfrac{\tilde{\Gamma}_{11}}{2ik_1} & \dfrac{\tilde{\Gamma}_{12}}{2ik_1}\dfrac{\beta_1}{\beta_2} \\ 0 & 1 & \dfrac{\tilde{\Gamma}_{21}}{2ik_2}\dfrac{\beta_2}{\beta_1} & \dfrac{\tilde{\Gamma}_{22}}{2ik_2} \end{bmatrix} \begin{bmatrix} C_1^+(x_i-\epsilon) \\ C_2^+(x_i-\epsilon) \\ C_1^-(x_i+\epsilon) \\ C_2^-(x_i+\epsilon) \end{bmatrix} \tag{B28}$$

which we write more simply as follows

$$[I+B]\begin{bmatrix} C_1^+(x_i+\epsilon) \\ C_2^+(x_i+\epsilon) \end{bmatrix} = [I,-B]\begin{bmatrix} C_1^+(x_i-\epsilon) \\ C_2^+(x_i-\epsilon) \\ C_1^-(x_i+\epsilon) \\ C_2^-(x_i+\epsilon) \end{bmatrix} \tag{B29}$$

I being the unit matrix and the matrix B is given by

$$B = \begin{bmatrix} \dfrac{-1}{2ik_1}\tilde{\Gamma}_{11} & \dfrac{-1}{2ik_1}\tilde{\Gamma}_{12}\dfrac{\beta_1}{\beta_2} \\ \dfrac{-1}{2ik_2}\tilde{\Gamma}_{21}\dfrac{\beta_2}{\beta_1} & \dfrac{-\tilde{\Gamma}_{22}}{2ik_2} \end{bmatrix} \tag{B30}$$

From eq. (B29), we then deduce

$$\begin{bmatrix} C_1^+(x_i+\epsilon) \\ C_2^+(x_i+\epsilon) \end{bmatrix} = [I+B]^{-1}\ [I,-B] \begin{bmatrix} C_1^+(x_i-\epsilon) \\ C_2^+(x_i-\epsilon) \\ C_1^-(x_i+\epsilon) \\ C_2^-(x_i+\epsilon) \end{bmatrix} \tag{B31}$$

or equivalently

$$\begin{bmatrix} C_1^+(x_i+\epsilon) \\ C_2^+(x_i+\epsilon) \end{bmatrix} = [(I+B)^{-1}, -(I+B)^{-1}B] \begin{bmatrix} C_1^+(x_i-\epsilon) \\ C_2^+(x_i-\epsilon) \\ C_1^-(x_i+\epsilon) \\ C_2^-(x_i+\epsilon) \end{bmatrix} \tag{B32}$$

Similarly, by subtracting eqs. (B25) and (B26) and following a similar derivation, we obtain

$$\begin{bmatrix} C_1^-(x_i-\epsilon) \\ C_2^-(x_i-\epsilon) \end{bmatrix} = [-(I+B)^{-1}B, (I+B)^{-1}] \begin{bmatrix} C_1^+(x_i-\epsilon) \\ C_2^+(x_i-\epsilon) \\ C_1^-(x_i+\epsilon) \\ C_2^-(x_i+\epsilon) \end{bmatrix} \tag{B33}$$

Grouping the results (B32) and (B33), we obtain the final relation

$$\begin{bmatrix} C_1^-(x_i-\epsilon) \\ C_2^-(x_i-\epsilon) \\ C_1^+(x_i+\epsilon) \\ C_2^+(x_i+\epsilon) \end{bmatrix} = \begin{bmatrix} -(I+B)^{-1}B & (I+B)^{-1} \\ (I+B)^{-1} & -(I+B)^{-1}B \end{bmatrix} \begin{bmatrix} C_1^+(x_i-\epsilon) \\ C_2^+(x_i-\epsilon) \\ C_1^-(x_i+\epsilon) \\ C_2^-(x_i+\epsilon) \end{bmatrix} \tag{B34}$$

where the square matrix in the second member is the required Scatter matrix by definition.

For the general case of M modes, we can easily generalize the 2×2 matrix given in eq. (B30). The general expression of the $\alpha\beta$ element of the matrix B can be written

$$B_{\alpha\beta} = \frac{i}{2} \frac{\tilde{\Gamma}_{\alpha\beta}}{\sqrt{k_\alpha k_\beta}}. \tag{B35}$$

Appendix C: Ohm's Law

In section 2.2.2 of chapter 2, we argued that the semi-classical conductance of a random array of scatterers could be deduced by replacing the amplitude scatter matrix [s] by the probability scatter matrix [S]. In this appendix, we prove on a specific example that this will lead to the semi-classical result for the conductance of a random array of scatterers, i.e, Ohm's law.

Let us consider a random array of scatterers, all characterized by the same scatter matrix (see eq. (2.25) of chapter 2), and with independent spacing $\{d_n\}$ between them. For each of those scatterers, the probability scatter matrix can be written

$$[S] = \begin{bmatrix} T & R \\ R & T \end{bmatrix} \tag{C1}$$

Where the matrices [R] and [T] have the explicit form

$$[R] = \begin{bmatrix} \delta \ldots & \delta \\ \vdots & \\ \delta \ldots & \delta \end{bmatrix} \quad \text{and} \quad [T] = \begin{bmatrix} 1-(2M-1)\delta & & \delta \ldots & \delta \\ & \delta & & \vdots \\ & \vdots & & \cdot \\ & \delta & \ldots & 1-(2M-1)\delta \end{bmatrix} \tag{C2}$$

$$\text{where} \qquad \delta = \left| \frac{e^{2iM_\alpha} - 1}{2M} \right|^2 \tag{C3}$$

For the propagation between scatterers, the probability scatter matrix is given by (see eqs. (2.6) and (2.7) of chapter II)

$$[S^{(n)}] = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \tag{C4}$$

Now we cascade the probability scatter matrices in the same way as the amplitude scatter matrices. It is apparent eqs. (A3) to (A6) of appendix A that $[S^{(n)}]$ cascaded with any matrix [S] yields back [S]. In other words, the lengths $\{d_n\}$ will not appear in the final expression of the conductance as it should be in a semiclassical treatment. Therefore, the overall scatter matrix is obtained by cascading N identical sections with scatter matrices [S]. Consider first the result of cascading two sections

$$R_2 = [R] + [T][R](1-[R]^2)^{-1}[T] \tag{C5a}$$

We can write [R] and [T] as,

$$[R] = \delta[a] \tag{C5b}$$

$$[T] = (1-2M\delta)1 + \delta[a] \tag{C5c}$$

where,

$$[a] = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{bmatrix} \tag{C5d}$$

Equation (C5a) is simplified to yield,

$$R2 = [R] + (\frac{1}{1-\delta^2 M^2})[T][R][T] \tag{C6}$$

Here we have used the relation

$$R^2 = \delta M[R] \tag{C7}$$

Equation (C6) can be simplified further using eqs. (C5a-d).

$$[T][R][T] = [T](1-M\delta)[R]$$

$$= (1-M\delta)^2[R] \tag{C8}$$

Hence,

$$R_2 = \frac{2}{1+M\delta}[R] \tag{C9}$$

Thus R_2 can be written as,

$$R_2 = \delta_2[a] \tag{C10}$$

Where

$$\delta_2 = \frac{2\delta}{1+\delta} \tag{C11}$$

Note that,

$$\frac{M\delta_2}{1-M\delta_2} = 2\left(\frac{M\delta}{1-M\delta}\right) \tag{C12}$$

Similarly, by cascading two sections each having two scatterers we can show that,

$$R_4 = \delta_4[a] \tag{C13}$$

where

$$\frac{M\delta_4}{1-M\delta_4} = 2\,\frac{M\delta_2}{1-M\delta_2} = 4\,\frac{M\delta}{1-M\delta} \tag{C14}$$

we can continue this process indefinitely to get,

$$R_N = \delta_N[a] \tag{C15}$$

where

$$\frac{M\delta_N}{1-M\delta_N} = N\left(\frac{M\delta}{1-M\delta}\right) \tag{C16}$$

so that, the conductance g_N of N sections is obtained from eq. (2.4) in the text

$$g_N = 2M(1-M\delta_N)$$

$$= 2M\left(\frac{1-M\delta}{1+(N-1)M\delta}\right)$$

$$\simeq \frac{2M}{N\delta} \quad \text{if} \quad N >> \frac{1}{M\delta} \tag{C17}$$

Which is the classical result.

Appendix D: A Simple Derivation of Weak Localization

For a device length L satisfying the condition,

$$\Lambda < L < \Lambda_{Loc} \tag{D1}$$

where Λ and Λ_{Loc} are the elastic mean free path and localization length respectively, the normalized conductance of the sample can be derived from the approximate formula (see eq. (2.4) of chapter II)

$$g = G/(e^2/h) = 2tr(t^+ t) = 2\sum_{i,j}^{M} T_{ji}(E_F) \tag{D2}$$

Alternatively, we could write the normalized conductance in terms of the reflection amplitudes R_{ji}. Indeed, current conservation implies

$$\sum_{j=1}^{M}(T_{ji}+R_{ji}) = 1 \tag{D3}$$

This allows us to rewrite the normalized conductance as follows

$$g = 2[M-\sum_{i,j}^{M} R_{ji}] \tag{D4}$$

The weak localization phenomenon is due to the importance of the coherent backscattering which states for every incident mode i, we have

$$R_{ii} \sim \frac{2}{M} \tag{D5}$$

whereas

$$R_{ji} \sim \frac{1}{M} \quad \text{if } i \neq j \tag{D6}$$

Therefore, we expect the normalized quantum conductance g_Q to be different from its classical counterpart g_{cl} by an amount given by

$$g_Q - g_{cl} \simeq 2M\left(\frac{1}{M}\right) \simeq -2 \tag{D7}$$

LIST OF REFERENCES

[1] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, Phys. Rev. Lett. 42, 673 (1979).

[2] 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).

[3] 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).

[4] H. Van Houlen, B. J. Van Wess, M. G. J. Heijman and J. P. Andre, Appl. Phys. Lett. 49(26), (1986) 1781-1783. (1986).

[5] S. Y. Chou, D. A. Antoniadis, H. I. Smith and M. A. Kastner, Solid St. Comm. Vol 61, N0 9, pp 571-572 (1987).

[6] G. P. Whittington, P. C. Main, L. Eaves, R. P. Taylor, S. Thoms, S. P. Beaumont and C. P. W. Wilkinson, Int. Conf. " Superlattice Microstructures and Microdevices " Gothenberg, August 1986 (to be published in Superlattice and Microstrutctures).

[7] B. Ricco and M. Ya. Azbel, Phys. Rev. B 29(4), 1970 (1984).

[8] S. C. Feng, P. A. Lee and A. D. Stone, Phys.Rev.Lett. 56 (1986) 1960; Erratum 56 (1986) 2772.

[9] P. A. Lee, A. D. Stone and H. Fukuyama, Phys. Rev. B 35 (1987) 1039.

[10] B. L. Altshuler, JETP Lett. 41, 64 (1985).

[11] Y. Imry, Europhys. Lett. 1, 249 (1986).

[12] P. A. Lee, Physica A, 140 (1986).

[13] S. Washburn and R. A. Webb, Advances in Physics, 1986, Vol.35, No.4, 375-422. (and references therein); S. Datta, M. R. Melloch, S. Bandyopadhyay, R. Noren,

M. Vaziri, M. Miller, R. Reifenberger, Phys. Rev. Lett. 55, 2344 (1985).

[14] P. W. Anderson, Phys. Rev. 109, 1492 (1958); Phys. Rev. B, (1981) 4828-4836.

[15] A. D. Stone, Phys. Rev. Lett. 54(25), 2692 (1985).

[16] N. Giordano (preprint 1987).

[17] J. Sak and B. Kramer, Phys. Rev. B 24, 1761 (1981).

[18] A. Fuchs and G. Mahler, Solid State Commun. 55, 1035 (1985).

[19] C. J. Lambert and M. F. Thorpe, Phys. Rev. B 27, 715 (1983).

[20] P. Eordos and R. C. Henderson, Adv. Phys. 31, 65 (1982).

[21] A similar Scatter matrix approach has been used recently to characterize wave propagation in microstrip step discontinuities, T. S. Chu and T. Itoh, IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-34, 2, 280-284 (1986).

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

[23] R. Landauer, Internal memo, IBM Thomas J. Watson Research Center, (1966); IBM J. Res. Dev. 1, 223 (1985).

[24] M. Buttiker, Y. Imry, R. Landauer, and S. Pinhas, Phys. Rev. B 31, 6207 (1985).

[25] M. Ya. Azbel, J. Phys. C 14, L225 (1981).

[26] B. Shapiro, Phys. Rev. B 35, 8256 (1987).

[27] L. I. Schiff, Quantum Mechanics, Mc Graw Hill, New York (1968).

[28] E. N. Economou, Green's Functions in Quantum Physics, Springer Series in Solid-State Sciences 7 (1979).

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

[30] S. Bandyopadhyay, B. Das, R. Reinfenberger, S. Datta, W. P. Hong an P. K. Battacharya ; March Meeting of the American Physical Society at New-York, 1987.

[31] S. Datta, M. Cahay, M. Mclennan (to appear in Phys. Rev. B).

3. INFLUENCE OF IMPURITY SCATTERING ON THE PERFORMANCE OF AHARANOV-BOHM DEVICES

3.1 Introduction

Quite recently, a new class of interesting semiconductor devices based on quantum interference effects has been proposed involving electron transport parallel to the heterolayers. A typical example of those *Aharanov-Bohm* devices is shown in Fig. 3.1. The two wells are only separated by a few hundred angstroms and are modulation doped to reduce the effects of scattering. The quantum interference occurs not between multiple reflected waves but between two alternative paths provided by the two contiguous GaAs layers separated by an AlGaAs barrier. The interference pattern between the two paths is then modified by placing the structure in an external magnetic field parallel to the interfaces of the heterostructure. This leads to a current modulation by an external magnetic field which is periodic in flux with a period equal to the quantum flux, *i.e*, $\phi_0 = h/e$. This is the semiconductor analog of the Aharanov-Bohm effect (hereafter referred as A-B) observed in vacuum over twenty-five years ago [1].

The possibility to observe those oscillations in semiconductor heterostructures was quite questionable since, contrary to the case in vacuum, the presence of impurities and various kind of defects (resulting from the imperfection of the interfaces during growth) were supposed to destroy any interference effect and wash out the A-B oscillations. However, in 1984, the first observation of A-B oscillations in semiconductor heterostructures (AlGaAs-GaAs) were reported (see Fig. 3.2) [2]. Back then, the size of the conductance fluctuations was about 5 e^2/h corresponding to a relative conductance modulation $\Delta G/G$ of about .5 %. (e^2/h is the usual unit of conductance in all the experimental and theoretical work concerned with the A-B oscillations). More recently, experiments performed by the Purdue group with different materials have shown conductance modulation (peak to peak) as big as 70 e^2/h [3] corresponding to $\Delta G/G$ of about 10%. The barrier layers are of InAlAs while the wells are made of InAs-GaAs monolayer superlattices to eliminate alloy scattering. Oscillatory magnetoresistance due to the A-B effect has also observed in small metallic gold and silver rings [4] but the size of the relative conductance modulation was only about 0.001 % in those devices.

The origin of the h/e oscillations in both metals and semiconductor structures is now well understood following the extensive theoretical work of a number of researchers [5-7]. They show that all that is required for the observation of the A-B effect is that the electrons retain *phase memory* while crossing the two channels. The length over which memory is maintained (the phase coherence length) can be much longer than the random walk step length Λ_{el} (the mean free path). In other words, elastic scattering doesn't cause the electrons to loose phase memory; only inelastic processes

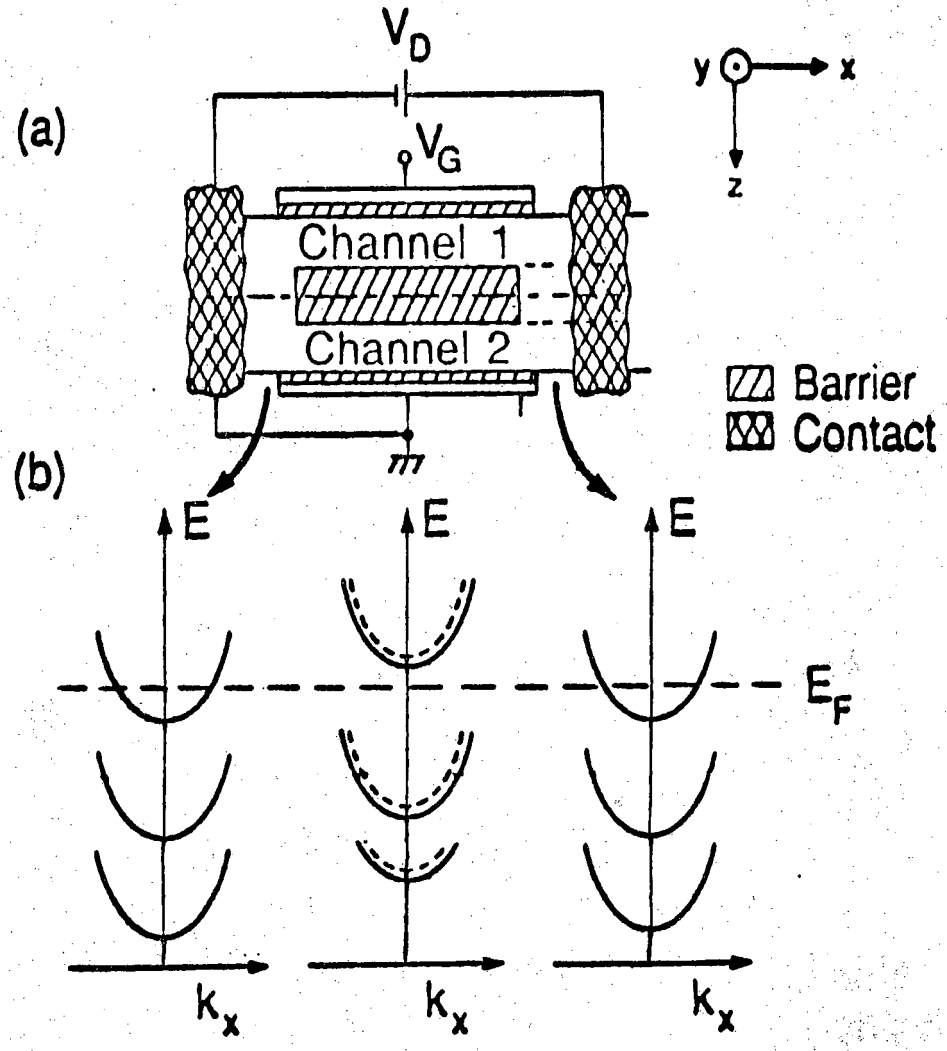


Figure 3.1 (a) Typical Aharanov-Bohm semiconductor heterostructure. The device is uniform along y with a width W_y. (b) Dispersion relations for different transverse modes in the channels and outside regions.

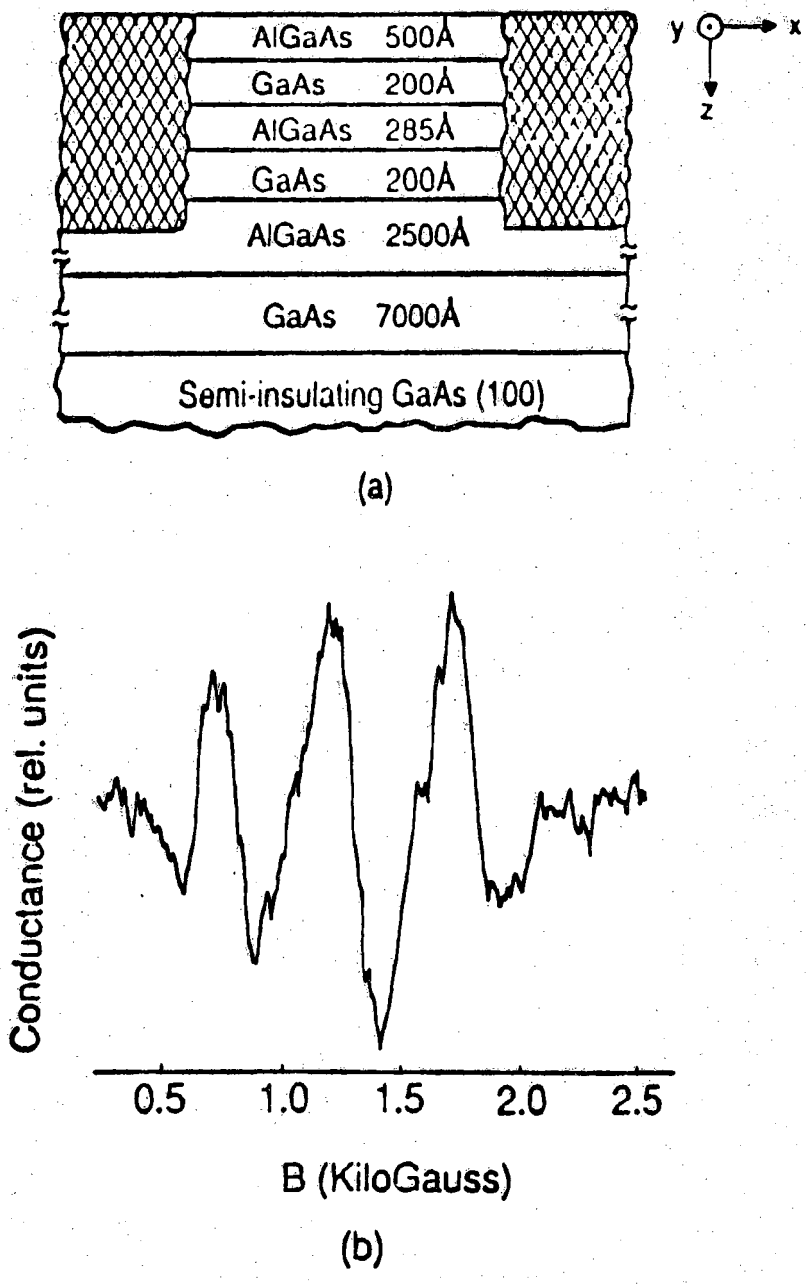


Figure 3.2 (a) Experimental device used by Datta *et al.* [2]. In this GaAs double quantum well structure, the length of the channels L $\sim$ 2 μm and W_y $\sim$ 2 μm. (b) Measured conductance modulation of the structure shown in Fig. 3.2.(a). The peak to peak conductance modulation $\Delta G \sim 5\ e^2/h$.

lead to undeterministic phase randomization of the electron wavefunction. In all those experiments, it is therefore important to keep the length of the arms of the A-B devices shorter than the inelastic diffusion length. This can always be achieved by working at sufficiently low temperatures.

The difference in size of the conductance modulation observed in metals and semiconductors has been shown to be related to the difference in transport mechanism prevailing in the two type of materials. In a metal, the electron's trajectory is a random walk with an average step length Λ_{el} (One may instead assume an average collision rate $1/\tau$, and that the mean free path length is the product of the Fermi speed and the mean time between collision, $\Lambda_{el} = v_f\tau$). Mean free paths in common metals depend strongly on the cleanliness of the material and on the quality of the crystal that forms the wire. In materials that have many lattice defects, for instance, chemical impurities, vacancies or grains, the mean free path is usually of the order of the average distance between scattering sites. Typically in metal prepared by standard methods, this length is only a few nanometers or a few tens of nanometers. The motion of the electron in a metal is therefore interrupted by frequent scattering events since the size of device is large compared with the mean free path length, $\Lambda_{el} << L$.

In semiconductor heterostructures, the situation is quite different [8]. Advances in semiconductor microtechnology have made it possible to fabricate extremly high-mobility conductive channels isolated from the surface by epitaxial insulators. In GaAs, an electron with a velocity 10^7cm/sec has an elastic (plus inelastic) mean free path as long as 4μm, if the mobility is 10^6cm^2/V—sec. Channel lengths of .25 μm or less can be fabricated by present-day technology so that an electron can travel from one contact to another ballistically with essentially no scattering - elastic or inelastic. In this ballistic regime, as we will review in this chapter, it may be possible to attain large (approaching 100%) conductance modulation in a magnetic field [8] even if the transverse dimension (y direction in Fig. 3.1.(a)) is large, *i.e*, when the structure is multi-moded. Furthermore, the number of propagating modes in semiconductor heterostuctures can be around a few tens only. This is another striking difference with the case of small metallic rings where the number of propagating channels can be as large as 10^5.

Finally, a new device concept has also been suggested recently in which the quantum interference between the two arms of the A-B structure would be controled by an external gate voltage [9] rather than by an external magnetic field. Even though there has yet not been any experimental demonstration of this so called *Electrostatic* A-B effect, the fact that the phase difference between the two paths can be changed by π with a very small gate voltage (1 mV) suggests that the Quantum Interference Transistor (QUIT) based on this concept should have high transconductances and low power dissipation. Additional work to design better structures and demonstrating the potentiality of the QUIT device is therefore necessary. This will be the subject of the

last part of this chapter where we will also investigate the influence of impurity scattering on the conductance modulation of A-B devices by an external gate voltage.

This chapter is organized as follows. In sections 3.2 and 3.2.1, we describe the typical A-B heterostructures and show explicitely that, in realistic semiconductor devices, the transport is due to a few (< 100) propagating channels only. This is done by calculating the number of propagating modes available in each channel at the Fermi energy. This calculation takes into account the contact potentials existing between the different regions of the device due to the space charge effect resulting from the specified doping densities in the device.

In sections 3.2.2 and 3.2.3, we present a simple theory to calculate the conductance of A-B devices based on the Landauer approach following the discussion given in ref. [8]. However, we extend the theory in order to include the multiple reflections at the ends of the two channels and the influence of impurity scattering on the performance of the A-B devices [the impurity potential is modeled as discussed in detail in sections 2.3.1 and 2.3.2 of chapter 2].

Section 3.3 is devoted to the analysis of the magnetic A-B effect and study the influence of impurity scattering on the size of the conductance modulation. We make the important distinction between the case where the impurity scattering into both arms is either *correlated* or not.

Section 3.4 consists in a theoretical treatment of the *Electrostatic* A-B effect which has not yet been observed experimentally. Our theoretical treatment is useful in designing better structures for experimental demonstration of this effect. Some numerical examples are used to show that the size of the conductance modulation by an external gate voltage can still be substantial even in the presence of impurity scattering into both arms. Finally, section 3.5 contains our discussion and conclusions.

3.2 Theory

3.2.1 Device Structure

Consider a structure (Fig. 3.1.(a)) in which a conductive channel splits into two isolated channels and then merges into one. The structure is assumed uniform in y having a width W_y. In each of the two end regions ($x < 0$ and $x > L$) and in the middle region ($0 < x < L$) we can calculate a set of transverse modes (or subbands) in the z-direction by solving the eigenvalue problem

$$\left[-\frac{\hbar^2}{2m^*}\frac{d^2}{dz^2} + E_c(z) \right] |n> = \epsilon_n |n> \tag{3.1}$$

The potential energy $E_c(z)$ includes the conduction band discontinuities as well as any band bending due to an applied electric field E_z or due to space charge effects. Since the function $E_c(z)$ is different in the three regions, the transverse modes obtained by solving eq. (3.1) are also different. Each subband has a parabolic dispersion given by

$$E(k_x,k_y) = \epsilon_n + \frac{\hbar^2 k_x^2}{2m^*} + \frac{\hbar^2 k_y^2}{2m^*} \tag{3.2}$$

The dispersion curves are sketched in Fig. 3.1.(b). Hereafter, we will assume that the device (that is, $E_c(z)$) is perfectly symmetric about $z = 0$ as any interferometer should be and that only the lowest subband is occupied everywhere in the device (*i.e*, $n = 1$). The importance of this assumption will be discussed later.

It will be noted that there will be contact potentials between the different regions due to space charge that will cause a vertical displacement of the dispersion curves in one region relative to other regions. The relative positioning of the subband in the three regions at equilibrium is fixed by the requirement of a constant Fermi level for the specified doping densities and will be considered in detail shortly.

The criterion for single-modedness can be satisfied with lithographically defined structures but they are difficult to fabricate. Alternative schemes utilizing film growth techniques such as MBE to define the channels have been discussed in ref. [10]. On the other hand, it is quite difficult practically to reduce litographically the dimension of the device in the y-direction to make it single-moded along this axis. Therefore, we allow multiple modes with different k_y values to propagate throughout the structure. These allowed k_y eigenstates strongly depend on the shape of the confining potential in the y direction. In all our numerical examples, the confinement in the y direction will be of the *particle in a box* type, the size of the box being the effective channel width along the y axis after etching the device in that direction (the theory is however valid for any arbitrary potential energy profile in that direction). Generally, we will be concerned with low temperature behavior of the device so that the subbands fill up to the Fermi level. The energy dispersion relation is then given by eq. (3.2) where

$$\epsilon_1 = \frac{\pi^2\hbar^2}{2m^*W^2} \quad \text{and} \quad k_y = \frac{m\pi}{H} \tag{3.3}$$

if the potential confinement is of the *particle in a box* type in both the y and z directions (we have set n=1 in eq. (3.2) since the structure is assumed to be single moded

in the z direction). The A-B structure is then similar to a rectangular waveguide for electrons such as the one shown on Fig. 3.3 in which the current flows in the x direction.

The density of states per unit length inside the rectangular waveguide can then be written

$$\frac{dn}{dE} = \frac{m^*}{\pi\hbar^2} \sum_{m}^{M} \frac{\theta(E - E_m^y - \epsilon_1)}{\sqrt{\frac{2m}{\hbar^2}(E - E_m^y - \epsilon_1)}} \tag{3.4}$$

where

$$E_m^y = \frac{m^2\hbar^2\pi^2}{2m^*H^2} \tag{3.5}$$

Equation (3.4) is seen to diverge at each new subband energy. At zero degree Kelvin, this density of states, when integrated up to the Fermi level gives the total number of occupied states per unit length. Performing the integration analytically, we obtain

$$n_S H = \frac{2}{\pi} \sum_{m=1(n=1)}^{M} k_x^m \tag{3.6}$$

n_S being the surface charge density in the portion of the electron waveguide with width W and height H; k_x^m is the longitudinal part of the wavevector for electron with total incident energy E_f. Finally, M is the total number of propagating modes necessary to include in the summation appearing in eq. (3.4) to make the two sides of the equation equal to each other.

In order to calculate the built-in potential in the channel region, we use eq. (3.6) with n_S set equal to $n_S(W/W')$ in order to keep the carrier density the same everywhere inside the device. In this region, the different longitudinal wavevectors k_x^m are deduced from eq. (3.2) after replacing E_f by $E_f - V_0$, V_0 being the built-in potential in the channel region. The Fermi level and built-in potential were calculated for the device parameters listed in Table 3.1. The A-B devices were assumed modulation doped so that the surface charge density is $10^{12}cm^{-2}$ everywhere inside the electron waveguide. This charge density is low enough to insure the single-modedness in the z-direction throughout the device. Notice that V_0 is actually negative, *i.e* , the conduction band energy profile is actually bending downwards in the channels region. As can be seen from table 3.1, for device dimensions within the reach of present day technologies (such as MBE) and litography techniques, there is only a few propagating modes (< 100) throughout the structure, *i.e*, with positive longitudinal wavevectors k_x^m available at the Fermi energy. This is a completely different regime than the

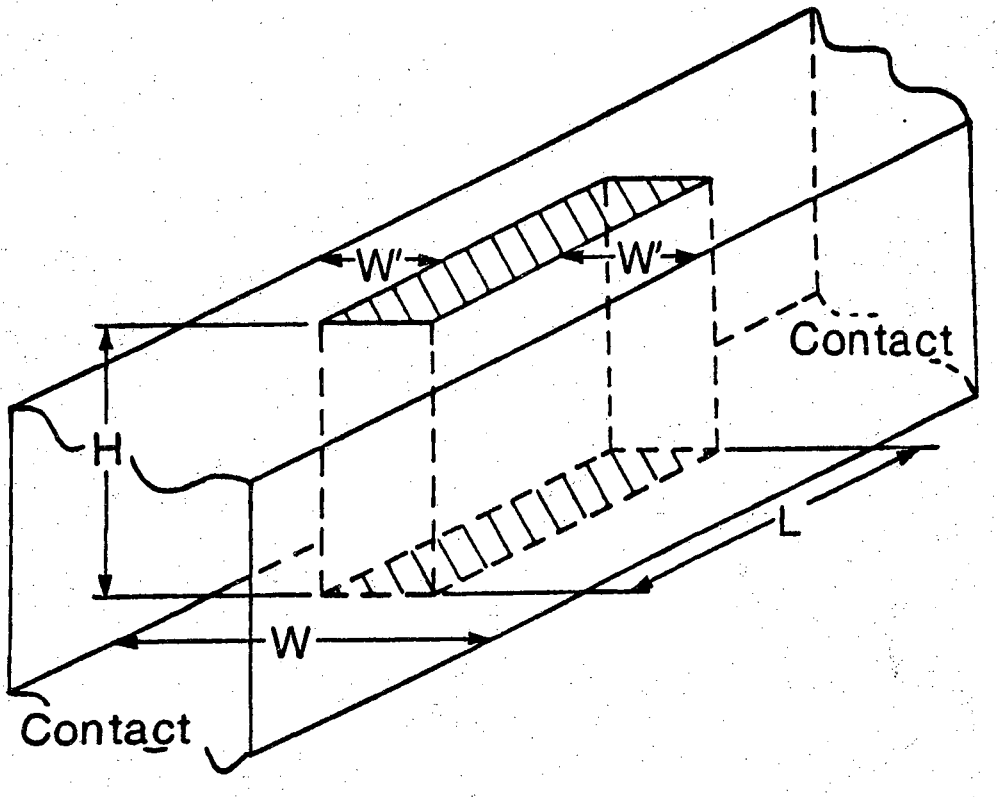


Figure 3.3 With a *particle in a box* confinement in the y and z directions, The A-B structure is similar to a rectangular waveguide for electrons. W, W′ and H are the widths of the outside, channel regions and the height of the electron waveguide respectively.

diffusion regime needed to describe the transport properties of metallic rings in which the number of propagating modes can be as high as 10^5.

Having characterized the device parameters of the A-B semiconductor heterostructures, we will next discuss the Landauer approach to calculate the conductance modulation of those devices. Then, in sections 3.3 and 3.4, we will consider in detail the case of the magnetic and electrostatic A-B effects respectively.

Table 3.1 - Device Parameters

Parameters	Fermi level (E_F)	Built-in potential	Number of prop. modes in contacts (channels)
W = 150 Å W' = 50 Å H = 0.5 μm	0.061228 eV	-0.17549 eV	40 (23)
W = 150 Å W' = 50 Å H = 1 μm	0.060956 eV	-0.17561 eV	80 (46)

3.2.2 Conductance Formula

For the typical A-B device shown in Fig. 3.1, The current I for an applied drain voltage V_D is given by [8],

$$I = \frac{2e}{h} \int dE\, [f(E) - f(E + eV_D)] \sum_{\substack{k_y', k_y'' \\ n', n''}} \left| (T_{tot})_{n'', k_y''; n', k_y'}(E) \right|^2 \tag{3.7}$$

where $(T_{tot})_{n'', k_y''; n', k_y'}(E)$ is the transmission coefficient from wavevector k_y' in subband n' at the left end region to wavevector k_y'' in subband n'' at the right end region. Starting with eq. (3.7) and under extremely small drain bias, the conductance of the device is given by

$$G = \frac{2e^2}{h} \int dE(-\frac{df}{dE}) \sum_{\substack{k_y', k_y'' \\ n', n''}} \left| (T_{tot})_{n'', k_y''; n', k_y'}(E) \right|^2 \tag{3.8}$$

At low temperatures, we can replace $-\frac{df}{dE}$ by a delta function at the Fermi level so that the normalized conductance g becomes

$$g = \frac{G}{(\frac{2e^2}{h})} \sum_{k_y', k_y''} \left| (T_{tot})_{k_y'', k_y'}(E) \right|^2 \tag{3.9}$$

in which we have dropped the indices n′ and n′′ since the structure is assumed to be single moded in the z direction. This (Landauer) formula was used in chapter 2 to study the problem of localization in extremly small 2-D resistors. The evaluation of the total transmission matrix T_{tot} in eq. (3.9) is more involved in the case of the A-B devices. Indeed, we must include the effect of multiple reflections occuring at both ends of the channels (points P and Q of Fig. 3.1). This will be explained in detail in the next section.

