INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE CENTRATOM TRIESTE

SMR.459 - 31

SPRING COLLEGE IN CONDENSED MATTER ON:
"PHYSICS OF LOW-DIMENSIONAL STRUCTURES"

(23 APRIL - 15 JUNE 1990)

BACKGROUND MATERIAL FOR COURSE OF LECTURES ON:

QUANTUM TRANSPORT IN ULTRASMALL DEVICES

S. DATTA
Purdue University
School of Electrical Engineering
In 47909
West Lafayette
U.S.A.

---

141 229 15 40 17 Galley 1

I.O.P. — R.O.P. — Paper 229/1

Rep. Prog. Phys. (1990)000-000 Printed in the UK

Quantum transport in ultrasmall electronic devices

Supriyo Datta and Michael J McLennan

School of Electrical Engineering, Purdue University, West Lafayette, IN 47097, USA

Abstract

Device analysis has traditionally been based on the semiclassical Boltzmann transport equation. Despite its impressive successes, this approach suffers from an important limitation; it cannot describe transport phenomena in which the wave nature of electrons plays a crucial role. A variety of such quantum effects have been discovered over the years, such as tunnelling, resonant tunnelling, weak and strong localisation, and the quantum Hall effect. Since 1985, experiments on ultrasmall structures (dimensions ≤ 100 nm) have revealed a number of new effects such as the Aharanov–Bohm effect, conductance fluctuations, non-local effects and the quantised resistance of point contacts. For ultrasmall structures at low temperature, these phenomena have clearly shown that electron transport is influenced by wave interference effects not unlike those well known in microwave networks. New device concepts are being proposed and demonstrated that are based on these wave properties.

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

This review was received in November 1989.

0034-4885/89/000000+00$00.00

S Datta and M J McLennan

Contents

0097'''00681'''02731

1. Introduction

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

$$J = e\mu(\mathscr{E})n\mathscr{E} + eD(\mathscr{E})\nabla n. \tag{1.1}$$

Here J is the current density, n is the electron density, $\mathscr{E}$ is the electric field and μ and D are the basic transport parameters called the mobility and the diffusion coefficient respectively. For simplicity, we restrict ourselves to one type of carrier, namely electrons. In deriving equation (1.1) one makes two main assumptions:

(i) Electrons are particles with an effective mass that move in an external field according to Newton's law with occasional scattering by phonons and impurities.

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

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

$$v(k)\cdot\nabla f + \frac{e\mathscr{E}}{\hbar}\cdot\nabla_k f = S_{op}f \tag{1.2}$$

where $f(r, k)$ is the semiclassical distribution function that tells us the number of electrons at r having the wavevector k; v is the velocity of an electron with wavevector k; S_{op} is the scattering operator which is usually evaluated by applying Fermi's golden rule to the individual scatterers. In deriving equation (1.2) one needs only the *first* of the two assumptions listed after equation (1.1). Consequently, hot-electron effects are accounted for.

As devices shrink to dimensions comparable with the wavelength of electrons, it is expected that the wave nature of electrons will play an increasingly important role and even the first assumption will no longer be valid. On such small length scales the semiclassical distribution function is no longer a valid concept, due to the uncertainty relation between r and k. To analyse and design devices on a sub-100 nm scale it will be necessary to go beyond the Boltzmann equation (1.2) and develop simulation techniques based on quantum kinetic equations (figure 1). The development of an appropriate kinetic equation is an active topic of current theoretical research that has recently gained impetus from the surge of experimental activity in the area of quantum transport.

The development of molecular beam epitaxy since the late 1960s has made it possible to grow ultrathin (~20 Å) layers of different materials with atomically sharp interfaces. This has led to the development of *vertical* quantum devices where the current flows *perpendicular* to the layers (figure 2). Changes in the material composition gives rise to variations in the conduction band edge, which the electrons feel as an

0470'''03502

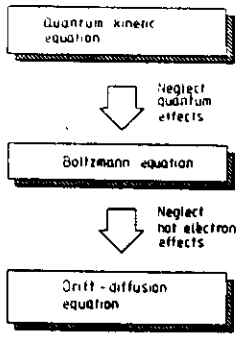


Figure 1. Hierarchy of transport theories.

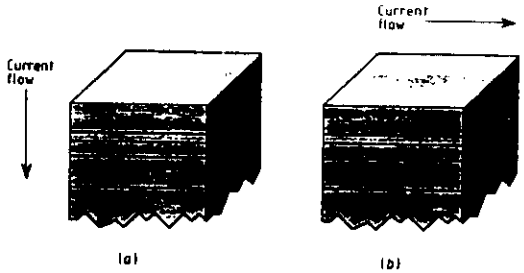


Figure 2. (*a*) Vertical and (*b*) lateral quantum devices.

effective potential. Alternatively, modulations in the doping profile during film growth can be used to tailor the electrostatic potential through space-charge transfer. Some of these vertical structures, such as resonant tunnelling devices, have now reached a high level of maturity and have emerged as potentially useful practical devices. By contrast *lateral* quantum devices, with current flowing *parallel* to the layers, are still in their infancy. In these structures, the potential can be defined through patterned electrodes which deplete selected regions. Such structures have only recently been made possible by the advances in nanolithographic techniques. It will probably be many years before useful devices based on such effects become practicable. However, since 1985, there has been a flood of experiments revealing novel quantum effects at low temperatures, causing great excitement among both basic and applied physicists. On the one hand, such experiments open up new ways to study fundamental questions of physics; for instance, the role of dissipation in microscopic phenomena. On the other hand, these studies raise the possibility of radically new electronic devices that operate by controlling the phase of the wavefunction, rather than by controlling the carrier density, as in present-day devices. The last few years have seen the emergence of a new research area that has been given a variety of names such as 'mesoscopic physics', 'nanostructure physics' and 'nanoelectronics'.

The outline of this review is as follows. In section 2, we will discuss the basic conceptual framework that one uses to describe electron transport in ultrasmall devices.

We will then review various quantum effects that have been observed, and discuss possible device implications in section 3. Our current theoretical understanding of quantum transport phenomena is reviewed in section 4. We conclude in section 5 by discussing future prospects of quantum devices. The purpose of this article is to review our current understanding of quantum transport in ultrasmall structures. Due to the rapid developments in this area it is inevitable that, despite the authors' best efforts, some aspects may not be covered adequately.

2. Background

All of the phenomena that we will discuss in this article are essentially *one-electron* phenomena, although it is possible that many-body effects will play a more significant role in the electronic properties of small structures. Most of the experimental observations to date are well explained, at least qualitatively, in terms of the simple one-particle picture described below. A notable exception is the fractional quantum Hall effect, which is outside the scope of this review.

2.1. Current-voltage formula

An electronic device is typically connected to two contacts across which a voltage is applied (figure 3(*a*)). Each of these contacts launches a steady stream of electrons onto the device, of which a fraction is transmitted to the other contact. At equilibrium with both contacts having the same electrochemical potential, the current $I_{1\to 2}$ transmitted from contact 1 to contact 2 is exactly balanced by the current $I_{2\to 1}$ transmitted from contact 2 to contact 1. An applied voltage shifts the local chemical potential μ_1 in contact 1 with respect to the local chemical potential μ_2 in contact 2, making $I_{1\to 2}$ different from $I_{2\to 1}$ and causing a net current flow through the device. The currents $I_{1\to 2}$ and $I_{2\to 1}$ may be evaluated as follows.

The incident flux I_1^+ from contact 1 is written as

$$I_1^+ = e\sum_n \int \frac{dk_n}{2\pi}\frac{\hbar k_n}{m^*} f(E - e\mu_1) \qquad (2.1)$$

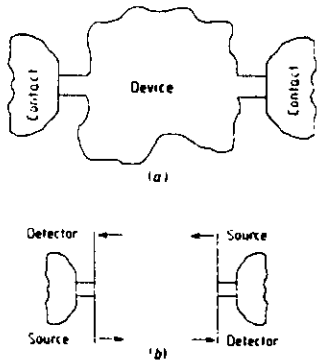


Figure 3. (*a*) A device with two contacts. (*b*) The two contacts in (*a*) act as source and detector with the device as the intervening medium.

where $f(E)$ is the Fermi–Dirac function, m^* is the effective mass, k_n is the longitudinal wavenumber, and n denotes the transverse modes or sub-bands (including spin), all in lead 1. The energy E is equal to the energy ε_n at the bottom ($k_n = 0$) of sub-band n plus the longitudinal kinetic energy $\hbar^2 k_n^2/2m^*$

$$E = \varepsilon_n + (\hbar^2 k_n^2/2m^*). \tag{2.2}$$

The sub-band energy ε_n is the sum of the potential energy and the transverse kinetic energy. For a large area contact the allowed energies ε_n are essentially continuous, while for a small contact they form a discrete set. Using the definition for the total energy (2.2) we express the incident flux I_1^+ (2.1) as

$$I_1^+ = \frac{e}{h} \sum_n \int_{\varepsilon_n}^{\infty} dE f(E - e\mu_1). \tag{2.3}$$

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

$$I_{1\to 2} = \frac{e}{h} \sum_{n,m} \int_{\varepsilon_n}^{\infty} dE f(E - e\mu_1) \tau_{21}^{nm}(E)$$

$$= \frac{e}{h} \int dE f(E - e\mu_1) T_{21}(E) \tag{2.4}$$

where

$$T_n(E) = \sum_{m,m} \tau_n^{mm}(E) \theta(E - \varepsilon_n) \tag{2.5}$$

and $\theta(E)$ is the Heaviside step function. Similarly, we can show that

$$I_{2\to 1} = \frac{e}{h} \int dE f(E - e\mu_2) T_{12}(E). \tag{2.6}$$

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

$$I_1 = -I_2 = I_{1\to 2} - I_{2\to 1}$$

$$= \frac{e}{h} \int dE [T_{21}(E) f(E - e\mu_1) - T_{12}(E) f(E - e\mu_2)]. \tag{27}$$

The approach described above has been widely used in tunnelling problems (Frenkel 1930, Duke 1969, Tsu and Esaki 1973). In these problems, it is usually assumed that there are no phase-breaking processes within the device so that the transmission coefficients may be obtained from the one-electron Schrödinger equation (this is discussed further in section 2.3). Indeed, to the knowledge of the authors, equation (2.7) has only been applied to such phase-coherent transport problems. In this limit, it is straightforward to show that equation (2.7) leads to the correct equilibrium condition; namely, that the current I_1 is zero with $\mu_1 = \mu_2 = \mu_0$. This requires that

$$\int dE f(E - e\mu_0)(T_{21}^{(0)}(E) - T_{12}^{(0)}(E)) = 0. \tag{2.8}$$

The superscript '0' is added to indicate that the transmission coefficients are evaluated at equilibrium with a constant electrochemical potential μ_0 everywhere. In the case of phase-coherent transport, we will later show from the symmetry properties of the S-matrix that $T_{21}(E) = T_{12}(E)$. Consequently, the validity of equation (2.8) is obvious.

From the above derivation of equation (2.7), however, it seems that this equation could be applicable more generally, provided one knows how to compute the transmission coefficients in the presence of phase-breaking processes. For example, in the extreme limit of incoherent transport, one could compute the transmission coefficients from a semiclassical Monte Carlo simulation and use them in equation (2.7) to obtain the I–V characteristics. But in the presence of inelastic scattering, it is difficult to show that equation (2.7) leads to the correct equilibrium condition (2.8). In the presence of inelastic scattering $T_{21}(E) \neq T_{12}(E)$, as we can see by considering the simple two-probe device in figure 4. We have a single inelastic scatterer on the right of a potential barrier. An electron incident with energy E from probe 1 crosses the barrier, loses energy to the scatterer and exits into probe 2. But an electron incident with energy E from probe 2 loses energy to the scatterer before crossing the barrier and cannot cross the barrier into probe 1. Clearly $T_{21}(E) > T_{12}(E)$ in this case, and in general there is no obvious relationship between $T_{21}(E)$ and $T_{12}(E)$. Consequently, the validity of equation (2.8) is not obvious, and the authors are not aware of a general proof. Much more is understood in the limit of smalll applied biases, which we consider below.

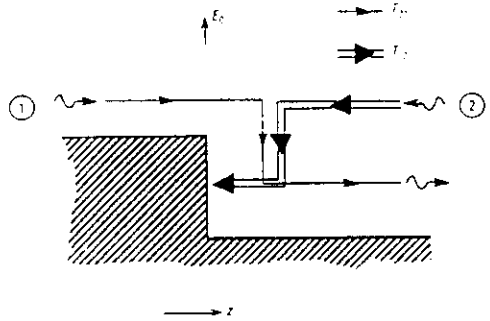


Figure 4. A simple example to show that $T_{21}(E) \neq T_{12}(E)$ in the presence of phase-breaking processes.