3.2.3 Derivation of the Overall Transmission Coefficient Through the Aharanov-Bohm Device

In order to calculate the total transmission coefficient appearing in eq. (3.9), we need to consider the multiple reflections which occur at the edges of the two channel regions (Points P and Q on Fig. 3.1). To that extent, we divide the A-B device into three sections, two of which are the junctions at the ends of the channels (Points P and Q on Fig. 3.1) and the channels region. By cascading the scatter matrices associated to those individual sections as discussed in appendix A, we can then deduce the total transmission coefficient from one side of the device to the other. This calculation is now outlined in detail.

First, we need to calculate the scatter matrix at each of the two junctions (x = 0 and x = L). At those points, there is no coupling between modes since the potential is supposed to be uniform along the y-direction at the two splitters. The only coupling between modes comes from the impurities located in the two channels of the device and will be considered later.

At the left junction (x=0.) in Fig. 3.1 we can write a scatter matrix connecting the wave amplitudes for a given E and k_y at $x = 0^-$ (A) to those at $x = 0^+$ (B)

$$\begin{pmatrix} B_1^+ \\ B_2^+ \\ A^- \end{pmatrix} = \begin{pmatrix} t_1 & r_{11}' & r_{12}' \\ t_2 & r_{21}' & r_{22}' \\ r & t_1' & t_2' \end{pmatrix} \begin{pmatrix} A^+ \\ B_1^- \\ B_2^- \end{pmatrix} \tag{3.10}$$

where the subscripts '+' and '-' denote the amplitude of waves travelling along positive and negative x, respectively, while the subscripts '1' and '2' refer to channels 1 and 2, respectively. Furthermore, the primed (unprimed) quantities appearing in the square matrix characterize the various reflection and transmission coefficients while

travelling from right to left (left to right) respectively. The amplitudes are defined as $\sqrt{k_x}$ times the wavefunction so that the current are proportional to the squares of the amplitudes; with this definition the scatter matrix must be unitary since the current is conserved. We can write (3.10) more compactly as,

$$\begin{pmatrix} B^+ \\ A^- \end{pmatrix} = \begin{pmatrix} t & r' \\ r & t' \end{pmatrix} \begin{pmatrix} A^+ \\ B^- \end{pmatrix} \tag{3.11}$$

This relation can be easily extended to the case of n propagating modes (differing by their k_y values). Since there is no coupling between modes at the splitters, the quantities (r,r',t,t') appearing in eq. (3.11) become then diagonal matrices whose different diagonal elements are the respective reflection and transmission coefficients of the various modes at the splitters. In that case, A^+,A^- and B^+,B^- are $M\otimes 1$ and $2M\otimes 1$ column vectors respectively, M being the total number of allowed k_y values resulting from the confining potential in the y direction; r,r',t,t' are matrices with dimensions $M\otimes M$, $2M\otimes 2M$, $2M\otimes M$ and $M\otimes 2M$ respectively. For the case of a perfectly symmetrical structure, the various blocks of the scatter matrices at the right junction (x=L) can be obtained from the scatter matrix at the left junction (3.11) by making use of the following substitution

$$(r,t) \Rightarrow (r',t') \tag{3.12}$$

Therefore, at the right junction, we can write the scatter matrix connecting the wave amplitudes at $x=L^-$ (C) to those at $x=L^+$(D) as follows

$$\begin{pmatrix} D^- \\ C^+ \end{pmatrix} = \begin{pmatrix} t' & r \\ r' & t \end{pmatrix} \begin{pmatrix} C^+ \\ D^- \end{pmatrix} \tag{3.13}$$

Since the two channels are isolated, the wave amplitudes at B an C are simply related as follows

$$\begin{pmatrix} C^- \\ B^+ \end{pmatrix} = \begin{bmatrix} T & R' \\ R & T' \end{bmatrix} \begin{pmatrix} B^+ \\ C^- \end{pmatrix} = \begin{pmatrix} t_1 & 0 & r_1' & 0 \\ 0 & t_2 & 0 & r_2' \\ r_1 & 0 & t_1' & 0 \\ 0 & r_2 & 0 & t_2' \end{pmatrix} \begin{pmatrix} B^+ \\ C^- \end{pmatrix} \tag{3.14}$$

where

$$s_1 = \begin{bmatrix} t_1 & r_1' \\ r_1 & t_1' \end{bmatrix} \quad \text{and} \quad s_2 = \begin{bmatrix} t_2 & r_2' \\ r_2 & t_2' \end{bmatrix} \tag{3.15}$$

are the scatter matrices characterizing the propagation of the different modes in channel 1 and 2 respectively. Both r_i and r_i' (i=1,2) are non-zero in the presence of impurity scattering in the channels. Furthermore we assumed the two channels to be independent. The two scatter matrices in eq. (3.45) can then be generated independently by taking into account the impurity configuration in each of the two arms as was described in the previous chapter. For the special case of ballistic transport in the two arms, eq. (3.14) must be replaced by

$$\begin{pmatrix} C^- \\ B^+ \end{pmatrix} = \begin{bmatrix} P & 0 \\ 0 & P' \end{bmatrix} \begin{pmatrix} B^+ \\ C^- \end{pmatrix} \tag{3.16}$$

P is a diagonal matrix whose diagonal elements describe the phase-shifts in the two channels., *i.e*,

$$P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix} \tag{3.17}$$

where

$$(P_1)_{mn} = \exp[ik_x^{m_I}]\delta_{mn} \tag{3.18}$$

and

$$(P_2)_{mn} = \exp[ik_x^{m_{II}}]\delta_{mn} \tag{3.19}$$

For a particular energy, $k_x^{m_I}$ and $k_x^{m_{II}}$ are the wavenumbers in the x-direction in channels 1 and 2 respectively and are deduced from eq. (3.2). As long as there are no magnetic or electric field $P' = P$. The presence of either one of them will change the wavevectors of the different propagating modes in the two channels. This leads to a phase-shift of the different modes between the two arms responsible for the A-B oscillations. This will be explained in detail in sections 3.3 and 3.4.

We now have scatter matrices (3.11,3.13,3.14) connecting A and B, B and C and C and D, respectively. What we need is the transmission matrix T from A^+ to D^+, so that we can calculate the conductance from eq. (3.9). The easiest way to calculate this transmission matrix is to get the overall scatter matrix of the structure according

to the law for cascading scatter matrices described in appendix A. The total transmission matrix across the device after cascading the scatter matrices (3.11),(3.13) and (3.14) in that order is found to be

$$T_{tot} = t' \left[I - (R' + T[I - r'R]^{-1} r' T')r'\right]^{-1} T \left[I - r'R\right]^{-1} t \tag{3.20}$$

where we made use of eqs. (A3) to (A6) in appendix A of chapter II. In the case of ballistic transport, both R and R' are identically zero and eq. (3.20) reduces to

$$T_{tot} = t' \left[I - Tr' T' r'\right]^{-1} Tt \tag{3.21}$$

which was analyzed in detail in ref. [8]. Neglecting the multiple reflections at the junctions is equivalent to making the following substitution in eq. (3.20)

$$[I - \cdots]^{-1} \rightarrow I \tag{3.22}$$

In that case, the total transmission matrix can then be written

$$T_{tot} = t' \; T \; t = t' \begin{pmatrix} t_1 & 0 \\ 0 & t_2 \end{pmatrix} t \tag{3.23}$$

t_1, t_2 being the transmission matrices for channel 1 and 2 respectively. For the special case where the junction scatter matrix is approximated by a Shapiro scatter matrix [11]

$$S = \begin{pmatrix} \sqrt{\epsilon} \times I & a \times I & b \times I \\ \sqrt{\epsilon} \times I & b \times I & a \times I \\ -(a+b) \times I & \sqrt{\epsilon} \times I & \sqrt{\epsilon} \times I \end{pmatrix} \tag{3.24}$$

where

$$a = \frac{1}{2} \left(\sqrt{1-2\epsilon} - 1\right) \tag{3.25}$$

and

$$b = \frac{1}{2} \left(\sqrt{1-2\epsilon} + 1\right) \tag{3.26}$$

and I is the M$\otimes$ M identity matrix. Equation (3.23) can then be written

$$T_{tot} = \epsilon(t_1 + t_2) \tag{3.27}$$

A more rigorous approach for the junction scatter matrix could be deduced by requiring that the wavefunction and its derivative be continuous across the interfaces at x=0. and x=L following the technique described in ref. [12].

3.3 Magnetic Aharanov-Bohm Effect

We first consider the case where the structure shown in Fig. 3.5 is placed in a magnetic field directed along the positive y axis. Ideally, in order to observe *pure* A-B oscillations, we would like the magnetic field to be located in the shaded region of Fig. 3.5 only. This is obviously not realizable in practice since both the contacts and device feel the presence of the external magnetic field. Rigorously, one should derive the scatter matrix of the two arms taking into account the fact that the magnetic field is piercing both channels. The problem of a single channel in a magnetic field has been extensively studied in the literature in conjunction with the observation of the Quantum-Hall effect [13]. However, this problem has mainly been concerned with the high magnetic field limit. Here, on the other hand, we are mainly interested with the low field limit. Indeed, several A-B oscillations have been observed in semiconductor heterostructures for field intensity lower than one Tesla [2].

Furthermore, as pointed out by Stone [14], the A-B oscillations are likely to be seen when the aspect ratio of the structure is as small as possible. The aspect ratio is defined as the ratio between the area of each channel and the area of the region separating them. In case the aspect ratio is poor, the h/e oscillations can still be present but are superimposed to aperiodic conductance oscillations fluctuations versus magnetic field similar to those discussed in chapter 2.

Since our main interest is to describe the influence of impurity scattering on the A-B oscillations with period h/e, we will assume an infinite aspect ratio and assimilate both channels to two two-dimensional resistors extending in the x and y directions only (see Fig. 3.1).

Let us thus consider two two-dimensional resistors located at z=0 and z=z_0, with a magnetic field B_y. The vector potential A_x is equal to $B_y z$. Thus, neglecting the extent of the resistors in the z-direction, we have $A_x = 0$ in resistor 1, and $A_x = B_y z_0$ in resistor 2. Consequently, the dispersion relation is modified from eq. (3.2) to

$$E = \epsilon_m + \frac{(\hbar k_m - eB_y z_{0)}^2}{2m^*} \tag{3.28}$$

for resistor 2. In this equation, we have

$$\epsilon_m = \frac{m^2 \pi^2 \hbar^2}{2m^* H^2} \tag{3.29}$$

where H is the width of the resistors in the y direction and the confining potential has been assumed to be of the *particle in a box* type in that direction. The dispersion relations are thus shifted horizontally by $\Delta k_m = eB_y z_0/\hbar$ for resistor 2. Since Δk_m is the same for every mode m and also for forward and reverse wave, the amplitude for every Feynman path in resistor 2 is phase-shifted by exactly the same amount θ_B.

$$\theta_B = eB_y z_0 L/\hbar \tag{3.30}$$

The conductance g can be deduced from eqs. (3.9) and (3.27)

$$g = 2\,\epsilon^2 \sum_{m=1}^{M} \sum_{m'=1}^{M} \left| t^{(1)}_{m',m} + t^{(2)}_{m',m} e^{i\theta_B} \right|^2 \tag{3.31}$$

where the superscripts 1 and 2 refer to resistors 1 and 2, respectively. Equation (3.31) neglects multiple reflections at the two ends, so that only the h/e oscillation is predicted. Detailed calculations taking into account the scatter matrices for the three-way splitters at the two ends show h/2e and higher order oscillations [15]. Here we use the simplified expression in eq. (3.31) which can be written as,

$$g = \epsilon^2 (g_1 + g_2) + g_c \cos\theta_B + g_s \sin\theta_B \tag{3.32}$$

where g_1 and g_2 are the individual conductances of resistors 1 and 2 respectively, while g_c and g_s are given by

$$g_c = 4\,\epsilon^2 \sum_{m=1}^{M} \sum_{m'=1}^{M} \mathrm{Re}(t^{(1)}_{m',m}\, t^{(2)}_{m',m}{}^*) \tag{3.34}$$

$$g_s = 4\,\epsilon^2 \sum_{m=1}^{M} \sum_{m'=1}^{M} \mathrm{Im}(t^{(1)}_{m',m}\, t^{(2)}_{m',m}{}^*) \tag{3.35}$$

The magnitude of the conductance modulation g_B in a magnetic field is therefore given by

$$g_B = \sqrt{g_c^2 + g_s^2} \tag{3.36}$$

g_c and g_s can be rewritten as follows

$$g_c = 4\,\epsilon^2 \sum_{m=1}^{M} \sum_{m'=1}^{M} |t^{(1)}_{m',m}|\,|t^{(2)}_{m',m}| \cos(\phi^{(1)}_{m',m} - \phi^{(2)}_{m',m}) \tag{3.37}$$

and

$$g_s = 4\,\epsilon^2 \sum_{m=1}^{M} \sum_{m'=1}^{M} |t^{(1)}_{m',m}|\,|t^{(2)}_{m',m}| \sin(\phi^{(1)}_{m',m} - \phi^{(2)}_{m',m}) \tag{3.38}$$

where $\phi^{(1)}_{m',m}$ and $\phi^{(2)}_{m',m}$ are the phases of the various transmission coefficients $t^{(1)}_{m',m}$ and $t^{(2)}_{m',m}$.

While considering many ring configurations, we follow an argument due to Lee [16] and assume that after averaging over many samples

$$< |t^{(1)}_{m',m}|\,|t^{(2)}_{m',m}| > \sim \frac{\Lambda_{el}}{ML} \tag{3.39}$$

Λ_{el} being the elastic mean free path (assumed to be the same in both arms); L is the length of both channels and M is the number of propagating modes in both channels (Here $< >$ denotes averaging over impurity configurations). The conductance modulation is then given by

$$<g_B> = \frac{4\epsilon^2 \Lambda_{el}}{ML} < \Bigg[\Big(\sum_{m=1}^{M} \sum_{m'=1}^{M} \cos(\phi^{(1)}_{m',m} - \phi^{(2)}_{m',m}) \Big)^2 + \Big(\sum_{m=1}^{M} \sum_{m'=1}^{M} \sin(\phi^{(1)}_{m',m} - \phi^{(2)}_{m',m}) \Big)^2 \Bigg]^{1/2} > \tag{3.40}$$

In the case where the two arms have exactly identical impurity configurations, we have

$$\phi^{(1)}_{m',m} = \phi^{(2)}_{m',m}$$

for all m and m′. From eq. (3.40), we therefore deduce

$$<g^c_B> = 4\epsilon^2 M \frac{\Lambda_{el}}{L} = 2\epsilon^2 g_1 \tag{3.41}$$

where g_1 is the conductance of one arm; the extra supscript *c* has been added since eq. (3.41) is valid in the case of *correlated* scattering only. This predicts that, for the case of *correlated* scattering, we should get 100 % conductance modulation even if the structure is multi moded (along the y direction). This is an important result since it predicts that any scattering potential that affects both channels equally has no effect on the interference even if it is inelastic or time-dependent. (For instance, long wavelength acoustic phonon should thus have no effect on the interference). Physically, this is a consequence of the fact that the interference pattern is not affected if

the entire interferometer is jiggled. It is thus important to have the two arms of the interferometer physically as close as possible (This is certainly of the semiconductor A-B devices where the separation between channels is a few hundred angstroms only and is another essential difference with the case of metallic rings with air gaps between the two conducting channels).

From eq. (3.41), the conductance modulation g_B is predicted to decrease as $1/N_I$ in the weak localization regime. Once the arm length is such that each resistor enters the strong localization regime, the conductance modulation should decrease exponentially with the length of the arm. As we will show below, our numerical simulations seem to be in agreement with those predictions.

Furthermore, as readily seen from eq. (3.41), the variance of the conductance modulation is predicted to be twice the variance of g_1, the conductance of each individual arm. From the universal conductance result, $\mathrm{var}(g_1) = 0.5\ e^2/h$. We therefore predict that, in the case of *correlated* scattering, the variance of the A-B conductance modulation is approximately given by

$$\mathrm{var}(g_B^c) = 2\epsilon\ \mathrm{var}(g_1) \sim \epsilon^2 e^2/h \tag{3.42}$$

In the case of *uncorrelated* scattering (*i.e*, when the two arms have completely different impurity configurations), the conductance modulation must be evaluated from eq. (3.40). If we assume that the phase differences $\phi^{(1)}_{m',m} - \phi^{(2)}_{m',m}$ are completely random over 2π, we obtain after averaging over ring configurations

$$\begin{aligned} <g_c^2> &= <\left[4\epsilon^2 \sum_{m=1}^{M}\sum_{m'=1}^{M} |t^{(1)}_{m',m}|\,|t^{(2)}_{m',m}| \cos(\phi^{(1)}_{m',m} - \phi^{(2)}_{m',m})\right]^2> \\ &= <g_s^2> = \frac{M^2}{2}\left(\frac{4\epsilon^2 \Lambda_{el}}{ML}\right)^2 \end{aligned} \tag{3.43}$$

Substituting those results back into eq. (3.41), the (average) conductance modulation g_B in the case of *uncorrelated* scattering into both arms is then given by

$$<g_B^u> = 4\epsilon^2\ \frac{\Lambda}{L} \tag{3.44}$$

where the extra supscript added as a reminder of the validity of this result for the case of *uncorrelated* scattering only. Taking the ratio of eqs. (3.41) and (3.44), we obtain the very simple result

$$\frac{<g_B^c>}{<g_B^u>} = M \tag{3.45}$$

for the ratio of the conductance modulation in the presence of *correlated* or *uncorrelated* scattering into both arms. We should emphasize that eqs. (3.41) through (3.45) were derived assuming the validity of eq. (3.39). Therefore, we expect eqs. (3.41-3.45) to be valid wherever eq. (3.39) holds, *i.e,* in the weak localization regime only.

We checked all our qualitative arguments on a specific numerical example. The two arms of the A-B devices were modeled as purely 2-D resistors such as those considered in chapter 2. We considered the specific case where as many as 30 modes are allowed to propagate into each channel. The impurity potential was modeled using the impurity model discussed in section 2.2.4.2 of chapter 2. We calculated g_B using eq. (3.36) in which the transmission coefficients were calculated as discussed in the numerical examples of chapter 2. The number of impurities N_I in both channels was progressively increased so that each resistor has a conductance behavior going from the quasi-ballistic to the strong localization regime. The parameters of the impurity model potential were chosen such that the elastic mean free path Λ_{el} in each resistor is equal to 33 impurities (see eq. (2.30) of chapter 2).

Figure 3.4.a) shows the conductance modulation g_B obtained in the ballistic regime both in the case of *correlated* and *uncorrelated* scattering. Figure 3.4.b) shows g_B in the region going from the weak to the strong localization regime. We notice in both figures a steadily decrease of g_B versus the number of impurities crossed into both arms in agreement with our qualitative discussion above. g_B is smaller in the case of *uncorrelated* scattering (see the triangles in Figs. 3.4.a) and 3.4.b) as one would expect by comparing eqs. (3.40) and (3.41). This also tells us that we no longer have 100 % conductance modulation in the case where the impurity configurations in the two arms are completely different.

Figure 3.5 compares the variance of the conductance modulation obtained in the case of *correlated* and *uncorrelated* scattering. for the former, var(g_B) is in very good agreement with the prediction (3.42) over the entire weak localization range ($33 \ll N_I < 1000$). for the latter, var(g_B) is considerably below the universal value e^2/h. This can be understood by using eq. (3.40). Indeed, there should be less spread (smaller variance) while averaging over ring configurations with different arms. Since the phase differences $\phi^{(1)}_{m,m} - \phi^{(2)}_{m,m}$ are presumably randomly distributed over 2π.

Finally, Fig. 3.6 shows a plot of the ratio defined in eq. (3.45). In the weak localization regime ($33 \ll N_I < 1000$), the ratio between $< g_B^u >$ and $< g_B^c >$ is indeed close to M, the number of propagating modes into both channels.

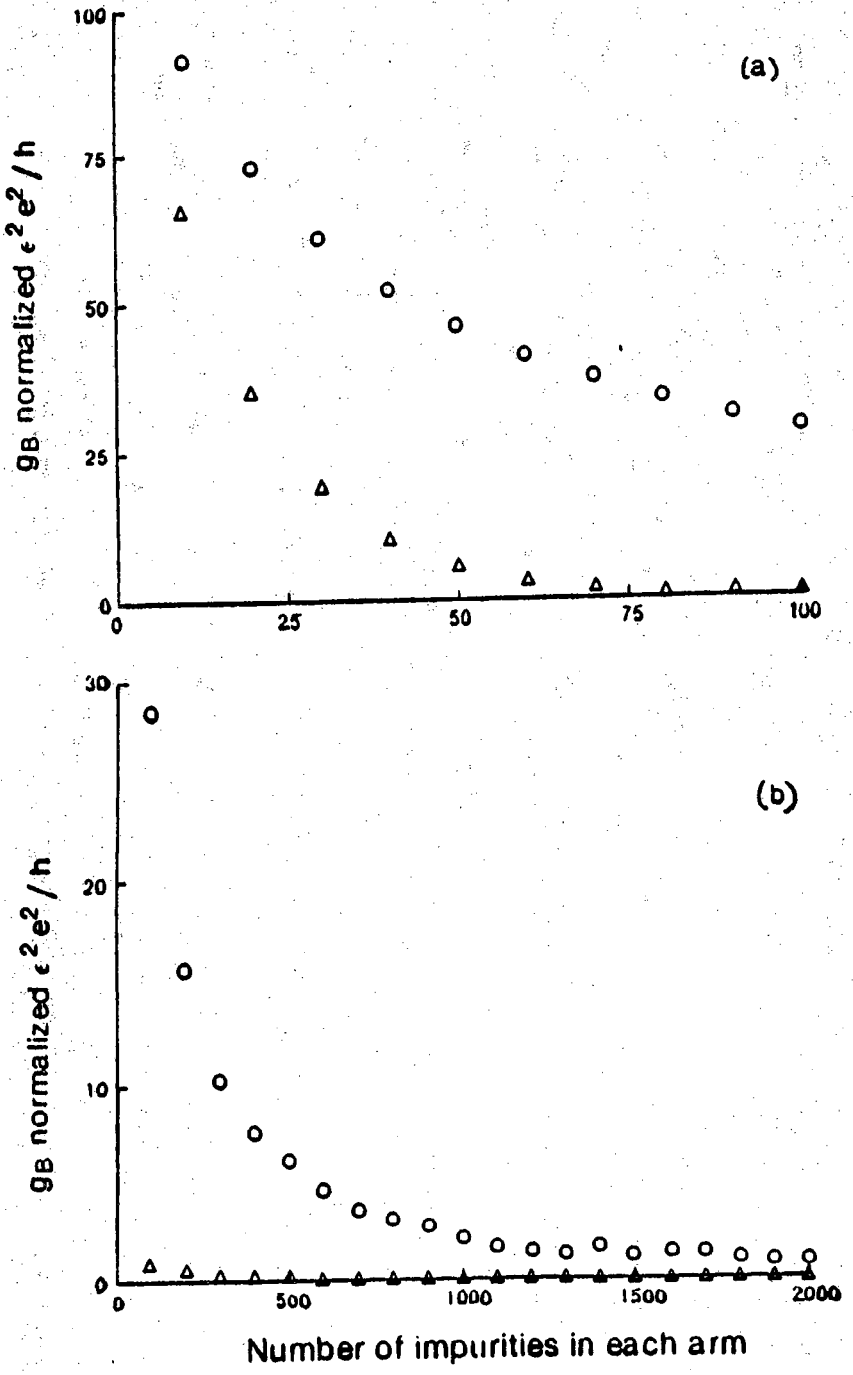


Figure 3.4 (a) Conductance modulation g_B versus number of impurities into both arms in the quasi-ballistic regime ($\Lambda_{el} = 33$ in this specific example). The circles (triangles) show the conductance modulation assuming identical (different) impurity configurations into both arms of the A-B device. (b) Same as Fig. 3.4.(a) for arm length extending from the weak to strong localization regime. In both cases, an average was taken over 45 ring configurations.

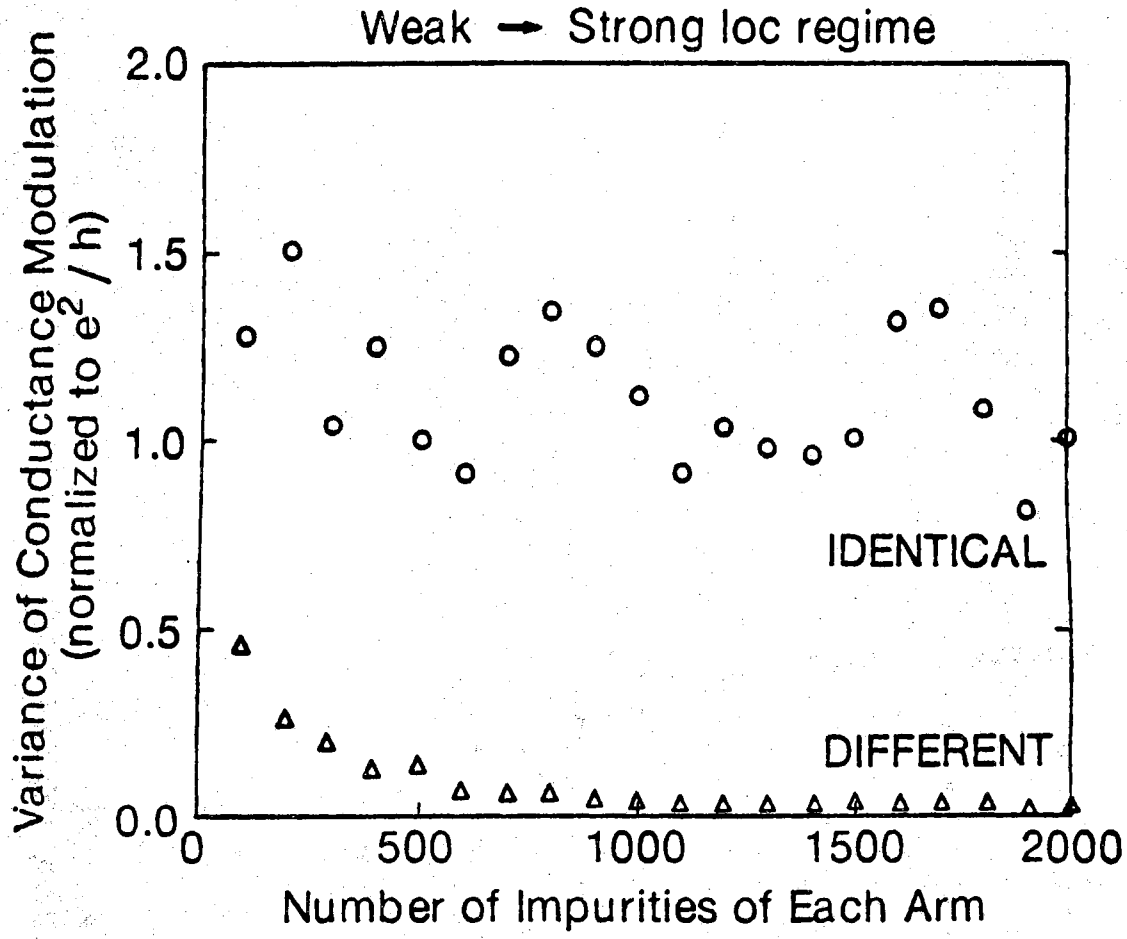


Figure 3.5 Variance of the conductance modulation shown in Fig. 3.4. The circles (triangles) correspond to the case of identical (different) impurity configurations into both arms. The variance was calculated by considering 45 ring configurations.

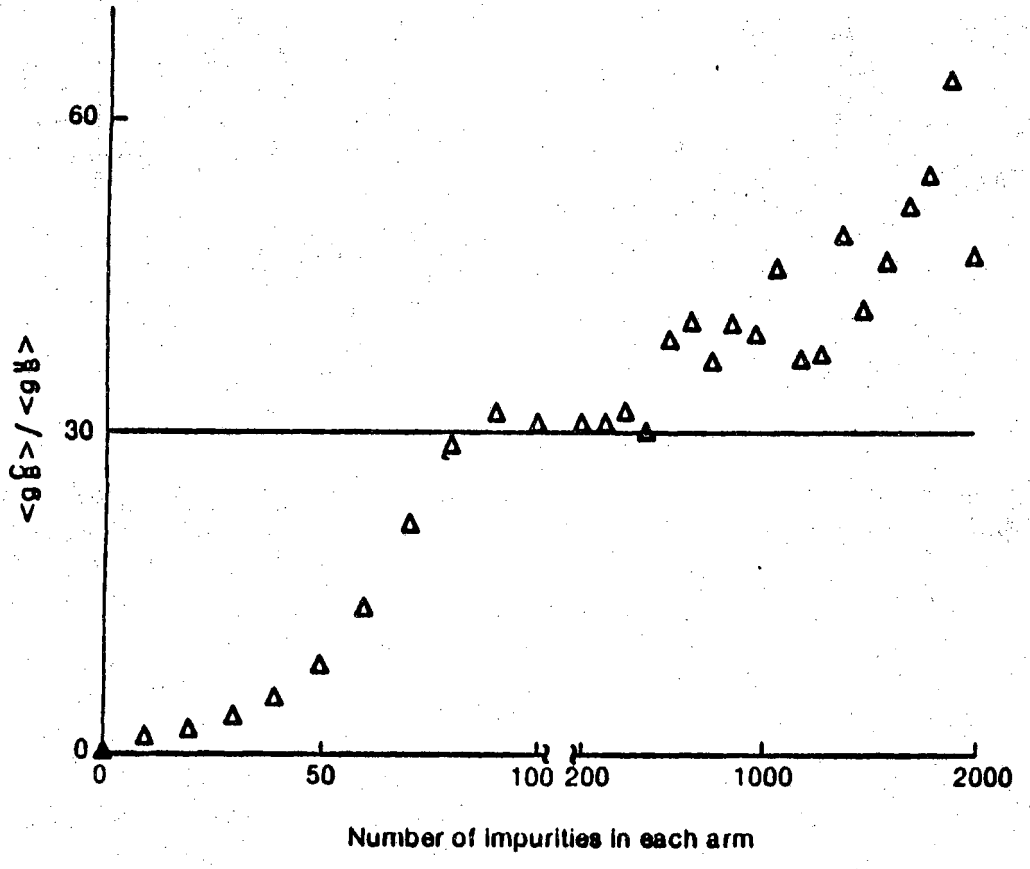


Figure 3.6 Ratio of the conductance modulation (see eq. (3.45)) obtained in the case of *correlated* (and *uncorrelated*) scattering versus number of impurities into both arms.

3.4 Electrostatic Aharanov-Bohm Effect

Since conductance fluctuations in a single resistor can be observed either by changing the potential or the magnetic field, it is natural to ask whether an Aharonov-Bohm effect can be observed as well by changing the potential in resistor 2 relative to resistor 1. Assuming that the scalar potential $V = 0$ in resistor 1 and $V = V_0$ in resistor 2, we find that the dispersion relation for resistor 2 is changed from eq. (2) to

$$E = \epsilon_m - eV_0 + \frac{\hbar^2 k_m^2}{2m^*} \tag{3.46}$$

Thus there is a vertical shift in the dispersion relations by $-eV_0$ for resistor 2, leading to a change in the wave number k_m for a given E by approximately,

$$\Delta k_m \simeq \frac{eV_0}{\hbar v_m} \tag{3.47}$$

where $v_m = \hbar k_m/m^*$ is the x-directed velocity of an electron in subband m with energy E. The difference with the magnetic field is that Δk_m is different for each mode. Ignoring the effect of multiple reflections, the conductance of the device can be derived as follows:

$$g = 2\epsilon^2 \sum_{m=1}^{M} \sum_{m'=1}^{M} \left| t^{(1)}_{m',m} + t^{(2)}_{m',m}(V_0) \right|^2 \tag{3.48}$$

Fig. 3.7 shows the conductance g calculated as a function of V_0 for different numbers of impurities (N_I) in each resistor (M=30, $\Lambda_{el} = 33.3$). There is a clear conductance modulation in the ballistic regime ($N_I = 0$) that quickly dies out as V_0 is increased or as we enter the diffusive regime. This can be understood as follows. The phase-shift can be written as $eV_0\tau/\hbar$ where τ is the transit time. The minima in the conductance correspond to values of V_0 such that $eV_0\tau/\hbar = (2n+1)\pi$. Even in the ballistic regime there is some spread in τ because different modes have different velocities. Conductance oscillations cannot be observed if the spread in the phase-shift $eV_0\Delta\tau/\hbar$ exceeds π. That is why the oscillations die out as V_0 is increased. Moreover, $\Delta\tau$ increases quickly in the diffusive regime due to multiple reflections. An electrostatic A-B effect can only be observed if the spread in transit times $\Delta\tau$ is less than the mean value $<\tau>$. In the magnetic A-B effect, the phase shift depends only on the flux enclosed; $\Delta\tau$ for the electrostatic effect thus plays a role analogous to the aspect ratio for the magnetic effect. The percentage conductance modulation can be increased by reducing the number of modes M, since it reduces $\Delta\tau$; however, the absolute value of the conductance decreases. We find that the electrostatic effect dies out much slower with increasing N_I if the two resistors are identical. This shows that this effect will

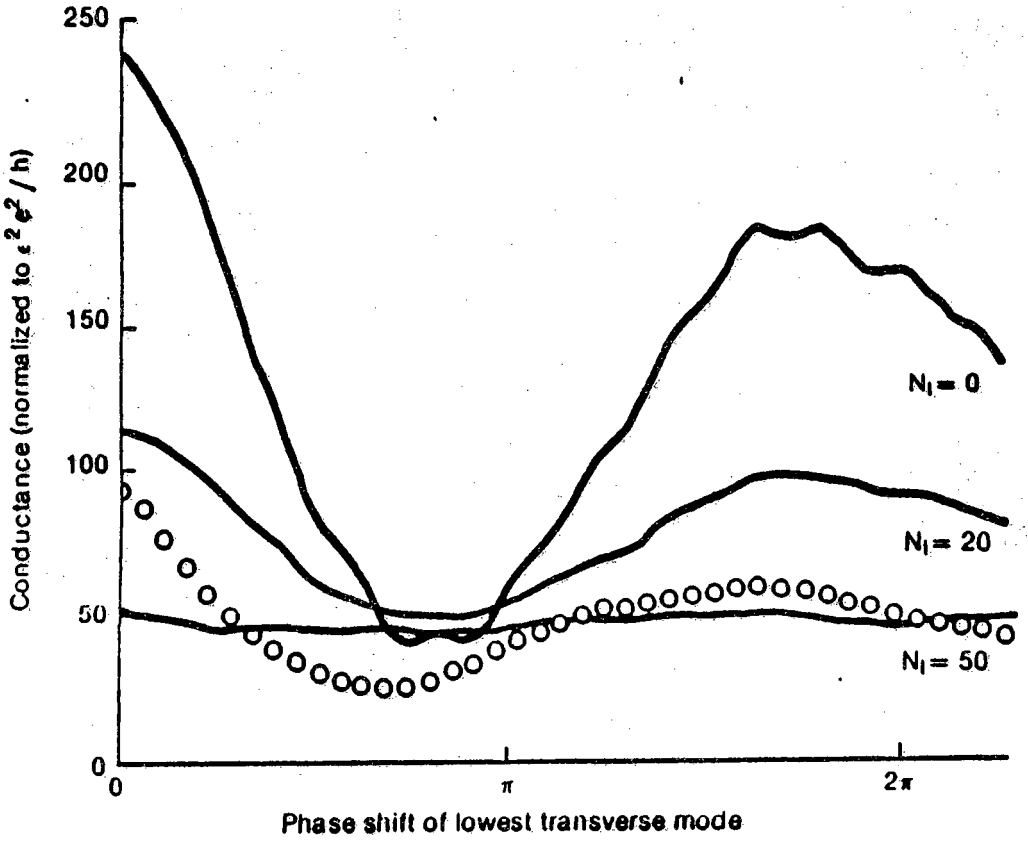


Figure 3.7 Conductance g vs. phase difference $\Delta k_1 L$ of the lowest transverse mode due to a potential difference V_0, for different values of N_I. The circles correspond to the case when both resistors have identical impurity configurations, with $N_I = 50$.

be easier to observe if the two arms are in close physical contact so that their scattering potentials are correlated. A typical device (quantum interference transistor) which could show strong conductance modulation under the application of an external gate voltage has been considered in reference [9].

3.5 Discussion and Conclusions

In this chapter, we have described a simple theory to calculate the conductance of Aharanov-Bohm semiconductor devices. The theory includes the effect of multiple reflections at the ends of the channels and also the influence of impurity scattering on the size of the conductance modulation.

An important distinction was made between the case of *correlated* and *uncorrelated* impurity scattering. The latter prevails in the case of metallic rings in which there is an air gap between the two channels. The former is more likely to happen in semiconductor structures due to the proximity of the channels.

In the case of *correlated* scattering, the conductance modulation of the magnetic A-B effect is predicted to still be **100%** even if the structure is multi-moded in the direction parallel to the interfaces of the structure (we stressed the importance for the structure to be single moded in the direction perpendicular to the interfaces).

For the *Electrostatic* A-B effect, we showed that non negligible conductance modulation could still be obtained even in the presence of impurity scattering. The size of the modulation is however drastically reduced in the case of *uncorrelated* scattering. This seems to rule out the possibility to observe such oscillations in metallic rings. However, the semiconductor heterostructures are certainly potential candidates to observe the *Electrostatic* A-B effect. Indeed, the screening length in semiconductors can be bigger than the separation between channels leading to some degree of correlation between the scattering potential into both arms of the device. This seems to be an imperative requirement to observe the conductance modulation in the recently proposed QUIT device.

LIST OF REFERENCES

[1] K. Von Klitzing, G. Dorda and M. Pepper, Phys. Rev. Letters 45, 494 (1980).

[2] S. Washburn and R. A. Webb, Advances in Physics, 1986, Vol.35, No.4, 375-422. (and references therein); S. Datta, M. R. Melloch, S. Bandyopadhyay, R. Noren, M. Vaziri, M. Miller, R. Reifenberger, Phys. Rev. Lett. 55, 2344 (1985).

[3] R. G. Chambers, Phys. Rev. Lett. 5, 3 (1960).

[4] R. A. Webb, S. Washburn, C. P. Umbach, and R. B. Laibowitz, Phys. Rev. Lett., 54, 2696 (1985).; V. Chandrasekar, M. J. Rooks, S. Wind, and D. E. Prober, Phys. Rev. Lett., 55, 1610 (1985).

[5] S. Bandyopadhyay, B. Das, R. Reinfenberger, S. Datta, W. P. Hong an P. K. Battacharya ; March Meeting of the American Physical Society at New-York, 1987.

[6] R. Landauer, Internal memo, IBM Thomas J. Watson Research Center, (1966); IBM J. Res. Dev. 1, 223 (1985).

[7] L. Gunther, and Y. Imry, Solid St. Commun., 7, 1391 (1961).

[8] For a review see Büttiker, SQUID-85: Superconducting Quantum Interference Devices And Their Applications, edited by D. A. Halbohm and H. Lübbig (New York: Walter-de Gruyter), 529 (1985).

[9] S. Datta and S. Bandyopadhyay, Phys. Rev. Lett. 58, 717 (1987).

[10] A. D. Stone, Phys. Rev .Lett 54, 2692 (1985)

[11] S. Bandyopadhyay, S. Datta, and M. R. Melloch, Superlattices and Microstructures, 2, No.6, 539 (1986).

[12] P. A. Lee, Physica A, 140 (1986).

[13] S. Datta, *Physics of Quantum Electron Devices*, ed. F. Capasso, to be published by Springer-Verlag in 1987.

[14] B. Shapiro, Phys. Rev. Lett.**50**, 747 (1983).

[15] A. M. Kriman and P. Ruden, Phys. Rev. B32, 8013 (1985).

[16] M. Cahay and S. Datta (unpublished).

4. ELECTRON TRANSFER ACROSS A JUNCTION BETWEEN TWO REGIONS WITH VERY DIFFERENT CONFINING POTENTIALS

4.1 Introduction

An important problem in ultrasmall submicron devices is to understand electron transport across a junction between two regions with very different confining potential profiles in the transverse direction, so that the electron wave function in the transverse plane changes significantly from one region to the other. Consider, for example, a transition from a narrow channel of width W into a wide channel of width W′ (Figure 1). This structure can be viewed as an idealized model for the transition from a two-dimensional FET channel to a wide contact region. Classically the electrons would simply pour out of the channel with unity transmission probability; quantum mechanically there is a significant probability for the electron to be reflected at the junction and this affects the contact resistance. A more intriguing problem is the following; suppose there are two channels emptying into the same contact (Figure 3). Normally, we assume that the current flowing out of one channel is independent of the other channel. However, if the two channels are spaced very close compared to a De Broglie wavelength, then the current will be greater if the wave function in the two channels are in phase than if they are out of phase. This interference between adjacent channels is of interest in understanding the Aharonov-Bohm effect in double quantum well structures and can play a significant role in ultrasmall devices of the future.

The problem of transition between regions with different subband structure has not yet received much attention in the literature. The model presented here is intended to compliment the pioneering work of Kriman and Ruden.[1] This method makes the practical solution of a larger variety of problems possible by eliminating difficult analytical steps at the expense of additional computer time. Solving for the scatter matrix involves in principle working on an infinite set of linear equations. It is common to truncate this set to those involving only the lowest order modes (those which oscillate least in the transverse direction). When the electron transfer takes place between regions of widely different transverse dimension the mode structure on either side of the interface is very different as well. The amplitude of the modes in the wide region will be rather uniform over the entire interface, but the amplitude of the modes in the narrow region will be large where the channel(s) is(are) defined, and small elsewhere. This makes it impossible to reasonably meet boundary conditions making the wave function and its normal derivative at the interface continuous without using a very large number of modes.

Kriman and Ruden got around this problem by analytically projecting the operator equations onto the basis set consisting of the modes on only one side of the interface. Essentially this employs all the modes (an infinite number) on the other side of

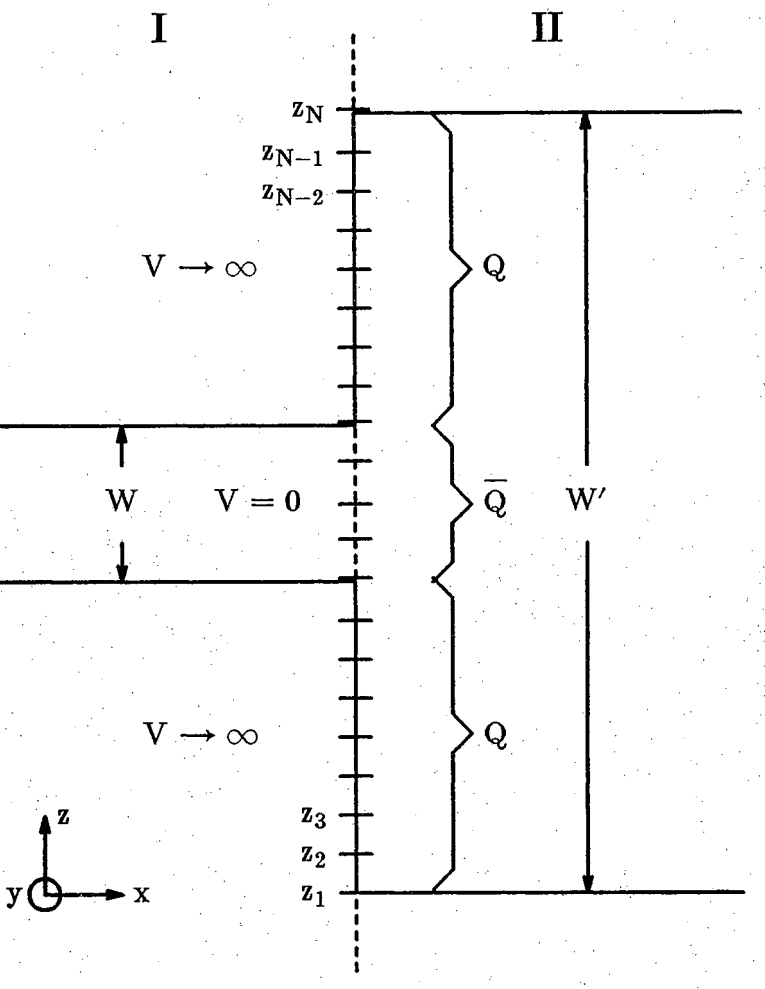


Figure 1. The boundary conditions are applied to the average of the wave function and its normal derivative over each interval $[z_n, z_{n+1})$. In the regions Q, $\langle x|\Psi\rangle = 0$. In $\bar{Q}$, $\langle x|\Psi\rangle$ and $\frac{d}{dx}\langle x|\Psi\rangle$ are continuous.

the interface. In all but the most simple cases this involves computing a difficult infinite sum or integral. It does, however, have several advantages if the projection can be done, those being that the truncation approximation conserves current exactly, the number of linear equations to solve is roughly half that of the case when the projection cannot be made and the solution is slightly more straight forward. In their paper Kriman and Ruden very cleverly picked as an application of their method the harmonic oscillator potential on one side of the interface and let the potential be zero on the other. This choice allowed them to make the projection very easily. Unfortunately this is the only problem we know of that can be easily solved using their method.

This paper describes another way to solve the problem of the mismatched modes by satisfying the most appropriate set of boundary conditions on the average of the

wave function and the average of its normal derivative over a subdivision of the interface. In regions of large potential the magnitude of the wave function is small and as a result the physics of the device under scrutiny doesn't depend strongly on knowing the exact wave function in these regions. It is convenient to define the wave function to be zero in these regions which is equivalent to allowing the potential to approach infinity. This saves the trouble of solving for information that is really unnecessary. In the other regions where the the potential is small and the magnitude of the wave function is large, the boundary conditions used make the wave function and its normal derivative continuous. Another advantage is that in the model for multi-channel devices the solution for the modes in the transverse direction can be simplified by decoupling the various conducting channels with infinite potential barriers.

In Section 2 the theory is discussed, including the model and the details of the solution. In Section 3 the model is applied to two important examples, a single channel emptying into a reservoir and then two channels emptying into a reservoir with special attention paid to the effects of the wave nature of the electron in both cases. Section 4 is a summary of results and conclusions.

4.2 Theory

We start with the single particle single band effective mass Hamiltonian neglecting variations in effective mass. This model is most appropriate to electron transport in the conduction band rather than the valence band since the conduction band is more correctly approximated by a single band calculation than the valence bands which are more tightly coupled together. We focus on the interfacial region where the mode structure changes. To the left of the interface we assume there is some non-zero distance in the propagation direction (here after denoted as the x direction) that the confining potential varies only in the transverse directions. Denote this region by I. Likewise to the right of the interface we assume that there is some non-zero distance in the x direction where the other confining potential is constant in x. Denote this region by II. In order to simplify the discussion we consider only one transverse dimension (z). The extension to both transverse dimensions is very simple. The interfacial plane is given by $x = 0$. Our goal is to obtain the scatter matrix for the interface.

$$H \mid \psi> = E \mid \psi>$$

The Hamiltonian may be expressed

$$H = -\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial x^2} + H_z,$$

$$H_z \equiv \begin{cases} -\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2} + V_I(z) & \text{for } x \in I \\ -\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2} + V_{II}(z) & \text{for } x \in II \end{cases}$$

The normalized eigenstates or eigenmodes of H_z in region I are labeled with α so that

$$H_z \mid \alpha> = E_\alpha \mid \alpha>$$

and similarly in region II the eigenmodes of H_z are labeled with α' so that

$$H_z \mid \alpha'> = E_{\alpha'} \mid \alpha'>$$

The total Hamiltonian H is separable in each region so that we may take product state solutions of the form

$$<x \mid \alpha,\sigma> \equiv \frac{1}{\sqrt{2\pi k_\alpha}} \exp(i\sigma k_\alpha x)\, s(x,I) \mid \alpha>$$

where $\sigma = \pm 1$ and s(x,I) is one if $x \in I$ and zero otherwise. Similarly for region II

$$<x \mid \alpha',\sigma> \equiv \frac{1}{\sqrt{2\pi k_{\alpha'}}} \exp(i\sigma k_{\alpha'} x)\, s(x,II) \mid \alpha'>.$$

The wave vectors are given by

$$k_\alpha \equiv \sqrt{\frac{2m}{\hbar^2}(E - E_\alpha)}$$

and

$$k_{\alpha'} \equiv \sqrt{\frac{2m}{\hbar^2}(E - E_{\alpha'})}$$

The factor $1/\sqrt{k_\alpha}$ and $1/\sqrt{k_{\alpha'}}$ are included so that each mode carries the same amount of current. These solutions form a complete set and a superposition of them form a natural way to describe electron transport in the wave picture. The correct superposition is determined by the continuity relations

$$(<x \mid \psi>)_{x=0^-} = (<x \mid \psi>)_{x=0^+} \tag{1}$$

always, and

$$\left(\frac{d}{dx}<x \mid \psi>\right)_{x=0^-} = \left(\frac{d}{dx}<x \mid \psi>\right)_{x=0^+} \tag{2}$$

provided the potential V(z), is bounded on both sides of the interface.

For an incident wave in mode $\mid \beta,+>$ ($\mid \beta',->$) from the left (right), the wave function can be expressed as the incident wave plus a sum of reflected waves into all

modes $|\ \alpha,->$ ($|\ \alpha',+>$) on the left (right) and a sum of transmitted waves into all the modes $|\ \alpha',+>$ ($|\ \alpha,->$) on the right (left).

$$|\ \psi_\beta> = |\ \beta,+> + \sum_\alpha |\ \alpha,->R_{\alpha\beta} + \sum_{\alpha'} |\ \alpha',+>T_{\alpha'\beta} \tag{3}$$

The $R_{\alpha\beta}$ are reflection coefficients from incident mode $|\ \beta,+>$ into modes $|\ \alpha,->$. Similarly the $T_{\alpha'\beta}$ are transmission coefficients from incident mode $|\ \beta,+>$ into transmitted mode $|\ \alpha',+>$.

We do not wish to solve the Schrodinger equation in regions that are not of great interest to us such as in the barrier regions because this presents practical problems by increasing the solution time and memory requirements on the computer. The solutions in these regions contribute little to the physics. We, therefore, model the potential in these regions as being infinite. Another advantage of this is the ease with which one may define separate channels in a device without having to resort to potentials that produce complicated eigenstates $|\ \alpha>$ or $|\ \alpha'>$. At this point it is natural to divide the interface into two sets, Q and $\overline{Q}$. Q is the set of all z values such that either $V_I(z)$ or $V_{II}(z)$ but not both are infinite. $\overline{Q}$ is the set of all z such that neither $V_I(z)$ nor $V_{II}(z)$ is infinite. The wave function is zero over the compliment of $Q \cup \overline{Q}$ and so that part of the interface is uninteresting. Using this model, the continuity of the derivative does not hold where the wave is impinging on a barrier region (for $z \in Q$). Instead in Q we have

$$(<x\ |\ \Psi>)_{x=0} = 0. \tag{4}$$

Partition $Q \cup \overline{Q}$ as is done in defining Riemann sums

$$[z_1,z_2),[z_2,z_3),[z_3,z_4),\ \cdots,[z_n,z_{n+1}),\ \cdots,[z_N,z_{N+1})$$

where no interval $[z_n,z_{n+1})$ contains points from both Q and $\overline{Q}$ as in Figure 1. Define the ket $|\ n>$ such that

$$<z\ |\ n> = \frac{s(z,[z_n,z_{n+1}))}{\sqrt{z_{n+1} - z_n}}.$$

Note

$$<n\ |\ m> = \delta_{nm}$$

and operating with $<n|$ takes a spatial average over the interval $[z_n, z_{n+1})$ and multiplies it by $\sqrt{z_{n+1} - z_n}$. Let $\Delta \equiv \max_n \{z_n - z_{n-1}\}$. The basis $|\ n>$ spans the space $Q \cup \overline{Q}$ in the limit as Δ approaches zero. Inserting (3) into (1), (2), and (4) and operating with the bra $<n|$ leads to

$$\left.\begin{aligned} \frac{1}{\sqrt{k_\beta}}\langle n \mid \beta\rangle + \sum_\alpha \frac{1}{\sqrt{k_\alpha}}\langle n \mid \alpha\rangle R_{\alpha\beta} &= \sum_{\alpha'} \frac{1}{\sqrt{k_{\alpha'}}}\langle n \mid \alpha'\rangle T_{\alpha'\beta} \\ \sqrt{k_\beta}\langle n \mid \beta\rangle - \sum_\alpha \sqrt{k_\alpha}\langle n \mid \alpha\rangle R_{\alpha\beta} &= \sum_{\alpha'} \sqrt{k_{\alpha'}}\langle n \mid \alpha'\rangle T_{\alpha'\beta} \end{aligned}\right\} \text{for } [z_n, z_{n+1}) \in \bar{Q}$$

$$\left.\frac{1}{\sqrt{k_\beta}}\langle n \mid \beta\rangle + \sum_\alpha \frac{1}{\sqrt{k_\alpha}}\langle n \mid \alpha\rangle R_{\alpha\beta} = 0\right\} \text{for } [z_n, z_{n+1}) \in Q$$

The above set of equations assumed an incident mode from the left $\mid \beta, +\rangle$. Similar equations can be written for incident modes from the right $\mid \beta', -\rangle$. If these equations are collected they may be written in a single matrix equation $\mathbf{AS} = \mathbf{B}$.

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} R & T' \\ T & R' \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

Each of these matrix elements are themselves matrices whose elements are given by

$$A_{11_{n\alpha}} = -B_{11_{n\alpha}} \equiv \frac{1}{\sqrt{k_\alpha}}\langle n \mid \alpha\rangle$$

$$A_{12_{n\alpha'}} = -B_{12_{n\alpha'}} \equiv \frac{-1}{\sqrt{k_{\alpha'}}}\langle n \mid \alpha'\rangle$$

$$A_{21_{n\alpha}} = B_{21_{n\alpha}} \equiv \sqrt{k_\alpha}\,\langle n \mid \alpha\rangle$$

$$A_{22_{n\alpha'}} = B_{22_{n\alpha'}} \equiv \sqrt{k_{\alpha'}}\,\langle n \mid \alpha'\rangle$$

In writing the matrix form of these equations it was noted that $\langle n \mid \alpha'\rangle = 0$ and $\langle n \mid \alpha\rangle = 0$ for $[z_n, z_{n+1}) \in Q$.

The scatter matrix for the junction **S** may then be solved for.

$$\mathbf{S} = \mathbf{A}^{-1}\mathbf{B}$$

In practice the parameter Δ is chosen small enough so that further reduction in its value has no effect on the elements of the scatter matrix out to a desired decimal place. Operating with the bra $\langle n \mid$ has the effect of demanding that the boundary conditions be met on the average values of the wave function and its normal derivative over each interval of the partition $[z_n, z_{n+1})$. This provides a convenient method of approximating the solution to the infinite number of equations for **S** by the solution to a finite number of equations. The properties of the solution that depend on the symmetry of the problem are not affected by this approximation, but current conservation is not preserved exactly. This, however, can be used to obtain an idea of how well converged the solutions are. As the reflection and transmission coefficients for the propagating modes converge to the correct solution (as Δ is decreased) the current is conserved exactly.

4.3 Applications

This section applies the previously described method to two specific problems of electron transport in semiconductor devices. The first is the transition from a channel into a contact region which is common in many devices today. The second is the transition from two parallel channels into a contact. This transition occurs in parallel channel quantum interference transistors employing the Aharonov Bohm effect.

The model employed for both examples is three dimensional. On both sides of the interface the device is homogeneous in the large y dimension so that k_y is conserved. The junction at the interface is considered to be abrupt, both in geometrical features and in the built in potential. The effect of potential grading may be quite significant and will be investigated in a later paper.[2,3] The built in potential is calculated from charge neutrality in the device. Far into either side of the device the states are filled up to an unknown energy level using the appropriate two dimensional density of states and doping densities. The built in potential is the difference between the energy levels on the left and right.

The first application is of interest because of the light it sheds on how the quantum mechanical reflections influence the contact resistance. This is the problem that Kriman and Ruden studied. Our results are very similar and indicate that the details of the confining potential used in the model make little difference.

The second application highlights a principle which is expected to play an important role in devices of the future, quantum mechanical wave interference between two parallel channels emptying into a single contact. This interference can be used to modulate the current through the device and has already been demonstrated a number of times experimentally.[4,5]

4.3.1 Coupling from a Single Channel to a Single Reservoir

Metal-oxide-semiconductor field-effect transistors, high electron mobility transistors, heterojunctions, and thin p—n—p doping layers are all examples where electrons are squeezed tightly in one dimension. At the channel contact interface this squeezing relaxes and the electron behavior becomes three dimensional. In this section the ejection from the channel into the drain contact is studied. The complimentary injection from the source contact into the channel is similar.

The drain contact is modeled as a large channel region of width W' in the z direction and the channel as a small channel region of width W in the z direction. This is portrayed in Figure 1. Reflection symmetry about $z = 0$ guarantees conservation of parity so that symmetric eigenmodes will not couple to antisymmetric eigenmodes and vice versa.

The eigenmodes in the channel are given by

$$<z \mid \alpha> = \sqrt{\frac{2}{W}} \sin(\frac{\alpha \pi z}{W}) \, s(z,[-W/2,W/2])$$

and those in the contact region are given by

$$<z \mid \alpha'> = \sqrt{\frac{2}{W'}} \sin(\frac{\alpha' \pi z}{W'}) \, s(z,[-W'/2,W'/2]).$$

Current reflection coefficients $\mid R_{\alpha\beta} \mid^2$, from incident modes in the channel $\mid \beta>$ into reflected modes in the channel $\mid \alpha>$ as a function of incident energy are seen in Figure 2 for two different values of W′. As the energy is increased, the number of populated sub-bands increases and reflection back into these sub-bands becomes possible. The only material specific parameter in this theory is the effective mass. The energy scale given in Figure 2 is for GaAs, however this may be multiplied by $\frac{.067}{m^*}$ to apply the figures to another material. The temperature influences the incident energy through the Fermi Dirac factor. For a uniform doping of $2x10^{17}cm^{-3}$ in both the channel and the contact with contact width $W' = 1100Å$ and an incident energy of k_BT above the Fermi energy at room temperature, about ten percent of the current is reflected back into the channel. At low temperatures the incident energy is near the Fermi energy and about thirty percent of the current is reflected. It is very interesting that changing the value of W′ has little effect on the reflections back into the channel when it is much greater than W as can be seen by examining Figure 2. This means that the size of the contact has little effect on the device if it is much larger than the channel width. Although the model used here is very different from the parabolic potential detailed by Kriman and Ruden, it gives very similar results. The major difference between the parabolic potential and the square well potential seems to be at what energy the subbands appear.

4.3.2 Parallel Channel Structure

Recently the feasibility of using quantum mechanical electron wave interference between two parallel semiconductor channels to modulate the current entering a contact reservoir has been demonstrated.[4,5] By By varying a magnetic flux through the region between the parallel channels the difference in phase accumulated by the electron wave function in each channel can be changed. If the two parallel channels can be brought together into a single channel with a single propagating mode, the current may be modulated by 100 percent. Unfortunately it is difficult to construct a device where this happens. Some of the devices constructed had channel regions made of GaAs separated by barrier layers of AlGaAs defined by Molecular Beam Epitaxy (MBE).[4] The channel region was connected between two contact regions heavily doped n type by the Ge of a Au-Ge alloy implant. The contact regions where the

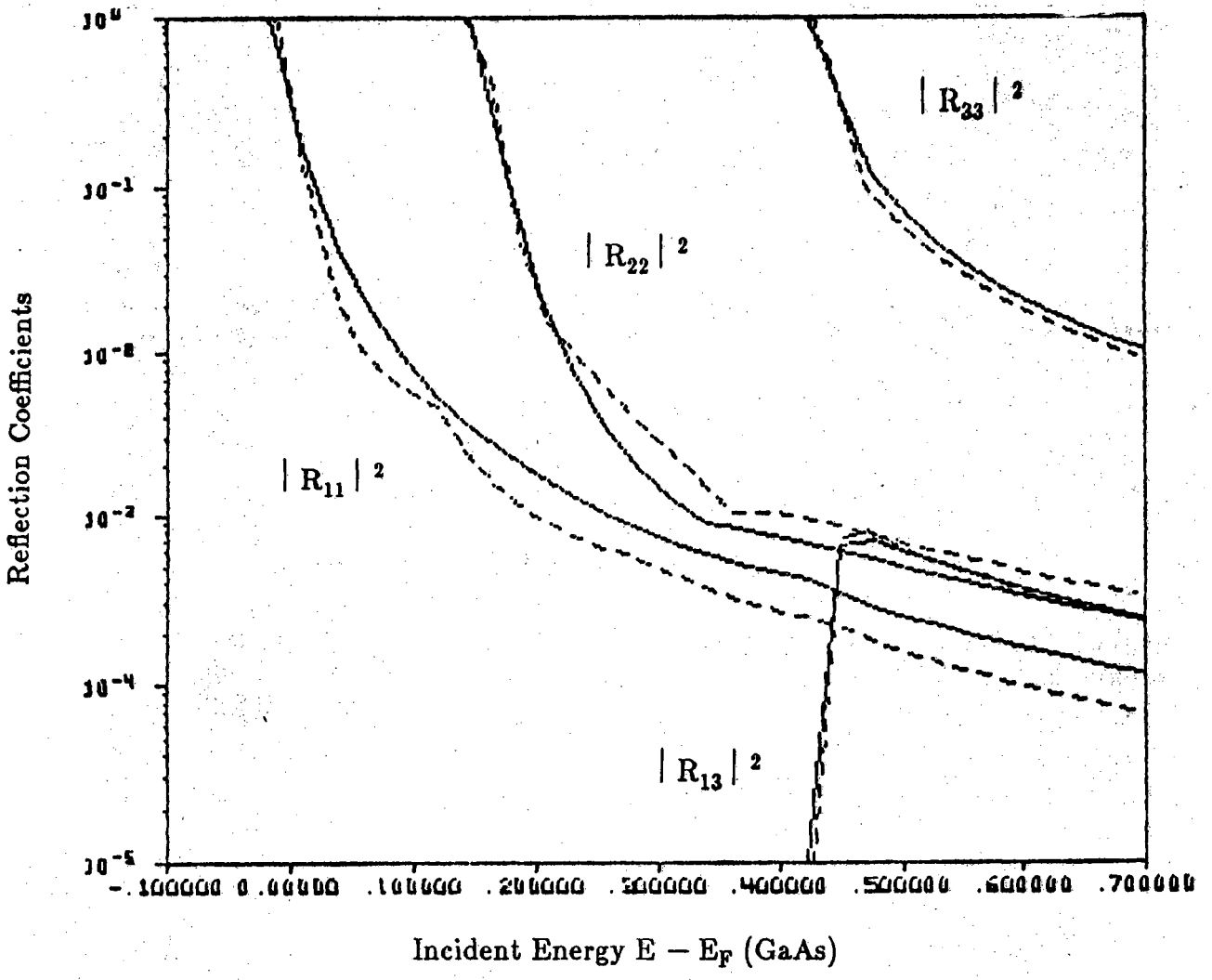


Figure 2. Current reflection coefficients as a function of incident energy. Dashed line denotes: $W' = 300$Å. Solid line denotes: $W' = 1500$Å. The doping density is held constant at $10^{17}cm^{-3}$ and W is set to 100Å.

channels join were multimoded. This is an undesirable situation because the the current into one mode may be least when the current into another is greatest. The important question is how devastating is this effect?

The answer is summarized in Table 1. Table 1 shows that the doping is an important parameter. As the doping is increased the number of occupied states increases so that eventually a new subband is populated which means that another mode becomes propagating. This is generally not a problem in the channel where the spacing between subbands is large because the channels are narrow. However, in the contact the subbands are closely spaced because of its large width. In addition to this, contacts are generally heavily doped in order to make good electrical contact with the outside world. In the contacts, therefore, many modes are likely to be propagating causing the modulation to be significantly lowered.

Doping		Number of Propagating	Modulation
Contact	Channel	in the Contact	Percentage
(cm^{-3})			
10^{16}	10^{16}	3	91
$3x10^{16}$	$3x10^{16}$	4	67
10^{17}	10^{16}	7	22
10^{17}	10^{17}	7	14
10^{18}	10^{16}	15	11
10^{18}	10^{18}	15	5

Table 1.

The geometry for the model is detailed in Figure 3. The dopings used for the channel region in the model allowed only one propagating mode in each channel, but many propagating modes in the contact. The ejection current is proportional to the sum over the propagating modes in the contact of the magnitude squared of the sum of the transmissions from the channels into these modes.[6]

$$I_{eject} \sim \sum_{\alpha'} | T_{\alpha' 1} + e^{i\phi} T_{\alpha' 2} |^2$$

Here $T_{\alpha' 1}$ is transmission from the first channel into the α'th mode in the contact and $T_{\alpha' 2}$ is transmission from the second channel into the α'th mode in the contact. The phase difference between the wave function in channel 1 and channel 2 is included explicitly as ϕ. In the case studied with symmetry about $z = 0$ the transmission coefficients for the two channels are related by $T_{\alpha' 1} = (-1)^{\alpha'+1} T_{\alpha' 2}$ where α' even is for antisymmetric modes and α' odd is for symmetric modes. Using this

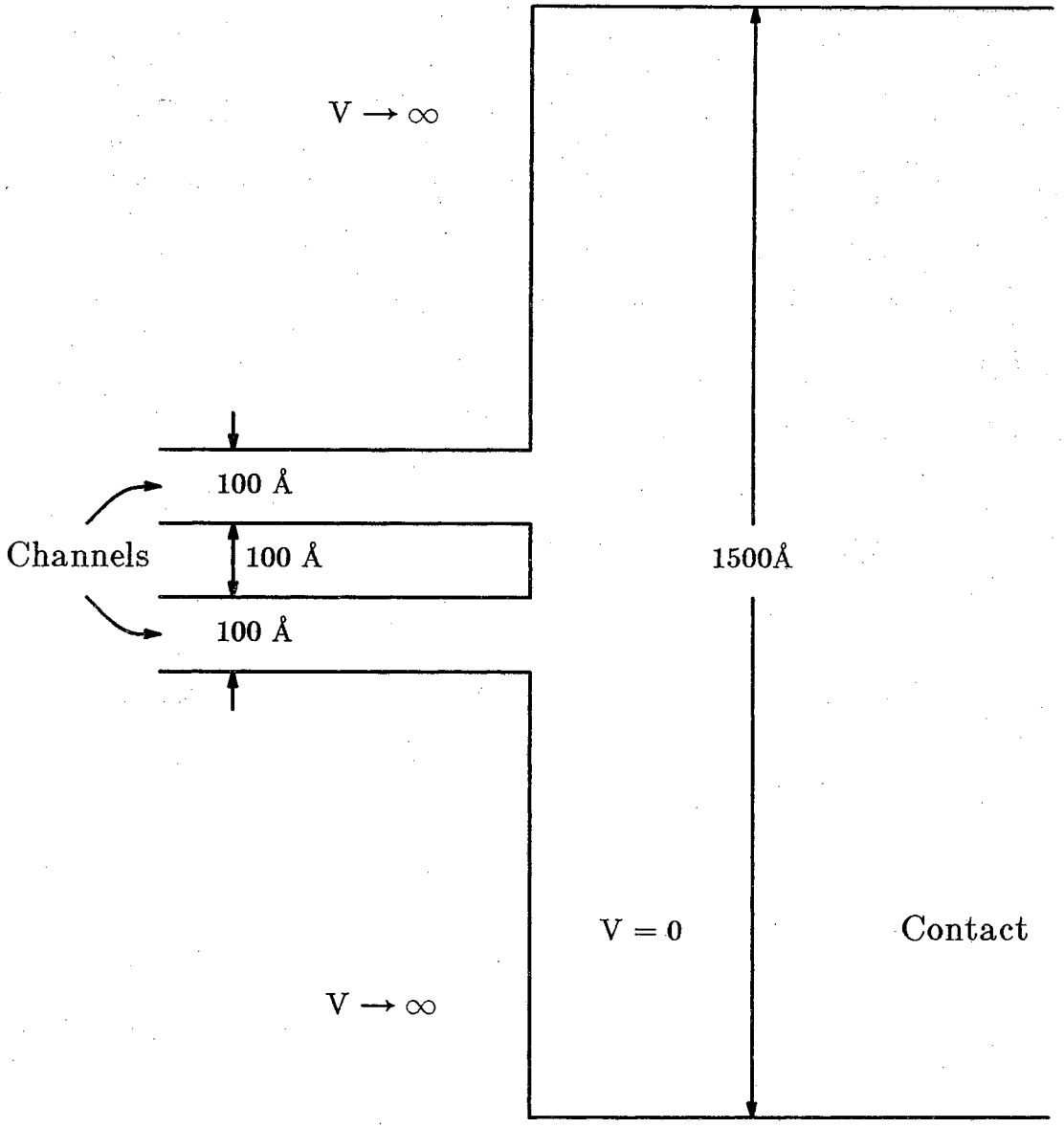


Figure 3. The model used for two channels entering a reservoir.

$$I_{eject} \sim 2\left\{ \sum_{\alpha' \text{ even}} | T_{\alpha' 1} |^2 (1 - \cos\phi) + \sum_{\alpha' \text{ odd}} | T_{\alpha' 1} |^2 (1 + \cos\phi) \right\}$$

so that the ejection current may be modulated by changing ϕ as long as either the symmetric or the antisymmetric modes dominate.

The amount of modulation possible is determined by the unbalance in transmission coefficients between the symmetric and antisymmetric propagating modes. The current transmission coefficients from one channel into each of the propagating modes is shown in Figure 4 for $k_y = 0$ and several different doping densities. It was found that the relative transmissions into the different modes have a characteristic "footprint" as seen in Figure 4. This footprint can be predicted by a simple calculation of the overlap of the different modes (Figure 5). The relative coupling is determined primarily by the overlap integrals. This is because the geometry becomes the common denominator when the doping density is changed.

For the geometry studied there are three different domains. The first is where the doping is light enough in the contact region so that "bumps" of the modes extended over the entire 300 angstrom channel region. In this case symmetric modes dominate. Second is the domain where approximately two bumps extend over the channel region. In this case the antisymmetric modes dominate. The last domain is where the modes oscillate many times over the channel region. In this case there is little transmission.