2.2. *Linear response*

Equation (2.7) is suitable for computing the current flowing through the circuit in response to an applied potential difference $\mu_1 - \mu_2$. If this difference is 'small', then one can simplify equation (2.7) as follows. At equilibrium with $\mu_1 = \mu_2 = \mu_0$, I_1 is zero, as we just discussed. Now, if we assume that the electrochemical potentials μ_1 and μ_2 deviate only slightly from the equilibrium value μ_0 then we can expand the Fermi–Dirac functions in equation (2.7) in a Taylor series about $e\mu_0$:

$$f(E - e\mu_{1,2}) \simeq f_0(E) + e\left(-\frac{\partial f_0}{\partial E}\right)(\mu_0 - \mu_{1,2}). \tag{2.9}$$

Here $f_0(E)$ stands for $f(E - e\mu_0)$. Substituting this expansion (2.9) into the expression

for the current (2.7), and using the identity found in equilibrium (2.8), we obtain

$$I_1 = \frac{e^2}{h}(\tilde{T}_{21}\mu_1 - \tilde{T}_{12}\mu_2) \tag{2.10}$$

where

$$\tilde{T}_{ij} = \int dE \left(-\frac{\partial f_0}{\partial E}\right) T_{ij}^{(0)}(E). \tag{2.11}$$

We have assumed that in the limit of linear response the transmission coefficients $T_{ij}(E)$ are well approximated by the equilibrium values $T_{ij}^{(0)}(E)$. Since the current in equation (2.10) must be zero with $\mu_2 = \mu_1$, we have $\tilde{T}_{21} = \tilde{T}_{12}$; this is also proven from more general considerations, below. Hence, we rewrite our expression for the current (2.10) as

$$I_1 = \frac{e^2}{h}\tilde{T}_{12}(\mu_1 - \mu_2). \tag{2.12}$$

This is one form of the *Landauer formula*, which suggests that the effective conductance connecting two contacts is equal to $(e^2/h)\tilde{T}_{12}$. However, it should be noted that this is not the conductance of the device itself (see figure 3), since we do not know *a priori* what fraction of the applied potential $\mu_1 - \mu_2$ is actually dropped across the device, and what fraction is absorbed by contact resistance. The question of how the actual conductance of the device can be obtained was raised by Landauer (1957) in his pioneering paper, and has since been addressed by numerous authors (Büttiker *et al* 1985, Hu 1987, Eränen and Sinkkonen 1987, Jain and Kivelson 1988). It seems that there is no unique answer to the above question, for it depends on how the potential drop across the device is actually measured. This ambiguity has led to different versions of the Landauer formula (for a review, see Imry 1986a, Landauer 1970, 1987, 1988).

Experimental measurements of the conductance are usually performed using four-probe structures, rather than two-probe structures, in order to minimise the effect of contacts. The current is fed in through two probes and the voltage is monitored through a pair of probes in the middle (figure 5). For a while it was not clear how this four-probe conductance could be computed theoretically because of the ambiguity regarding what

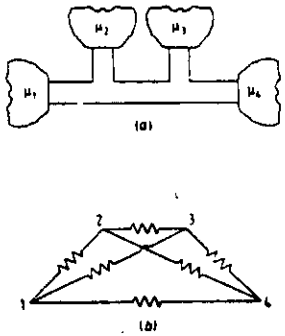


Figure 5. (*a*) A four-probe Hall bridge. (*b*) Equivalent resistor network *in the absence of magnetic fields.* (After Datta 1989a.)

the voltage probes measure. Büttiker (1988a, 1988b) found a simple and elegant solution to this problem. He noted that since there is really no qualitative difference between the current probes and the voltage probes in a Hall bridge, one could treat all the probes on an equal footing and simply extend the two-probe formula (2.10) by summing over all the probes

$$I_i = \frac{e^2}{h}\sum_{j\neq i}(\tilde{T}_{ji}\mu_i - \tilde{T}_{ij}\mu_j). \tag{2.13}$$

Note that for $i = j$, the summand is identically zero. We can thus drop the restriction from the summation and write,

$$I_i = \frac{e^2}{h}\sum_j(\tilde{T}_{ji}\mu_i - \tilde{T}_{ij}\mu_j). \tag{2.14}$$

We interpret $\tilde{T}_{ii}$ as the probability that an electron incident in probe i will be reflected back into the same probe (usually written as R_{ii}).

If there are no magnetic fields ($B = 0$), it can be shown that $\tilde{T}_{ij} = \tilde{T}_{ji}$. The Landauer–Büttiker formula (2.13) is then precisely what one obtains by applying Kirchhoff's laws to a network of resistors connecting each contact i and contact j through a conductance G_{ij} given by

$$G_{ij} = G_{ji} = \frac{e^2}{h}\tilde{T}_{ij} = \frac{e^2}{h}\tilde{T}_{ji}. \tag{2.15}$$

Thus, in the absence of magnetic fields, one can visualise mesoscopic systems in terms of an equivalent resistor network as shown in figure 2.3(b).

In the presence of a magnetic field, the Landauer–Büttiker formula is still valid, as verified by the derivation of Baranger and Stone (1989c). In general, the coefficients $\tilde{T}_{ij}$ have the following properties:

$$\tilde{T}_{ij}|_B = \tilde{T}_{ji}|_{-B} \tag{2.16}$$

and

$$\sum_i \tilde{T}_{ij} = \sum_i \tilde{T}_{ji} = 2M \tag{2.17}$$

where M is the number of transverse modes in the contacts, and the factor of 2 accounts for spins. The property (2.17) ensures that the currents are all zero when the electrochemical potentials are all equal. These properties are easy to prove if we assume phase-coherent transport throughout the device. One can then invoke S-matrix reciprocity (based on time reversibility arguments) to write

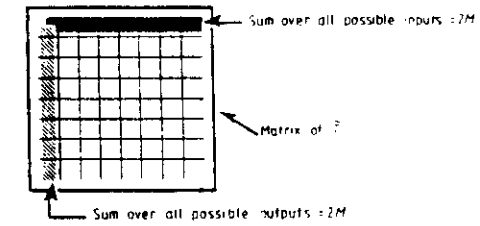
$$\tau_{ij}^{mn}(E)|_B = \tau_{ji}^{nm}(E)|_{-B}. \tag{2.18}$$

Using the definition of $T_{ij}(E)$ (2.5), we obtain

$$T_{ij}(E)|_B = T_{ji}(E)|_{-B}. \tag{2.19}$$

The desired symmetry property (2.16) follows readily by integrating out the energy dependence (2.11).

The first part of the normalisation property (2.17) is trivial—it merely states that the current entering lead j must be accounted for at all possible outputs. Consequently, if we write $\tilde{T}_{ij}$ in the form of a matrix, the columns must sum to $2M$ (figure 6(*a*)). The second part of equation (2.17), however, is rather subtle. It states that the total



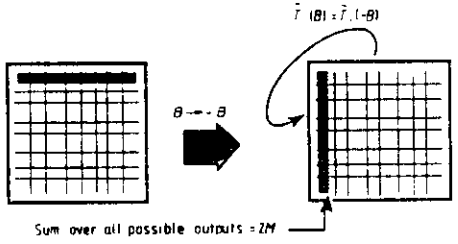


Figure 6. (*a*) Rows and columns of the matrix $\tilde{T}_{ij}$ must sum to $2M$. (*b*) Reversal of the magnetic field merely transposes the matrix.

current *entering* all of the various leads must equal that which finally exits through lead j; that is, the rows must also sum to $2M$. This is because under reversal of magnetic field, the rows transpose into columns (figure 6(*b*)), and the columns sum to $2M$ as before, due to current conservation. Proof of this property can be shown algebraically, as follows. We note that due to current conservation, which holds regardless of the magnetic field,

$$\sum_{i,m} \tau_{ij}^{mn}(E)|_B = \sum_{i,m} \tau_{ij}^{mn}(E)|_{-B} = 1. \tag{2.20}$$

Using the definition of $T_{ij}(E)$ (2.5)

$$\sum_i T_{ij}(E)|_B = \sum_i T_{ij}(E)|_{-B} = 2M \tag{2.21}$$

and the symmetry property shown above, (2.19), we obtain

$$\sum_i T_{ij}(E) = \sum_i T_{ji}(E) = 2M. \tag{2.22}$$

Again, the normalisation property (2.17) follows readily by integrating out the energy dependence (2.11). Note that this property has an interesting implication for two-lead geometries: $\tilde{T}_{12} = \tilde{T}_{21}$, regardless of the magnetic field.

These relationships, however, are not as easy to prove if we allow phase-breaking processes to occur within the device. In fact, when phase-breaking processes are included, our intermediate symmetry property (2.19) is no longer valid,

$$T_{ij}(E)|_B \neq T_{ji}(E)|_{-B},$$

as discussed at the end of section 2.1. However, Büttiker (1988a) has shown that the final symmetry property (2.16),

$$\tilde{T}_{ij}|_B = \tilde{T}_{ji}|_B,$$

is still valid, even in the presence of phase-breaking processes within the device. This is an important point, since phase-coherent transport is a theorist's idealisation that is never realised precisely. The symmetry property (2.16), on the other hand, can be used to prove general relationships like Onsager reciprocity in mesoscopic systems (Büttiker 1986a).

2.3. Transmission coefficients

To use any of the current equations ((2.7), (2.12) or (2.13)), we need to know the transmission coefficients. The problem of current flow is thus reduced to a scattering problem not unlike that encountered in, say, nuclear physics. It is as if the two contacts act as source and detector, with the device as the intervening medium (figure 3(b)); the problem then is to compute the scattering characteristics of this medium.

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

$$\left(\frac{(p - eA)^2}{2m^*} + eV\right)\Psi(r) = E\Psi(r). \tag{2.23}$$

Here $A(r)$ and $V(r)$ are the vector and scalar potentials within the device. The scalar potential $V(r)$ includes band-bending due to externally applied fields and *space charge effects*, band-edge discontinuities due to heterojunctions, and microscopic fields due to elastic scatterers such as defects or impurities. In the absence of magnetic fields, the vector potential A may be set equal to zero, so that we can simplify the Schrödinger equation (2.23) as follows:

$$\nabla^2\Psi(r) = -\frac{2m^*}{\hbar^2}(E - V(r))\Psi(r). \tag{2.24a}$$

This equation is very similar to Maxwell's equation used in integrated optics (assuming $\mathscr{E} \cdot \nabla\varepsilon = 0$ for simplicity),

$$\nabla^2\mathscr{E} = -\omega^2\mu\varepsilon(r)\mathscr{E}. \tag{2.24b}$$

Here $\mathscr{E}$ is the electric field, ω is the radian frequency, μ is the permeability and ε is the spatially varying dielectric constant. Comparing these two equations (2.24*a*, *b*) it is evident that electron waves moving through a medium with a spatially varying potential $V(r)$ are analogous to light waves moving through a medium with a varying dielectric constant (or refractive index). The analogies between these two types of waves are listed in table 1. Most of the phenomena we discuss in section 3 (except those involving magnetic fields) have familiar optical analogies which we will mention as we go along.

One question that might bother the reader regarding this analogy is the fact that electrons are fermions while photons are bosons. This difference is not important, so long as we are discussing one-particle phenomena where every electron or photon *interferes with itself*. The electric field in Maxwell's equations can then be viewed as

Table 1. Analogies between electron waves and electromagnetic waves.

Electron waves	↔	EM waves
$\Psi(r,t) = \psi(r)e^{-iEt/\hbar}$	↔	$E(r,t) = \mathcal{E}(r)e^{-i\omega t}$
Energy E	↔	Frequency ω
Confined channels	↔	Waveguides
Sub-bands	↔	Transverse modes
Ψ	↔	E
$\nabla\Psi$	↔	H
Charge density	↔	Energy density
Current density	↔	Poynting vector
$\nabla^2\psi = -\frac{2m^*}{\hbar^2}(E-V)\psi$	↔	$\nabla^2\mathcal{E} = -\omega^2\mu\varepsilon\mathcal{E}$

the wavefunction of a single photon, and the analogy with the Schrödinger equation seems complete. But if we view the electric field (as we usually do) as the macroscopic field due to a coherent state with billions of photons, then there is no analogous state known with normal electrons. (However, the superconducting state is analogous to the coherent state of light, and the Josephson effect is a well known manifestation of the macroscopic wavefunction of superconducting electrons; we will not discuss this further.)

In view of these analogies between electrons and photons, one might wonder why the Boltzmann picture works at all. Why aren't quantum interference effects more common? One of the chief reasons is the existence of *phase-breaking* scattering processes that destroy interference phenomena. (In fact, a Boltzmann-like transport equation is also used to describe the radiative transfer of photons through stellar atmospheres.) A phase-breaking scattering process is one in which the scatterer changes its internal state. As a result, successive electrons, encountering the scatterer in different states, suffer different phaseshifts, thus wiping out any stationary interference patterns. Another way to view a phase-breaking interaction is as a *measurement* process. By monitoring the state of a scatterer, one can gain information regarding the path of the electron between the source and the detector (Stern *et al* 1989). A well known principle in quantum interference phenomena is that any process yielding information regarding which of the various interfering alternatives was actually taken, tends to destroy the interference. In general, however, it may be difficult to distinguish between the 'system' and the 'environment', and there are subtle unresolved issues in providing a general definition for what constitutes a phase-breaking process (Leggett 1989).