It is evident that in order to maximize the modulation the doping density must be kept low insuring the number of propagating modes in the contact remains reasonable (preferably in the first domain). It is also evident that the device could be improved by making the 300Å channel region smaller. This is in fact observed experimentally as well.[7]

4.4 Summary and Conclusions

In this paper we have discussed a method of modeling electron transfer from confined channel regions into much less confined contact regions which eliminates difficult analytical steps, replacing them with more mundane procedures to be done on the digital computer. The solution gives the scatter matrix for an abrupt junction. These scatter matrices may be used as an integral part of the solution for a complete device by connecting solutions of different sections. The method is applied to two different structures highlighting two different physical effects due to the wave nature of the electron. The major assumption applied in each example is that the most important scattering is due to the large change in the potential cross section from the channel region to the contact.

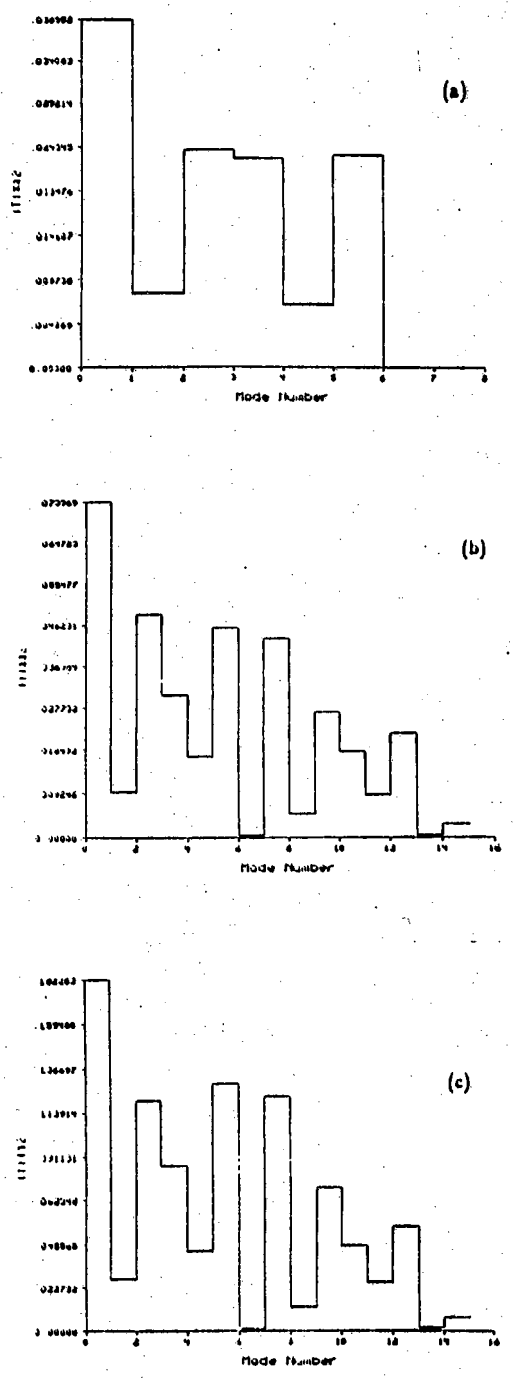


Figure 4. $|T_{\alpha' 1}|^2$ from one channel into the reservoir. (a) Channel doping is $10^{16}cm^{-3}$ and reservoir doping is $10^{17}cm^{-3}$. (b) Channel doping is $10^{16}cm^{-3}$ and reservoir doping is $10^{18}cm^{-3}$. (c) Channel doping is $10^{18}cm^{-3}$ and reservoir doping is $10^{18}cm^{-3}$.

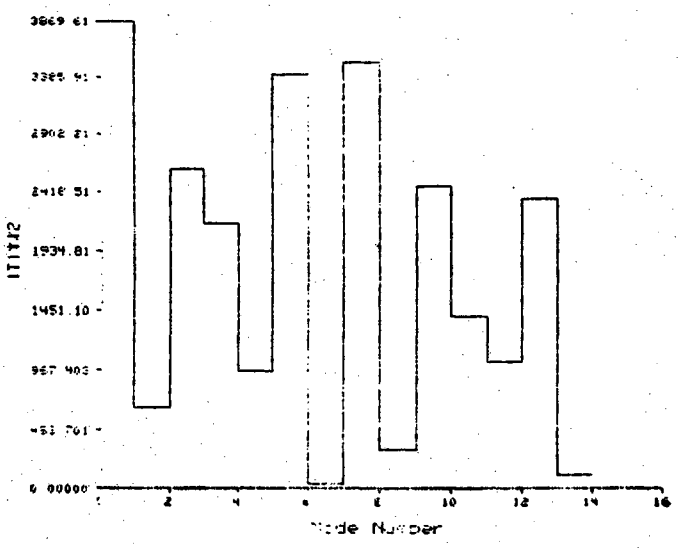


Figure 5. The approximation $| T_{\alpha' 1} |^2 \cong \sum_{[z_n z_{n+1}) \in \bar{Q}} | \langle n | \alpha' \rangle |^2$.

The first application modeled the transfer from a channel region into a contact as occurs in many types of field effect transistors as an abrupt junction. In this application the quantum mechanical reflection at the junction between the channel and the contact was studied as a function of the incident energy of the electron. Even though the junction model used was abrupt, the the contribution to the electrical resistance due to quantum mechanical reflection was not found to be very important.

In the second and most interesting application, the phenomena of interference between two parallel single moded channels emptying into a single contact was examined. The current may be modulated by changing the phase of the electron in each channel. The results indicate the amount of modulation is determined by the number of propagating modes in the contact and the separation and width of the channel region. In order to maximize the modulation the number of propagating modes in the contact, the separation between the channels and the channel width must all be minimized, the separation between the channels and the channel width must all be minimized.

REFERENCES

1. Alfred M. Kriman and P. Paul Ruden, Phys. Rev. B 32,
2. Jaroslaw Uher, Fritz Arndt, MTT-35, #6, pp. 552
3. Jens Bornemann and Fritz Arndt, MTT-35, #6, pp. 561

4. S. Datta, M. R. Melloch, S. Bandyopadhyay, R. Noren, M. Vaziri, M. Miller, and R. Reifenberger, Phys. Rev. Lett. Vol. **55**, No. 21, 2344 (1985)

5. G. Timp, A.M. Chang, J.E. Cunningham, T.Y. Chang, P. Mankiewich, R. Behringer, R.E. Howard, Phys. Rev. Lett., Vol. 58, No. 26, 2814 (1987)

6. S. Datta, M. R. Melloch, S. Bandyopadhyay, and M. S. Lundstrom, Appl. Phys. Lett **48** (7), 17 February 1986

7. S. Bandyopadhyay, private communication.

5. NUMERICAL STUDY OF EMITTER-BASE JUNCTION DESIGN FOR AlGaAs/GaAs HBT's.

5.1 Introduction

In recent years, Heterojunction Bipolar Transistors (HBT's) have become a subject of extensive experimental and theoretical research. HBT's have a number of advantages over conventional homojunction transistors. The most important one derives from the use of a wide gap emitter which permits increased base doping to lower base resistance without sacrificing emitter injection efficiency.[1]

For an AlGaAs/GaAs N-p emitter base heterojunction, the band line ups result in a conduction band spike and notch as illustrated in Fig 1a. Kroemer [1] and subsequently many researchers pointed out that this conduction band spike suppresses electron injection which lowers emitter injection efficiency (γ), which subsequently reduces common emitter current gain (β), so an appropriate grading of the heterojunction (HJ) to suppress the spike is suggested. The typical grading length ranges between 300Å to 500Å. The band diagram of a graded heterojunction is shown in Fig. 2. with a grading length of 300Å (In this chapter, graded HBT refers to fully graded HBT so that conduction band spike is almost absent). The conclusion that grading of the HJ leads to a large increase in the emitter injection efficiency across the HJ is largely based on the fact that the grading at the HJ removes the spike in the conduction band which otherwise suppresses electron injection in case of an abrupt HJ, but grading at the same time increases the recombination current in the emitter base depletion region of the graded HBT by an order of magnitude compared to an abrupt HBT. This particularly becomes a large percentage of E-B electron current at low bias. Also, in case of abrupt HJ, electrons will tunnel through the tip of the conduction band spike, thus the effective height of the conduction band spike will be much lower than that calculated using classical formulations. In fact it was found that the tunneling current is fifty to sixty times higher than the thermionic emission current. So it is not obvious that grading of a HJ will always lead to much higher emitter injection efficiency in a graded HBT as compared to an abrupt HBT.

Also, the common emitter current gain β is controlled by both γ and the base transport factor (α_t) [2], and when both γ and α_t are close to unity, the smaller of the two dictates the magnitude of β. So from this point of view, if an abrupt or graded heterojunction bipolar transistor has sufficiently high electron injection efficiency, then its common emitter current gain will be limited by the relatively smaller base transport factor rather than electron injection efficiency across the HJ. To summarize the above discussion, we can say that besides conduction band spike, R-G current in the E-B depletion region and base transport factor, α_t can greatly limit the performance of the HBT. The results of the experiments so far have not been conclusive since some experimental studies [3] have shown higher current gain in case of graded HBT, and some [2] have shown higher current gain in case of an abrupt HBT. These rather

	LAYER	THICKNESS	DOPING	Al FRACTION
Emitter	N AlGaAs	1000Å	5.0×10^{17}	0.3
Base	p GaAs	1000Å	1.0×10^{19}	
Collector	n GaAs	3000Å	1.0×10^{17}	

* Graded HBT has a linear grading of 300Å.

Table 1. The details of the HBT structure used in the simulation.

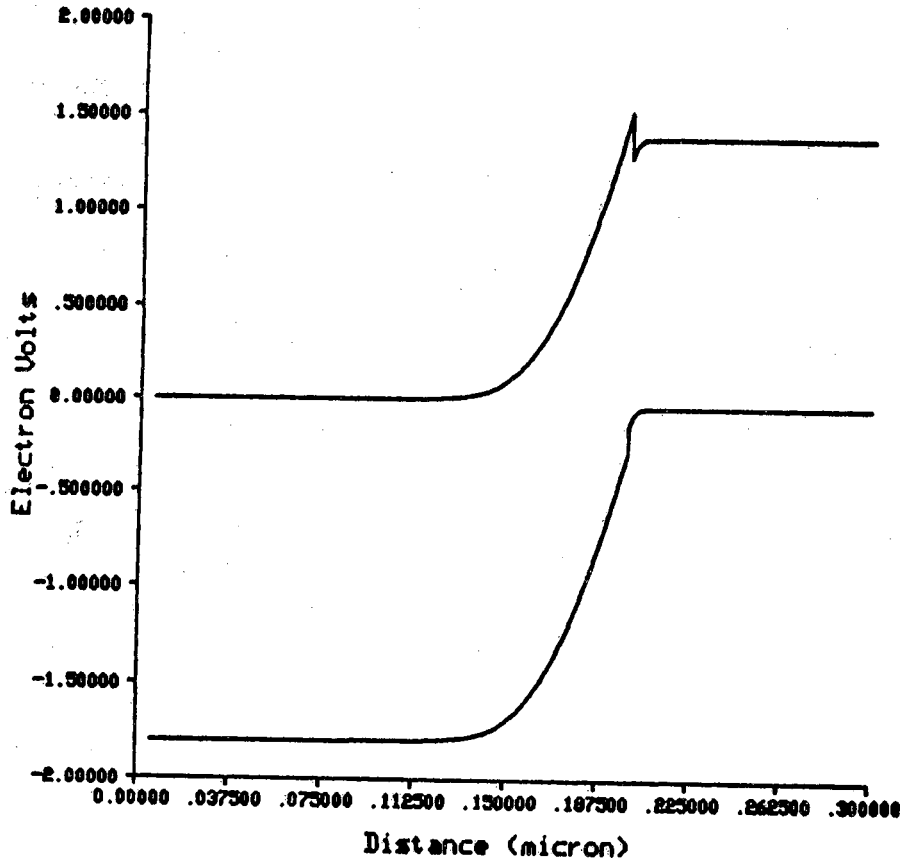


Fig. 1a. Equilibrium energy band diagram for abrupt Np heterojunctions. The device parameters are given in Table 1.

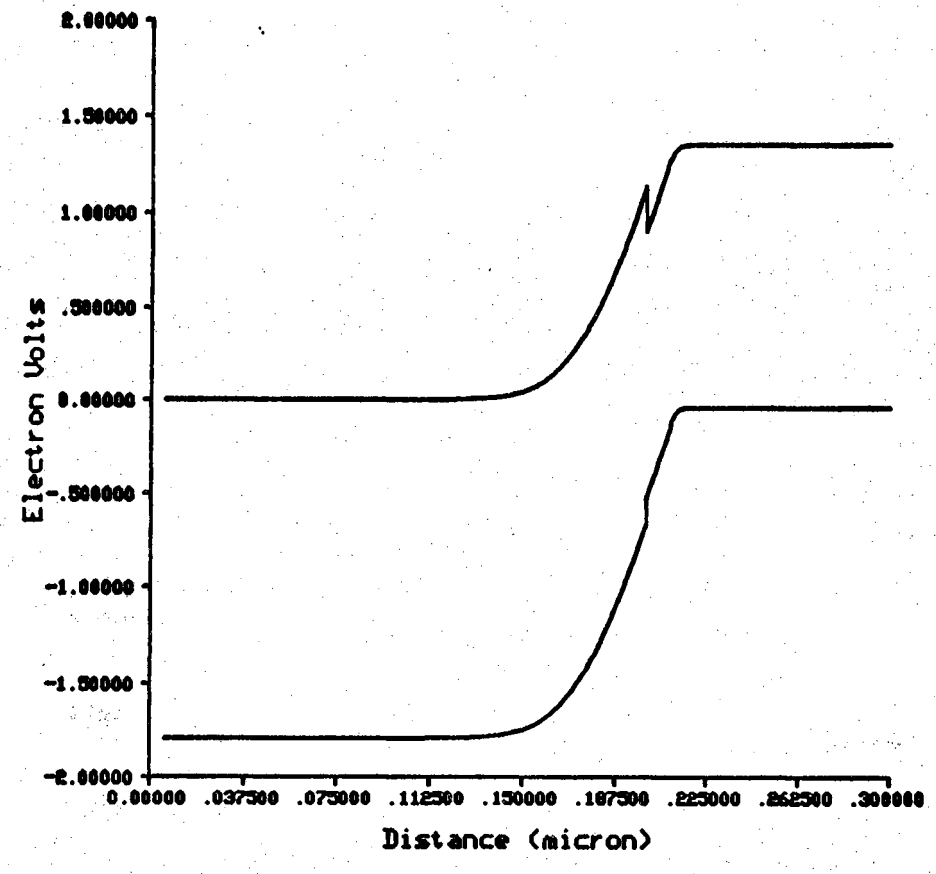


Fig. 1b. Equilibrium energy band diagram for abrupt Np heterojunctions with 100 Å intrinsic set-back layer. The device parameters are given in Table 1.

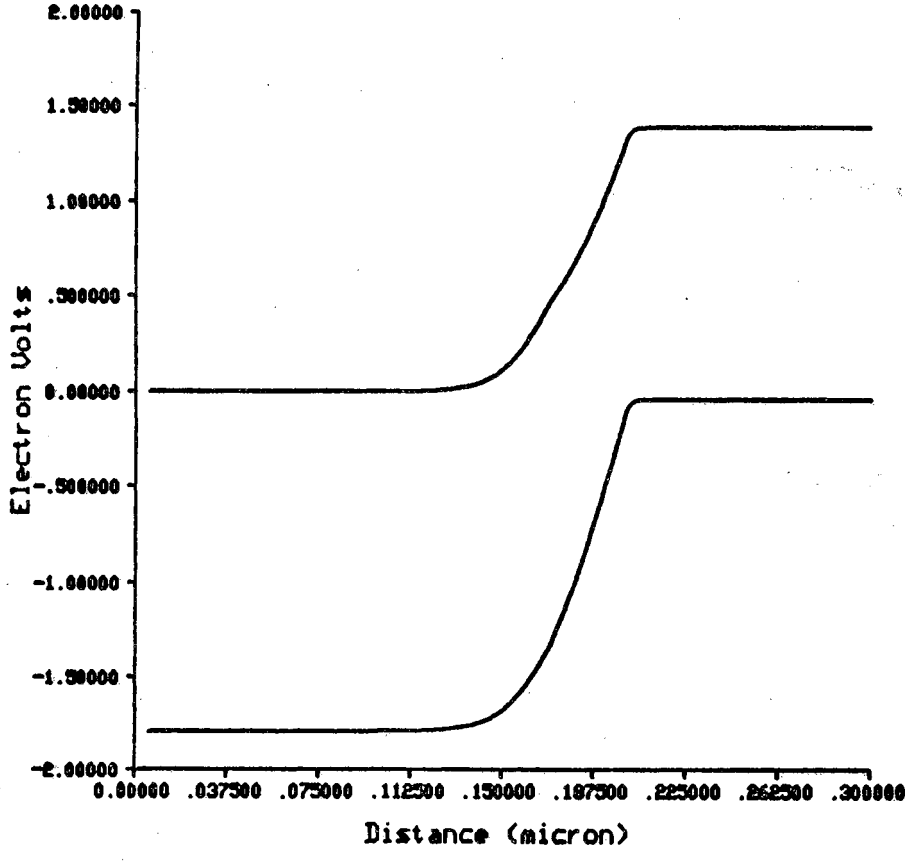


Fig. 2. Equilibrium energy band diagram for graded Np heterojunctions. The device parameters are given in Table 1. The grading is over a distance of 300 Å.

conflicting experimental results necessitates a model of HBT which can treat an abrupt HBT properly using quantum mechanical approach and also which includes the effect of grading on the recombination current at the E-B depletion region of an HBT.

In the past, quite a few models of HBT [4], [5] have been proposed for both the graded and abrupt HBT. But most of the models of HBT have treated electron injection across the HJ of an HBT using either thermionic emission or drift diffusion approach which are classical in nature. These models can treat a fully graded junction quite accurately, but it is severely limited when applied towards a study of an abrupt HBT, since these models fail to take into account the quantum mechanical tunneling current through the conduction band spike of an abrupt HJ which constitute the major portion of electron current across the emitter base junction of an abrupt HBT. Thus these models underestimate the magnitude of emitter injection efficiency across an abrupt HJ. The study reported in [5] treated the transport process across an HJ, using both a thermionic emission model and a thermionic field diffusion model which takes into account the tunneling of electrons through the conduction band spike. But in their quantum mechanical treatment of electron tunneling through the conduction band spike, they used an approximate WKB type method [6] which assumes the shape of the barrier to be parabolic. Further they neglected the variation of effective masses of electrons across the HJ and the reflection of carriers above the potential barrier. Also their model doesn't include the effect of grading on the recombination current at the E-B depletion region of an HBT which we found to be very important.

To remove some of the shortcomings and approximations of the previously proposed models of HBT's, we propose in this chapter a model which treats the electron injection in case of an abrupt HBT properly by using quantum mechanical approach. Our approach includes no assumption about the shape of the barrier and takes into account the variation of effective masses across the HJ. The elctron current across the E-B junction in case of graded HBT and hole current across the E-B junction in case of both the abrupt and graded HBT's are calculated using the conventional drift-diffusion approach which are expected to be quite accurate in these cases. The recombination current in the E-B depletion region is taken into account by using Shockley-Hall-Read model of recombinations of carriers through trap levels in the semiconductor. This model doesn't take into account the effect of recombination due to interface states and also the recombination along the perimeter of the emitter.

We then used quantum mechanically calculated electron current across the HJ of an abrupt HBT and classically calculated hole current given by (1), to calculate the emitter injection efficiency of an abrupt HJ (3). For the fully graded junctions, we

used the conventional drift diffusion approach to calculate both the electron (2) and hole (1) current.

$$J_p = \frac{qD_p}{W_e}\frac{{n_{ie}}^2}{N_d}(e^{\frac{qV_{eb}}{kT}} - 1), \tag{1}$$

$$J_n = \frac{qD_n}{W_b}\frac{{n_{ib}}^2}{N_a}(e^{\frac{qV_{eb}}{kT}} - 1), \tag{2}$$

where D_p,D_n are the diffusion coefficients of holes in the emitter and electrons in the base , W_e, W_b are the lengths of the emitter and the base, n_{ie}, n_{ib} are the intrinsic carrier concentrations in the emitter and base , V_{eb} is the emitter base bias, N_d, N_a are the emitter and base dopings, and J_p, J_n are the hole and electron current flowing out of the base and emitter respectively. The emitter injection efficiency is

$$\gamma = \frac{J_n}{J_n + J_p + J_{rg}}, \tag{3}$$

where J_n is the electron current across the heterojunction and J_{rg} is the recombination-generation (RG) current in the emitter base depletion region.

Our results show that for an idealised case, when the recombination current in the emitter base depletion region and in the quasi neutral base (i.e. $\alpha_t = 1.0$) are neglected, the common emitter gain of graded HBT is two order of magnitude higher than that in an abrupt HBT, but the magnitude of β for both the HBT's are much higher than experimentally observed values suggesting that recombination current plays an important role in limiting the value of experimentally observed β. So to account for the effect of recombination current in E-B depletion region , we included SHR recombination current in the calculation of emitter injection efficiency and to account for the recombination of minority carriers in the quasineutral base, we modified the base transport factor by taking diffusion length of electrons in the p type GaAs base as 1 μm [3]. The base transport factor is calculated from

$$\alpha_t = 1 - \frac{1}{2}\left(\frac{W_b}{L_n}\right)^2 \tag{4}$$

where W_b is the width of the base and L_n is the electron diffusion length in the base. With the inclusion of recombination current, our results show that the common emitter current gain of a graded HBT is lower than abrupt HBT at low bias where higher recombination current in the E-B depletion region of graded HBT dominates over electron current. As the bias increases across the E-B junction , the β of the graded HBT increases more rapidly than an abrupt HBT and at higher bias both give about the same value of β. The reason for this kind of behaviour at higher current is that the current gain at that level of emitter base bias is limited by the base transport factor rather than emitter injection efficiency in both the graded and abrupt HBT's, otherwise the gain of the HBT's would have been much higher than

the experimentally observed values.

The effect of an intrinsic set-back layer, generally used between base and HJ to stop diffusion of dopants into the emitter region, is also studied using the same model. Since the doping profile in the set-back layer will be dependent on the diffusion rate of the dopants from the base as well as on various parameters during growth of the crystal, it would be incorrect to assume some particular doping profile of the set back layer. For this we studied the extreme cases of a purely intrinsic set-back layer and no set-back layer, so that the real experimental situation will represent something in between these two extreme cases. The band diagram of an HJ with an intrinsic set-back layer is given in Fig. 1b. From our study, we found that the intrinsic set-back layer increases the recombination current in both the graded and abrupt junction by an order of magnitude. Also, in case of abrupt HBT, the intrinsic layer pulls down the conduction band spike which increased the electron current by an order of magnitude, while in case of graded HBT eletron current is not very much increased by the intrinsic layer. So as a result the inclusion of the intrinsic set-back layer deteriorated the performance of graded HBT more than compared to an abrupt one.

5.2 Quantum Mechanical Treatment of Electron Injection

The electron current across the abrupt E-B junction is obtained by analyzing the collisionless propagation of electron waves across the conduction band profile [7]. The energy band profile was obtained from a classical, numerical solution to Poisson's equation [8]. The formulation allows for arbitrary spatial variation of AlAs mole fraction and dopant density, includes Fermi-Dirac statistics for free carriers and treats shallow dopant deionization. From the resulting conduction band profile, we then compute the electron current by assuming that the contacts launch electron waves which propagate without collision through the structure. The numerical technique is essentially a quantum mechanical treatment of thermionic emission which includes the effects of carriers tunneling through the tip of the barrier, reflections of carriers above the barrier, and is valid for arbitrary variation of the conduction band edge. The technique is briefly described below.

Each of the two contacts of the one-dimensional junction is assumed to be in local thermodynamic equilibrium. The Fermi-levels of these contacts are separated by the applied bias. The contacts launch electrons into the device with a spectrum of momenta, $\vec{p}$. The electron wavefunction in the device,

$$\Psi(\vec{r}) = \psi(z)\exp(i\frac{\vec{p_t}}{\hbar}\cdot\vec{r_t}) , \tag{5}$$

is determined by solving

$$\frac{d}{dz}\left[\frac{1}{\theta(z)}\frac{d\psi(z)}{dz}\right]+\frac{2m^*(z_C)}{\hbar^2}\left[E_p+E_t(1-\theta(z)^{-1})-E_C(z)\right]\psi(z)=0 \qquad (6)$$

for the envelope function $\psi(z)$ [7]. In (6) $\theta(z)=m^*(z)/m^*(z_C)$ describes the spatial variation of the effective mass with respect to that in the contact $m^*(z_C)$. E_t is the transverse energy, $p_t^2/2m^*(z_C)$, and E_p the longitudinal energy, $p_z^2/2m^*(z_C)$. The contact is located at $z = z_C$.

The electron current is obtained by summing the contributions due to the various $\vec{p}$ for each of the two contacts. The net electron current is the difference between the currents injected from the two contacts

$$J_n=-q\int\frac{d^3\vec{p}}{4\pi^3\hbar^3}p_z\left[\frac{1}{m^*(z_L)}f_L(\vec{p})T_{LR}(\vec{p})-\frac{1}{m^*(z_R)}f_R(\vec{p})T_{RL}(\vec{p})\right], \qquad (7)$$

where $f_L(\vec{p})$ is the Fermi-Dirac factor with the Fermi level of the left contact, $m^*(z_L)$ is the effective mass in the left contact, and $T_{LR}(\vec{p})$ is the magnitude of the *current* transmission coefficient.

When $m^*(z)$ is position-independent, $\psi(z)$ as determined from (6) is independent of $\vec{p_t}$ and the integral over p_x and p_y in (7) can be performed analytically. The result is

$$J_n=-q\int\frac{dp_z}{2\pi\hbar}p_z\left[\frac{1}{m^*(z_L)}\sigma_L(p_z)T_{LR}(p_z)-\frac{1}{m^*(z_R)}\sigma_R(p_z)T_{RL}(p_z)\right] \qquad (8)$$

where $\sigma_L(p_z)$ is given by

$$\sigma_L(p_z)=\frac{m^*(z_C)\,k_BT}{\pi\hbar^2}\ln\left[1+\exp((E_{FL}-E_{CL}-E_p)/k_BT)\right]; \qquad (9)$$

$E_p = {p_z}^2/2m^*(z_C)$ and E_{FL} is the Fermi level in the left contact. Equation (8) shows that to evaluate the electron current, the wave equation (6) must be solved to find $T_{LR}(p_z)$ and $T_{RL}(p_z)$.

An examination of (6) shows that $\psi(z)$ depends on the transverse energy when the effective mass varies with position. The result is that the integral over p_x and p_y cannot be performed analytically and a rigorous evaluation of J_n would require a calculation of $\psi(z)$ over a grid of transverse energies. To avoid this complication, we follow Vassell [9] and replace E_t in (6) by its thermal average, k_BT.

The computational procedure consists of incrementing the longitudinal momentum from zero to some maximum value. For each p_z, the wavefunction is computed, and the contribution to J_n for electrons between p_z and $p_z + dp_z$ is evaluated. The contributions for each p_z are then summed to evaluate the integral (8) numerically. This procedure is then repeated for the other contact.

In addition to evaluating the electron current, it is also straightforward to evaluate the electron density n(z) from the solution of the wave equation. This carrier density can then be inserted in Poisson's equation and a new conduction band profile computed. The process can be repeated to iteratively obtain the self-consistent energy band profile. While such computations have been reported [7], they were not necessary for the problem under study because the junction region is depleted of carriers.

In the present version of our quantum mechanical model, we didn't consider the nonparabolicity of conduction bands and also neglected higher energy conduction bands (X_1, X_2 etc.). Since the doping level in the emitter of an HBT is not very high, the error introduced by the first assumption should be small. The second approximation is not very good when aluminium fraction in AlGaAs is greater than 0.45 since after that AlGaAs transforms to an indirect gap semiconductor from direct gap semiconductor and we need to consider both the Γ and X_1 valleys of AlGaAs. But generally in the case of an HBT, the aluminium fraction in AlGaAs emitter is below 0.45 which makes the emitter material a direct gap semiconductor and in that case our assumption should introduce negligible error [10].

To verify the accuracy of the above numerical technique, it is applied to study the transport of electrons across some simple potential barriers. The simple potential barrier used is a triangular barrier for which an analytic expression of transmission coefficient is obtained using Airy functions and tunneling current is calculated accurately from the analytic expression of transmission coefficient. Then the numerical technique discussed above, is applied to the triangular barrier and a result with less than 1.0% error is obtained in the calculation of transmission coefficient and tunneling current.

5.3 Electron Injection Current across the HJ of the HBT

The quantum mechanical method discussed in the previous section was applied to study a typical abrupt HBT structure. First we didn't include any set-back layer at the E-B junction of the HBT, and then we included a 100Å purely intrinsic set-back layer at the E-B junction of both the graded and abrupt HBT's studied first. The details of the structures are given in Table. 1. Figures 1a and 1b show the band diagrams of the abrupt HBT without and with a 100Å intrinsic set-back layer respectively. The band diagram of the graded structure without a set-back layer is given in Fig. 2. Our subsequent discussion is mostly centred around an HBT with no set-back layer unless the presence of an intrinsic set-back layer is specifically mentioned.

First, we studied an abrupt HBT. In Fig. 3a, we plotted the transmission probability versus incident energy of electrons across the HJ of the HBT under study, using our numerical technique and the analytical technique described by Grinberg et. al.[5]. As can be seen from the plot, the transmission coefficient calculated by numerically solving Schrodinger's equation (our method) is higher than calculated by the analytical technique of [5] at low energies and it also doesn't approach to unity immediately above the barrier due to reflection of carriers above the barrier. The transmission coefficient, calculated by the approximate method in [5] increases rapidly after certain energy and approaches to unity at the tip of the barrier. The overall effect of this type variation of approximate transmission coefficient from the actual one is that when it is integrated over the full spectrum of tunneling energy states, we get a compensating effect and the approximate method gives a fairly accurate result in terms of tunneling current. But the energy spectrum of the injected electrons given by this method would be inaccurate and for this reason this approximate method of [5] can not be used in cases where an energy spectra of the injected electrons is needed for example in order to study the transport of electrons across the quasi neutral base. For comparison purpose, a plot of injected current density across the HJ vs incident energy of the electrons calculated using two methods is given in Fig. 3b. The same kind of behaviour of the approximate method in [5] is also observed in case of the simple triangular barrier discussed earlier.

To estimate the effects of quantum mechanical tunneling across the abrupt HJ of an HBT, the electron current across the HJ was calculated both by quantum mechanical method and thermionic emission theory. The results are plotted in Fig. 4 along with the experimentally observed characteristics reported in [11]. As can be seen from the figure, the quantum mechanically calculated current is an order of magnitude higher than themionic emission current, which establishes the fact that thermionic field emission is the dominant mechanism of carrier transport across the emitter base junction of an abrupt HBT. It is also found that at higher biases as the potential barrier becomes thicker, thermionic emission current starts to dominate the overall magnitude of electron current. The same view is also shared by [5]. For comparison purpose, we calculated the tunneling current across the HJ using the technique described in [5],[6] and the results by the latter method, though an approximate one agrees reasonably well with our rigorous calculations (Our calculations show a 20-30% more tunneling current.). The reasons for this good agreement of the approximate method with ours are already discussed. As can be seen from Fig. 4, we also plotted the experimentally observed current voltage characteristics of a similar HBT reported in [11]. The thermionic emission current is found to be much lower than the experimentally observed value of current. The quantum mechanically calculated current agrees better with the experimental results compared to thermionic emission current but it is still lower than the experimentally observed value. The possible reasons for this discrepancy are discussed in the next section.

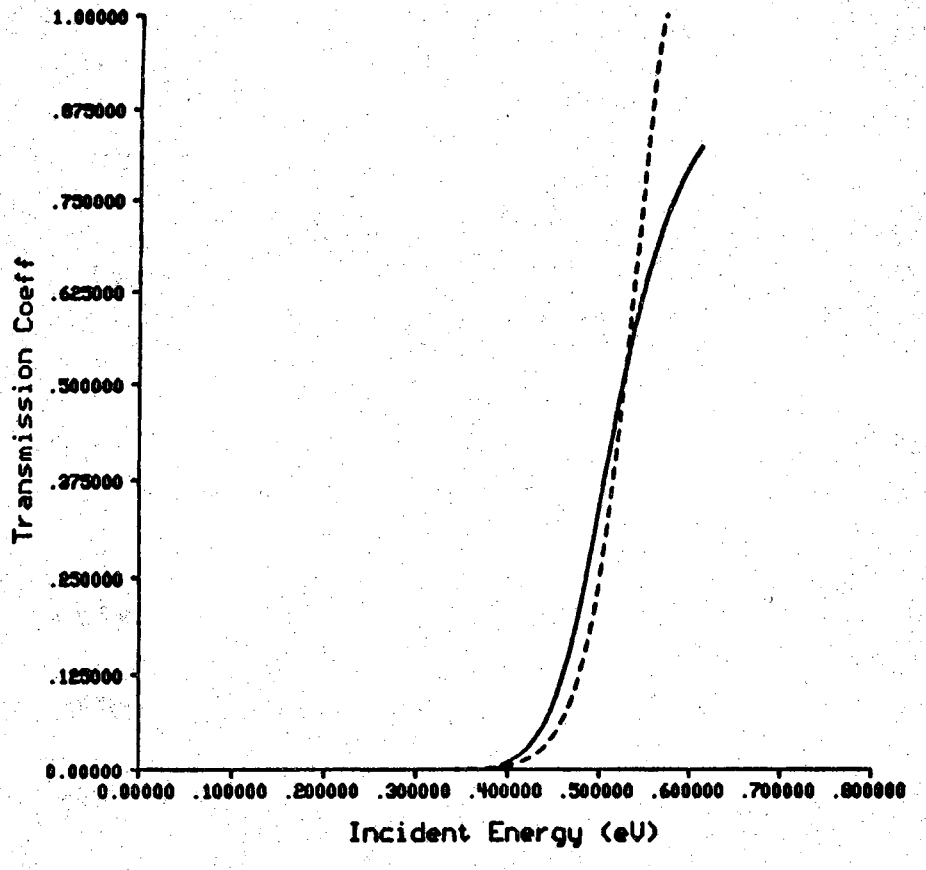


Fig. 3a. Transmission Coefficient vs. Incident Energy of electrons across the conduction band spike. Solid line and dashed line respectively indicate the rigorous and the analytical method [5] used for calculation.

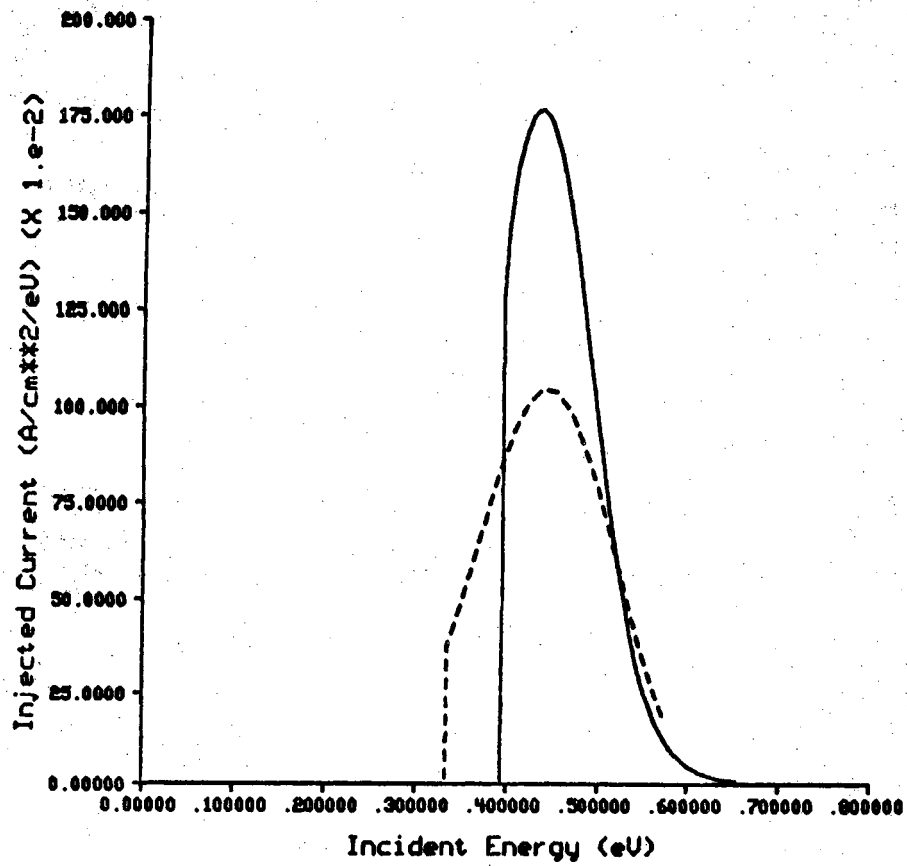


Fig. 3b. Injected Current Density (A/cm**2/eV) vs. Incident Energy (eV) across the conduction band spike. Solid line and dashed line respectively indicate rigorous and analytical methods [5] used for calculation.

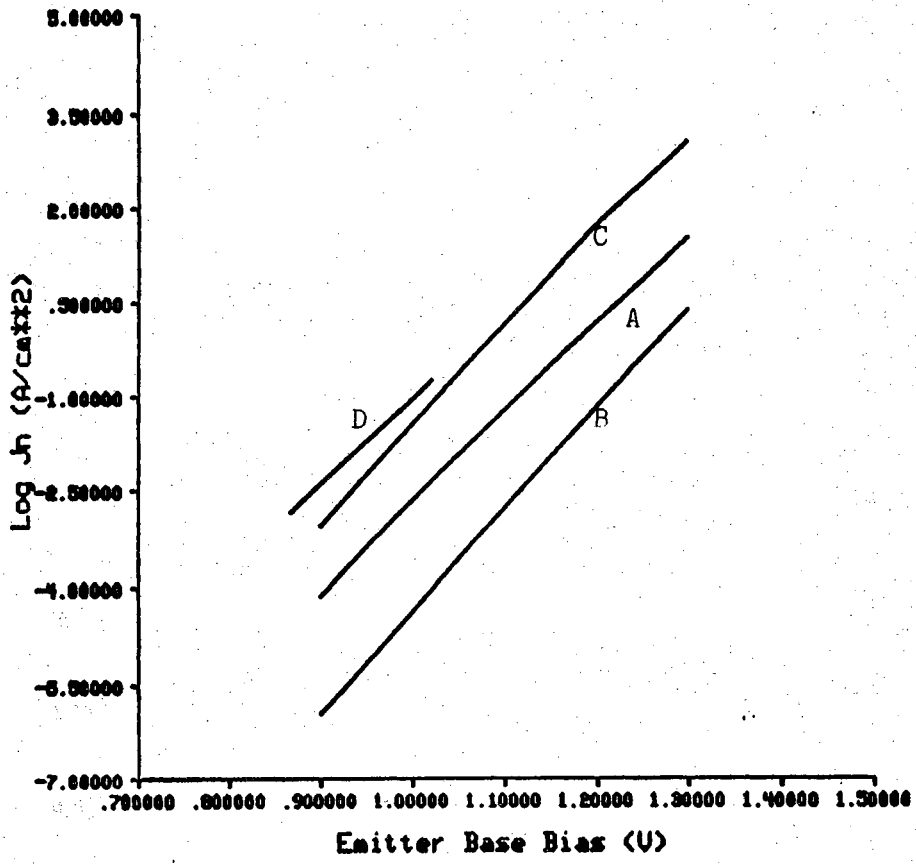


Fig. 4. Electron Current vs. E-B voltage across an abrupt heterojunction in the abrupt HBT, calculated using A) Quantum mechanical method B) Thermionic emission theory, C) Quantum mechanical method with a set-back layer, D) Experimentally observed in [11].

Next we included the intrinsic set-back layer in the structure and calculated the electron current using the same quantum mechanical technique (Fig 4.) and our results show that an intrinsic set-back layer increases the electron current by an order of magnitude and it is much closer to the experimentally observed value. The increase in the electron current comes from the lowering of the conduction band spike at the HJ due to the intrinsic layer of GaAs (Fig. 1b). Of course in an experimental situation, the set-back layer won't be purely intrinsic and the increase in the electron current will depend on the doping profile of the intrinsic region. So our calculation with the intrinsic layer basically gives the upper limit of the electron current. So we see that the electron current calculated using our model is still lower than that observed by Ito et.al. We believe the reason for this discrepancy mailnly arises from the fact that in realistic devices a junction can't be perfectly abrupt in doping (see Fig. 1 of [11]) as modeled in our program. Also there will be recombination due to interface states and parasitic recombinations along the perimeter of the emitter contact which we neglected in our calculation. Also it is noted that the experimental data for two very similar structures grown by two different methods and reported by two different researchers vary widely between themselves [3], [11]. For this reason no attempt is made in this chapter to precisely match the experimental data with our calculated data, instead the nature of variation of different characteristics of HBT is discussed with the aid of our model.

In case of the graded HBT, the electron current is calculated using the conventional drift and diffusion approach [12] and as expected the magnitude of the electron current is found higher than in the similar abrupt structure at the same E-B bias (Fig. 4). But the effect of set-back intrinsic layer on the magnitude of electron current across the E-B junction in case of graded HBT is found very nominal as compared to the case of an abrupt HBT. This is because in case of a graded HBT there is no conduction band spike at the HJ that is pulled down due to inclusion of the intrinsic layer, so the barrier to the electrons more or less remain the same. The effect of intrinsic set-back layer is much more profound in the recombination current which is discussed in the next section.

5.4 Recombination-Generation current at Heterojunctions

The recombination-generation (RG) current in an HBT mostly comes from the emitter base depletion region and from the recombination of minority carriers in the quasi-neutral base. The recombination of minority carriers in the quasi-neutral base is taken into account by taking the diffusion length of minority carriers in the base to be 1 μm and then calculating the base transport factor as discussed before.

The RG current in the forward biased E-B depletion region of an HBT around the operating bias comes mainly from the recombinations of carriers through the

midgap states. The recombination rate is computed numerically using Schockley-Hall-Read model (SHR) as given in the following equation

$$R(x) = \frac{p(x)n(x) - n_i^2(x)}{\tau_p(n(x)+n_i(x)\exp((E_t-E_i)/kT)+\tau_n(p(x)+n_i(x)\exp(-(E_t-E_i)/kT))} \quad (10)$$

where τ_p and τ_n are minority carrier life times and are assumed uniform and equal to 1 nsec throughout the device. E_t is the trap level, E_i, the intrinsic fermi level; k and T are respectively Boltzmann constant and temperature.

The position dependent recombination rates for both an abrupt and graded HJ are presented in Fig. 5. From the plot, we can see that the rate of recombination is higher in a graded junction as compared to an abrupt one. This is because most of the recombinations occur in the larger bandgap emitter and in the case of graded junctions the intrinsic carrier concentration is increased as compared to the abrupt one in the emitter side of the depletion region due to grading. Experimental observation of the base current in both the graded and abrupt HBT supports this view [3]. In this calculation of recombination current, we neglected recombination due to interface states present at HJ and also the recombination current that arises due to recombinations along the perimeter of the emitter.

The inclusion of the intrinsic set-back layer in both the graded and abrupt HBT increased the recombination current in the emitter base depletion region by an order of magnitude. This increase in the recombination of minority carriers in the E-B depletion region is mainly the result of an additional band bending on the base side of the E-B junctions due to the inclusion of the set-back layer which traps electrons.

As can be seen from our studies, the effect of the intrinsic layer is more profound in case of a graded junction than an abrupt junction since in the former the electron current remains about the same whereas recombination current increases, but in case of abrupt HJ both the electron and recombination current increases with the inclusion of the set-back layer. Thus the set-back layer reduces the magnitude of γ more in case of graded HBT than in an abrupt one, particularly at low bias. So the inclusion of the set-back layer in one hand improves the performance of the HBT's by stopping diffusion of dopants from the p type base into the N type emitter, on the other hand it also increases the RG current in the E-B depletion region. It should be made clear that our study takes the case of two extreme situation, one with no set-back layer and one with purely intrinsic set-back layer. The actual situation will be some what in between thses two extremes.

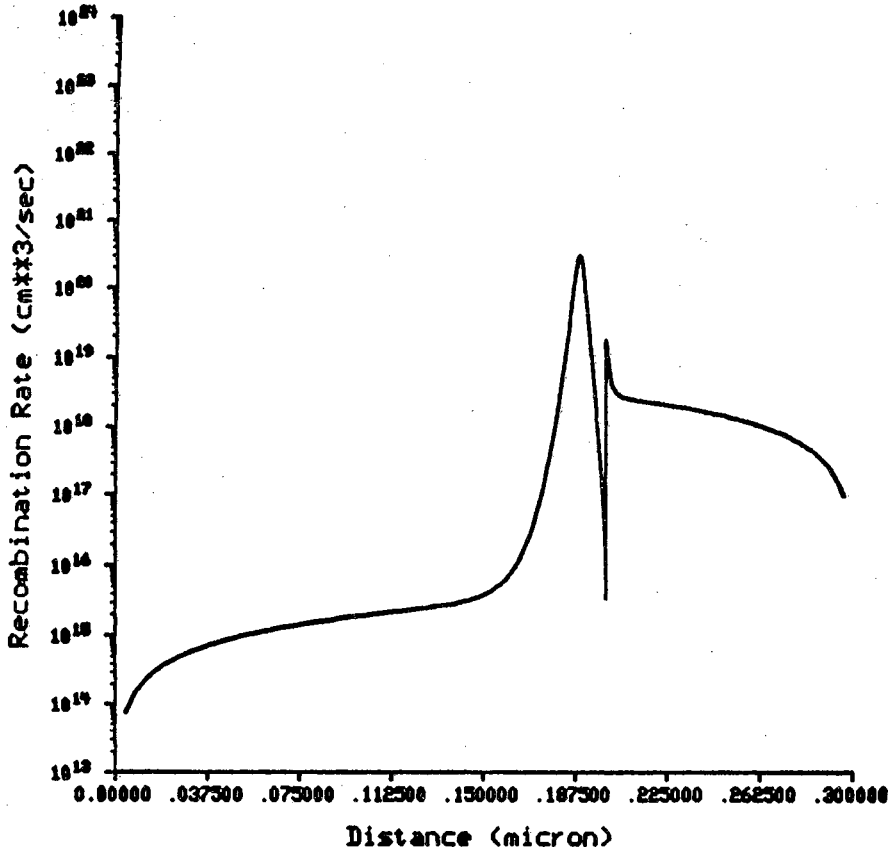


Fig. 5a. Recombination rate in the E-B depletion region of an abrupt HBT.

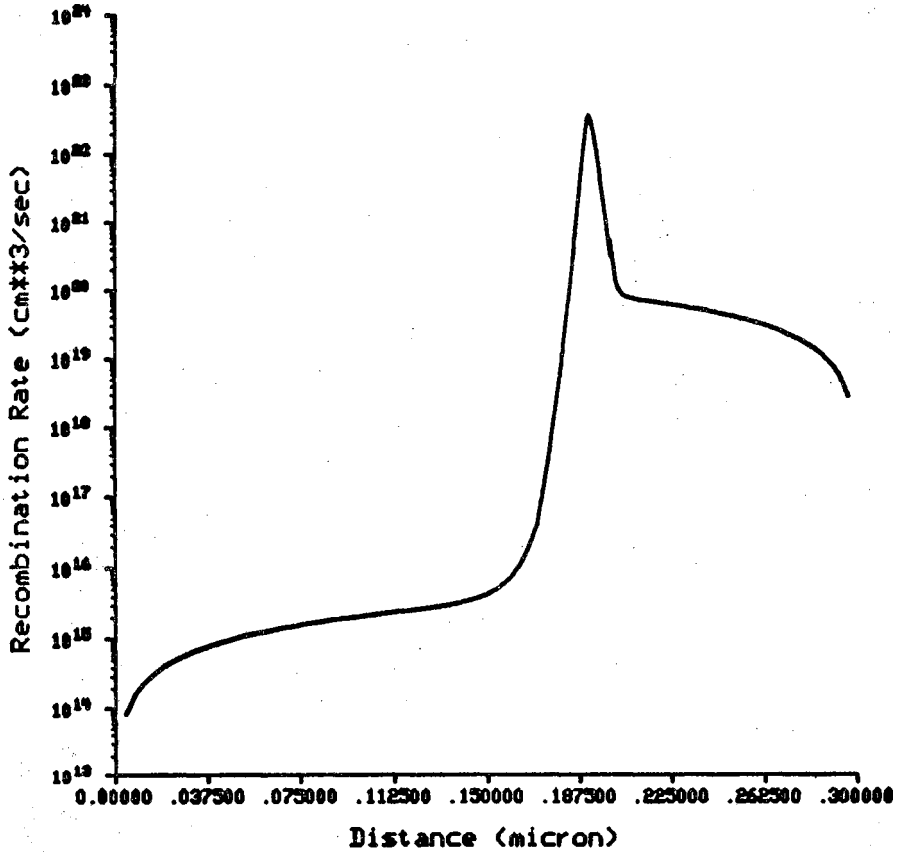


Fig. 5b. Recombination rate in the E-B depletion region of a graded HBT.

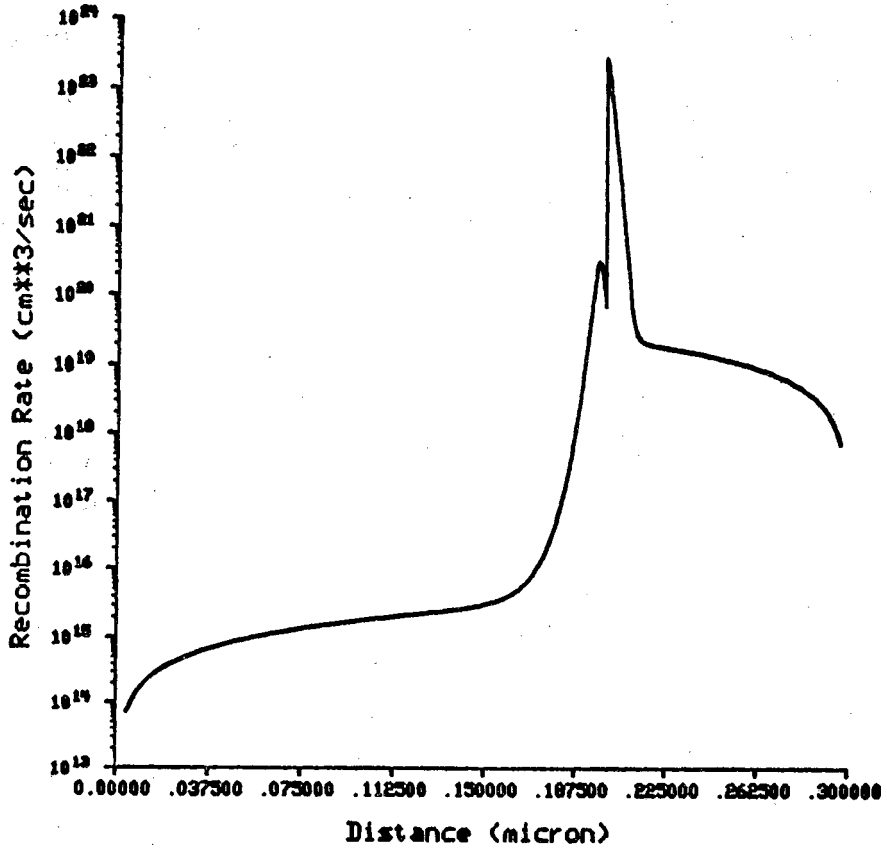


Fig. 5c. Recombination rate in the E-B depletion region of an abrupt HBT with an 100 Å set-back layer.

5.5 Common Emitter Current Gain of Abrupt and Graded HBT's

In this section, first we present our calculation of common emitter current gain β for both the graded and abrupt HBT with no set-back layer (see Table 2). In the later part of this section, the effect of inclusion of the set-back layer on the magnitude of β is discussed.

The β of the abrupt HBT is calculated using the quantum mechanical method discussed in Sec. II. The hole current and RG current are calculated using classical formulations. In case of the graded HBT, the conventional drift diffusion approach is used. The base transport factor in both the graded and abrupt HBT's are taken to be 0.995 (Calculated by taking the diffusion length of electron in the base to be 1 μm). The results of the calculations are presented in Table. 2 and in Figs. 6a,6b.

As can be seen from the plot of β versus V_{eb}, the common emitter current gain β for the graded HBT is much higher than the abrupt HBT in case of an ideal situation where there is no recombination current in either quasi neutral base or emitter base depletion region. The values of β for both the graded and abrupt HBT's in this ideal case are much higher than the experimentally observed values. This confirms that the recombination current plays a major role in limiting the magnitude of β in both the graded and abrupt HBT. As can be seen from the other plot of β versus V_{eb} which includes the effect of recombination current, the common emitter current gain, β, of a graded HBT is lower than the abrupt HBT at lower bias. The reason for this kind of behaviour is that the recombination current in case of graded HBT is relatively large compared to an abrupt one and at low value of electron current, this higher recombination current in case of graded HBT reduces the emitter injection efficiency of a graded HBT more than that of an abrupt one. At higher bias, the β of a graded HBT increases more rapidly than abrupt HBT but both saturates around 170 to 200, the current gain of the graded HBT being slightly higher than the abrupt one. These comparable values of current gains of the graded and abrupt HBT suggests that abrupt HBT has sufficient electron injection efficiency to have a comparable current gain of a graded one.

To investigate this further, the emitter injection efficiency is calculated for both abrupt and graded heterostructures. It is found that at higher values of injected electron current though γ for graded junctions is higher than that in an abrupt junction, still both are much higher than the magnitude of α_t. This means, the base transport factor which is lower than such a high value of γ, will primarily control the common emitter current gain. In that case, the emitter injection efficiency obtained in the case of abrupt heterojunctions is adequate enough for a high current gain and it is the transport of electrons through the base which determines whether an abrupt HBT or a graded HBT will give higher β. It should also be mentioned that modelling of base transport factor is not adequate at present, and an accurate modelling of transport of electrons across the base will result in a higher base transport factor for an abrupt

Device	Bias (V)	Em. Inj. Effy.		Common Emitt. Curr. Gain			
		Includes R-G		No R-G α_t=1.0		Includes R-G α_t=0.995	
		Quant	Thermo	Quant	Thermo	Quant	Thermo
Abrupt HBT	0.9	0.9654	0.29405	2×10^4	309	25	1
	1.0	0.99118	0.687	1.24×10^4	249	72	3
	1.1	0.9976	0.9198	0.74×10^4	203	135	11
	1.2	0.999	0.9796	0.38×10^4	179	165	39
Graded HBT	0.9	0.898		1.0×10^6		9	
	1.0	0.9766		8.8×10^5		35	
	1.1	0.9956		3.5×10^5		106	
	1.2	0.9992		3.4×10^5		171	

Table 2. The results of simulation.

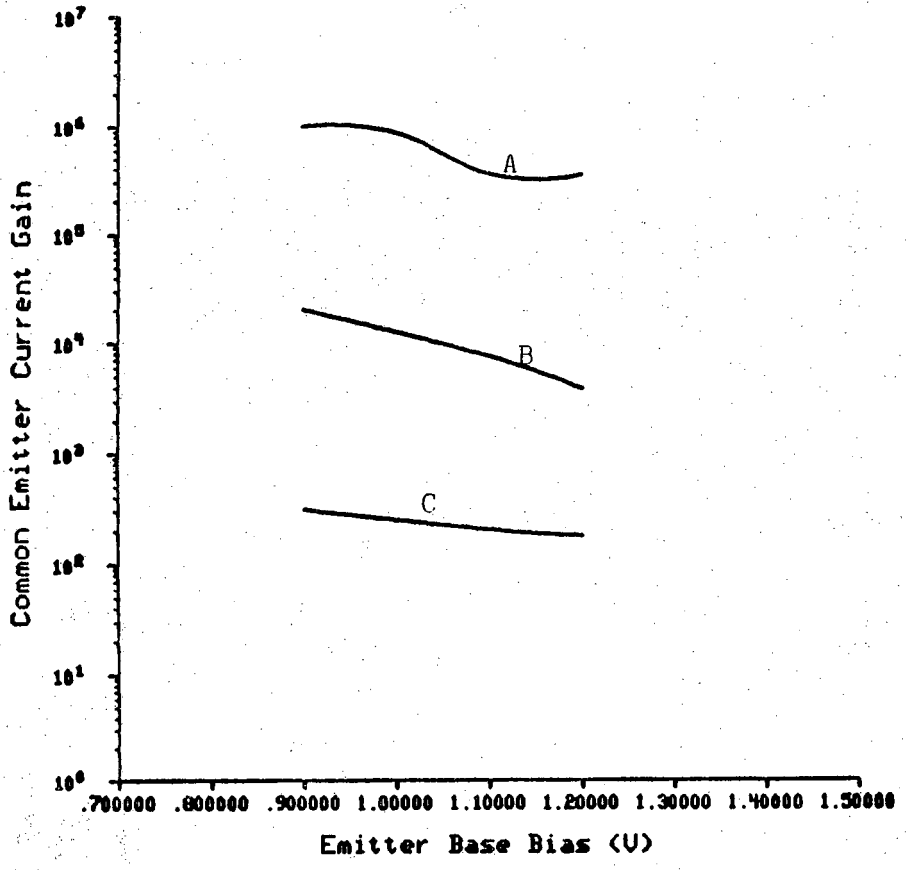


Fig. 6a. Common emitter current gain vs electron current across the HJ, calculated using A) Drift-diffusion model for graded HBT, B) Quantum mechanical model for abrupt HBT, C) Thermionic emission model for abrupt HBT. The RG current in the depletion region is neglected and $\alpha_t = 1.0$.

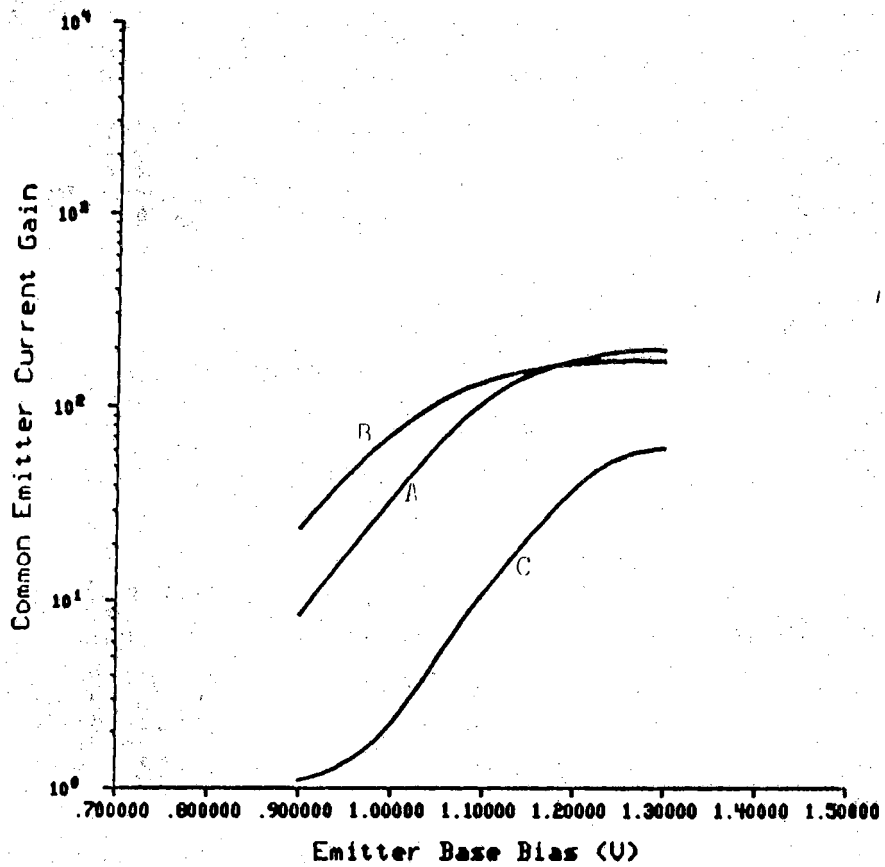


Fig 6b. Common emitter current gain vs electron current across the HJ, calculated using A) Drift-diffusion model for graded HBT, B) Quantum mechanical model for abrupt HBT, C) Thermionic emission model for abrupt HBT. The RG current in the depletion region is included and α_t = 0.995.

heterojunction, since electrons are injected at a higher energy into the base compared to that in a graded heterojunction. These hot electrons spend much less time in the base and consequently their probability of recombination is much less compared to diffused electrons as in the graded heterojunction. In our calculations we have taken the base transport factors for both the graded and abrupt HBT to be the same. A higher base transport factor in case of abrupt HBT will eventually explain a higher common emitter current gain in an abrupt HBT as observed in [2].

Another interesting observation from Fig. 6 can be made about the use of thermionic emission model for abrupt HBT's. Since thermionic emission model underestimates the electron current across an abrupt HBT to a great extent, the emitter injection efficiency obtained from the use of this model is less than the actual one. Because of the lower emitter injection efficiency predicted by it, the thermionic emission model predicts the common emitter current gain of an abrupt HBT to be much lower than graded HBT and thus this model fails to bring out the correct picture of the behaviour of an abrupt HBT.

Next we calculated β for the case where an intrinsic set-back layer of 100 Å exists at the E-B junction of both the graded and abrupt HBT's. The results are plotted in Figs 7a and 7b for both the abrupt and graded HBT's. From the plot, we can see that the effect of the intrinsic layer on the performance of the graded HBT is to reduce the magnitude of β whereas in case of the abrupt HBT, the magnitude of β remains about the same in both the cases. This kind behavior is observed due to a relatively higher reduction of γ in case of graded HBT as compared to an abrupt one. The reason for lower γ in case of graded HBT is discussed in the earlier section. It should be mentioned that the presence of the set-back layer is necessary to prevent the diffusion of dopants across the heterojunction which otherwise will shift the location of the N-p junction from the compositional HJ. This will also degrade the performance of the HBT. Moreover as mentioned earlier our study treats the two extreme cases of an intrinsic setback layer and no set-back layer. So the actual situation will be somewhat in between these two extremes. So this study gives some qualitative effects of inclusion of a set-back layer to help the experimentalists decide on the design trade-offs involved in this matter.

5.6 Conclusion

In this chapter, we described a quantum mechanical method to analyse transport of electrons across any arbitrary shape of potential barrier. The numerical technique discussed in the earlier section is used to study the transport of electrons across the emitter base junction of HBT. From our study we found that the WKB type method discussed in [5] has reasonable accuracy to be used for the calculation of tunneling current in case of abrupt HBT, but the method is limited to a particular shape of the

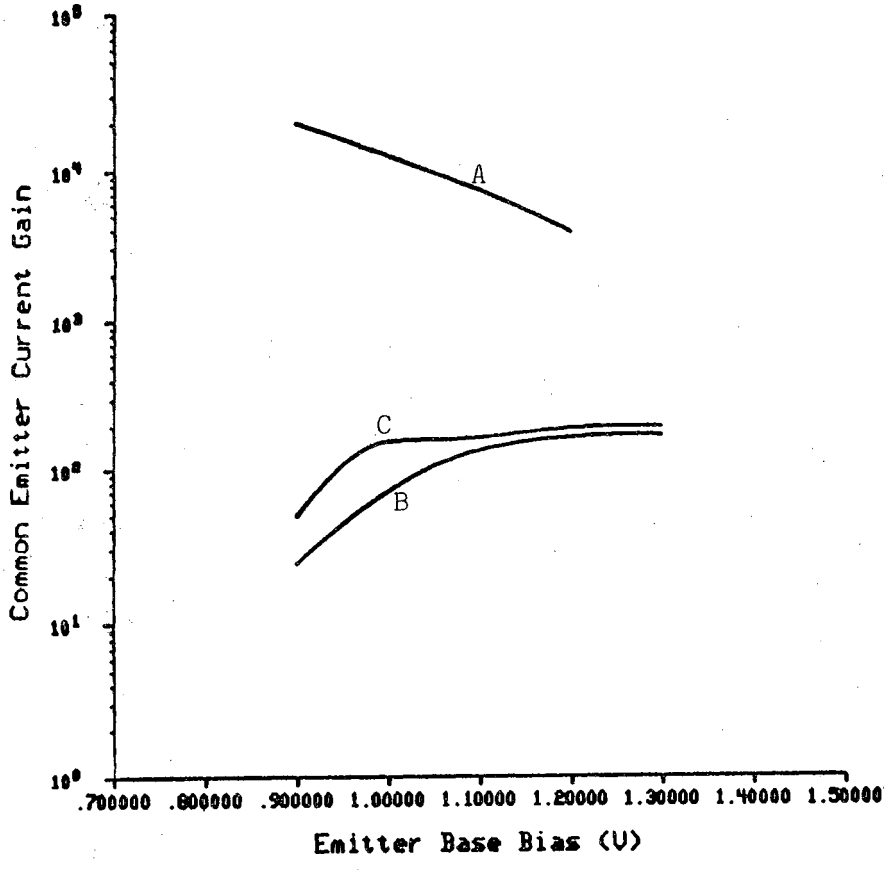


Fig. 7a. The effect of set-back layer on the current gain of the abrupt HBT. A) The ideal case, B) When RG in the depletion region is included and α_t = 0.995 C) When the 100 Å intrinsic set-back layer is included.

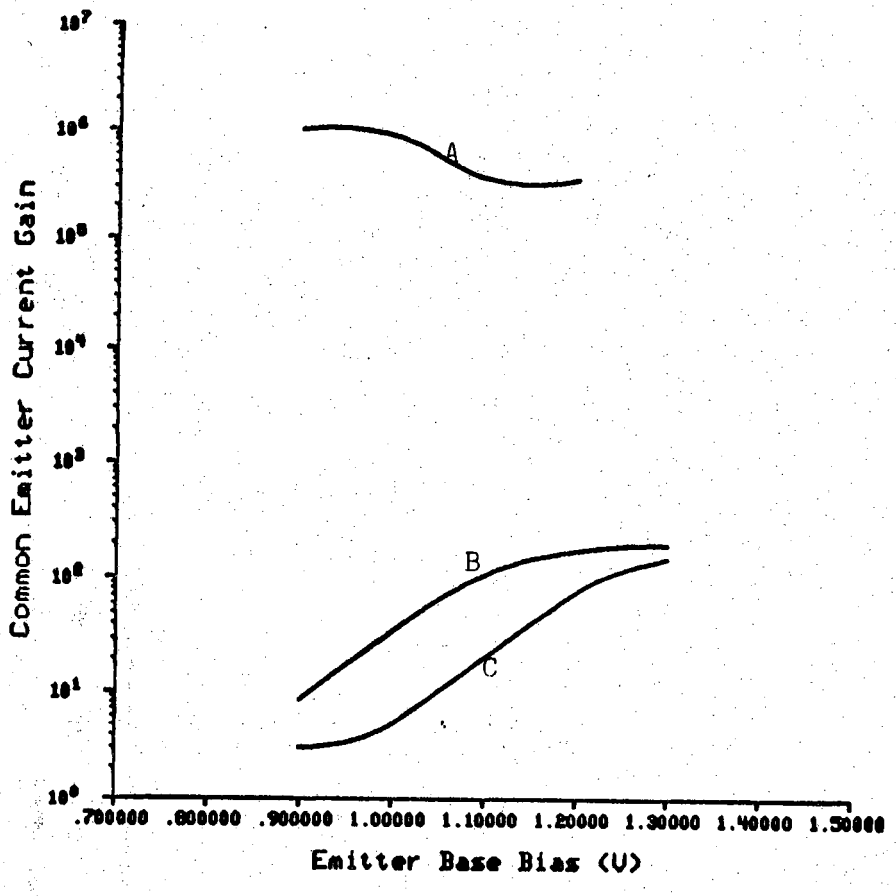


Fig. 7b. The effect of set-back layer on the current gain of the graded HBT. A) The ideal case, B) When RG in the depletion region is included and $\alpha_t = 0.995$ C) When the 100 Å intrinsic set-back layer is included.

potential barrier. The method discussed in this chapter is perfectly general and can be used for any arbitrary shape of the potential barrier.

The treatment of electron transport by the quantum mechanical method shows that the thermionic emission theory gives an underestimated value of γ in case of the abrupt HBT leading to an incorrect estimate of its performance.

Our study shows that RG current plays an important role in limiting the performance of an HBT. It is also found that RG current in case of the graded HBT is an order of magnitude higher than that in case of abrupt HBT. This reduces the the magnitude of γ, consequently the magnitude of β in case of the graded HBT more than in an abrupt HBT. The inclusion of the intrinsic set-back layer further deteriorates the performance of the graded HBT.

We also found that at low value of collector current, RG current plays an important role in deciding the magnitude of γ and subsequently β in case of both the graded and abrupt HBT's. But the effect on the graded HBT is more pronounced due to a higher RG current. At higher value of collector current, it is found that β will be limited by α_t rather than γ, which leads to the conclusion that it is possible to achieve higher current gain in case of abrupt HBT in contrary to the earlier studies. The experimental observations in [2] supports this.

Our quantum transport model has some limitations. The most important one is that it can't treat collisions in the device at present. Also in this model we need to consider higher energy bands other than Γ valley, if we want to treat indirect gap AlGaAs, in other words the present model gives best results when aluminium concentraton in AlGaAs is low enough to make it indirect i.e. $x<0.45$. In our study of HBT, we neglected the recombination current due to interface states. Also the recombination current around the edge of the emitter is neglected in this study. To obtain a quantitave agreement of the theoretical results with the experimental ones, these effects have to be incorporated in this model.

References

[1] H. Kroemer, "Heterostructure Bipolar Transistors and Integrated Cirsuits," *Proc. IEEE,* Vol. 71, p. 13, 1982.

[2] P.M.Enquist, L.P.Ramberg and L. F. Eastman "Comparison of compositionally graded to abrupt emitter-base junctions used in the heterojunction bipolar transistor" *J. Appl. Phys.,* Vol. 61, pp. 2663-2669, 1987.

[3] K. Taira, C. Takano, H. Kawai and M. Arai "Emitter grading in AlGaAs/GaAs heterojunction bipolar transistor grown by metalorganic chemical vapor deposition" *Appl. Phys. Letters,* Vol. 49, pp. 1278-1280, 1986.

[4] J. Yoshida, M. Kurata, K. Morizuka, and A. Hojo, "Emitter-Base Bandgap Grading Effects on GaAlAs/GaAs Heterojunction Bipolar Transistor Characteristics," *IEEE Trans. Electron Dev.,* Vol. ED-32, pp. 1714-1721, 1985.

[5] A.A. Grinberg, M.S. Shur, R.J. Fischer, and H. Morkoc, "An Investigation of the Effect of Graded Layers and Tunneling on the Performance of AlGaAs/GaAs Heterojunction Bipolar Transistors," *IEEE Trans. Electron Dev.,* Vol. ED-31, pp. 1758-1765, 1984.

[6] F. A. Padovani and R. Stratton "Field and Thermionic-Field emission in Schottky barriers" *Solid-State Electron.,* Vol. 9, pp. 695-707, 1966.

[7] M.S. Lundstrom and R.J. Schuelke, "Numerical Analysis of Heterostructure Devices," *IEEE Trans. Electron Dev.,* Vol. ED-30, pp. 1151-1159, 1983.