Elastic scattering by the sample boundaries or by defects and impurities plays an important role in determining mobility; but it is not phase-breaking, since the scatterer has no internal degree of freedom and remains unaffected by the process. But inelastic scattering by phonons or by other electrons is phase-breaking. The phase-coherence time τ_ϕ usually increases significantly as we go to lower temperatures, because electron–electron scattering processes are suppressed. Such processes do not contribute to the mobility, since the momentum of the electronic ensemble is unchanged; any momentum lost by one electron is picked up by another. Consequently, the mobility is nearly constant at temperatures below, say, 10 K—once the phonons are frozen out. But the phase-breaking time is orders of magnitude larger at 0.1 K than it is at 10 K. Thus, although high-mobility films are certainly desirable for quantum devices, the mobility is in general not a reliable indicator of the phase-coherence time.

A phase-coherence length as long as 1–10 μm is not uncommon at a temperature of 1 K, but it gets significantly shorter at higher temperatures and for hot electrons. Phase-breaking processes are thus inevitably present in most devices at reasonable temperatures and bias levels, and the assumption of phase-coherent transport (section 2.3) is often inaccurate.

A more compelling reason to allow for dissipative processes has recently been explained by Landauer (1989b). Recall that the chemical potentials in the Landauer–Büttiker formula (2.13) are measured deep within each contact reservoir, where electrons have relaxed to local equilibrium. Strictly speaking, the transmission probabilities $\hat{T}_{ij}$ should be computed between these points—between points deep within each reservoir. The transmission probabilities, however, are usually computed over a small portion of the device where dissipation can be neglected. Thus, the transition regions connecting the deepest points of the reservoir to the active device area are completely ignored. Inherently, these regions must contain dissipation, so that electrons entering the reservoir will finally relax to local equilibrium.

The question of how phase-breaking processes can be included in the transmission coefficients $\hat{T}_{ij}$ is at the forefront of current research. If transport is perfectly coherent, then the transmission coefficients can be computed from elementary quantum mechanics. At the other extreme, if phase-breaking processes are so frequent that one can assume totally incoherent transport, then semiclassical Monte Carlo simulation can be used to compute the transmission coefficients; this is equivalent to solving the Boltzmann equation (1.2). However, in the middle ground where transport is partially coherent, there are no simple answers, as yet.

3. Quantum effects in electron transport

In this section, we will briefly survey various quantum effects that have been observed in semiconductor microstructures. These effects can broadly be divided into two categories: those involving devices whose transverse dimensions are either much longer or much shorter than the phase-coherence length. In devices belonging to the first category, the sub-band energies ε_n are nearly continuous, and one can describe electron transport in terms of plane waves. In those belonging to the second category, it is more appropriate to view transport in terms of discrete waveguide modes. Vertical devices (figure 2(*a*)) usually have cross sectional dimensions on the order of several micrometres, so that transport can be described in terms of plane waves; however, with advances in nanolithography, vertical devices with submicrometre cross sections are also being investigated (Reed *et al* 1988). Lateral devices (figure 2(*b*)) are formed by lithographically patterning a modulation-doped heterostructure, where current flow is confined to a two-dimensional electron gas (2DEG). Therefore, one of the transverse dimensions is always sufficiently small that only a few modes (commonly referred to as 'sub-bands') are involved. We will classify such devices under plane wave transport if the other transverse dimension is large, and under guided wave transport if it is small.

3.1. Plane wave transport

3.1.1. Single-barrier devices. The classic example of a quantum device is the tunnel diode, which exhibits negative differential resistance (NDR) due to Zener tunnelling of electrons through the band gap in a heavily doped p–n junction (Esaki 1958, 1974).

Recently, NDR has also been reported in single-barrier heterostructures using material systems HgCdTe-CdTe-HgCdTe (Chow *et al* 1988) and InAs-AlGaSb-InAs (Beresford *et al* 1989, Söderström *et al* 1989). NDR arises in these material systems because within the barrier, the tunnelling electrons have energies closer to the valence band edge than to the conduction band edge (figure 7). Initially the tunnelling probability is high due to the proximity of the valence band states. As the bias is increased, the energy of the tunnelling electrons shifts towards the middle of the band gap where the tunnelling probability is lower; the current therefore decreases. NDR has also been observed in GaAs-AlAs-GaAs heterostructures due to a very different mechanism (Mendez *et al* 1987, Landheer *et al* 1989). AlAs appears as a potential barrier for electrons in the Γ-valley in GaAs, but as a potential well for electrons in the X-valley in GaAs. NDR arises due to electrons in the three-dimensional contacts tunnelling into and out of the quantised two-dimensional energy levels in the X-valley in AlAs. This is similar to the resonant tunnelling effect, which is described in section 3.1.2.

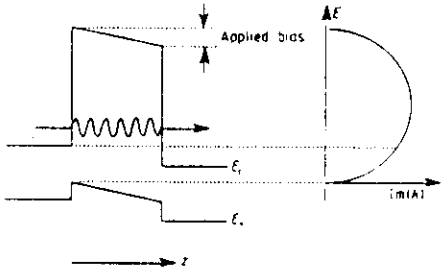


Figure 7. Close proximity of electrons to the valence band states can produce a large probability of tunnelling which decreases at higher applied biases. (After Söderström *et al* 1989.)

Most of the effects that we address in this review arise from interference phenomena due to the wave nature of electrons. A somewhat different type of quantum effect has recently attracted much attention. This effect, known as 'Coulomb blockade', arises from the discrete particulate nature of the electronic charge. The transfer of a single electron through a small tunnel junction drastically reduces the voltage applied to the junction. As a result, current is inhibited at small biases, since the change induced in the voltage would be larger than the existing bias (for a review, see Likharev 1988).

3.1.2. Double-barrier devices. Next to the tunnel diode, the most well known quantum device is the resonant tunnelling diode (Esaki and Tsu 1970, Tsu and Esaki 1973, Chang *et al* 1974, Esaki 1986, Capasso *et al* 1986). This device consists of two barriers in series as shown in figure 8(*a*). It is often compared to the Fabry-Perot interferometer used in optics; the two barriers play the role of partially silvered mirrors that form a resonant cavity. The transmission coefficient shows sharp peaks as a function of the longitudinal kinetic energy, as one would expect for a resonant cavity. The device thus acts as an energy filter which only allows incident electrons with certain discrete values of the longitudinal kinetic energy to go through to the other contact. An applied

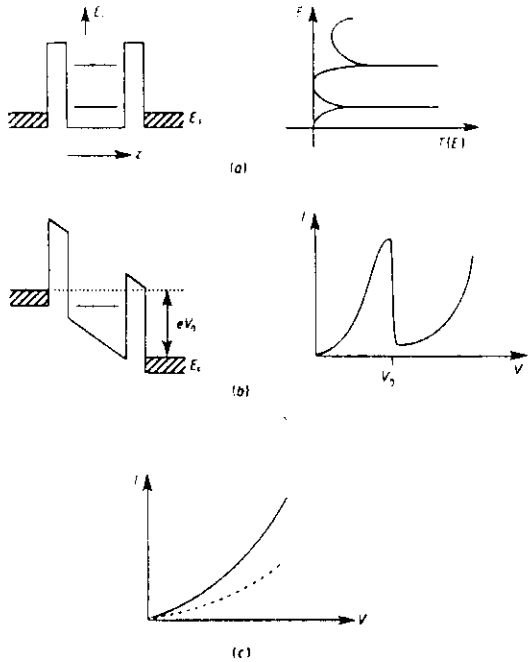


Figure 8. Resonant tunnelling diode: (*a*) energy band diagram and transmission coefficient versus energy; (*b*) band diagram under bias and *I-V* characteristics; (*c*) expected *I-V* for one barrier (full curve) and for two barriers in series (broken curve).

bias lowers the resonant energy relative to the energy of the incident electrons. When the resonant energy falls below the conduction band edge in the emitter, there is a sharp drop in the current leading to negative differential resistance (NDR) as shown in figure 8(*b*).

When connected to a resistive load, a resonant tunnelling diode will settle into one of two stable states found by drawing a load line onto the I-V characteristic in figure 8(*b*); this *extrinsic* bistability is useful for the design of logic circuits based upon this family of devices. Additionally, there is the phenomenon of *intrinsic* bistability (Zaslavsky *et al* 1989, Alves *et al* 1989), which gives rise to a hysteresis in the *I-V* characteristics. This can be understood as follows. When the voltage is swept up, one approaches the NDR region with a filled well; the stored charge in the well tends to raise the resonant energies with respect to the emitter, so that a higher applied bias is needed to shut off transmission through the lowest level. When the voltage is swept down, however, one approaches the NDR region with an empty well, and the transition occurs at a lower applied bias. Thus, the *I-V* characteristic shows a hysteresis around the NDR region. This is analogous to the well known bistability in non-linear optical Fabry-Perot interferometers, which arises from the build-up of light intensity between the two mirrors. This result has also been observed in theoretical simulations that

solve the Schrödinger equation (2.24*a*) self-consistently with the Poisson equation (Kluksdahl *et al* 1989, Mains *et al* 1989).

It will be noted that the resonant tunnelling NDR is a quantum effect that cannot be understood within a semiclassical framework. It is easy to show that for a single tunnel barrier the current increases monotonically with voltage, as shown in figure 8(*c*). If we view electrons as particles, we would expect a double barrier to act like two single barriers in series. We would thus expect the current to be half that of a single barrier for a given voltage as shown by the broken curve in figure 8(*c*), in sharp contrast to what is actually observed (figure 8(*b*)). Vertical structures, where the potential barriers are formed by layers of materials with different band gaps, have reached a high level of maturity, reporting peak-to-valley ratios as high as 30:1 (Broekaart *et al* 1988). By contrast, lateral structures, with the potential barriers imposed through a dual gate, have only recently been reported (Chou *et al* 1989, Ismail 1989a). Because the height of the barriers can be controlled simply by changing the gate voltage, these structures are promising tools for the study of resonant tunnelling phenomena.

From an applied point of view, resonant tunnelling devices have generated a great deal of interest as possible candidates for high-frequency and fast switching applications. Resonant tunnelling diodes have been used as mixers up to 1.8 THz and as oscillators up to 400 GHz (Sollner *et al* 1983, 1988). Switching operations with a rise time as short as 2 ps have been reported (Whitaker *et al* 1989). For practical applications it is often more convenient to have a three-terminal device, and a number of schemes have been proposed and demonstrated for building transistors based on resonant tunnelling phenomena. The extra terminal can be used to control the resonance spectrum of the quantum well. In lateral structures, this is done simply by changing the barrier height through the applied gate voltage (Chou *et al* 1989, Ismail 1989a). In vertical structures, this has been accomplished through the Stark effect (Beltram *et al* 1988) and through direct contact to the quantum well (Reed *et al* 1989). Resonant tunnelling structures have also been incorporated into the bases and emitters of bipolar transistors, where they modify the properties of the conventional device to include negative transconductance (Capasso 1989a).

Resonant tunnelling as a phenomenon is also useful as a probe in studying basic transport phenomena. For instance, the energy filtering characteristic of double-barrier structures has been used to do a spectroscopic analysis of transport through the base of a bipolar transistor (Berthold *et al* 1989). Sharp peaks have been observed in the conductance of narrow metal-oxide-semiconductor field-effect transistor (MOSFET) channels, as the Fermi level is adjusted through a gate voltage. This structure in the conductance has been ascribed to resonant transmission through localised levels, and thus reveals information regarding the energy spectrum of such levels (Azbel 1983, Fowler *et al* 1988).

3.1.3. Multiple-barrier devices. It would seem that one could come up with some interesting structures by considering more than two barriers. However, electron transport in multiple-barrier structures has been investigated far less that in single- or double-barrier devices. One interesting aspect of such structures is the possibility of engineering strongly asymmetric device characteristics. For example, the structure shown in figure 9(*a*) has two quantum wells whose resonant energy levels are misaligned under zero bias. Under forward bias, the levels align (figure 9(*b*)), and one would expect a large transmission through the structure. Under reverse bias, on the other

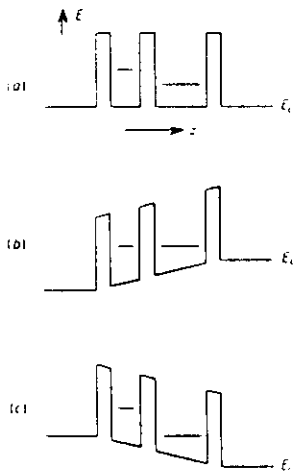


Figure 9. (*a*) In equilibrium, the resonant levels of two neighbouring quantum wells are misaligned. The levels are (*b*) aligned and (*c*) separated as the device is biased with different polarities.

hand, the levels are separated (figure 9(*c*)), and transmission would be greatly reduced. Indeed, diode-like characteristics have been demonstrated using a series of quantum wells with graded widths (Kirchoefer *et al* 1985).