[8] M. Cahay, M.McLennan, S. Datta and M.S. Lundstrom, "Importance of space charge effects in resonant tunneling devices," *Appl. Phys. Letters* Vol. 50, pp. 612-614, 1987

[9] M.O. Vassell, J. Lee, and H.F. Lockwood, "Multibarrier Tunneling in GaAlAs/GaAs Heterostructures," *J. Appl. Phys.,* Vol. 54, pp. 5206-5213, 1983.

[10] A. C. Marsh "Electron Tunneling in GaAs/AlGaAs Heterostructures" *Journal of Quantum Electronics,* Vol. QE-23, pp. 371-376, 1987.

[11] H. Ito "Generation Recombination current in the Emitter-Base Junction of AlGaAs/GaAs HBT," *Japanese Journal of Applied Physics* Vol. 25, pp. 1400-1404, 1986.

[12] S.M. Sze, *Physics and Technology of Semiconductor Devices,* 2nd Ed., John Wiley and Sons, New York, 1984.

SEQUAL 2.0

A Guide for the User

Michael J. McLennan
Purdue University
Department of Electrical Engineering
West Lafayette, IN 47907

Preface

SEQUAL is a device simulation program which performs a quantum mechanical analysis of electron transport. The user need not be an expert in quantum mechanics to obtain useful results from this program, although an understanding of the subject is an invaluable aid in the interpretation of output. For this reason, the theoretical formulation of the analysis is presented in some detail, following a brief introduction. The knowledgeable (or anxious) user may turn directly to the discussions of input and output in Chapters 3 and 4. The reader is cautioned, however, not to skip Chaper 2 entirely. At the heart of many a program "error" lies a misinterpretation of output.

Table of Contents

List of Tables

List of Figures

1

Introduction

SEQUAL is a device simulation program, computing Semiconductor Electrostatics by QUantum AnaLysis. Given the characteristics of a particular device, SEQUAL will compute the electron density and the current density, using a quantum mechanical description of electrons. To enhance the versatility of this program, it was designed to be a post-processor for classical simulation programs; as such, it does not calculate any material parameters, so it can be used with any material system. Instead, SEQUAL relies upon the output from classical analysis programs such as FISH1D, PUPHS, or SEDAN, to provide the description of a device. A direct comparison of classical and quantum mechanical results is then possible.

Often (for devices which merit the use of this program) the difference between classical and quantum mechanical results is significant. In such cases, the difference in electron density places the classical description of the electrostatic potential in doubt. The quantum mechanical distribution of space-charge demands a solution for the electrostatic potential which is self-consistent. When a self-consistent solution is desired, SEQUAL allows for iteration between calculations of the electron density and the electrostatic potential. Hence, SEQUAL can provide a correct solution of quantum mechanical electrostatics.

Figure 1.1 depicts the input/output structure of SEQUAL. A numerical description of the device, obtained from the output of a classical analysis program, is the primary source of input. In addition, SEQUAL accepts a list of commands called the "input deck," which allows for selection of program options. A summary of the analysis is written to standard output, in a form well suited to printing. SEQUAL does not produce any plotted output, since undoubtedly the user has a program favored for this purpose. Instead, SEQUAL can generate output files in a variety of formats, to interface with many plotting programs. Quantities available for writing to output files include the input device description, the electron density, the transmission coefficient, and the squared magnitude of the electron wavefunction, output as a surface in both position and energy spaces.

The following chapters are intended not only to describe the operation of SEQUAL, but also to provide some understanding of the calculations performed. In

Chapter 2, the theoretical formulation of the analysis is discussed. Some of the difficulties encountered in a numerical implementation of the theory are noted, and an approach is outlined, which is used by SEQUAL to insure accurate results. Chapter 3 details the input structure of SEQUAL, specifying both the syntax of the input deck and the format of the device description. Program output is discussed in Chapter 4, and an example calculation is presented in Chapter 5. Finally, Chapter 6 examines some practical considerations, for the implementation and use of SEQUAL.

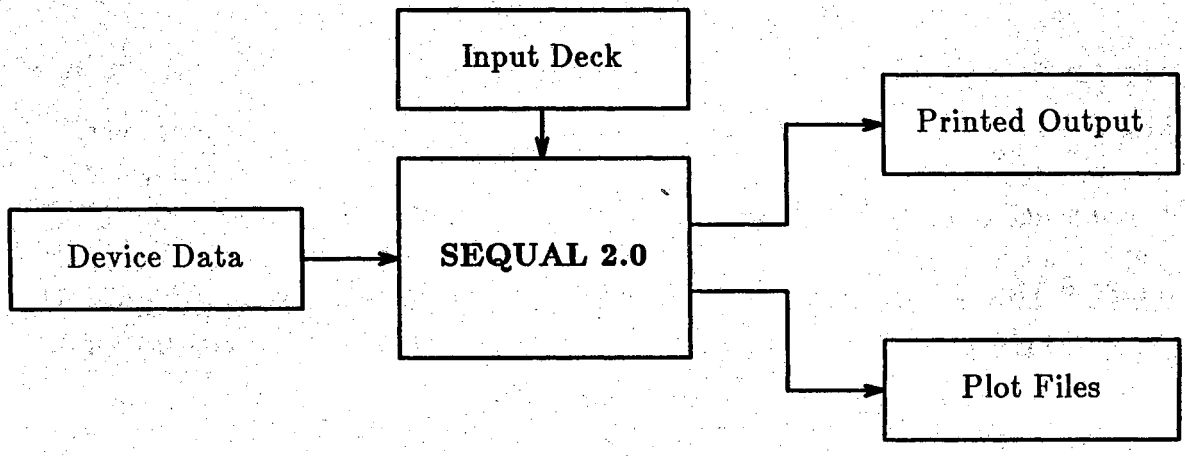


Figure 1.1: Structure of input/output in SEQUAL

2

Theoretical Formulation

The computations performed in SEQUAL are based upon a number of simplifying assumptions:

- Electrons are assumed to be majority carriers; holes are completely ignored in the analysis.
- Profiles of doping density and material composition assumed to vary in only one dimension (hereafter referred to as the "longitudinal direction").
- Effects of the crystal potential are parameterized by an effective mass, which is constant in each material region, and which changes abruptly at a material interface.
- Solutions reflect steady-state conditions, and are independent of time.
- The energy dispersion relation is assumed to be parabolic.
- Electron-electron interactions are neglected.
- Transport is assumed to be ballistic; overall device dimensions are assumed to be sufficiently short, so that scattering can be neglected.

A pictorial representation of the analysis is presented in Figure 2.1(a). Contacts, assumed to be in local thermodynamic equilibrium, inject electrons into a device, with a spectrum of wavevectors $\vec{k}$. These electrons interact with the potential profile E_C; some are transmitted across the structure, and the remainder are reflected back to the injecting contact. Contacts are assumed to supply a continuum of electron energies, and these energies are referred to as "propagating states." If inelastic scattering were included in the model, some of the injected electrons could scatter into states of lower energy. In particular, some electrons could scatter into states which could not be populated by either contact. Such energies, which fall below the conduction-band

edge of either contact, are referred to as "bound states." Solutions of the Schrödinger equation for both propagating states and bound states are described below. Because the primary importance of SEQUAL is its application to transport problems, the solution for propagating states is considered in more detail.

2.1. Propagating Electronic States

2.1.1. Solution of the Schrödinger Equation

Quantum mechanically, an electron is represented by a wavefunction $\psi_{\vec{k}}(\vec{r})$, which is obtained by solving the (time-independent) Schrödinger equation:

$$\frac{\partial}{\partial z}\left(\frac{m_c^*}{m^*(z)}\frac{\partial}{\partial z}\psi_{\vec{k}}(z)\right)+\frac{2m_c^*}{\hbar^2}\left[E_z+E_t\left(1-\frac{m_c^*}{m^*(z)}\right)-E_C(z)\right]\psi_{\vec{k}}(z) = 0 \quad (2.1)$$

where

$$E = \frac{\hbar^2 k_z^2}{2m_c^*}$$

is the energy of the injected electron in the direction of propagation, and

$$E_t = \frac{\hbar^2}{2m_c^*}(k_x^2 + k_y^2)$$

is the energy in transverse directions. In the form presented above, the Schrödinger equation accounts for spatial variations in the effective mass $m^*(z)$, relative to the effective mass of the injecting contact (m_c^*). The conduction-band profile $E_C(z)$ is assumed to be known for a particular device.

An arbitrary potential profile can be described by segmenting a device into a number of tiny intervals, as shown in Figure 2.1(b). Each interval is delineated by points in position space called "nodes." Within each interval, the potential is approximated by its average value. Of course, intervals should be sufficiently small to accurately represent a potential profile. In an region of constant potential, the solution of the Schrödinger equation is some linear combination of plane-waves; hence, the wavefunction for an entire device is deduced by matching boundary conditions between plane-wave solutions in each interval.

In this respect, the solution of the Schrödinger equation is equivalent to a transmission line analysis. Each interval of constant potential acts like a small

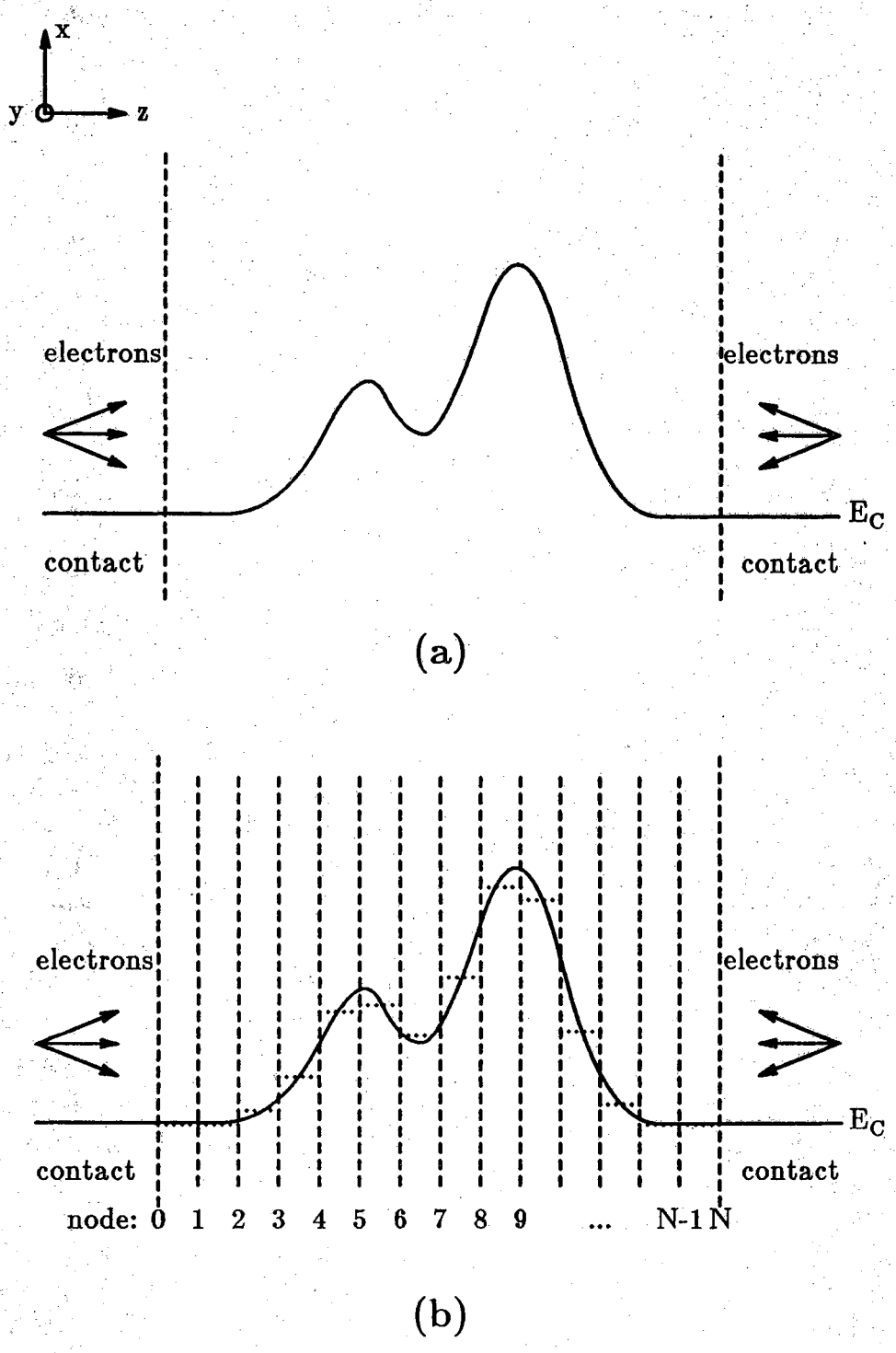


Figure 2.1: (a) Electrons are injected from contacts into a device with an arbitrary potential profile; and (b) any potential profile can be represented by a series of tiny intervals, in which the potential is approximately constant.

section of transmission line with a constant impedance. At the junction between two transmission lines, voltage and current must be continuous. Similarly, proper solutions of the Schrödinger equation require continuity of:

$$\psi_{\vec{k}}(z) \quad \text{and} \quad \frac{1}{m^*(z)}\frac{\partial}{\partial z}\psi_{\vec{k}}(z)$$

at the boundaries between each interval of constant potential. Just as a voltage plane-wave would be reflected by a mismatch in impedance, the electron wavefunction is reflected by changes in the conduction-band profile. When interpreting the results of an analysis, it is useful to keep this analogy at the back of one's mind: Many quantum mechanical effects (*e.g.*, tunneling, resonance states, *etc.*) can be understood in the general context of wave phenomena.

2.1.2. Calculation of Electron Density and Current Density

Because each contact injects electrons into a device, the electron density can be resolved into two components:

$$n(z) = n^{l\to r}(z) + n^{r\to l}(z) \tag{2.2}$$

Electrons injected from the left contact (node 0) are labeled $l\to r$; those injected from the right contact (node N) are labeled $r\to l$. Each component is obtained by integrating the squared-magnitude of the wavefunction over the entire spectrum of wavevectors:

$$n^{l\to r}(z) = \int_0^\infty \frac{dk_z}{2\pi} \left| \psi_{\vec{k}}^{l\to r}(z) \right|^2 \sigma^{l\to r}(k_z) \tag{2.3}$$

$$n^{r\to l}(z) = \int_0^\infty \frac{dk_z}{2\pi} \left| \psi_{\vec{k}}^{r\to l}(z) \right|^2 \sigma^{r\to l}(k_z) \tag{2.4}$$

where the factors $\sigma(k_z)$ represent the integration over transverse wavevectors k_x and k_y:

$$\sigma^{l\to r}(k_z) = \frac{m_c^* k_B T}{\pi\hbar^2} \ln\left[1 + \exp\left[(E_F - E_C(0) - \frac{\hbar^2 k_z^2}{2m_c^*}) / k_B T \right]\right]$$

$$\sigma^{r\to l}(k_z) = \frac{m_c^* k_B T}{\pi\hbar^2} \ln\left[1 + \exp\left[(E_F - E_C(N) - \frac{\hbar^2 k_z^2}{2m_c^*}) / k_B T \right]\right]$$

Here, E_F is the Fermi-level in the injecting contact; m_c^* is the effective mass in the injecting contact; $E_C(0)$ and $E_C(N)$ are the conduction band energies at nodes 0 and N

(left and right contacts, respectively); T is the temperature; and k_z is the incident wavevector, in the direction of propagation.

Notice that, in the Schrödinger equation (2.1), the wavefunction $\psi_{\vec{k}}(\vec{r})$ is dependent on the transverse energy E_t. It was assumed in the derivation of (2.3) and (2.4) that this dependence is weak, so that the wavefunction is approximately constant over all transverse wavevectors. The wavefunction is then evaluated at a single, representative E_t, and removed from the transverse integration. Hence, the integration over transverse wavevectors was performed analytically, and the result is embodied in the factors $\sigma(k_z)$. Normally, the representative transverse energy is assumed to be the thermal average energy, k_BT. Although this assumption is reasonable in many cases, it is strictly an approximation. A rigorous calculation would require an evaluation of the wavefunction at all (longitudinal and transverse) wavevectors. Such a calculation would increase execution time tremendously, to the point where computations are no longer tractable.

Current density for electrons can also be resolved into two components, due to the two, oppositely flowing streams of electrons:

$$J = J^{l\rightarrow r} - J^{r\rightarrow l} \tag{2.5}$$

where

$$J^{l\rightarrow r} = \frac{-q\hbar}{m_c^*} \int_0^\infty \frac{dk_z}{2\pi}\, k_z\, T^{l\rightarrow r}(k_z)\, \sigma^{l\rightarrow r}(k_z) \tag{2.6}$$

$$J^{r\rightarrow l} = \frac{-q\hbar}{m_c^*} \int_0^\infty \frac{dk_z}{2\pi}\, k_z\, T^{r\rightarrow l}(k_z)\, \sigma^{r\rightarrow l}(k_z) \tag{2.7}$$

The functions $T^{l\rightarrow r}(k_z)$ and $T^{r\rightarrow l}(k_z)$ are the transmission coefficients from left to right, and from right to left, respectively. Each represents the fraction of injected electrons which propagate across the entire device. Each is proportional to the squared-magnitude of the wavefunction, at the proper end of the device:

$$T^{l\rightarrow r}(k_z) \propto \left| \psi_{k_z}^{l\rightarrow r}(N) \right|^2 \qquad T^{r\rightarrow l}(k_z) \propto \left| \psi_{k_z}^{r\rightarrow l}(0) \right|^2$$

In essence, therefore, both the electron density and the current density are determined by integrating the squared-magnitude of the wavefunction.

2.1.3. Numerical Integration Concerns

By using the formulas presented above, calculation of the electron density and current density is straightforward: A grid of wavevectors k_z is chosen for the injection of electrons, and the squared-magnitude of the wavefunction is integrated over that grid. The remaining difficulty lies in determining a suitable grid of wavevectors. Naively, one might assume that a uniform mesh (if sufficiently dense) would provide adequate resolution; this is not the case. Consider, for example, electrons experiencing a transmission resonance. For certain device geometries, it is possible to achieve a large transmission of electrons at particular (resonant) wavevectors. In this case, the wavefunction peaks sharply at resonance, and is nearly zero at all other wavevectors. If resonances are sufficiently sharp, they could be missed entirely by a uniform mesh. Even if a series of successively finer meshes were used, there is no guarantee that the wavefunction would be properly resolved. Aside from being an inefficient solution, therefore, the use of a uniform mesh is a little like playing Russian Roulette.

To insure proper resolution of the wavefunction, the following approach is used in SEQUAL. For each direction of propagation, electrons are injected at uniform intervals in k_z-space. In each interval of k_z-space, the squared-magnitude of the wavefunction is examined for a local maximum. If a maximum is found to exist somewhere in the interval, then the exact wavevector corresponding to the maximum is calculated. The process by which a maximum is detected and located in k_z-space is somewhat complicated, and is unimportant for this discussion. By isolating wavefunction maxima, it is assured that all features of the wavefunction will be resolved clearly.

The integrations required by equations (2.3-2.6) could be performed by a variety of numerical techniques. In SEQUAL, Gauss-Legendre quadrature was chosen, because it offers high accuracy with a comparatively small number of wavefunction evaluations. The "order" of the integration process refers to number of mesh points in an interval; mesh points are chosen by mapping zeroes of Legendre polynomials into a given interval. An estimate of the error obtained for a particular interval is calculated by comparing the results from different orders of integration.

An overview of the entire process of injection and integration is as follows. A uniform grid of wavevectors is used to isolate maxima in the squared-magnitude of the wavefunction. The wavevectors corresponding to maxima define a grid of integration intervals, in which the wavefunction is smooth and well-behaved. Within each integration interval, successive orders of Gauss-Legendre quadrature are applied, until the accuracy of the result is acceptable. In principle, the process of isolating maxima in k_z-space could be performed at each node. Realistically, however, it need only be

performed at those nodes for which the wavefunction varies most rapidly in k_z-space. By default, SEQUAL performs isolation of maxima for the first and last nodes (nodes 0 and N) of a device. Because of "standing-wave" patterns produced by reflection, the wavefunction varies most rapidly at the contact nodes. Other nodes for which rapid variations are expected (*e.g.*, resonance states) can be specified at the time of execution.

2.2. Bound Electronic States

In contrast to the propagating electronic states considered in the preceding sections, bound states cannot be populated by either contact. In a real device, electrons injected from the contacts could inelastically scatter into states of lower energy; however, inelastic scattering has been neglected in SEQUAL, thereby removing the link between propagating states and bound states. In many cases, the density of electrons residing in bound states can be significant. This contribution is particularly important if a self-consistent solution of the electrostatic potential is required. For this reason, SEQUAL allows for the consideration of bound states.

Bound state energies are determined by solving for the eigenvalues of the Schrödinger equation (2.1), written in finite-difference form. It is assumed that the wavefunction vanishes at both ends of the device; hence, the spatial location of bound states should be sufficiently removed from the contacts. The actual solution of the eigensystem is performed by an International Mathematical and Statistical Library (IMSL) routine *eigrf*. Given the eigenfunctions $\psi_i(z)$ and the energy eigenvalues ϵ_i, the electron density for each state is determined by:

$$n_i(z) = \frac{m^*(z)k_BT}{\hbar^2\pi} \mid \psi_i(z) \mid^2 \ln\left[1 + \exp\left[(E_F - \epsilon_i)/k_BT\right]\right] \tag{2.8}$$

where E_F is the Fermi-level, which is assumed to be constant across the entire device. This result was obtained by assuming that states are populated according to Fermi-Dirac statistics, an assumption which is valid only in equilibrium. For each bound state, a two-dimensional electron density can be calculated by integrating the electron density $n_i(z)$ over the length of the device:

$$n_i^{2-D} = \frac{k_BT}{\hbar^2\pi} \ln\left[1 + \exp\left[(E_F - \epsilon_i)/k_BT\right]\right] \int_0^{z_N} dz\, m^*(z) \mid \psi_i(z) \mid^2 \tag{2.9}$$

Of course, a finite-difference formulation with nodes 0 to N in position-space will have $N+1$ eigenvalues; of these, only the states which cannot be populated by either contact are considered to be bound states. That is, states are "bound" if the energy

eigenvalue ϵ_i is below the conduction-band edge of either contact. The total (three-dimensional) density of bound-state electrons, therefore, is the sum of contributions from each bound state:

$$n_b(z) = \sum_i n_i(z) \tag{2.10}$$

where the index i is restricted according to:

$$\{ i \mid \epsilon_i < E_C(0), \epsilon_i < E_C(N) \}$$

2.3. Iteration for Self-Consistency

In the preceding sections, a method was described for the calculation of electron density, assuming that the conduction-band profile was known. In general, the conduction-band profile is determined by two components: a static component $\Delta E_C(z)$, describing the offset of the band edge in different material layers, and the electrostatic potential $\Phi(z)$:

$$E_C(z) = \Delta E_C(z) - q\Phi(z)$$

Of course, the electrostatic potential can be determined, given the electron density, by solving the Poisson equation:

$$\frac{d}{dz}\left(\kappa_r(z)\,\epsilon_0\,\frac{d}{dz}\Phi(z)\right) = -q\left[N_D^+(z) - n(z)\right] \tag{2.10}$$

where $\kappa_r(z)$ is the (position-dependent) relative dielectric constant, and $N_D^+(z)$ is the density of ionized donors.

For a proper analysis of any device, the solutions for both the electron density and the electrostatic potential should be self-consistent. In SEQUAL, a fully self-consistent solution can be obtained by iterating between the solutions for each quantity; an overview of the iteration process is presented in Figure 2.2. After an initial guess of the electrostatic potential has been input, the electron density $n^0(z)$ is calculated. If iteration is not allowed, the process terminates, yielding a solution of electron density for the given potential profile. If iteration is specified, however, a better guess for the electrostatic potential is determined, by solving the Poisson equation. The corresponding electron density is then calculated, and the process is continued. Iteration is terminated when the calculations converge to a certain number of significant figures, or when the iteration counter i reaches some maximum value.

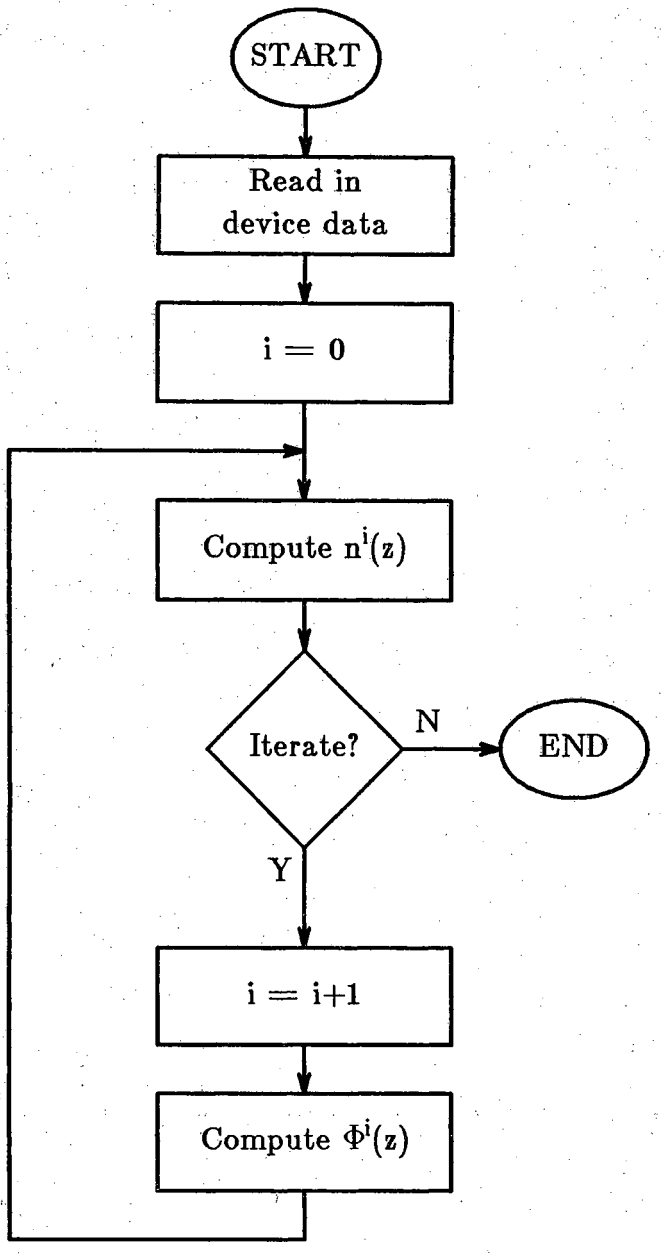


Figure 2.2: Flowchart for a self-consistent solution

3

Program Input

SEQUAL accepts two kinds of input: the device description, and the input deck (see Figure 1.1). Output from a classical analysis program constitutes the device description, a file containing columns of numbers which characterize the device in one dimension. One column is a list of "nodes," or positions in space; all other columns list the value of a quantity (*i.e.*, the conduction band energy, the effective mass, *etc.*) at each node. The second source of input, the input deck, is a list of commands which control the operation of SEQUAL. These commands specify the format of input and output files, modify the default values of input variables, and select a variety of program options. Descriptions of both types of input are presented in detail, below.

3.1. Device Data

The numerical representation of a device is obtained primarily from a file containing several columns of floating-point data; each column represents a different input quantity. Data can be stored in either ASCII or (Fortran 77) binary format. Files containing the necessary information are typically generated as output from a classical analysis program. To provide a flexible interface for many different programs, the quantity associated with each column can be specified in the input deck; therefore, data columns can appear in any order. Moreover, the units of input quantities can be specified, if they differ from the default units in SEQUAL. (For details of the input deck, see section 3.2, below.)

Possible input quantities are listed in Table 3.1. Notice that it is not necessary to have all six quantities specified in the device description file. If it is more convenient, profiles of the donor doping density, the effective mass, and the dielectric constant can be specified in the input deck. Because these quantities are usually constant over large regions (*e.g.*, in a particular material layer), it is doubtful that all classical analysis programs would provide their value at each node. If tabular output of these quantities is unavailable, specifying the values in the input deck is a more convenient approach than modifying output to include the extra columns. Notice also that unless a self-consistent solution is required, profiles of the electrostatic potential

"Device Data" Input Quantities		
Quantity	**Units**	**Necessary for Program Execution**
position-space grid	cm	•
conduction-band profile	eV	•
electrostatic potential	V	‡
ionized donor doping density	cm^{-3}	†
effective mass	(m_0)	†
dielectric constant	(ϵ_0)	†, ‡

† May be specified alternatively in the input deck.
‡ Necessary only for a self-consistent solution.

Table 3.1: Summary of possible input quantities for device data input file.

Input quantities represented by the columns below are in the default order:

position	conduction band profile	electrostatic potential	ionized donor doping density	effective mass	dielectric constant
0.000000e+00	0.000000	0.000000	2.000000e+18	0.0670	12.8464
1.000000e-07	0.000000	0.000000	2.000000e+18	0.0670	12.8464
2.000000e-07	0.000000	0.000000	2.000000e+18	0.0670	12.8464
3.000000e-07	0.000000	0.000000	2.000000e+18	0.0670	12.8464
4.000000e-07	0.000000	0.000000	2.000000e+18	0.0670	12.8464
4.990000e-07	0.000000	0.000000	2.000000e+18	0.0670	12.8464
5.000000e-07	0.365000	0.000000	2.000000e+18	0.08953	11.4705
6.000000e-07	0.365000	0.000000	2.000000e+18	0.08953	11.4705
7.000000e-07	0.365000	0.000000	2.000000e+18	0.08953	11.4705
8.000000e-07	0.365000	0.000000	2.000000e+18	0.08953	11.4705
9.000000e-07	0.365000	0.000000	2.000000e+18	0.08953	11.4705
9.990000e-07	0.365000	0.000000	2.000000e+18	0.08953	11.4705
1.000000e-06	0.000000	0.000000	2.000000e+18	0.0670	12.8464
1.100000e-06	0.000000	0.000000	2.000000e+18	0.0670	12.8464
1.200000e-06	0.000000	0.000000	2.000000e+18	0.0670	12.8464
1.300000e-06	0.000000	0.000000	2.000000e+18	0.0670	12.8464
1.400000e-06	0.000000	0.000000	2.000000e+18	0.0670	12.8464
1.500000e-06	0.000000	0.000000	2.000000e+18	0.0670	12.8464

Figure 3.1: Example listing of a device-description input file in ASCII format

and the dielectric constant are not required. These two quantities are used only in the solution of Poisson's equation, and are therefore unnecessary if iteration is not allowed. Commonly, the device description file will contain only the position-space grid and the conduction-band profile; all other quantities will be specified in the input deck, or will be unnecessary.

From the above discussion, it would seem that the conduction-band profile must always appear in the data description file. This is not exactly the case. The conduction-band profile is composed of two parts:

$$E_C(z) = \Delta E_C(z) - qV(z)$$

where $V(z)$ is the electrostatic potential, and $\Delta E_C(z)$ specifies the offset of the conduction band in different material layers. For a device fabricated with a single material, $\Delta E_C(z)$ is zero (or some arbitrary constant) everywhere. The alignment of energy bands for differing materials, however, produces some offset in the conduction-band edge, described by $\Delta E_C(z)$. An alternative to specifying the conduction-band profile, therefore, is specifying both the electrostatic potential $V(z)$ and the conduction-band offset $\Delta E_C(z)$. In fact, SEQUAL requires a knowledge of all three quantities for a self-consistent solution, so at least two must be specified at the start of execution. In a self-consistent solution, the conduction-band profile is calculated at the beginning of each iteration by adding contributions from the (constant) offset potential and the (updated) electrostatic potential.

An example of a simple device description file in ASCII format appears Figure 3.1. The device which it represents is composed of three material layers--two layers of GaAs sandwiched around a layer of AlGaAs. Each layer is 5.0×10^{-7} cm (50 Å) thick, so the overall device length is 1.5×10^{-6} cm (150 Å). The GaAs layers have an electron effective mass of $m^* = 0.067m_0$, and a dielectric constant of $\epsilon = 12.8464\epsilon_0$. The AlGaAs layer has an electron effective mass of $m^* = 0.08953m_0$, and a dielectric constant of $\epsilon = 11.4705\epsilon_0$. An offset of 0.365 eV in the conduction-band edge of AlGaAs can be seen clearly in the conduction-band profile, since the electrostatic potential was assumed to be zero. Note that the nodes on either side of a material interface are closely spaced, so that the change in the conduction-band edge at the interface is abrupt. Finally, the entire device was assumed to have a ionized donor density of 2×10^{18} cm^{-3}.

3.2. Input Deck

In days of old, when "small" computers fit neatly in a warehouse, and their operators spoke a language no one else understood, the preferred method of input was the punched card. Each line of input was recorded on a card; each file was created by stacking cards in order. Although card-punch machines have become obsolete, their associated terminology has not. The following description of SEQUAL's input "deck" is but one example of jargon refusing to die.

In SEQUAL, the user controls program operation via the input deck, a short file of commands read from standard input. Essentially, the input deck is a list of assignment statements, specifying particular values for SEQUAL's input parameters. Each input parameter is represented by a key word, and logically related "keys" are grouped together on a command line or "card." Figure 3.2 illustrates the input deck syntax. Each card begins with the card name, and is followed by a list of assignments to various keys. A single card can be continued on any number of lines by specifying the continuation character "+" in place of the card name. Note that space is not allowed around the equal sign in an assignment to a key, although anywhere else the use of white space (including spaces, tabs, and commas) is encouraged, to promote legibility. Blank lines and comment lines are also allowed. A comment line is defined as any line beginning with a character which is neither alphabetic nor the "+" continuation character. Notice from Figure 3.2 that a single key can be assigned multiple values (as many as 10 values), by separating each value with a slash; white space is not allowed between values and slashes. A multiple-value assignment can also be continued on any number of lines, as shown in Figure 3.2.

The use of multiply-valued keys will become clear as the details of each input card are explained. Consider, however, the following card, which specifies the electron effective mass in a number of material layers:

```
matter  nodes=5/11/17  emass=0.067/0.08953/0.067
```

This is the input deck specification of effective mass, for the example considered in Figure 3.1. The key *nodes* is assigned the ending nodes for three different material layers. The first layer (GaAs) is defined by nodes 0-5; the middle layer (AlGaAs), by nodes 6-11; and the final layer (GaAs), by nodes 12-17. The key *emass* is assigned the electron effective mass in each of the three material regions. The other use of multiply-valued keys is illustrated by the following card, which specifies the temperature of a device:

```
device  temp=77.0/300.0  bias=0.0
```

Since two temperatures are specified for the key *temp*, two separate calculations will

Typical Input Deck Card:

cardname *key1=value* *key2=value*

Continuation of a Single Card:

cardname *key1=value*
\+ *key2=value*
\+ *key3=value*

Assigning Multiple Values to a Single Key:

cardname *key1=value1/value2/value3*

Continuation of Assignment of Multiple Values:

cardname *key1=value1/value2/value3*
\+ *key1=value4/value5*
\+ *key1=value6/value7*

Comment Line:

? If the first character on a line is neither alphabetic
? nor a "+" the rest of the line is ignored.

Figure 3.2: Syntax of cards in the input deck

be performed--one at 77 ° K, and the other at 300 ° K. Notice that although two temperatures are specified, only one bias is given. Both calculations, therefore, are performed at the stated bias. To perform one calculation at 77 ° K with zero bias, and another at 300 ° K with a bias of 0.1 V, one would use:

```
device  temp=77.0/300.0  bias=0.0/0.1
```

In general, every key is allowed to have multiple values. Certain keys (such as *temp* and *bias*) direct SEQUAL to perform several different calculations; other keys (such as *nodes* and *emass* considered above) specify information which remains the same for all calculations. Given an input deck, SEQUAL first extracts all of the data from "global information" keys. Remaining keys ("control" keys) are then examined to determine the number of calculations to be performed. The first value of each control key is used in the first calculation; the second value, in the second calculation, and so on. When a list of values for a particular key has been exhausted, the last value of that key is used in any subsequent calculations. The following card,

```
device  temp=77.0/200.0/300.0  bias=0.0
```

is equivalent to

```
device  temp=77.0/200.0/300.0  bias=0.0/0.0/0.0
```

although it demands less typing.

Hence, the use of multiply-value keys can greatly increase the power of the input deck in SEQUAL. Global information keys can be used to specify those device parameters which are not available in tabular form. Control keys can be used to direct SEQUAL to perform a number of similar calculations in a single program execution. The remainder of this chapter is devoted to a detailed explanation of each program key. Further examples of program operation will be presented in Chapter 5.

input — specify format of device description file

Default State:

input file=seq.in format=zevdmk ascii=true

file The name of the device description file is specified by the *file* key of the *input* card. Since *file* is a control key, several different device descriptions can be processed in a single execution of SEQUAL, by assigning *file* to each of the file names. Note that, in SEQUAL, a file name (or the value assigned to any key, for that matter) is limited to 15 characters.

format To enhance the flexibility of SEQUAL's interface with the output of other programs, data columns in the device description file can appear in any order. The order is simply specified in the *format* key of the *input* card. The quantity associated with each column of the data description file is represented by a single character, as follows:

Character	Quantity
z	Position-space grid
e	Conduction-band profile
v	Electrostatic potential
d	Ionized donor doping density
m	Electron effective mass
k	Relative dielectric constant
?	(column is ignored)

Notice that it is possible for a device description file to contain extraneous columns, which do not correspond to any input quantities. These columns, referred to by a question mark in the format specification, are read and ignored. For example, suppose a classical analysis program outputs data in tabular form, with the columns: position-space grid, electron density, hole density, ionized donor density, electric field, and conduction-band profile. Of these quantities, only the position-space

grid, the ionized donor density, and the conduction-band profile are desired for input. The format specification for such an input file would be:

```
input    format=z??d?e
```

Note that space is not allowed between the different characters in a format specification.

ascii

A device description file can be stored in either ASCII or (Fortran 77) binary format. The storage type is specified with the *ascii* key of the *input* card:

Value	Implication
true	File is in ASCII format
false	File is in (Fortran 77) binary format

Files in ASCII format contain floating-point data represented by ASCII characters. This is the normal method of output, since data can be listed easily. Files in binary format contain floating-point data represented in the computer's internal format. Acceptable binary-format files must have been produced by a Fortran 77 program, and quantities written must have been single-precision (real), floating-point numbers.† Files in binary format have the advantage of demanding less storage space than equivalent files in ASCII format.

† Of course, if the precision of floating-point variables is somehow doubled by the compiler, then input quantities will be expected to be double precision.

scale

change the units of input quantities in the device description file

Default State:

scale cm=1.0 ev=1.0 v=1.0 cm**-3=1.0

If quantities in the device description file are not in the units required by SEQUAL, they can be scaled automatically by specifying an appropriate scale factor. Only values in the device description file are converted. Values in the input deck are assumed to have the appropriate units.

cm The length scale of the position-space grid is altered by the *cm* key. For example, if the position of each node is in units of microns (1 μm = 10^{-6} m), the appropriate scale factor would be:

```
scale cm=1.0e4
```

ev The energy scale of the conduction-band profile is altered by the *ev* key. For example, if the conduction-band edge at each node is in units of meV (1 meV = 10^{-3} eV), the appropriate scale factor would be:

```
scale ev=1.0e3
```

v The scale of the electrostatic potential is altered by the *v* key. For example, if the potential at each node is in units of mV (1 mV = 10^{-3} V), the appropriate scale factor would be:

```
scale v=1.0e3
```

cm-3** The scale of the donor doping density is altered by the *cm**-3* key. For example, if the doping density at each node is in units of m^{-3}, the appropriate scale factor would be:

```
scale cm**-3=1.0e6
```

matter specify material layers and properties

Default State:

matter

(In the default state, profiles of material parameters are specified in the device description file.)

Normally, material properties are specified in the device description file. When a tabular listing of material properties is difficult to obtain, however, the *matter* card provides an alternative method of input. As many as ten different material layers can be defined, each with differing material properties.

nodes In the data description file, a single line of input (in other words, a single position in space) constitutes a "node." Nodes are referred to by number, starting from zero at the beginning of the file. Material layers, therefore, are defined in the *matter* card by specifying the nodes which are endpoints of material regions. For example, the device description file presented in Figure 3.1 contains 18 lines; the range of node numbers, then, is 0 to 17. To define a single material region with an effective mass of 0.067, the *matter* card would read:

```
matter  nodes=17  emass=0.067
```

To define three material layers (as in the example of Figure 3.1) the matter card would read:

```
matter  nodes=5/11/17  emass=0.067/0.08953/0.067
```

Note an equal number of values must be assigned to *nodes*, *delec*, *emass*, and *krel*.

delec The conduction-band offset ΔE_C can be specified for each material layer defined by *nodes*. The offset is understood as the conduction-band discontinuity at a material interface, and is specified in electron-volts. Note that, for an abrupt interface, the nodes on either side of the interface should be closely spaced.

emass The electron effective mass can also be specified for each material layer defined by *nodes*. The value is input as a dimensionless quantity, in terms of the free electron mass, m_0.

krel The relative dielectric constant can also be specified for each material layer defined by *nodes*. The value is input as a dimensionless quantity, in terms of the dielectric constant of free space, ϵ_0.

doping specify the ionized donor doping density

Default State:

doping ni=1.79e6

(In the default state, the doping-density profile is specified in the data description file.)

nodes Just as the *nodes* key of the *matter* card defines different material regions, the *nodes* key of the *doping* card defines regions with a constant density of ionized donors. Unless a self-consistent solution is desired, the donor density is needed only at the contacts (first and last nodes), to determine the Fermi-level of injected electrons. In this case, the specification could be as simple as:

```
doping  nodes=0/19/20  nd+=5.0e17/0/5.0e17
```

for a device with nodes 0 to 20, with an ionized donor density of 5.0×10^{17} cm^{-3} in each contact. Because the donor density at interior nodes is irrelevant (unless a self-consistent solution is required), the example above could be equally expressed as:

```
doping  nodes=20  nd+=5.0e17
```

For specifying the contact doping, or for some crude approximations to real doping profiles, the *doping* card provides a convenient method of input.

nd+ The density of ionized donors (in units of cm^{-3}) is assigned to *nd+* for each region defined by the *nodes* key. Notice that SEQUAL requires the density of *ionized* donors. Partial ionization of donors, therefore, must be calculated by the classical analysis program contributing input.

ni If the ionized donor density in either device contact (*i.e.*, the first and last nodes) is identically zero, then the intrinsic electron density assigned to *ni* is assumed, for the calculation of the Fermi-level. SEQUAL assumes that charge neutrality exists in the contacts,

between electrons and ionized donors; therefore, the ionized donor density in the contacts, unless it is identically zero, is used to determine the Fermi-level for injection of electrons. Notice that if an extremely small (but non-zero) donor density is specified, a Fermi-level will be deduced to correspond to this concentration. This provides a trick for specifying the electron concentration in the case of a *p-n* junction. Holes are completely ignored in SEQUAL, and hence, there is no way of specifying a doping density for acceptors. In order to obtain the proper electron concentration, the "ionized donor density" on the *p*-side of the junction should be set equal to the minority carrier (electron) concentration. If, instead, it were set to zero, the value assigned to *ni* would be assumed for the minority carrier (electron) concentration.

device specify macroscopic device parameters

Default State:

device temp=300.0 area=1.0 bias=0.0

temp The device temperature (in °K) for a calculation is assigned to the *temp* key.

area The cross-sectional area (in cm^2) of a device is assigned to the *area* key. It is used as a multiplicative constant in the calculation of total current. The default area is 1 cm^2, so that the total current (in A) is the same as the current density (in A/cm^2).

bias An applied bias can be added to the device, in addition to any existing bias in the output from a classical analysis program. The amount of bias (in V), assigned to the *bias* key, is applied as a linear potential drop across the lightly-doped (middle) region of the device. For a self-consistent solution, several iterations should be allowed after the application of bias, so that the electrostatic potential will return to the self-consistent state. If the internal nodes of the device are no more lightly doped than the contacts, the bias is applied as a linear drop across the entire device. In this case, obtaining convergence for self-consistent calculations might be difficult. Note that it is possible to have a device description file in which bias is built into the electrostatic potential and the conduction-band profile. In some cases, it may be more convenient (or accurate) to apply bias in the classical analysis program, and use SEQUAL to analyze the resulting potential profile.

solve — specify parameters controlling the analysis

Default State:

solve prec=3 itmax=9 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 (see Figure 2.2). Convergence is achieved when the number of significant figures in the current density (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).

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 (see section 2.2). 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 (see section 2.1.1). 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.

maxima specify parameters for the isolation of wavefunction maxima

Default State:

maxima prec=2 kscale=1.0

(In the default state, the first and last nodes are "watched" for wavefunction maxima; interior nodes are ignored.)

prec In the process of stepping through k_z-space, SEQUAL looks for maxima in the squared-magnitude of the wavefunction (see section 2.1.3). This insures both the proper integration of the wavefunction, and the proper resolution of the transmission coefficient. The number of significant figures requested for the calculations related to maxima isolation is assigned to the *prec* key of the *maxima* card. Normally, the default value demands sufficient precision. Particularly sharp transmission resonances, however, may require greater precision to be properly resolved. When a doubt arises, results from two different requested precisions should be compared.

kscale SEQUAL steps through wavevector-space for the purpose of isolating maxima in the electron wavefunction (see section 2.1.3). Intervals of k_z-space between successive maxima are then integrated, to determine the electron density and the current density. Normally, the default k_z-step is adequate for isolation of wavefunction maxima. The k_z-step can be scaled to any size, however, according to the value assigned to *kscale*. Notice that the size of the k_z-step (if it is sufficiently small) has no effect on integration of the wavefunction. Integration intervals are defined by local maxima in the wavefunction. A smaller k_z-step can provide better resolution in the output of quantities, such as the transmission coefficient and the wavefunction magnitude. A larger k_z-step should be used with extreme caution: If wavefunction maxima are improperly resolved, the calculations for electron density and current density will be in error.

watch The isolation of k_z-space maxima in the wavefunction could be accomplished at all position-space nodes. This would increase execution time tremendously, however, without adding to the accuracy of the calculation. It is necessary to isolate k_z-space maxima only at those nodes for which the wavefunction varies rapidly. By default, the first and last nodes of any device are "watched" for maxima in k_z-space (see section 2.1.3). Additional nodes can be considered, if the node numbers are assigned to the *watch* key.

integ — specify parameters for the integration of the wavefunction

Default State:

integ orders=2/4/6/8/10/12/16/20/24 kbt=10.0 et=1.0

orders For the calculation of electron density and current density, the magnitude of the wavefunction must be integrated in k_z-space (see section 2.1.2). Integration is performed using Gaussian quadrature with Legendre polynomials (see section 2.1.3). To achieve the precision requested (via *prec* of the *solve* card), successive orders of integration are applied to an integration interval, until the precision is obtained, or until the list of integration orders is exhausted. Orders available for the integration process are assigned to the *orders* key. In the default state, all available orders are assigned to *orders*. For cases in which the reduction of execution time is more important than the accuracy of results, the list of integration orders may be abbreviated. Indeed, a single integration order could be specified (*e.g.*, the highest order available), if an estimation of the solution accuracy is unnecessary.

kbt In theory, integration of the wavefunction should be performed over the range of all wavevectors from zero to infinity (see section 2.1.2); in practice, integration must be truncated at some large but finite wavevector. The point of truncation is determined in SEQUAL by considering the number of significant figures in integration results. Integration is performed up to some minimum wavevector. Beyond this, integration is continued (if necessary) to achieve the requested precision, for both the electron density and the current density. The minimum wavevector for integration corresponds to an energy which is some multiple of k_BT higher than the maximum energy in the conduction band. The value assigned to *kbt* specifies the multiple of k_BT.

et In the derivation of formulas for both the electron density and the current density, the wavefunction was assumed to be weakly dependent on the transverse energy E_t (see section 2.1.2). This allowed the wavefunction, evaluated at a representative transverse energy, to be

removed from the integration over transverse momenta. Hence, the integration over transverse momentum could be performed analytically. The particular transverse energy at which the wavefunction is evaluated is determined by the value assigned to *et* (in units of k_BT). A reasonable assumption for the representative transverse energy is the thermal average energy, k_BT.

title — specify a title for printed (standard) output

Default State:

title

(In the default state, no title appears on output pages.)

The *title* card is unique for two reasons: It is the only card in the input deck which does have any keys, and which cannot be continued on multiple lines. Any text following the card name is taken as the title of the execution. This title appears in the heading of each page of printed output.

print — specify the form of printed (standard) output

Default State:

print tcoeff=* format1=zevn format2=zdmk verbose=true

tcoeff For propagating electrons, a listing of the transmission coefficient versus incident energy can be obtained in printed output, according to the value assigned to *tcoeff*:

Value	Implication
l-to-r	Print the transmission coefficient for electrons propagating left-to-right
r-to-l	Print the transmission coefficient for electrons propagating right-to-left
both	Print transmission coefficients for both directions of propagation
*	Do not print transmission coefficients

For self-consistent solutions, a listing of the transmission coefficients is provided only for the final iteration.

format1

format2 A tabular listing of both input parameters and calculated results can be obtained in printed output. Two sections of output, each with a maximum of four columns, can be defined with the output format keys, *format1* and *format2*. The quantity associated with a particular column is represented by a single letter, as follows:

Character	Quantity
z	Position-space grid
e	Conduction-band profile
v	Electrostatic potential
n	Electron density
d	Ionized donor doping density
m	Electron effective mass
k	Relative dielectric constant
*	(No output in a section)

The definitions shown above, except for the addition of *n* and *, are identical to those presented in conjunction with input format. Unlike the input format specifier, *format1* or *format2* (or both) can be assigned to *, which causes an output section to be suppressed. For example, to obtain a single output section listing position, conduction-band energy, and electron density, the *print* card could be either of the following:

```
print    format1=zen  format2=*

print    format1=*  format2=zen
```

verbose

Because SEQUAL can provide so many sections of output in a single run, the user is allowed the option of suppressing superfluous output, with an assignment to *verbose*:

Value	Implication
true	All output sections are supplied
false	Superfluous output sections are suppressed

Although some of the "verbose" output may seem unnecessary, it is included to aid the user in understanding the problem description, and in evaluating the performance of SEQUAL.

output specify the form of (plotting) output files

Default State:

output file=seq data=* columns=3 ascii=true

For each calculation, data presented in printed output are also available for storage in files. Such files can be used as input to plotting programs, or as input (for further processing) to SEQUAL.

file The argument of the *file* key on the *output* card is used as a basis in forming all output file names. For each file created, an extension is added to the root file name specified, to identify both the type and format of stored data. (For an understanding of different output files and their respective extensions, see section 4.2.)

data Different types of data can be requested for output, according to a string of single-character keys assigned to *data*:

Character	Type of Data Stored
d	Device data (quantities *vs.* position)
b	Bound-state data
t	Transmission coefficient data
w	Surface plot of the wavefunction magnitude
*	(No data is stored)

In an assignment to *data*, the characters above may appear in any order. For example, the following two cards are equivalent:

```
output    data=dtw

output    data=wdt
```

Notice that space is not allowed between the characters.

columns For compatibility with many different plotting programs, data can be written to output files in single-column, paired-column, or multiple-column formats, according to the value assigned to *columns*:

Value	Implication
1	Single-column format
2	Paired-column format
≥ 3	Multiple-column format

ascii Output files can be stored in either ASCII format or (Fortran 77) binary format, according to the value assigned to *ascii* on the *output* card:

Value	Implication
true	File is in ASCII format
false	File is in (Fortran 77) binary format

Summary of Input/Output Keys			
Card	**Key**	**Value**	**Key Type †**
input	file	text (*filename*)	control
	format	(z, e, v, d, m, k, ?)	control
	ascii	logical	control
scale	cm	real	control
	ev	real	control
	v	real	control
	cm**-3	real	control
matter	nodes	integer	global information
	delec	real (eV)	global information
	emass	real (m_0)	global information
	krel	real (ϵ_0)	global information
doping	nodes	integer	global information
	nd+	real (cm^{-3})	global information
	ni	real (cm^{-3})	control
title	--	--	--
print	tcoeff	(l-to-r, r-to-l, both, *)	control
	format1	(z, e, v, d, m, k, n, *)	control
	format2	(z, e, v, d, m, k, n, *)	control
	verbose	logical	control
output	file	text (*filename*)	control
	data	(d, b, t, w, *)	control
	columns	integer	control
	ascii	logical	control

† For a explanation of "control" keys and "global information" keys, see section 3.2.

Table 3.2: Summary of keys which control the input/output of SEQUAL

Summary of Execution Keys			
Card	**Key**	**Value**	**Key Type †**
device	temp	real (°K)	control
	area	real (cm^2)	control
	bias	real (V)	control
solve	prec	integer	control
	itmax	integer	control
	states	(prop, bound, all)	control
	inject	(l-to-r, r-to-l, both)	control
maxima	prec	integer	control
	kscale	real	control
	watch	integer	global information
integ	orders	integer	global information
	kbt	real (K_BT)	control
	et	real (K_BT)	control

† For a explanation of "control" keys and "global information" keys, see section 3.2.

Table 3.3: Summary of keys which control the execution of SEQUAL

4

Program Output

SEQUAL provides two kinds of output: printed output, and output files. For each execution, a record of input data and output results is written to standard output. Because this record is formatted with Fortran line-printer codes (providing page ejection, *etc.*) it is intended to serve as a printed report of all calculations. Pages are numbered, and headings are printed at the top of each page. Output files, on the other hand, are not well suited to examination; rather, they are provided to serve as input for plotting programs. Because both forms of output can include a wide variety of different results, each form is presented in detail, in the following sections.

4.1. Printed Output

Figure 4.1 depicts the general form of printed output in SEQUAL. For each execution of the program, a number of different output sections could appear in the printed record. Each section will only appear, however, if it is necessary. Moreover, some of the output is verbose, and can be suppressed at the request of the user. In Figure 4.1, the sections grouped together in a dashed box present output for each calculation. Since SEQUAL can perform several calculations in a single program execution, these sections may appear several different times in the printed record. (See discussion of "control" keys in section 3.2.) Although the example output pages presented in Chapter 5 are easily worth a thousand words, a brief description of each section is given below:

Input Deck:
Printed output begins with an echo of the input deck. Syntax errors (if any are found) are pointed out under each offending line.

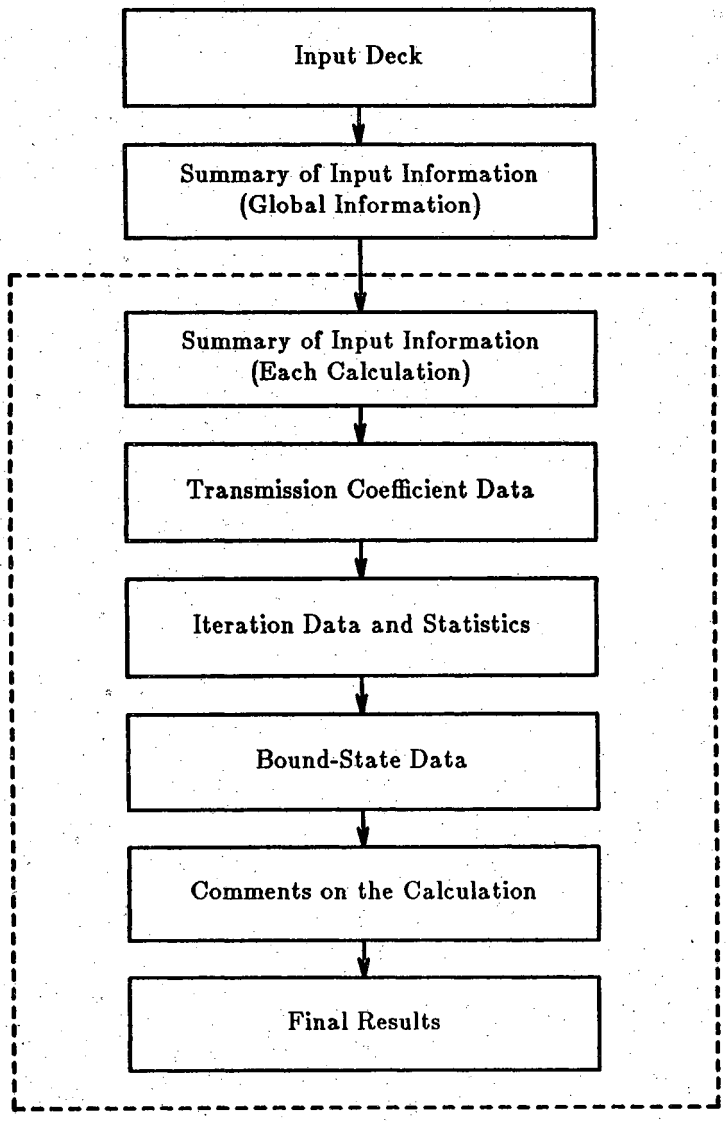


Figure 4.1: Diagram of printed output. Items in the dashed box are presented for each calculation (see discussion of "control" keys in section 3.2).

Summary of Input Information (Global Information):
Device data which has been specified on input cards *matter* or *doping* is summarized in a pictorial fashion. Because this information is global to all calculations, it appears only once, near the start of printed output.

Summary of Input Information (Each Calculation):
For each calculation, SEQUAL presents a page summarizing all important input parameters. Although this page is helpful in clarifying the details of a particular calculation, it is considered "verbose," and it can be suppressed.

Transmission Coefficient Data:
A table of the transmission coefficient versus energy can be requested for each direction of propagation. For iterative calculations, a table is generated only for the results of the final iteration.

Iteration Data and Statistics:
For iterative calculations, the degree of convergence obtained for each iteration is summarized in tabular form. In addition, statistics characterizing the propagating-state solution are presented. Because most of SEQUAL's execution time is spent in the analysis of propagating states, these statistics can be used to evaluate the overall performance of the program. Although propagating-state statistics are sometimes useful, they are considered "verbose," and can therefore be suppressed.

Bound-State Data:
For bound-state calculations, SEQUAL presents a table of bound-state energies and two-dimensional electron densities. If a propagating-state solution was also requested, a breakdown of the electron density is printed, showing the contributions due to propagating states and bound states.

Comments on the Calculation:
All warnings, cautions, and error messages encountered in a particular calculation are collected onto a single page. A brief explanation is presented for each problem encountered, along with suggestions for relieving the difficulty.

Final Results:
A section of final results begins by listing the current obtained for the bias across the device. Following this, device data is presented in two sections, as specified by *format1* and *format2* on the *print* card.

4.2. Output Files

In addition to the printed record, SEQUAL can create a number of data files for the storage of results. Because these files contain raw data, they are well suited for use with plotting programs. Of course, different plotting programs require different formats for input data. Some programs demand x-axis data and y-axis data in separate files; some require a paired listing of x and y values. To accommodate the majority of programs, SEQUAL can output data in three formats, according to the value assigned to *columns* on the *output* card. Data can be written in single-column, paired-column, or multiple-column formats. Obviously, many data files will be created by the single-column or paired-column option; files are distinguished by an extension added to the specified file name. Extensions were chosen to indicate at a glance the contents of a particular data file. For instance, the files ending in ".z" contain a single column, listing the position-space grid; files ending in ".zv" contain two columns of data--the position-space grid and the electrostatic potential. A complete listing of extensions is presented in Tables 4.1-4.4.

Table 4.4 shows that, for output files describing the wavefunction surface, the format is independent of the number of columns specified. Files ending in ".wlr-z" or ".wrl-z" contain a list of position-space nodes; similarly, files ending in ".wlr-e" or ".wrl-e" contain a list of energy-space nodes. The squared-magnitude of the wavefunction, weighted by $\sigma/2\pi$, is listed in files ending with ".wlr-m" or ".wrl-m" extensions. Data appear sequentially: For each energy, the quantity is listed across the entire device, from the first node to the last. Plotting programs reading the data, therefore, will read a matrix of values. In the process of reading data, the index for position-space entries will vary more rapidly.

Output files in multiple-column format can also be used as input to SEQUAL. Consider, for instance, the computation of a self-consistent current-voltage characteristic. Using a multiple-valued assignment to the *bias* key, the user can instruct SEQUAL to calculate current at a number of different biases. For a self-consistent calculation, a bias is applied to an input potential, and solutions for the electron density and electrostatic potential are performed iteratively. One would expect the self-consistent potential obtained for one bias to be quite close to the solution for the

next bias. Rather than apply biases to the same input file, the clever user will feed output results back in, as input.

In Figure 4.2, an example input deck is shown, to perform the feedback. The file *classical* contains output from a classical analysis program, with data in the default column-format *zevdmk*. A self-consistent calculation is performed at zero bias, and the results are stored in the file *out1.z...k*. This file is then used as the device description for the next calculation. Because the output from SEQUAL includes three columns describing electron density (propagating, bound, and total), the input format becomes *zev???dmk* for the three calculations with feedback. For each calculation, an additional bias of 0.1 V is applied to the potential profile. Hence, current is calculated for biases of 0 V, 0.1 V, 0.2 V, and 0.3 V.

```
                **** I/O Feedback:
                *      Results of one calculation are stored in output
                *      files, and read back as input.
                ****

input           file=classical/out0.z...k/out1.z...k/out2.z...k
  +             format=zevdmk/zev???dmk

device          bias=0.00/0.10/0.10/0.10
solve           itmax=15

output          file=out0/out1/out2/out3  data=d  columns=3
```

Figure 4.2: Input deck for feedback of output files

Summary of Output Files (*data*=d)		
columns=	File Extension	Quantity
1	.z	position
	.e	conduction-band energy
	.v	electrostatic potential
	.p	propagating electron density
	.b	bound-state electron density
	.n	total electron density
	.d	ionized donor density
	.m	effective mass
	.k	relative dielectric constant
2	.ze	position conduction-band energy
	.zv	position electrostatic potential
	.zp	position propagating electron density
	.zb	position bound-state electron density
	.zn	position total electron density
	.zd	position ionized donor density
	.zm	position effective mass
	.zk	position relative dielectric constant
≥ 3	.z...k	position conduction-band energy electrostatic potential propagating electron density bound-state electron density total electron density ionized donor density effective mass relative dielectric constant

Table 4.1: Summary of output files created for storage of device data

Summary of Output Files (*data*=t)		
columns=	File Extension	Quantity
1	.tlr-e	energy (propagation left-to-right)
	.tlr-c	transmission coefficient (propagation left-to-right)
	.trl-e	energy (propagation right-to-left)
	.trl-c	transmission coefficient (propagation right-to-left)
2, ≥3	.tlr-ec	energy transmission coefficient (propagation left-to-right)
	.trl-ec	energy transmission coefficient (propagation right-to-left)

Table 4.2: Summary of output files created for storage of transmission coefficient data

Summary of Output Files (*data*=b)		
columns=	File Extension	Quantity
1	.bs-e	bound-state energy
	.bs-n	two-dimension electron density
2, ≥3	.bs-en	bound-state energy two-dimensional electron density

Table 4.3: Summary of output files created for storage of bound-state data

Summary of Output Files (*data*=w)		
columns=	File Extension	Quantity
1, 2, $\geq$3	.wlr-z	position (propagation left-to-right)
	.wlr-e	energy (propagation left-to-right)
	.wlr-m	$\lvert\psi^{l\to r}\rvert^2\,\sigma^{l\to r}\,/\,2\pi$
	.wrl-z	position (propagation right-to-left)
	.wrl-e	energy (propagation right-to-left)
	.wrl-m	$\lvert\psi^{r\to l}\rvert^2\,\sigma^{r\to l}\,/\,2\pi$

Table 4.4: Summary of output files created for storage of wavefunction surface plots

5

Example Calculations

To illustrate some of the abilities of SEQUAL, example calculations are presented in this chapter, for a resonant tunneling device fabricated by Ray, *et. al.* † The device, pictured in Figure 5.1, is composed of two $Al_{0.45}Ga_{0.55}As$ barriers sandwiched around a GaAs potential well.† Classically, electrons injected from the contacts at energies below the top of the barrier are completely reflected; quantum mechanically, however, it is possible for electrons to be transmitted. Electrons, after tunneling through one potential barrier, can experience multiple reflections inside the quantum well, before tunneling through the other barrier. Because of the wave-nature of electrons, these multiple reflections can constructively interfere, producing a large transmission across the entire device. In summary, electrons injected at particular "resonant" energies will experience unity transmission; those injected at energies off-resonance will be strongly reflected. Resonant energies are often referred to as "quasi-bound states," since in the process of multiple reflection, electrons are effectively bound to the well.

SEQUAL can be used to graphically illustrate the resonance condition described above. Figure 5.2 presents a surface plot of the wavefunction versus position and incident electron energy. The quantity $\log_{10}(|\psi^{r\to l}(z)|^2 \, \sigma^{r\to l}(k_z) \,/\, 2\pi)$ plotted can be interpreted loosely as the probability of finding an electron at a particular position, or at a particular energy. It is plotted on a logarithmic scale, so that important features can be seen clearly. Electrons, injected from the right-hand contact, are propagating from right to left. At low energies, electrons are strongly reflected, and the wavefunction exhibits a pattern of standing waves, near the right-hand contact. At the resonant energy, however, the wavefunction peaks sharply within the GaAs-well, and a ridge of transmitted electrons can be seen extending to the left contact. Note that the wavefunction peak is localized to the GaAs well in position-space, and it is extremely narrow in energy-space.

† S. Ray, P. Ruden, V. Sokolov, R. Kolbas, T. Boonstra, and J. Williams, "Resonant Tunneling Transport at 300 K in GaAs-AlGaAs Quantum Wells Grown by Metalorganic Chemical Vapor Deposition," *Applied Physics Letters*, 48(24), pp. 1666-1668, 1986.

The wavefunction surface plot presented in Figure 5.2 was generated from the output of SEQUAL. A position-space grid, for the structure shown in Figure 5.1, was created and stored in the file *rtd*. Positions were written in angstrom units (1 Å $= 10^{-8}$ cm), and the electrostatic potential was taken as zero everywhere. The input deck, along with the rest of printed output, is shown in Figure 5.3. To insure proper resolution of the wavefunction, an additional node in the middle of the quantum well was specified for "watching." Injection of electrons was specified as right-to-left. Data files used for the wavefunction surface plot were obtained by specifying "*data*=w" on the *output* card. Notice that the electron density, shown in the section of final results, is smaller at the left-hand contact (near z=0). This is expected, because the electron density listed is not the total electron density; rather, it is the component due to electrons injected only from the right-hand contact.

SEQUAL can also be used to determine the current-voltage characteristic for this device, by assigning a number of voltages to the *bias* key. In Figure 5.4, an input deck is shown, which has been modified for this purpose. Notice that separate output-file names are specified for each bias, so that files are not overwritten. For the purposes of this example, a self-consistent solution was not necessary; hence, the maximum number of iterations was set to zero. As an initial guess, biases were specified in increments of 0.05 V; the input deck was then modified again, to include a few additional biases. Values of current, taken from the listings in the sections of final results, were stored in a separate file, and plotted. The resulting current-voltage characteristic appears in Figure 5.5.

Negative differential resistance, apparent in Figure 5.5, is a characteristic feature of resonant tunneling devices. For small biases, the dominant current component is supplied by electrons tunneling through the first quasi-bound state, from the right-hand contact. As larger biases are applied across the device, the quasi-bound state is lowered in energy, with respect to the conduction-band edge in the right-hand contact. More electrons are available for tunneling at energies near the conduction-band edge in the contact. Therefore, current increases to a maximum value. When the quasi-bound state is pulled below the range of injection energies of the contact, however, current is abruptly reduced. This effect is depicted graphically in Figure 5.6, which shows the conduction-band profile of the resonant tunneling device, at the bias of maximum current. Notice that the quasi-bound state energy is quite close to the conduction-band edge in the right-hand contact. Any additional bias lowers the quasi-bound state below the range of (right-hand) injection energies, and current is abruptly cut-off.