3.1.4. Superlattices. In the preceding section, we showed that the irregularity of a multiple-barrier structure could provide a diode-like *I–V* characteristic. At the other extreme, we might consider a perfectly regular array of identical wells and barriers, called a 'superlattice' (Esaki and Tsu 1970). If we consider only three periods of such a structure, we observe that the peaks in transmission (associated with the resonant energies of each individual quantum well) are split (figure 10(*a*)). For three barriers, each transmission peak is split into two; for four barriers, each peak is split into three, and so on (Tsu and Esaki 1973). Assuming that the barriers are thin enough so that electrons can easily tunnel from one well to the next, the finely split energy levels in an infinite superlattice broaden to form minibands, as shown in figure 10(*b*). Note that it is the periodicity of a superlattice that leads to the formation of minibands and minigaps—the detailed shape of the potential merely determines the final band structure. Indeed, by cleverly designing the potential, the characteristics of the band structure can be modified (Peeters and Vasilopoulos 1989).

Electrons moving in a superlattice miniband are decelerated by the applied field as they reach the negative mass region at the top of the and; this results in NDR. Of course, electrons will only reach the negative mass region if the scattering time is moderately long. An interesting phenomenon occurs when the scattering time is extremely long: electrons accelerated to the top of a miniband ($k = +\pi/a$) wrap around to the opposite side ($k = -\pi/a$) of the mini-Brillouin zone (figure 10(*b*)), where their velocity changes sign; the electrons thus oscillate back and forth in real space. The concept of a 'Bloch oscillator' is based on the possibility of extracting RF power from

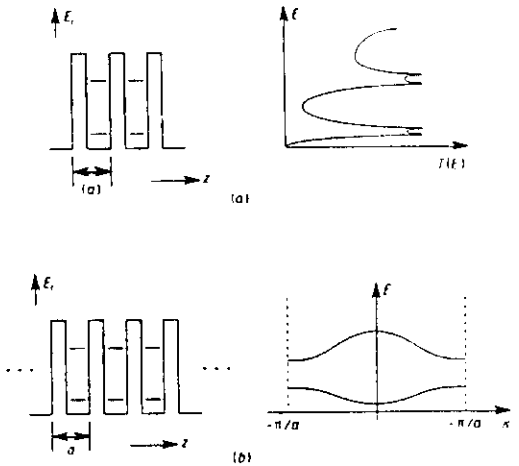


Figure 10. (*a*) Transmission through a three-barrier structure shows a split peak at the resonant energies of the individual quantum wells. (*b*) In a superlattice, the finely split transmission peaks broaden to form minibands.

these oscillations (Esaki and Tsu 1970). Although this is an intriguing concept, the effect is difficult to demonstrate, since for realistic scattering times the electric field required is rather large.

The presence of a strong electric field leads to other complications which can destroy the effect. For instance, electrons may acquire enough energy to tunnel into the next highest miniband, an effect known as Zener tunnelling. More importantly, the entire band picture of transport breaks down at sufficiently strong fields. The resonant energies of neighbouring quantum wells become misaligned, and electrons become localised within the wells. This in itself can lead to NDR, since current must then be carried by electrons hopping from well to well, a process which has a much lower mobility than miniband transport (Dohler *et al* 1975).

Lateral surface superlattice devices have recently been demonstrated, which attempt to harness the effects described above. Originally proposed by Sakaki *et al* (1976) and Bate (1977), these devices consist of a patterned gate which induces the superlattice potential on the surface of a 2DEG, as shown in figure 11. Current flows between a source and a drain positioned on either side of the superlattice; thus, the experimental geometry is simply a modulation-doped field-effect transistor (MODFET) with an unusual gate. The 'washboard transistor', whose gate is patterned with a 1-D grating, is one implementation of this (Tsubaki *et al* 1989). Experiments performed on Si metal-oxide-semiconductor (MOS) devices (Warren *et al* 1985) and GaAs/AlGaAs MODFET devices (Ismail *et al* 1988) have shown a modulation of the transconductance. Stronger effects, including negative differential resistance and negative transconductance, have been reported for structures having 2-D grids patterned into the gate (Bernstein and Ferry 1987, Ismail *et al* 1989). Presumably, this is due to the improvement in minigap widths caused by the additional degree of confinement.

Novel oscillations in the magnetoresistance, superimposed on the usual Shubnikov-de Haas oscillations, were recently reported in a lateral structure with a one-dimensional

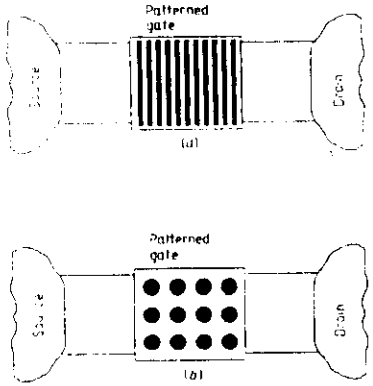


Figure 11. (*a*) One-dimensional and (*b*) two-dimensional lateral surface superlattices, imposed through a patterned gate electrode.

periodic potential (Gerhardts *et al* 1989, Winkler *et al* 1989). Shubnikov-de Haas oscillations in the magnetoresistance are periodic in $1/B$, having a period proportional to the carrier density. The novel oscillations recently observed are also periodic in $1/B$; however, the period is proportional to a/k_F, where a is the period of the superlattice potential and k_F is the Fermi wavevector.

3.1.5. Weak localisation. An interesting question to ask is whether one can observe any quantum interference effects from a random array of elastic scatterers, such as an ordinary resistor, at low temperatures. The intuitive answer is that due to the randomness, any interference effect would cancel out on the average. This, however, is not true. Quantum interference leads to enhanced backscattering, which can be demonstrated by computing the conductance of an array of scatterers from the Landauer-Büttiker formula (2.12). This calculation has been performed using two different models to obtain the transmission coefficient $\tilde{T}_{12}$, to clearly illustrate the effects of phase coherence (Cahay *et al* 1988). One is the coherent or the quantum model in which the *amplitude* scattering matrices of the individual scatterers are combined, taking phases into account. The other is the incoherent or the semiclassical model in which the *probability* scattering matrices are combined, taking no account of the phases.

Figure 12 shows the results obtained for both the quantum and the semiclassical model as the location of one scatterer in the array (the middle one) is changed. The semiclassical result is unaffected by this change, but the quantum result shows fluctuations due to the changing interference patterns. These conductance fluctuations have been observed experimentally in mesoscopic samples; however, in larger samples such fluctuations cannot be observed. This is because a large sample is basically an ensemble of uncorrelated units each having dimensions of the order of a phase-coherence length. When making measurements on large samples, one measures ensemble-averaged quantities due to this self-averaging feature.

But the interesting point to note from figure 12(*b*) is that the mean (or ensemble-averaged) value of the quantum conductance is *less* than the semiclassical conductance. This shows that interference causes an enhancement in the average backscattering from

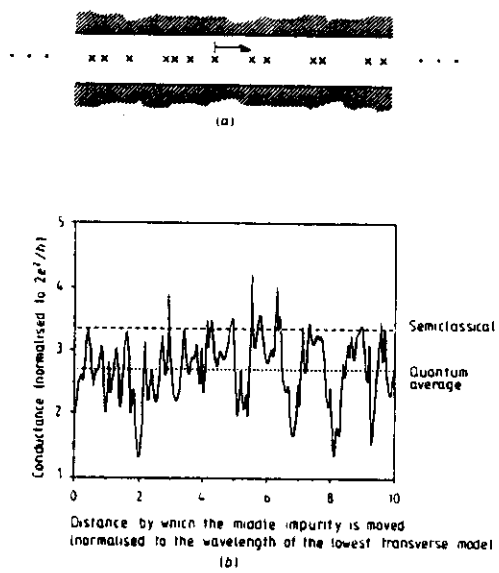


Figure 12. (*a*) Part of a random sample with fixed arrangement of static scatterers. (*b*) Conductance of the sample computed as a function of the position of the middle scatterer, keeping the rest of the array fixed. (After Datta 1989a.)

an array of scatterers—a phenomenon that has been observed with electromagnetic waves as well. Weak localisation, as this effect has come to be known, was the subject of extensive investigation in the early 1980s (recent review articles include Al'tshuler and Lee 1988, Bergmann 1984, Lee and Ramakrishnan 1985). It can be understood simply as follows. For each diffusive path that a reflected electron can follow, there exists a time-reversed path with an identical amplitude and phase. Time-reversed pairs of paths interfere constructively to produce the enhanced backscattering. If the time-reversal symmetry is broken by a magnetic field (tens of gauss is sufficient), the enhanced backscattering is destroyed. Consequently, weak localisation is characterised by a ***negative magnetoresistance***—a magnetic field causes the resistance to decrease from its quantum to its semiclassical value. Because the effect depends on phase coherence, this magnetoresistance measurement is one of the common techniques for measuring the phase-coherence time.

3.1.6. Spin interference. The spin of an electron, which is analogous to the polarisation of electromagnetic waves, usually does not play any significant role in transport processes; it merely doubles the number of states leading to a multiplicative factor of 2. However, non-trivial effects can arise in materials with strong spin–orbit coupling. The phenomenon of weak antilocalisation is a well known example of such an effect (Bergmann 1982). In the presence of spin–orbit coupling, an electron following a diffusive path will experience some overall rotation of its spin. For strong spin–orbit coupling, the final orientation of the spin is statistical. Rotations for time-reversed paths, however, are exactly the opposite of one another, and on the average are

separated by 2π. As a result, the time-reversed paths interfere *destructively* (rotation by 4π leaves a spinor unchanged), so that the conductance is enhanced relative to the semiclassical value. In this case, an applied magnetic field destroys the enhanced transmission, so that the magnetoresistance is *positive*. Measurements of this effect have been used to determine the strength of spin–orbit coupling in semiconductors such as InAs (Yamaguchi 1985).

It is possible that spin phenomena could be harnessed for device applications. In a manner similar to weak antilocalisation, for instance, interference between two paths with different spin rotations could be used to modulate current flow. A single path could also produce current modulation if spins were preferentially injected and detected. Indeed, a device based upon spin injection and detection has recently been proposed (Datta and Das 1990). Although there is, as yet, very little experimental work in this area and the idea is purely speculative, we will explore this concept in greater detail below to emphasise the relationship between electron spin and light polarisation.

In the electro-optic light modulator, input light is polarised at 45° in the y–z plane (figure 13(*a*)), which can be represented as a linear combination of z- and y-polarised light,

$$\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}. \qquad (3.1)$$

(45°-pol.) (z-pol.) (y-pol.)

As this light passes through the electro-optic material, the two normal modes suffer different phaseshifts $k_1 L$ and $k_2 L$, since the electro-optic effect makes the dielectric

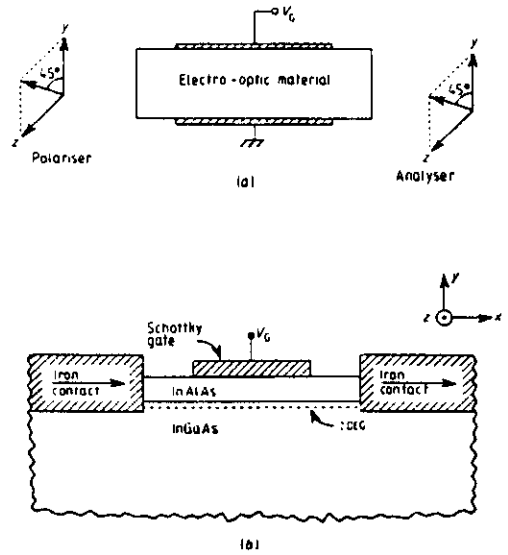


Figure 13. (*a*) An electro-optic modulator. (*b*) Proposed electron wave analogue of the optical device shown in (*a*). (After Datta and Das 1990.)

constant ϵ_{zz} slightly different from ϵ_{xx}. This causes the polarisation angle to precess as a function of distance travelled through the dielectric. Only the component along the original polarisation is allowed to pass through the analyser at the output, so that the output power P_0 is given by

$$P_0 \propto \left| (1\ 1) \begin{pmatrix} e^{ik_1 L} \\ e^{ik_2 L} \end{pmatrix} \right|^2 = \sin^2 \frac{(k_1 - k_2)L}{2}. \tag{3.2}$$

Thus, the light output is modulated with a gate voltage that controls the differential phaseshift $\Delta\theta = (k_1 - k_2)L$.