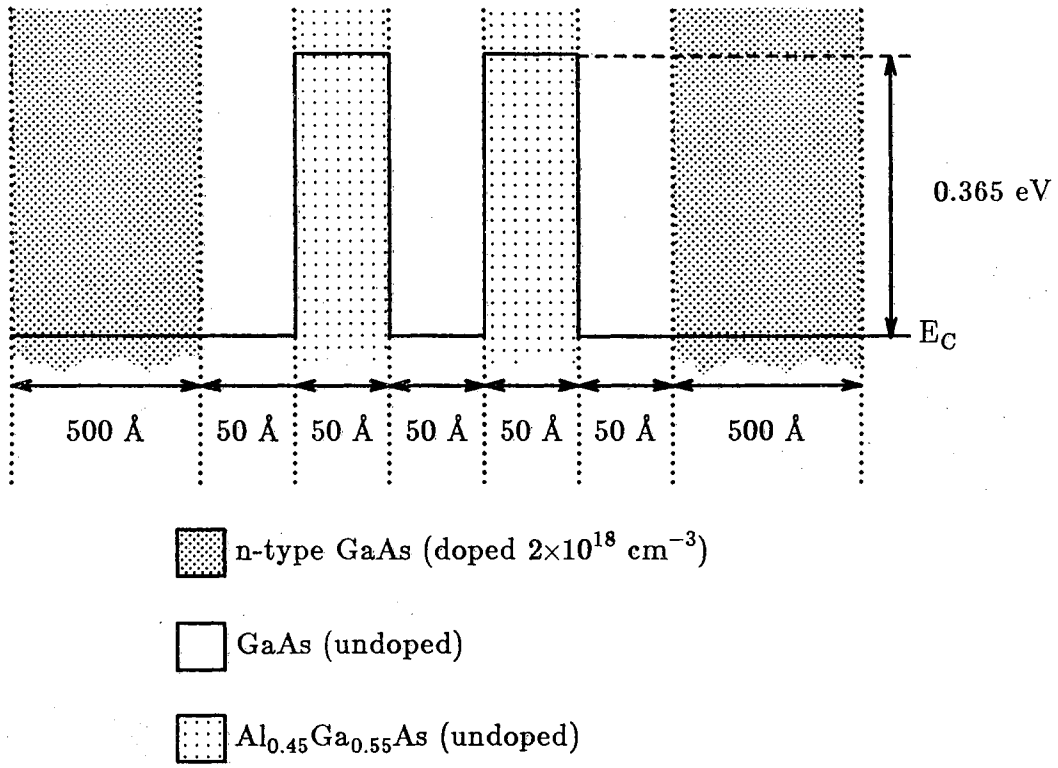


Figure 5.1: Structure of the resonant tunneling device examined in subsequent calculations

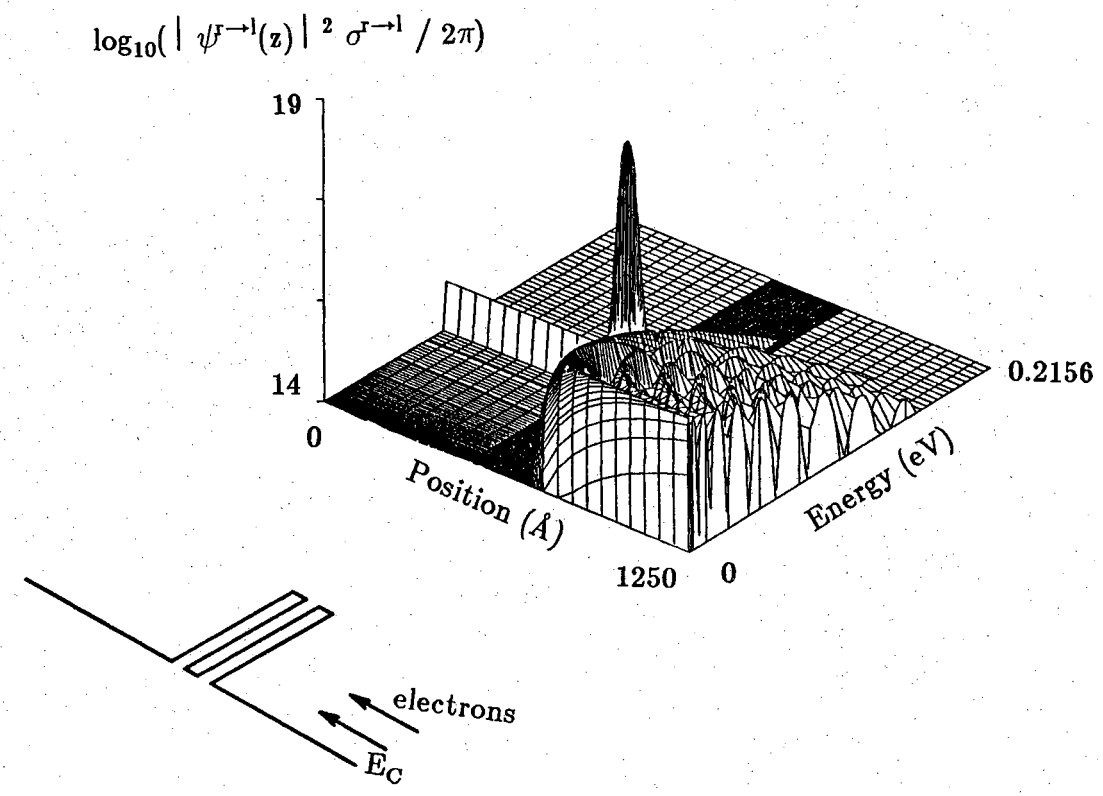


Figure 5.2: The wavefunction magnitude, weighted by the transverse integration $\sigma^{r\to l}(k_z)$, for electrons injected from the right contact of the resonant tunneling device in Figure 5.1

```
------- XXXX  XXXXX   XXXX   XX  XX   XXXX   XX -----------------------------
      XX     XX      XX  XX  XX  XX  XX  XX  XX                      SEQUAL 2.0
----- XXX   XXXX   XX  XX  XX  XX  XXXXXX  XX ------------- Purdue University
       XX  XX     XX  X   XX  XX  XX  XX  XX                       August 1987
-- XXXX   XXXXX   XXX X   XXXXX  XX  XX  XXXXXX -----------------------------

SEQUAL:  input deck
--------------------------------------------------------------------------------

        ************************************************************
        **  a resonant tunneling device fabricated by ray et. al.  **
        **    (applied physics letters, 48(24), p. 1666, 1986) is  **
        **    simulated.  a plot of the wavefunction surface in    **
        **    position-space and energy space is obtained, for     **
        **    electron injection from the right contact.           **
        ************************************************************

title   resonant tunneling device (appl. phys. lett., 48(24), p. 1666, 1986)


        -- input file "rtd" contains positions in angstroms --

input   file=rtd format=zv
scale   cm=1.0e8

matter  nodes=21/32/43/54/76
      + delec=0.0/0.365/0.0/0.365/0.0
      + emass=0.067/0.08953/0.067/0.08953/0.067

doping  nodes=10/65/76  nd+=2.e18/0./2.e18

device  temp=300.0  area=9.e-6
solve   itmax=0  prec=3  inject=r-to-l


        -- to insure proper resolution of the wavefunction,
        --  watch a node in the well-region (node 38)

maxima  watch=38

print   tcoeff=*  format1=zen  format2=*  verbose=true
output  file=rtd  data=w
--------------------------------------------------------------------------------

Executing a total of  1 calculation(s).
```

Figure 5.3: Example of printed output

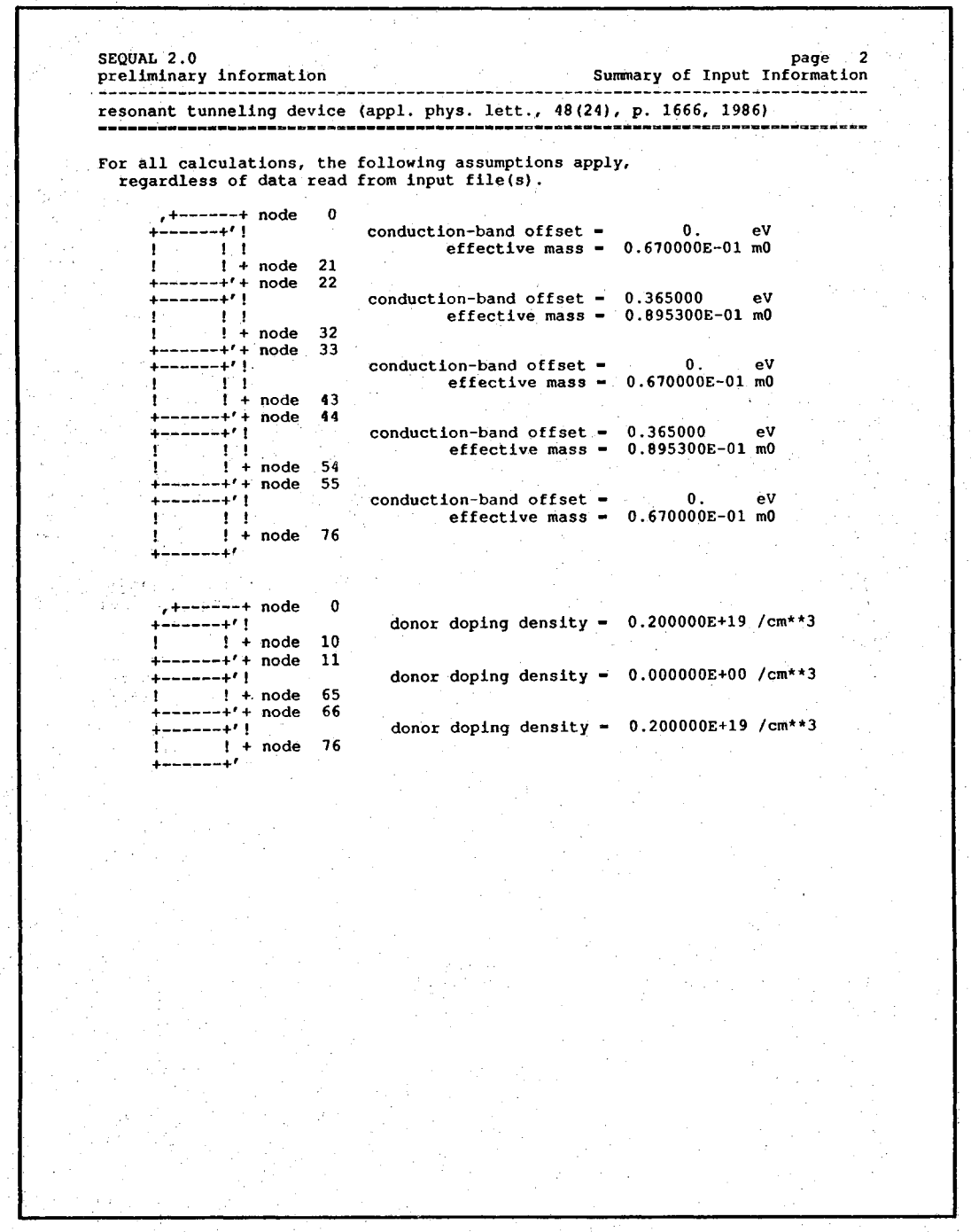


Figure 5.3: Example of printed output (continued)

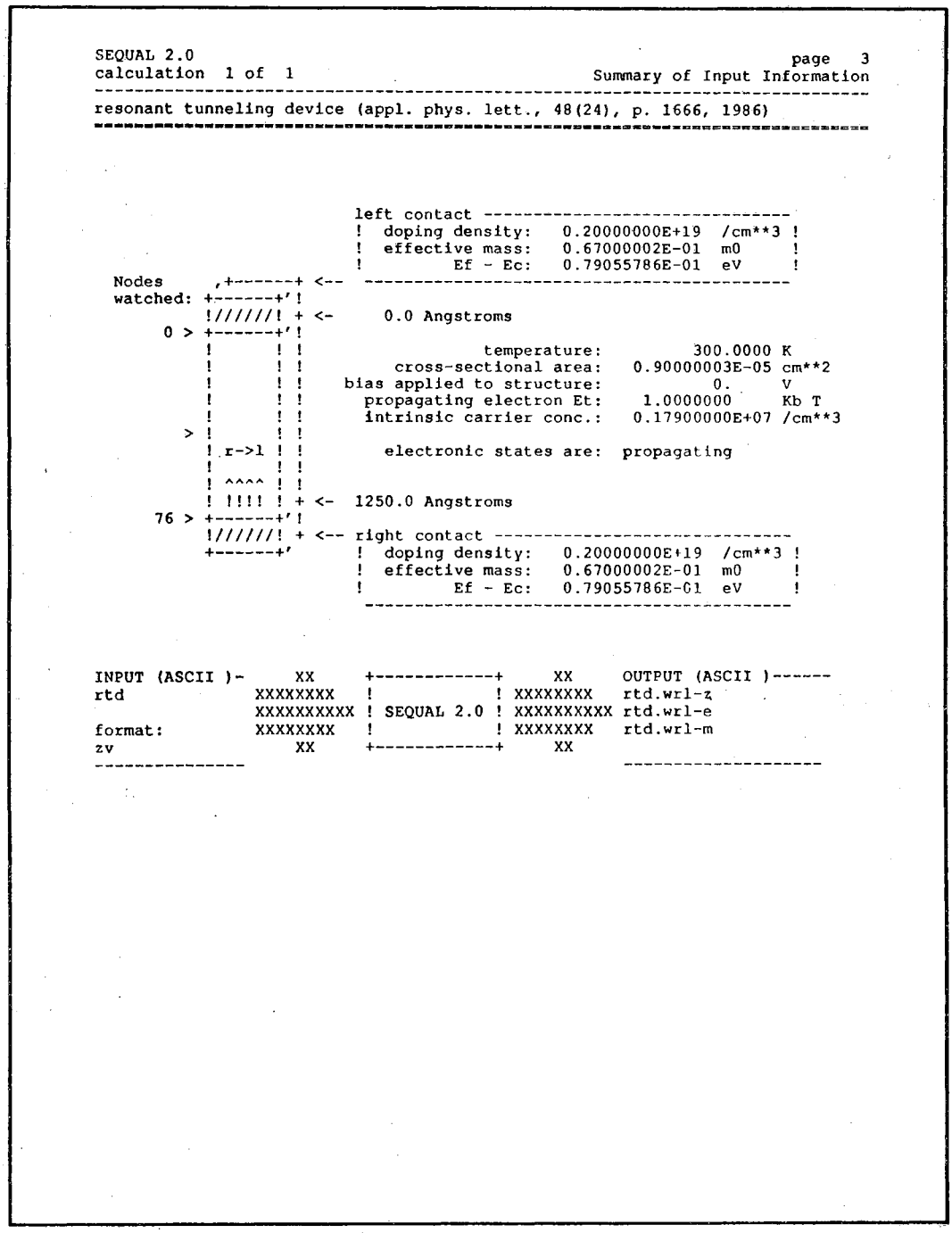
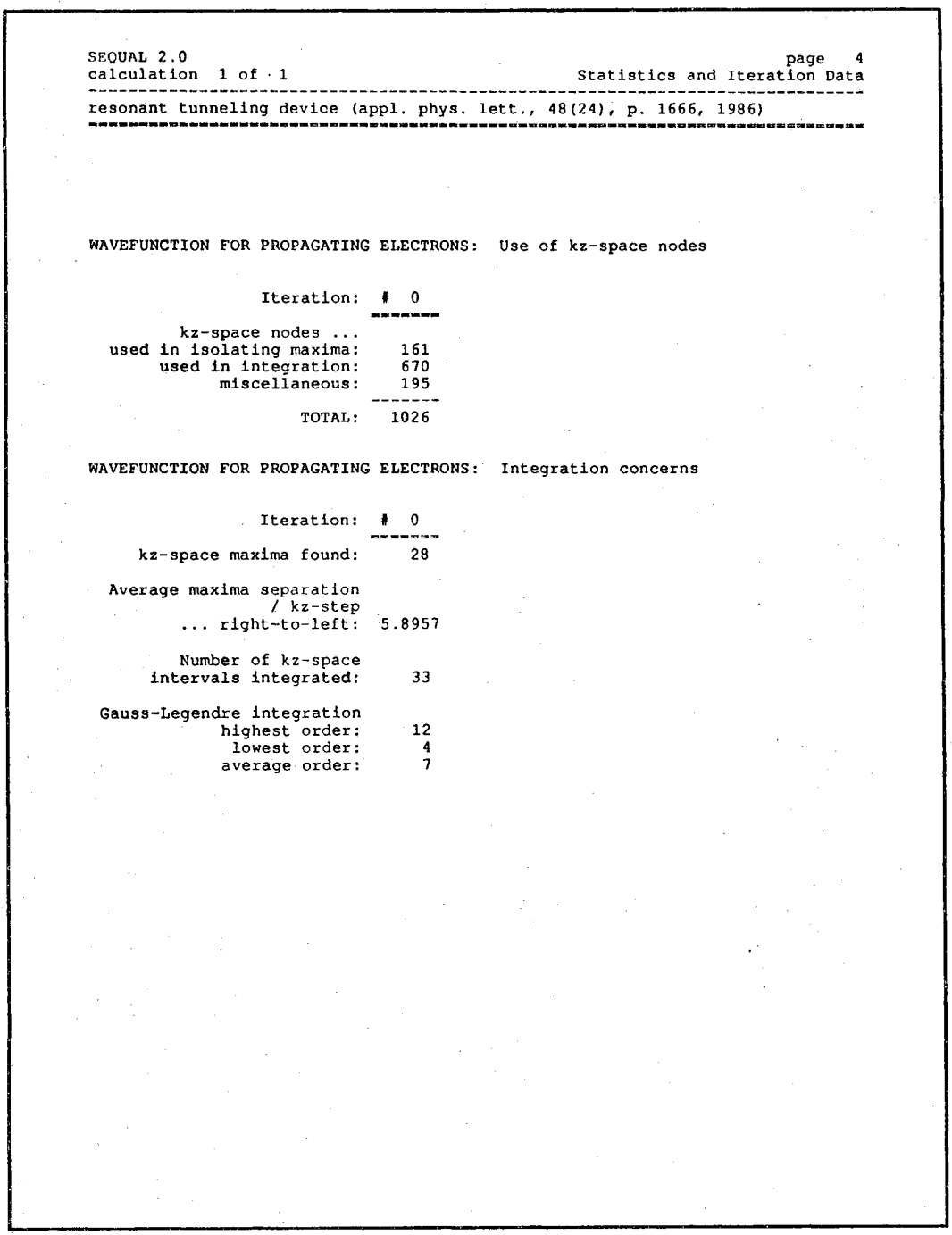


Figure 5.3: Example of printed output (continued)



```
SEQUAL 2.0                                                    page   4
calculation  1 of  1                    Statistics and Iteration Data
-------------------------------------------------------------------------
resonant tunneling device (appl. phys. lett., 48(24), p. 1666, 1986)
=========================================================================


WAVEFUNCTION FOR PROPAGATING ELECTRONS:  Use of kz-space nodes

                Iteration:  #  0
                            =======
        kz-space nodes ...
 used in isolating maxima:     161
      used in integration:     670
            miscellaneous:     195
                            -------
                   TOTAL:    1026

WAVEFUNCTION FOR PROPAGATING ELECTRONS:  Integration concerns

                Iteration:  #  0
                            =======
    kz-space maxima found:      28

  Average maxima separation
                 / kz-step
        ... right-to-left:  5.8957

       Number of kz-space
     intervals integrated:      33

 Gauss-Legendre integration
              highest order:     12
               lowest order:      4
              average order:      7
```

Figure 5.3: Example of printed output (continued)

```
SEQUAL 2.0                                                        page   5
calculation  1 of  1                                        Final Results
-------------------------------------------------------------------------------
resonant tunneling device (appl. phys. lett., 48(24), p. 1666, 1986)
===============================================================================


Current:                                  Voltage:
-----------------------------------       -----------------------------------
   LEFT-to-RIGHT:   0.000000E+00 A           In Input File:   0.000000E+00 V
   RIGHT-to-LEFT:  -0.137929E-01 A            Applied Bias:   0.000000E+00 V
                   --------------                            --------------
   Total Current:   0.137929E-01 A           Total Voltage:   0.000000E+00 V


Position          Conduction Band    Electron Dens.
(cm)              (eV)               (/cm**3)

        0.                     0.      0.13938995E+15
 0.50000000E-06                0.      0.13938993E+15
 0.10000000E-05                0.      0.13938994E+15
 0.15000002E-05                0.      0.13938995E+15
 0.20000000E-05                0.      0.13938994E+15
 0.24999999E-05                0.      0.13938995E+15
 0.30000003E-05                0.      0.13938995E+15
 0.35000000E-05                0.      0.13938994E+15
 0.40000000E-05                0.      0.13938995E+15
 0.45000002E-05                0.      0.13938994E+15
 0.49900000E-05                0.      0.13938994E+15
 0.49999999E-05                0.      0.13938994E+15
 0.50499998E-05                0.      0.13938994E+15
 0.51000002E-05                0.      0.13938994E+15
 0.51500001E-05                0.      0.13938995E+15
 0.52000000E-05                0.      0.13938994E+15
 0.52500000E-05                0.      0.13938994E+15
 0.52999999E-05                0.      0.13938994E+15
 0.53499998E-05                0.      0.13938994E+15
 0.54000002E-05                0.      0.13938994E+15
 0.54500001E-05                0.      0.13938994E+15
 0.54900001E-05                0.      0.13938995E+15
 0.55000000E-05       0.36500001       0.13994561E+15
 0.55500000E-05       0.36500001       0.18066795E+15
 0.55999999E-05       0.36500001       0.31577197E+15
 0.56499998E-05       0.36500001       0.63506427E+15
 0.57000002E-05       0.36500001       0.13511151E+16
 0.57500001E-05       0.36500001       0.29412975E+16
 0.58000001E-05       0.36500001       0.64667889E+16
 0.58500000E-05       0.36500001       0.14281898E+17
 0.58999999E-05       0.36500001       0.31607775E+17
 0.59499998E-05       0.36500001       0.70023107E+17
 0.59899994E-05       0.36500001       0.13236001E+18
 0.60000007E-05                0.      0.15169191E+18
 0.60500001E-05                0.      0.24259597E+18
 0.61000001E-05                0.      0.33233649E+18
 0.61500004E-05                0.      0.40712321E+18
 0.61999999E-05                0.      0.45546977E+18
 0.62499998E-05                0.      0.46995659E+18
```

Figure 5.3: Example of printed output (continued)

SEQUAL 2.0 page 6
calculation 1 of 1 Final Results

resonant tunneling device (appl. phys. lett., 48(24), p. 1666, 1986)

Position (cm)	Conduction Band (eV)	Electron Dens. (/cm**3)
0.63000002E-05	0.	0.44836270E+18
0.63500001E-05	0.	0.39400858E+18
0.63999996E-05	0.	0.31524703E+18
0.64500000E-05	0.	0.22417760E+18
0.64900000E-05	0.	0.15185733E+18
0.65000004E-05	0.36500001	0.13253833E+18
0.65499999E-05	0.36500001	0.60011229E+17
0.65999998E-05	0.36500001	0.27507440E+17
0.66500002E-05	0.36500001	0.13311273E+17
0.67000001E-05	0.36500001	0.79630943E+16
0.67499996E-05	0.36500001	0.79649460E+16
0.67999999E-05	0.36500001	0.13514750E+17
0.68499999E-05	0.36500001	0.28819871E+17
0.68999993E-05	0.36500001	0.65357604E+17
0.69500006E-05	0.36500001	0.15065666E+18
0.69900002E-05	0.36500001	0.29519630E+18
0.70000001E-05	0.	0.34111472E+18
0.70500000E-05	0.	0.56871785E+18
0.71000004E-05	0.	0.82325686E+18
0.71499999E-05	0.	0.10842372E+19
0.71999998E-05	0.	0.13329382E+19
0.72500002E-05	0.	0.15543195E+19
0.73000001E-05	0.	0.17382395E+19
0.73499996E-05	0.	0.18798473E+19
0.74000000E-05	0.	0.19792246E+19
0.74499999E-05	0.	0.20404274E+19
0.74899999E-05	0.	0.20663852E+19
0.74999994E-05	0.	0.20702057E+19
0.80000000E-05	0.	0.19887109E+19
0.85000001E-05	0.	0.20006154E+19
0.90000003E-05	0.	0.19999505E+19
0.94999996E-05	0.	0.19998500E+19
0.99999997E-05	0.	0.19998793E+19
0.10500000E-04	0.	0.19998764E+19
0.11000000E-04	0.	0.19998765E+19
0.11500000E-04	0.	0.19998763E+19
0.12000001E-04	0.	0.19998765E+19
0.12500000E-04	0.	0.19998761E+19

Figure 5.3: Example of printed output (continued)

```
      ************************************************************
      .    A resonant tunneling device fabricated by Ray et. al.
      .     (applied physics letters, 48(24), p. 1666, 1986) is
      .     simulated. Current is computed for several biases.
      .     Results are not self-consistent.
      ************************************************************

title           resonant tunneling device (appl. phys. lett., 48(24), p. 1666, 1986)


                -- input file "rtd" contains positions in angstroms

input           file=rtd format=zv
scale           cm=1.0e8

matter          nodes=21/32/43/54/76
  +             delec=0.0/0.365/0.0/0.365/0.0
  +             emass=0.067/0.08953/0.067/0.08953/0.067

doping          nodes=10/65/76  nd+=2.e18/0./2.e18


                -- compute current for biases:
                --   0 V, 0.05 V, 0.10 V, 0.15 V, 0.20 V, 0.25 V, 0.30 V

device          temp=300.0  area=9.e-6
  +             bias=0.00/0.05/0.10/0.15/0.20/0.25/0.30

solve           itmax=0  prec=3  inject=r-to-l
maxima          watch=38


print           tcoeff=*  format1=zen  format2=*  verbose=true
output          file=rtd00/rtd05/rtd10/rtd15/rtd20/rtd25/rtd30
  +             data=dt
```

Figure 5.4: Input deck for the application of several different biases

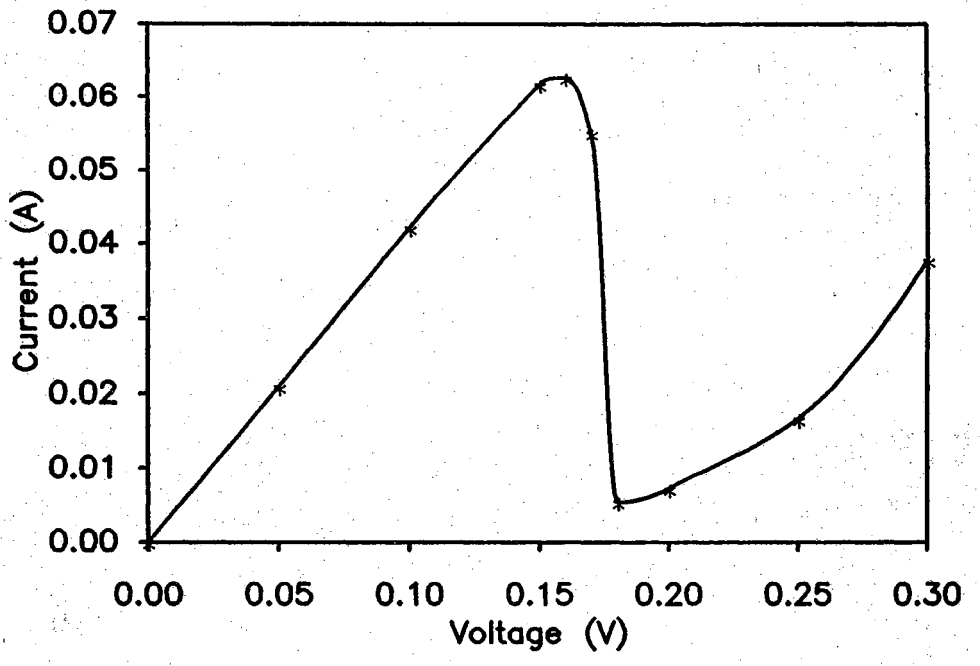


Figure 5.5: Current-voltage characteristic for the resonant tunneling device

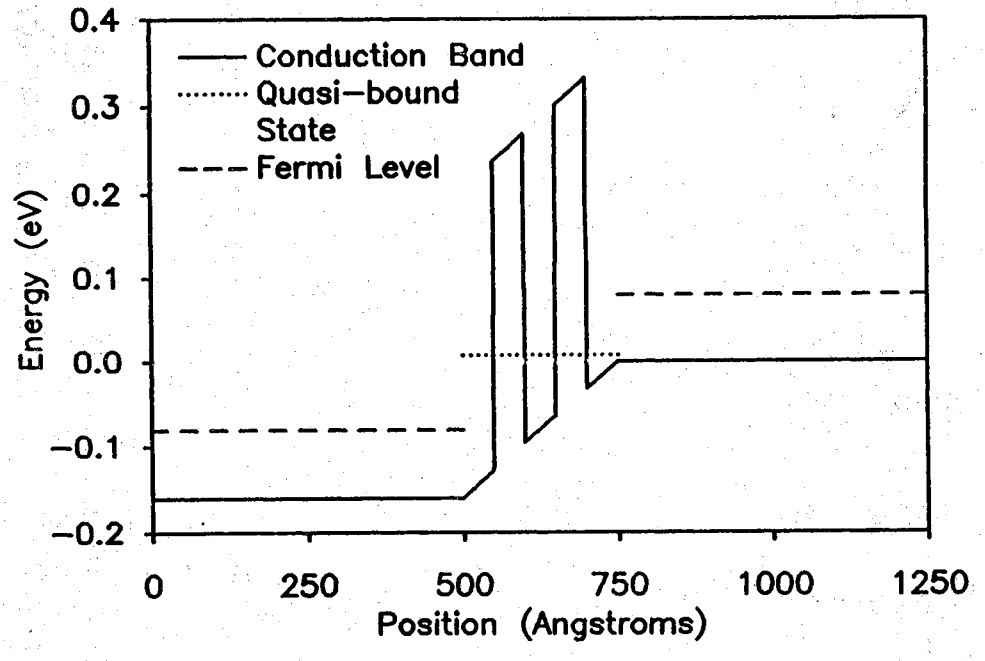


Figure 5.6: Conduction-band profile of the resonant tunneling device for the bias of peak current

6

Special Considerations

6.1. Installation Notes

SEQUAL is written in standard Fortran 77, and has been implemented at Purdue University on Sun™ workstations and the Dual VAX 11/780 machines. To ensure reasonable accuracy, floating-point representations should use at least 48 bits to describe the mantissa (fractional part). Therefore, typical installations must use double-precision complex variables. Although this is not allowed in standard Fortran 77, the *f77* compiler for the Dual VAX machines will automatically double the precision of all floating-point variables, when the argument "-r8" is specified. Furthermore, many compilers now recognize "double complex" as a data type. To account for these variations in Fortran 77 compilers, three versions of SEQUAL exist. One version, for machines with a large word-length, uses ordinary, single-precision complex variables. Another, for machines which recognize the "double complex" data type, uses double-precision complex variables. Finally, a third version exists for machines with insufficient precision and a strict compiler; this final version simulates the double-precision complex type by using pairs of double-precision variables.

For the analysis of bound states, SEQUAL relies upon *eigrf* in the International Mathematical and Statistical Library (IMSL). If this subroutine is not available to the user, it can be removed from the program code, and the bound-state analysis can be disabled. To accomplish these tasks, the user must comment-out the subroutine call, which appears in the subroutine *bstate*:

```
c
c
c  zzz zzz zzz  MACHINE DEPENDENT CODE  zzz zzz zzz
c
c  If IMSL routines are unavailable, comment out the
c    following line, so that SEQUAL can be compiled:
c
c     call eigrf(matrix,nodes,ia,ijob,val,vec,ia,wk,ier)
c
```

In addition, the constant *imsl* in the main program should be properly set, to disable the bound-state analysis:

```
c
c  zzz zzz zzz MACHINE DEPENDENT CODE zzz zzz zzz
c
c    imsl ... if IMSL routines are unavailable for a host
c             system, the parameter 'imsl' below should
c             be assigned the value '.false.'
c
      integer zimax, mxque, maxiter, maxii, errmax
      integer mxcard, mxkeys, mxarry, mxoind, mxord, nexts
      real maxreal
      logical imsl
      parameter ( zimax=251,
     +           maxiter=30,
                    .
                    .
                    .
     +            imsl=.false.   )
c
```

Notice that both of these corrections are clearly marked in the program code by the comment:

```
c  zzz zzz zzz MACHINE DEPENDENT CODE zzz zzz zzz
```

In SEQUAL, all machine-dependent constants are marked in this manner. For a proper installation, the user should search the program code for all appearances of this comment, and follow the instructions immediately following it. Since the vast majority of machine-dependent constants are used to check for overflow or underflow, the corrections should be obvious.

6.2. Pitfalls to Avoid

Among the most frustrating problems that a user might encounter are those which arise from the physics of a particular analysis. Suppose a (weary) user has created a device description and an input deck, has run SEQUAL, and has obtained results which appear to be in error. Resisting his impulse to burn this manual, the user would receive his reward in this section; it is devoted to relieving both confusion and misery.

At the heart of the propagating-state analysis lies the assumption that contacts are in local, thermodynamic equilibrium. Unless the potential profile is sufficiently flat near each contact, this assumption will be violated. To guarantee a proper analysis,

it is prudent to include "buffer" regions between the ideal contacts (nodes 0 and N) and the actual region of interest in the device. For example, in Chapter 5, the analysis of a resonant tunneling device included 500 Å buffer layers of GaAs, separating nodes 0 and N from the interior of the device. These buffer layers are merely extensions of the contacts, with the same material properties and the same density of ionized donors. They should be long enough to allow equilibrium to be restored: Electron density should return to the density of ionized donors, and the electrostatic potential should flatten out (*i.e.*, electric fields should decay to zero), within these regions. Because of the recursive dependence of both the electron density and the electrostatic potential, the use of buffer layers is especially important in self-consistent calculations. Unless equilibrium is restored near the contacts, errors in either the electron density or the electrostatic potential will feed back into the solution, making convergence an impossible goal.

Another source of feedback can further aggravate problems in convergence: If the contacts (or buffer layers) are lightly-doped with respect to the internal device structure, a self-consistent solution may be difficult to obtain. In this case, most of the band-bending (and most of the applied bias) will appear in buffer layers. Restoring equilibrium conditions near the contacts will be difficult. To some degree, the length of buffer layers can be increased; the analysis of long devices, however, is itself a difficult task, which will be described below.

In the same manner, non-equilibrium solutions may be difficult to obtain for conduction-band profiles which do not obstruct the flow of electrons. For instance, if the user attempts to apply a bias to a uniform block of semiconductor (for which the equilibrium conduction-band profile is completely flat), the results obtained will be nonsense. For a reasonable bias, the electron density will be higher at one contact, and lower at the other. In a real device, the flow of electrons is somewhat impeded by interactions with scattering mechanisms; in the limit of ballistic transport, however, there is nothing to impede the flow of electrons, aside from interactions with the potential. A conduction-band profile without reflective features might be thought of as representing a device with infinite conductance; because it is unable to support a bias, the user should not consider applying a bias.

In the analysis of bound states, the use of buffer layers is also important. As boundary conditions to the finite-difference solution of the Schrödinger equation, it was assumed that the wavefunction is zero at each end of the device (nodes 0 and N). If the wavefunction decays sufficiently within the buffer regions, this assumption is valid; otherwise, the solution of eigensystem may be in error.

From the preceding discussion, one might be tempted to include buffer layers of classical dimensions (say, ~ 1 μm) in every calculation. Unfortunately, this solution would create an even larger problem. As the overall device length increases, the wavefunction magnitude becomes a rapidly varying function of k_z. To insure an accurate integration of the wavefunction, SEQUAL integrates the k_z intervals between peaks in the wavefunction magnitude (see section 2.1.3). Therefore, the number of points in k_z space (and hence, the number of wavefunction solutions) increases dramatically with increasing device length. For devices with classical dimensions (~ 1 μm), the execution time of SEQUAL may be prohibitive. Typically, buffer regions 500 Å to 1000 Å long are sufficient to guarantee proper results. As a rule, it is more efficient to start with buffer layers that are too short, rather than too long.

Given the density of ionized donors, SEQUAL will compute the Fermi-level E_F for each contact (nodes 0 and N). Because the electron density calculated for the propagating-state solution is extremely sensitive to the Fermi-level, the proper E_F must be calculated within SEQUAL. A problem can arise, however, when using the output of classical analysis programs, which allow for non-parabolicity of the energy dispersion relation. For a device with contacts of differing materials, the Fermi-levels calculated by SEQUAL may not correspond to those found in the (non-parabolic) classical analysis. In this case, the device will appear to have a different bias in SEQUAL (*i.e.*, a different separation in the contact Fermi-levels) than it did in the classical analysis. This bias would be noted in the printed-output section of final results. When listing the bias across device, SEQUAL provides both the actual bias (*e.g.*, separation in contact Fermi-levels) and the applied bias. To remedy this situation, the user could apply a bias across the device which will counteract the erroneous Fermi-level separation.

Although (hopefully) the discussion above might be helpful, it is not intended to cover every conceivable problem. One final word of advice, from an author who has invested countless hours in the examination of program output: If SEQUAL says so, it is probably correct. More often than not, I have found a problem to stem from my own misconceptions and mistakes, rather than from the regimented computations of the program.

References

M. J. McLennan, *Quantum Ballistic Transport in Semiconductor Heterostructures*, M.S. Thesis, Purdue University, West Lafayette, IN (May 1987).

L. Esaki, "A Bird's-Eye View on the Evolution of Semiconductor Superlattices and Quantum Wells," *IEEE Journal of Quantum Electronics*, **QE-22**(9), pp. 1611-1624, 1986.

B. Ricco and M. Ya. Azbel, "Physics of Resonant Tunneling. The One-Dimensional Double-Barrier Case," *Physical Review B*, **29**(4), pp. 1970-1981, 1984.

F. Capasso, K. Mohammed, and A. Y. Cho, "Resonant Tunneling Through Double Barriers, Perpendicular Quantum Transport Phenomena in Superlattices, and Their Device Applications," *IEEE Journal of Quantum Electronics*, **QE-22**(9), pp. 1853-1869, 1986.

R. Tsu and L. Esaki, "Tunneling in a Finite Superlattice," *Applied Physics Letters*, **22**(11), pp. 562-564, 1973.

M. O. Vassell, Johnson Lee, and H. F. Lockwood, "Multibarrier Tunneling in $Ga_{1-x}Al_xAs$/GaAs Heterostructures," *Journal of Applied Physics*, **54**(9), pp. 5206-5213, 1983.

S. Ray, P. Ruden, V. Sokolov, R. Kolbas, T. Boonstra, and J. Williams, "Resonant Tunneling Transport at 300 K in GaAs-AlGaAs Quantum Wells Grown by Metalorganic Chemical Vapor Deposition," *Applied Physics Letters*, **48**(24), pp. 1666-1668, 1986.

H. Ohnishi, T. Inata, S. Muto, N. Yokoyama, and A. Shibatomi, "Self-Consistent Analysis of Resonant Tunneling Current," *Applied Physics Letters*, **49**(19), pp. 1248-1250, 1986.

M. Cahay, M. McLennan, S. Datta, and M. S. Lundstrom, "Importance of Space-Charge Effects in Resonant Tunneling Devices," *Applied Physics Letters*, **50**(10), pp. 612-614, 1987.