The analogous device based on electron waves is shown in figure 13(*b*). The polariser and analyser can be implemented using contacts made of a ferromagnetic material like iron. At the Fermi level in such materials the density of states for electrons with one spin greatly exceeds that for the other, so that the contact preferentially injects and detects electrons with a particular spin. Spin current polarisation up to -50% has been experimentally demonstrated using permalloy contacts (Johnson and Silsbee 1985), although further work in this area is needed. A contact magnetised in the x-direction preferentially launches and detects electrons spin-polarised along positive x which is represented as a linear combination of positive z-polarised and negative z-polarised electrons,

$$\underset{(+x\text{-pol.})}{\begin{pmatrix} 1 \\ 1 \end{pmatrix}} = \underset{(+z\text{-pol.})}{\begin{pmatrix} 1 \\ 0 \end{pmatrix}} + \underset{(-z\text{-pol.})}{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}. \tag{3.3}$$

Narrow-gap semiconductors like InGaAs provide the analog of an electro-optic material which will introduce a differential phaseshift between $+z$-polarised and $-z$-polarised electrons. It has been established both theoretically and experimentally that in 2DEGS in narrow-gap semiconductors there is an energy splitting between up-spin and down-spin electrons even when there is no magnetic field (Lommer *et al* 1988, Das *et al* 1989). The dominant mechanism for this 'zero-field spin-splitting' is believed to be the Rashba term in the effective mass Hamiltonian (Bychkov and Rashba 1984),

$$H_R = \eta(\sigma_z k_x - \sigma_x k_z) \tag{3.4}$$

which should be adjustable with a gate voltage through the spin-orbit coefficient η, though this has not been investigated experimentally. It should be mentioned that large effects can only be obtained if the width of the device in the z-direction is small, so that only a few transverse modes are involved, because the phaseshift is different for different modes. From this point of view, this device concept belongs in the next section along with other few-moded devices.

3.2. *Waveguide transport*

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

$$k_F = (2\pi n_s)^{1/2}. \tag{3.5}$$

Assuming a rectangular channel, we expect the transverse momentum to be quantised in multiples of π/W, so that the number of modes below k_F is approximately

$$M \sim \frac{k_F W}{\pi} \sim n_s^{1/2} W. \tag{3.6}$$

Thus a 0.5 μm wide channel with an electron density of 6.4×10^{11} cm^{-2} will have approximately 40 transverse modes. However, wires less than about 0.5 μm in physical width usually do not conduct, because the Fermi level is pinned near the exposed sidewalls, leading to fairly wide depletion layers which constrict the channel (Demel *et al* 1988). It is thus extremely difficult to control the number of modes in a wire without making it totally non-conducting.

The depletion layer width can be reduced by using a shallow etch, whereby the sidewalls of the modulation-doped GaAs channel are not exposed; only the top AlGaAs layer containing the dopants is partially etched. The shallow mesa also helps reduce surface effects from degrading the channel mobility (van Houten *et al* 1986). This is the technique used by Timp and co-workers in their pioneering work on electron waveguide transport (Timp *et al* 1987b). Since then, a variety of techniques for channel definition have been used by other groups, such as selective ion etch damage (Roukes *et al* 1988), helium ion beam damage (Cheeks *et al* 1988), strain-induced confinement (Kash *et al* 1989) and electrostatic confinement with a split gate (Wharam *et al* 1988, van Wees *et al* 1988a). These developments have led to tremendous activity in semiconductors since 1987, though most of the work on mesoscopic systems originated in metals (Washburn and Webb 1986, Webb and Washburn 1988).

3.2.1. Conductance fluctuations. Until the early 1980s it was thought that quantum effects could not be observed in samples with a large number of occupied modes; each mode was expected to contribute its own unique but independent part to the conductance, so that metals, for instance, were expected to be completely self-averaging. Simple, low-temperature resistance measurements on narrow metal wires, however, presented quite a different story (Washburn 1988 and references therein). The conductance was found to fluctuate in an applied magnetic field, and the fluctuations ($\sim e^2/h$) were completely *reproducible*. It is now understood (Stone 1985) that in regions smaller than a phase-coherence length the transmission coefficients of the various modes are indeed correlated within some characteristic energy range. For either high temperatures or large sample sizes, however, the effect is self-averaging: The overall conductance represents the independent contributions of small blocks the size of a phase-coherence length, or of energy intervals the size of an energy correlation length. For N such contributions, the effect will be reduced by $\sqrt{N}$.

In earlier discussion, we showed that the same conductance fluctuations ($\sim e^2/h$) can be obtained by moving a single impurity in an array of scatterers (figure 12). A similar result has been observed experimentally in silicon inversion layers. Fluctuations in the conductance were found to switch back and forth between two separate patterns as a single electron trap changed its occupancy (Skocpol *et al* 1986). In effect, the presence or absence of a single impurity completely changed the pattern of fluctuations. It is believed that changing the magnetic field is equivalent to changing the configuration of impurities (Lee *et al* 1987). Conductance fluctuations in a magnetic field have come to be known as 'magnetofingerprints' and can, in principle, be used to identify mesoscopic samples.

One of the puzzles in the early days of mesoscopic physics was the fact that the conductance fluctuations measured with a four-probe Hall bridge (see figure 5(*a*)) were not symmetric in the magnetic field (Benoit *et al* 1986). In a large rectangular Hall bridge, one measures ρ_{xx} directly, which *is* symmetric in a magnetic field ($\rho_{xx}(B) = \rho_{xx}(-B)$, Onsager relation). But in a mesoscopic Hall bridge the voltage drop is not uniform, and one measures some combination of ρ_{xx} and ρ_{xy}, which *is not* symmetric in B. Thus, one should regard mesoscopic samples as inhomogeneous conductors that obey the reciprocity relationship

$$R_{mn,kl}(B) = R_{kl,mn}(-B) \tag{3.7}$$

where $R_{mn,kl}$ is the resistance obtained by feeding a current between terminals m and n and measuring a voltage between terminals k and l. Using a well known symmetry property of the coefficients $\bar{T}_{ij}$,

$$\bar{T}_{ij}|_B = \bar{T}_{ji}|_{-B} \tag{3.8}$$

Büttiker (1986a) has shown that the four-probe resistances obtained from the Landauer–Büttiker formula (2.13) indeed obey this reciprocity relationship (3.7).

3.2.2. Quasi-ballistic wires. A study of quasi-ballistic channels—narrow Hall bridges having widths shorter than, and lengths much longer than, the elastic scattering length—has revealed novel features in the magnetoresistance spectra. These features, sketched in figure 14(*a*), arise from the influence of a narrow confining potential (defining the 'walls' of the waveguide), and so can be used to experimentally characterise the electrical widths of wires. In all measurements discussed below, we assume a magnetic field applied perpendicular to the plane of the 2DEG.

At low fields, a pronounced negative magnetoresistance appears due to the destruction of weak localisation (van Houten *et al* 1988, Taylor *et al* 1988). In quasi-ballistic channels, this effect persists to much higher fields than it normally would for a wide Hall bridge. It has been suggested (van Houten *et al* 1988) that this is because of specular scattering from the walls of the channel. Electrons in a quasi-ballistic wire move in a zig-zag fashion, being reflected back and forth by the confining potential, and travelling up and down the length of the wire. As a result, the time-reversed pairs of paths, which are central to the discussion of weak localisation, are much more difficult to visualise. These paths, envisioned as irregular loops in an unrestricted sample, are folded back onto themselves again and again, in order to fit within the narrow channel (figure 14(*b*)). Each of these conceptual folds decreases the total flux enclosed by the paths. Consequently, it requires higher fields to achieve an enclosed flux comparable to that which destroys weak localisation in a wide Hall bridge.

As the magnetic field is increased, the resistance first exhibits aperiodic fluctuations identical to those observed in metal wires, as described earlier. At some point, Shubnikov–de Haas (SdH) oscillations begin to appear. For a wide Hall bridge, SdH oscillations arise from the motion of Landau levels through the Fermi energy. Each trough in the resistance indicates that an integral number of Landau levels is filled. If the troughs are numbered and plotted versus $1/B$, the resulting plot is linear, and the slope of the line can be used to determine the carrier density (Ando *et al* 1982). This is because in a wide Hall bridge, the energies of each level are determined from the magnetic confinement alone—any influence of the experimental geometry can be ignored. In a quasi-ballistic channel, however, the influence of the narrow cross section cannot be ignored. At low fields, electrons feel an electrostatic confinement in addition

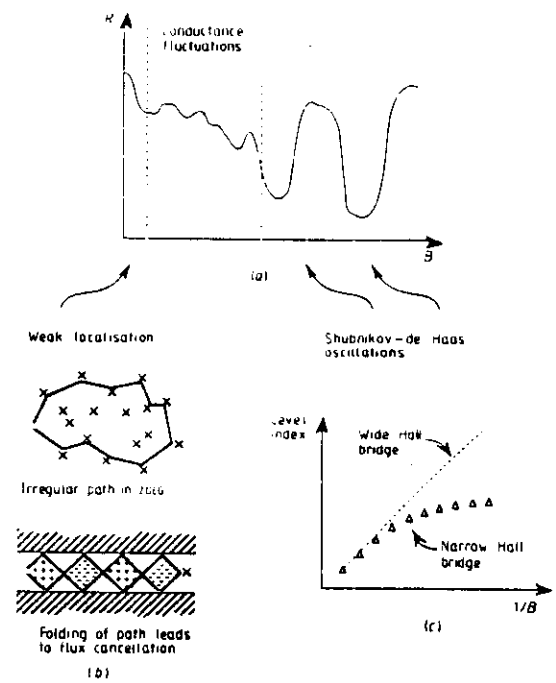


Figure 14. (*a*) Sketch of the magnetoresistance of a narrow Hall bridge. (*b*) Specular scattering from the walls of the channel causes flux cancellation in the time-reversed paths which produce weak localisation. (After van Houten *et al* 1988.) (*c*) Shubnikov–de Haas oscillations for a narrow Hall bridge are not periodic in $1/B$. (After Berggren *et al* 1988.)

to the magnetic confinement, which perturbs the level spacings. As a result, a plot of the level index (trough number) versus $1/B$ is highly non-linear (figure 14(*c*)). Only at very high fields is the linear behaviour recovered. Because this measurement is sensitive to the shape of the confining potential, it can be used to characterise the electrical width of fabricated channels (Berggren *et al* 1988).

To this point, we have focused on measurements of the longitudinal resistance ρ_{xx}; the transverse resistance ρ_{xy} also shows unexpected behaviour in the presence of a magnetic field. From elementary considerations, this Hall resistance is expected to increase linearly with magnetic field: $\rho_{xy} = B/n_s e$. Experiments performed on narrow Hall bridges have shown, however, that instead of being linear in the magnetic field, this resistance drops to zero within a small range around $B = 0$ (Roukes *et al* 1987, Ford *et al* 1988). Subsequent theoretical work (Baranger and Stone 1989b, Imry 1989, Kirczenow 1988) has shown that this 'quenching' of the Hall resistance is *not* an intrinsic property of narrow channels. Rather, it is caused by the inability of the voltage leads to couple to the sample.

Three different mechanisms have been identified which contribute to this poor coupling. The most important of these (Baranger *et al* 1990) is 'forward enhancement'

or 'collimation'. In a gently flared junction, electrons tend to remain in the same mode without scattering, so that as the channel widens, transverse momentum is gradually converted into the longitudinal direction. As a result, electrons form a collimated beam which simply passes the voltage lead without scattering into it. Such collimating behaviour has been demonstrated using both waveguide (Baranger and Stone 1989b) and billiard ball (Beenakker and van Houten 1989) models.

Thus, even though the Hall voltage exists, it cannot be measured by the flared voltage leads. If on the other hand, junctions with sharp corners could be fabricated—or if some other means of measuring the voltage were used—the usual Hall resistance would be observed. This underscores the impact that device geometry can have in the mesoscopic regime. More than that, it foreshadows a concern which we discuss further in section 3.2.5: every bend or kink within a phase-coherence length of the active device area can profoundly influence what is finally measured.

3.2.3. Ballistic wires. A striking demonstration of mode quantisation in electron waveguides was provided in a recent experiment. The conductance of a ballistic channel or 'point contact', whose width was reduced continuously through a split-gate structure (figure 15(*a*)), was seen to decrease by finite steps (Wharam *et al* 1988, van Wees *et al* 1988a). For a ballistic channel, an incident flux in any mode n is completely transmitted ($\tau^{mn}(E) = \delta_{mn}$) so that from the definition of the overall transmission (2.5)

$$T_{21}(E) = T_{12}(E) = 2M \tag{3.9}$$

where M is the number of modes and the factor 2 comes from the two spins. For linear response and low temperatures, the total transmission between contacts is that

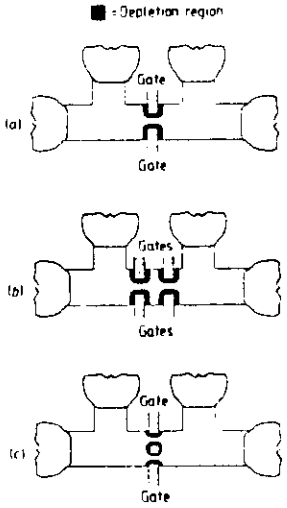


Figure 15. (*a*) A split-gate structure is used to constrict the width of an electron waveguide. Measurements have also been performed for two constrictions (*b*) in series and (c) in parallel.

evaluated at the Fermi energy (2.11)

$$\tilde{T}_{21} = \tilde{T}_{12} = 2M. \tag{3.10}$$

The conductance G_n of a ballistic channel is then obtained from the two-probe Landauer–Büttiker formula (2.12), using $\tilde{T}_{12}$ from the estimate (3.10) shown above:

$$G_n = \frac{2e^2}{h} M. \tag{3.11}$$

This simple estimate has been validated by more detailed considerations of transmission between the wide and narrow regions (Imry 1989, Szafer and Stone 1989). As the width of the channel is reduced, one expects the conductance to decrease linearly in large samples. But equation (3.11) shows that the conductance is quantised, since M is an integer. As the channel width is reduced, the sub-bands are cut off one by one so that the conductance decreases in discrete steps of $2e^2/h$. The same behaviour is observed in the presence of a fixed magnetic field, although fewer sub-bands are available below the Fermi level. This is because of the additional magnetic confinement, which increases the separation between sub-band energies. As a result, fewer quantised steps are observed over a comparable range of channel widths (van Wees *et al* 1988b, Wharam *et al* 1988).

This quantised conductance can also be achieved by raising the potential energy—rather than reducing the width—over some short region. The bias applied to a Schottky gate, for instance, can be used to shift the sub-band energies in the area below the gate. As each energy crosses the Fermi level, the sub-bands are again cut off one by one, so that the conductance decreases in discrete steps. Quantisation observed using this method (Hirayama and Saku 1989) is not as pronounced as that found in point contacts. Nevertheless, these experiments confirm the simple interpretation that the quantisation arises from ballistic transport through the available number of transverse modes.

Series and parallel arrangements of point contacts have also been investigated. The conductance of two point contacts in series (figure 15(*b*)), for instance, is determined by the conductance of the narrower constriction (Wharam *et al* 1988b, Timp *et al* 1989a). As the narrower constriction is widened, the conductance initially increases in quantised steps, and then levels off to the limiting value of the other constriction. The conductance of two point contacts in parallel (figure 15(*c*)) is quantised in steps of $4e^2/h$ (Smith *et al* 1989). At first glance, this appears to be the result expected if the conductances of the two constrictions are simply added. However, small differences in the fabricated widths of the point contacts should cause the modes in either channel to be cut off at slightly different voltages. Instead of changing by steps of $4e^2/h$, the conductance should change by two closely spaced steps of $2e^2/h$. In fact, this behaviour was observed experimentally when both constrictions neared pinch-off. Clearly, there must be some mechanism coordinating the behaviour of the two constrictions, causing both to act in unison away from pinch-off. At this time, however, the nature of this mechanism is unclear.

When driven to a high bias, point contacts have shown interesting non-linearities (Kouwenhoven *et al* 1989). Preliminary experimental evidence has recently been reported (Brown *et al* 1989) indicating that such structures could exhibit negative differential resistance, whose origin may be understood as follows. At a sufficiently large bias, the range of energies available for transport saturates to a maximum value. Higher biases can actually reduce the overall transmission by reflection of electron

waves within the constriction (as proposed by Brown *et al* (1989)) or by carrier heating and increased scattering. Any reduction in transmission would lead to negative differential resistance. This effect may find device applications (Kelly *et al* 1989).

3.2.4. Aharanov–Bohm effect. An obvious structure for observing quantum interference effects is a ring, because it provides two alternative paths from the input to the output (figure 16). The relative phase between the two paths can be controlled with a magnetic field perpendicular to the plane of the ring. By changing the interference condition, the conductance can be made to oscillate. This is named after Aharanov and Bohm (1959) who first proposed it as a means of showing the physical significance of potentials. Their intent was to show that effects of the magnetic field should be observed even if the field exists only in the centre of the ring, and not in the path of electrons. Recent solid state demonstrations of the Aharanov–Bohm (A-B) effect, however, do not really demonstrate their surprising prediction since the magnetic field in these experiments is nearly uniform everywhere. Nonetheless, the demonstration of the A-B effect in metal rings is a very important milestone in the development of mesoscopic physics (this work is reviewed in Washburn and Webb (1986)).

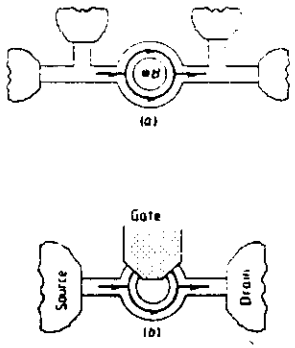


Figure 16. The Aharanov–Bohm effect causes the conductance of a ring structure to oscillate when relative phase between two paths is changed by either (*a*) a magnetic field or (*b*) an electric field.

It has been shown (Büttiker *et al* 1985) that if the transmission coefficients for different modes were totally uncorrelated then the conductance modulation $\Delta G/G$ due to the A-B effect would be of order $1/M$, where M is the total number of modes. (Experimentally, the conductance modulation turned out to be much larger than the $1/M$ estimate, indicating that the modes are not totally uncorrelated, as discussed in Washburn and Webb (1986).) In metals the observed conductance modulation is ~0.1%. Semiconductor rings are expected to show much larger interference effects, since M can be reduced to values less than 10. Indeed, experiments on semiconductor rings since 1987 have revealed conductance modulation by 1–20% (Timp *et al* 1987a, Ford *et al* 1989). To harness this effect in a three-terminal device, one could use an electric current to generate the required magnetic field; a device concept using superconducting electrodes to generate the magnetic field has been proposed (Yamamoto and Hohkawa 1988).

Semiconductor structures also hold the possibility of an electrostatic version of the A-B effect. This effect, which uses an electric field instead of a magnetic field, is actually much easier to understand, though it is much more difficult to observe. It is based upon the electron waveguide analogue of the Mach–Zender interferometer. This device, used as a modulator in integrated optics, consists of a single-input waveguide that splits and subsequently rejoins to form a single-output waveguide. In the optical interferometer, the phase difference between the two arms is controlled by changing the refractive index of one arm through the electro-optic effect. In view of the similarity between Schrödinger's (2.24*a*) and Maxwell's (2.24*b*) equations, we might expect that for electron waves, the phase-difference can be introduced by changing the potential V of one arm with respect to the other. Therefore, a gate voltage can be used to control the effect (figure 16(*b*)), as proposed independently by Fowler (1985) and Datta *et al* (1985, 1986). The applied potential shifts the sub-band energy ε_n in one arm with respect to that in the other. Since

$$E = \varepsilon_n + \frac{\hbar^2 k^2}{2m^*} \tag{3.12}$$

we have

$$|\Delta k| = |\Delta \varepsilon_n| \frac{m^*}{\hbar^2 k}. \tag{3.13}$$

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

$$|\Delta\theta| = |\Delta k| L = \frac{|\Delta \varepsilon_n|}{E_k} \frac{kL}{2} \tag{3.14a}$$

$$= |\Delta \varepsilon_n| \tau_t / \hbar, \tag{3.14b}$$

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

Because the phaseshift is proportional to the transit time (3.14*b*), the electrostatic effect is difficult to observe compared with the magnetic effect, where the phaseshift is nearly unique for thin rings. Recent experiments have shown that a potential applied to a gate covering one arm of a ring structure can be used to tune the conductance oscillations in a magnetic field (deVegvar *et al* 1989). However, no periodic conductance modulation was observed at a fixed magnetic field as the gate potential was changed. To observe the electrostatic effect, it seems important to design structures that minimise the spread in transit times. For example, an alternative structure using film growth rather than lithographic techniques to define the channels has been proposed (Datta *et al* 1986, Datta 1989a). However, this structure has not yet been fabricated.

3.2.5. Non-local effects. One of the important realisations of mesoscopic physics is that voltage probes can no longer be treated as non-invasive, classical objects which simply measure the voltage at a specific location. Recent voltage measurements illustrate this point (Skocpol *et al* 1987, Benoit *et al* 1987). Measurements were performed on Hall bridges with a number of variably spaced voltage probes. The measured voltage is expected to fluctuate with magnetic field due to conductance fluctuations, as discussed earlier. In a classical measurement, one would expect the

voltage fluctuations to disappear as the voltage probes are moved closer together. Instead, the experiments showed that the size of the fluctuations becomes constant. This result has also been verified theoretically (Büttiker 1987). In effect, the separation between the voltage probes cannot be reduced below a phase-coherence length, since an electron in one probe can 'see' this far into another.

A striking demonstration of this non-local behaviour is provided by conductance measurements performed on a metallic Hall bridge with a ring attached outside the classical current path (Umbach *et al* 1987). When embedded within the current path, such a ring is known to produce periodic oscillations in the conductance due to the A-B effect (figure 16(*a*)). In this experiment, however, the mere presence of the ring within a phase-coherence length of the current path (figure 17(*a*)) was sufficient to produce oscillations with precisely the period expected from estimates based upon the ring dimensions. Although such effects are intriguing in this context, they are the cause for serious concern when performing measurements in the mesoscopic regime. Because of the non-local nature of transport, even a simple bend in the current path is enough to affect a measurement. This was recently demonstrated by measurements performed on a Hall bridge designed so that the path of current could be bent at some point outside the 'measurement area' between voltage probes (Timp *et al* 1988). The resistance was found to increase as the path of current was bent; moreover, the change in the resistance decayed exponentially as the bend was moved away from the measurement region, indicating that interference effects were responsible for the anomaly.

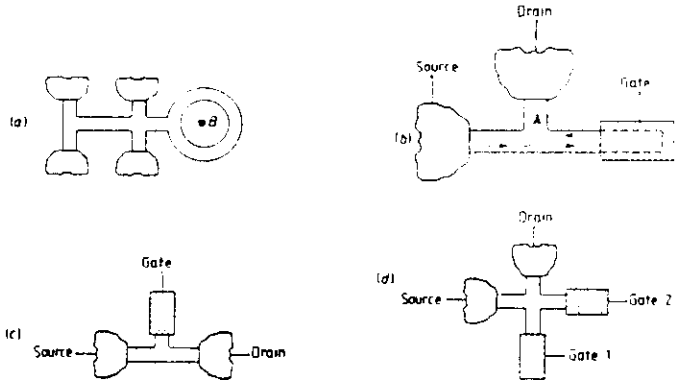


Figure 17. Geometries used to observe non-local effects: (*a*) A ring outside of the classical current path exhibits the Aharanov–Bohm effect (after Umbach *et al* 1987); (*b*) the conductance of a T-structure can be controlled by a remote gate (after Datta 1989a); (*c*) an alternative T-structure has the gate and drain interchanged (after Sols *et al* 1989); (*d*) the electronic analogue of the Michelson interferometer, using two gates to control the relative phase difference between interfering paths.

An implication of non-locality that could be important for device applications is that it is not necessary for a gate to be positioned between the source and the drain, as we are accustomed to expect. It can be located anywhere within a phase-coherence length. Consider the three-port network shown in figure 17(*b*). Ohmic contacts are made to two of the ports while a Schottky gate is used to change the phase of the reflection coefficient at the third port (Datta 1989a). The transmissivity from the source

0421'''02996

to the drain is determined primarily by the interference between the two paths shown in figure 17(*b*), and the gate controls the phase difference between these paths. One can view this structure too as an interferometer, with the T-junction acting like a partially silvered mirror that splits the incoming beam. Conductance oscillations as a function of the gate potential have been experimentally observed in this structure, though the underlying mechanisms have not yet been established conclusively (Miller *et al* 1989). An alternative structure with the drain and gate interchanged (figure 17(*c*)) was proposed independently (Fowler 1988, Sols *et al* 1989). Another possibility is a four-port structure shaped like a cross, having two gates; the differential voltage between these gates could control the interference in a manner reminiscent of the Michelson interferometer (figure 17(*d*)).

It seems that one can come up with new quantum device concepts by looking up a textbook on microwaves or optics. Unlike optical devices, however, electronic devices are not easily driven by a 'monochromatic' source. At large biases and high temperatures, a range of energies is available for transport. As a result, interference effects which are sensitive to the wavelength tend to wash out under practical operating conditions. Another difference is that electrons, unlike photons, obey the exclusion principle, so that the current carried by a single-moded quantum wire is limited to 80 nA per meV of bias. To achieve the currents carried by conventional devices, it seems that quantum devices must operate either at high biases or with multiple transverse modes. The challenge, therefore, is to design 'broadband' structures whose interference patterns survive the effects of energy averaging and mode averaging.

3.2.6. Edge states. To determine the coefficients $\bar{T}_{ij}$ for any arbitrary two-dimensional device is a difficult numerical task. But in the presence of a strong magnetic field, these coefficients can be deduced with a high degree of accuracy from a few simple arguments. In a strong magnetic field, current is carried around the boundaries of a device by 'edge states', which are understood to arise as follows (Büttiker 1988b). Under the influence of a magnetic field, electrons collapse into Landau levels. For sufficiently strong fields, one can treat the presence of the confining potential, which defines the boundaries of the device, as a perturbation on the level energy. Within a cyclotron radius of the potential walls, the Landau levels bend upward (figure 18(*a*)). A given Fermi energy usually intersects these states near the edges of the device, thereby defining the channels for current flow. Edge states on opposite sides of a device carry oppositely directed currents. If we assume that these states remain relatively isolated from one another (an assumption which is largely correct), then it is a simple matter to deduce the transmission coefficients for any arbitrary device. Current injected into any edge state is completely transmitted to the next contact along the path of the edge state. Therefore, the total transmission from one contact to another is obtained simply by counting the number of edge states which carry current in that direction.

From a semiclassical viewpoint, it appears that current is being carried by electrons that bounce along the potential walls in skipping orbits. Indeed, for weak magnetic fields, skipping orbits have been clearly demonstrated by electron focusing experiments (Beenakker *et al* 1989). For strong magnetic fields, the existence of edge states has been shown by a suppression of the A-B effect in ring geometries (Timp *et al* 1989b). Recall that the A-B effect arises from an electron interfering with itself as it traverses two possible paths around a ring. By changing the flux enclosed by the two paths, one can change the interference condition, and therefore change the conductance of

0585'''04103

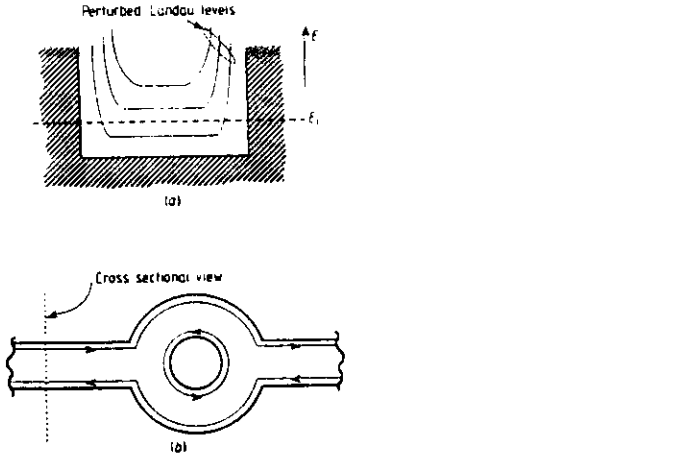


Figure 18. (*a*) Near the boundaries of an electron waveguide the Landau levels are bent upwards, forming 'edge states' where they intersect the Fermi level. (After Büttiker 1988b.) (*b*) For high magnetic fields, current is carried around the perimeter of a ring by oppositely directed edge states; this causes the Aharanov–Bohm effect to be suppressed (after Timp *et al* 1989b).

the ring. At some point, however, the magnetic field becomes strong enough that edge states form, and the path of an electron is no longer split. Instead, current is carried around the boundaries of the device by oppositely directed edge states (figure 18(*b*)). Since an electron can no longer interfere with itself, the A–B oscillations are suppressed.

Büttiker (1988b) has exploited these edge states to provide an elegant description of the integer quantum Hall effect (Prange and Girvin 1987). Consider the simple 'cross' geometry shown in figure 19(*a*). If we apply the Landauer–Büttiker formula

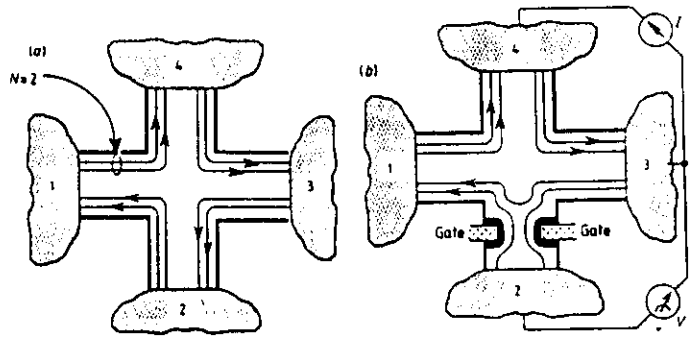


Figure 19. (*a*) Simple cross geometry used to explain the integer quantum Hall effect. (After Büttiker 1988b.) (*b*) Edge states can be forced to bypass a lead by constricting the width of a voltage probe. This leads to a suppression of SdH oscillations, as shown by van Wees *et al* (1988c).

(2.13) to each of the leads, we obtain a matrix equation relating currents and chemical potentials. In lead 1, for instance,

$$I_1 = \frac{e^2}{h}[(\tilde{T}_{21} + \tilde{T}_{31} + \tilde{T}_{41})\mu_1 - \tilde{T}_{12}\mu_2 - \tilde{T}_{13}\mu_3 - \tilde{T}_{14}\mu_4]. \tag{3.15}$$

Gathering similar results for leads 2, 3 and 4, we obtain,

$$\begin{pmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{pmatrix} = \frac{e^2}{h} \begin{pmatrix} \tilde{T}_1 & -\tilde{T}_{12} & -\tilde{T}_{13} & -\tilde{T}_{14} \\ -\tilde{T}_{21} & \tilde{T}_2 & -\tilde{T}_{23} & -\tilde{T}_{24} \\ -\tilde{T}_{31} & -\tilde{T}_{32} & \tilde{T}_3 & -\tilde{T}_{34} \\ -\tilde{T}_{41} & -\tilde{T}_{42} & -\tilde{T}_{43} & \tilde{T}_4 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} \tag{3.16}$$

where $\tilde{T}_i$ represents the total transmission out of lead i,

$$\tilde{T}_i = \sum_{j \neq i} \tilde{T}_{ji}. \tag{3.17}$$

We will now assume that current is carried by N edge states within the cross, as shown in figure 19(*a*). Thus, the only non-zero transmission coefficients are

$$\tilde{T}_{12} = \tilde{T}_{23} = \tilde{T}_{34} = \tilde{T}_{41} = N. \tag{3.18}$$

This greatly simplifies the matrix equation (3.16), to the form

$$\begin{pmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{pmatrix} = \frac{e^2}{h} \begin{pmatrix} N & -N & 0 & 0 \\ 0 & N & -N & 0 \\ 0 & 0 & N & -N \\ -N & 0 & 0 & N \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} \tag{3.19}$$

We are interested in computing the Hall resistance $R_{13,42}$. To accomplish this, we zero the currents I_2 and I_4 in the two voltage leads, and fix the driving currents to be $I_1 = -I_3 = I$. It is easy to see that a solution of equation (3.19) under these conditions is $\mu_1 = \mu_4$ and $\mu_2 = \mu_3$. Thus, the Hall resistance

$$R_{13,42} = \frac{\mu_4 - \mu_2}{Ne^2(\mu_1 - \mu_3)/h} = \frac{h}{e^2}\frac{1}{N} \tag{3.20}$$

is quantised in integer steps, since the number of edge states N is an integer. At each Hall plateau, the longitudinal resistance is known to vanish. This can be demonstrated by computing $R_{43,12}$ from the boundary conditions $I_1 = I_2 = 0$ and $I_4 = -I_3 = I$. A solution for the chemical potentials is $\mu_1 = \mu_2 = \mu_3$, and therefore the longitudinal resistance $R_{43,12}$ vanishes.

This vanishing resistance is often recognised as the extreme case of a SdH oscillation trough. In a weak magnetic field, the resistance is found to oscillate as each Landau level is moved through the Fermi energy. In particular, the resistance is a minimum when the Fermi energy is between Landau levels. For strong magnetic fields, this minimum resistance actually goes to zero, as shown above. The interleaving peaks in the resistance can be understood as follows. Each time a 'bulk' Landau level (i.e. that in the middle of a device) approaches the Fermi energy, the innermost edge states become strongly coupled. This is to be expected, since these edge states correspond to the same (perturbed) Landau level. The coupling introduces some small corrections to the matrix equation (3.19), so that all of the elements become non-zero. In this

case, the chemical potentials μ_1 and μ_3 will differ from μ_4—each will include a small influence from μ_4, as well. As a result, the resistance becomes non-zero. Once the bulk Landau level has passed through the Fermi energy, the innermost edge states disappear, and the remaining edge states are again uncoupled. Thus, the resistance returns to its zero (trough) value.

In all of this discussion, we have assumed a simple geometry having leads of equal width, so that all leads admitted the same number of edge states. When this assumption is violated, edge states can be selectively injected or detected. Using a point contact to constrict a voltage lead, for instance, van Wees *et al* (1988c) have shown that certain SdH oscillations can be suppressed. This can be understood by considering an equivalent geometry for their experiment, presented in figure 19(*b*). Note that the resistance measured in their configuration is $R_{43,23}$, as opposed to $R_{43,12}$ which is discussed above. As the width of lead 2 is reduced, some of the innermost edge states are forced to bypass the lead entirely. For simplicity, we assume that only one such edge state is affected. In this case, the transmission coefficients $\tilde{T}_{12}$, $\tilde{T}_{23}$, and $\tilde{T}_{13}$ are modified, so that the matrix equation (3.19) becomes

$$\begin{matrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{matrix} = \frac{e^2}{h} \begin{matrix} \begin{matrix} N & -(N-1) & -1 & 0 \end{matrix} \\ \begin{matrix} 0 & (N-1) & -(N-1) & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & N & -N \end{matrix} \\ \begin{matrix} -N & 0 & 0 & N \end{matrix} \end{matrix} \begin{matrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{matrix} \tag{3.21}$$

Again, this equation is valid in the absence of strong coupling between the innermost edge states. Using lead 2 as a voltage probe ($I_2 = 0$), the second row of this matrix equation requires that $\mu_2 = \mu_3$, so that $R_{43,23}$ goes to zero in the trough, as before. As the bulk Landau level approaches the Fermi energy, the innermost edge states will again become coupled. This coupling, however, will not affect the transmission coefficients $\tilde{T}_{2l}$ or $\tilde{T}_{l2}$, since the innermost edge states no longer transmit into lead 2. As a result, the second row of the matrix equation (3.21) will remain as before, although all other rows will be modified by the coupling. Because the second row requires $\mu_2 = \mu_3$, the resistance $R_{43,23}$ will remain zero, *even though we would have expected a peak*. Thus, the normal SdH oscillation is suppressed, simply because the voltage probe is blind to the relevant physics.

Of course, this simplified picture of transport assumes that electrons remain in a given edge state without scattering. The results of van Wees *et al* (1988c) provide a rather remarkable confirmation of this assumption. Electrons in their experiment traverse a distance on the order of 200 μm (the separation between contact 3 and the constriction) without scattering between edge states! The prospect of having such a long mean free path may spark interest in an entirely new class of devices, based upon ballistic transport through edge channels. However, the need for a high magnetic field to create the edge states is a drawback in this regard.

4. Current theoretical status

The last few years have seen an explosive increase in both experimental and theoretical activities on quantum transport in semiconductor microstructures. This work has led to an increased emphasis on sample-specific properties, as opposed to the ensemble-averaged properties that had dominated solid state physics in the past. A fairly good

qualitative understanding of the linear response has emerged, though a detailed quantitative understanding is still lacking. Also, much remains to be studied in the area of large-signal quantum transport far from equilibrium. In what follows, we review the current theoretical understanding of quantum transport, emphasising the issues which have yet to be resolved.

4.1. Linear response

A celebrated result of linear response theory is the so-called fluctuation-dissipation theorem, which relates the linear response of a system to the correlations of some equilibrium property (Kubo 1986). An example of this is the Kubo formula, which expresses the conductivity of a sample in terms of the correlations among its equilibrium current operators. These correlations are commonly evaluated by diagrammatic techniques (Doniach and Sondheimer 1974).

Much of the current theoretical work on quantum transport is based on the Landauer-Büttiker formula (2.13) derived in section 2; we rewrite it here for convenience:

$$I_i = \frac{e^2}{h} \sum_j (\tilde{T}_{ij}\mu_i - \tilde{T}_{ji}\mu_j). \tag{4.1}$$

Starting from a symmetry property of the coefficients $\tilde{T}_{ij}$ in a magnetic field

$$\tilde{T}_{ij}(B) = \tilde{T}_{ji}(-B) \tag{4.2}$$

Büttiker (1986a) has shown that this formula leads to the reciprocity relationship for an inhomogeneous conductor

$$R_{mn,kl}(B) = R_{kl,mn}(-B). \tag{4.3}$$

Here, $R_{mn,kl}$ is the resistance measured by feeding current through leads m and n and measuring the voltage between leads k and l. The establishment of this property was helpful in gaining the widespread acceptance that the Landauer-Büttiker formula (4.1) has received. This formula has also been connected to linear response theory. An initial derivation of the two-probe conductance (Fisher and Lee 1981) was later generalised to multiple probes (Stone and Szafer 1988) and shown to be valid even in the presence of strong magnetic fields (Baranger and Stone 1989c). It is now believed that any calculation based on the Landauer-Büttiker formula (4.1) should yield the same result as that obtained from the Kubo formula (Stone and Szafer 1988).

Consequently, the Landauer-Büttiker formula (4.1) is the starting point for much of the current theoretical work on mesoscopic structures. In the absence of dissipation, the coefficients $\tilde{T}_{ij}$ can be computed simply by solving the one-electron Schrödinger equation (2.23), using ordinary wave mechanics. A popular method for accomplishing this is the recursive Green function technique (Lee and Fisher 1981), which is based upon the tight-binding model. Part of the reason for the method's popularity is the ease with which disorder can be introduced—simply by randomising the energies of tight-binding sites. This feature has been useful, for instance, in numerical studies of conductance fluctuations (Stone 1985, Baranger *et al* 1988). Also, this method is easily applied to any two-dimensional geometry, and has therefore been used to compute the conductance of a constriction (Szafer and Stone 1989), as well as local and non-local bend resistances (Baranger and Stone 1989a) and the quenching of the Hall effect (Baranger and Stone 1989b). Abrupt constrictions or bends can also be handled by

matching the wavefunction and its derivatives across the junctions (Avishai and Band 1989a, b).

A different approach for solving the Schrödinger equation, which is useful for layered (vertical) geometries, is the scattering matrix technique. Individual layers of a device are characterised by scattering matrices, which are then combined to determine the overall transmission. An example of this is the study of conductance in disordered resistors (Cahay *et al* 1988). Samples were generated by randomly placing elastic scatterers along the length of a wire. Statistics gathered from a large number of samples demonstrated both weak and strong localisation, and conductance fluctuations of order e^2/h. An alternative to this type of direct numerical simulation of random samples is the study of the eigenvalue spectra of random transfer matrices which are related to the scattering matrices by a simple transformation (Imry 1986b, Muttalib *et al* 1987). Strong correlations are found in the eigenvalue spectra, giving a mathematical perspective to the nature of conductance fluctuations.

Alternatively, some authors have used diagrammatic techniques to evaluate the conductance. An early derivation of the two-probe Landauer–Büttiker formula (4.1) gave, as an intermediate step, a relationship between the conductance and the advanced and retarded Green functions (Fisher and Lee 1981). This defines a diagrammatic representation of the conductance, which was later used in studies of conductance fluctuations (Lee *et al* 1987, Serota *et al* 1987). Other authors (Maekawa *et al* 1987, Kane *et al* 1987, 1988, Hershfield and Ambegaokar 1988, DiVincenzo and Kane 1988) have evaluated the non-local conductivity tensor $\sigma(r, r')$ from the Kubo formulae, and then related this to the transmission coefficients $\tilde{T}_{ij}$ through the Fisher–Lee relation (Fisher and Lee 1981, Stone and Szafer 1988),

$$\tilde{T}_{ij} = \int \mathrm{d}(S_i)_\alpha \int \mathrm{d}(S_j)_\beta \sigma(r, r'). \tag{4.4}$$

Here, the integrals are performed over the cross sections of leads i and j, and the vectors S_i and S_j are normal to these surfaces. An advantage of diagrammatic techniques is that the phase-breaking time τ_φ can be included in the computation by introducing a cut-off in the diffusion propagator. Moreover, these techniques are particularly well suited to calculations of ensemble-average properties, since in this case the diagrams can be summed analytically. Although one is usually more interested in sample-specific solutions in the mesoscopic regime, this approach has contributed much to the understanding of disordered materials—in particular, to the phenomena of localisation and conductance fluctuations.

Diagrammatic techniques are not the only means of incorporating phase-breaking processes. Büttiker (1986b, 1988c) has shown that phase-coherence can be destroyed by connecting additional contact reservoirs to various points within a device. These fictitious reservoirs are not driven by external sources, so that the net current in each of the connecting probes is zero: electrons entering a reservoir are completely absorbed, and after being thermalised, they are injected back into the device with a random phase. Thus, although the transport between two reservoirs is phase-coherent and therefore reversible, the overall transport is *irreversible* due to the dissipation which occurs in the reservoirs. Büttiker (1988c) has used this technique to simulate the cross-over from coherent resonant tunnelling to incoherent tunnelling through two barriers.

If each additional reservoir in Büttiker's model acts as a localised inelastic scatterer, it seems that one could simulate distributed inelastic processes with a continuous

distribution of reservoirs

$$I(r) = \frac{e^2}{h} \int \mathrm{d}r' (\tilde{T}(r', r)\mu(r) - \tilde{T}(r, r')\mu(r')). \tag{4.5}$$

Calculations based upon this formula have been used to study the impact of inelastic scattering on localisation (D'Amato and Pastawski 1989, Datta and McLennan 1989). Although this simple picture of transport is physically appealing, sceptics would argue that it is purely phenomenological. However, the rigorous justification for such an approach has recently been presented (Datta 1989b). Starting from a microscopic Hamiltonian which allows electrons to interact with a bath of point-size inelastic scatterers, a transport equation was derived which reduced to equation (4.5) in the limit of linear response. As well as putting Büttiker's approach on firm ground, the derivation also provides a means of computing the coefficients $\tilde{T}(r, r')$ in the presence of dissipation. These coefficients are determined from the retarded Green function $G^R(r, r')$

$$\tilde{T}(r, r') = \frac{\hbar^2 |G^R(r, r')|^2}{\tau_\varphi(r)\tau_\varphi(r')}\Big|_{E=\varepsilon_F} \tag{4.6}$$

where $\tau_\varphi(r)$ is the phase-breaking scattering time, which is allowed to vary spatially. We have quoted the zero-temperature result (transport occurs only at the Fermi energy) for simplicity. Included in the definition of the Green function is an optical potential $i\hbar/2\tau_\varphi(r)$, computed from the electron self-energy. The presence of this imaginary potential causes $|G^R(r, r')|^2$ to decay as a function of distance from the injection point r'. In Büttiker's model, this corresponds to electrons which are gradually leaking out into the continuous distribution of probes. Thus, interference effects (which arise as fluctuations in $|G^R(r, r')|^2$) are gradually damped over a phase-breaking length. More importantly, this decay reproduces the self-averaging behaviour of large samples: only points within a phase-breaking length are strongly connected by the kernel $\tilde{T}(r, r')$, so that any large device can be thought of as a combination of uncorrelated units, each the size of a phase-breaking length.

4.2. *Harmonic generation*

The Landauer–Büttiker formula (4.1) can only predict the linear response of a device. This is because the transmission coefficients $\tilde{T}_{ij}$ are computed in equilibrium and remain fixed, so that current is simply proportional to the amount of applied bias. Obviously, the limitations of linear response are defined by the sensitivity of the transmission coefficients to an applied bias. An interesting example of non-linear behaviour is the rectifying nature of asymmetric quantum well diodes, discussed earlier in section 3.1.3. If the resonant energies in two neighbouring wells are slightly different, an applied bias will tend to align the energies for one polarity, and separate the energies for the opposite polarity (see figure 9). This results in a diode-like character, which can only be described by accounting for the changes in transmission due to the applied bias. Asymmetry has also been predicted (Al'tshuler and Khmel'nitskii 1985) and observed (Webb and Washburn 1988) in the voltage fluctuations of a disordered resistor. In the mesoscopic regime, ordinary ohmic behaviour is lost since the specific microscopic configuration of impurities makes a resistor inhomogeneous. Thus, quantum

mechanical interference effects can cause a pronounced asymmetry in the response of voltage to current, leading to harmonic generation. Harmonics as high as the tenth have recently been detected in the response of semiconductor ring structures (deVegvar *et al* 1982).

4.3. Large-signal response

To compute the response of a device to a large applied bias, we need to solve the more general equation (2.7), which is restated here for convenience,

$$I_1 = \frac{e}{h}\int dE[T_{21}(E)f(E-e\mu_1) - T_{12}(E)f(E-e\mu_2)]. \tag{4.7}$$

One effect of a large applied bias is to broaden the range of energies of electrons contributing to transport. Bagwell and Orlando (1989a) have shown that equation (4.7) can be represented as a convolution of the transmission coefficients with both a 'voltage windowing function' and a temperature broadening function. Of course, the transmission coefficients themselves also change with bias, so that the applied voltage does more than simply broaden the energy range. Neglecting all phase-breaking processes, the transmission coefficient $T_{12}(E) = T_{21}(E)$ can be calculated simply from the one-electron Schrödinger equation (2.23). In this manner, equation (4.7) has been widely used in the study of large-signal phenomena, including tunnelling (see, for example, Frenkel 1930, Duke 1969, Lenstra and Smokers 1988), resonant tunnelling (see, for example, Vassell *et al* 1983, Jogai and Wang 1985), transport in finite superlattices (see, for example, Tsu and Esaki 1973, Bagwell and Orlando 1989b) and transport through ballistic constrictions (Kelly *et al* 1989).

A major concern is how dissipative processes affect the results of these calculations. As electrons become heated by the applied bias, electron–electron scattering is significantly increased. As a result, phase-breaking processes are expected to become important for large signal response. For example, the effect of dissipation on tunnelling is a fundamental question that has been addressed by a number of authors (Caldeira and Leggett 1983, Bruinsma and Bak 1986, Caroli *et al* 1972). Another example is the valley current of a resonant tunnelling diode, which is believed to arise largely from electrons that are inelastically scattered. In the simplest models, inelastic scattering homogeneously broadens the resonance, so that the resonance is more difficult to switch off (Stone and Lee 1985, Büttiker 1988a). In addition, single-frequency phonons lead to a replication of the current peak, which appears as a shoulder on the I–V curve in the valley current region (Goldman *et al* 1987, Wingreen *et al* 1988).

Non-equilibrium Green function techniques provide a powerful approach to account for dissipative processes in large-signal quantum transport (recent reviews include Mahan (1987), Jauho (1989), Jauho and Ziep (1989) and Rammer and Smith (1986)). The fundamental quantity in this formalism is the Green function

$$G^<(r_1, r_2; t_1, t_2) = i\langle\psi^\dagger(r_2, t_2)\psi(r_1, t_1)\rangle \tag{4.8}$$

which contains information about the spatial and temporal correlations of the electron field. In particular, the electron density $n(r; t)$ is proportional to the diagonal element of the Green function, obtained by setting $r_1 = r_2$ and $t_1 = t_2$ in equation (4.8). An alternative form of the Green function $G^<(r, k, E, t)$ is obtained from $G^<(r_1, r_2; t_1, t_2)$

by transforming to centre-of-mass

$$r = \tfrac{1}{2}(r_1 + r_2) \qquad t = \tfrac{1}{2}(t_1 + t_2)$$

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

$$r_1 - r_2 \to k \qquad t_1 - t_2 \to E.$$

Quantum kinetic equations which describe the evolution of this non-equilibrium Green function are usually derived using the formalism developed independently by Kadanoff and Baym (1962) and Keldysh (1965); for a review see Langreth (1976).

Instead of the full correlation function $G^<(r, k, E, t)$ quantum kinetic equations are often formulated in terms of the Wigner function $f^W(r, k, t)$. The Wigner function is obtained by integrating $G^<(r, k, E, t)$ over the energy coordinate, and it is often preferred because it has many properties similar to the semiclassical distribution function $f(r, k, t)$. Specifically, it can be integrated over k to determine the electron density, and integrated over r to determine the current density (Wigner 1932). Kinetic equations for the Wigner function have also been derived starting from the quantum Liouville equation (see Kluksdahl *et al* (1989) and references therein). The Liouville equation is formulated in terms of the density matrix, which is related to the Wigner function by a transformation of variables. A number of authors have used the Liouville equation as a starting point for quantum kinetic theories (Iafrate and Krieger 1989, Barker 1982 and references therein). This formalism has been applied to the simulation of resonant tunnelling diodes, using a phenomenological relaxation term to describe dissipation (Ravaioli *et al* 1985, Frensley 1985, 1986, 1987, Kluksdahl *et al* 1989). The density matrix can also be computed using the path integral formulation (Feynman and Vernon 1963, Thornber 1978), which is particularly suited to handling strong electron–phonon interactions. In this formulation, a trace over the phonon states results in an effective potential which depends only on electronic coordinates. For numerical computations, a Monte Carlo sampling of electronic paths has been developed to compute the effective potential (Mason and Hess 1989).

Apart from accounting for dissipation, the Wigner function formalism is also suitable for time-dependent solutions. Although early attempts (Ravaioli *et al* 1985, Frensley 1985) had difficulty obtaining stable solutions, the problem was finally resolved by Frensley (1986, 1987). He noted that the boundary conditions appropriate for an open system correspond to the properties of a reservoir in the Landauer picture: reservoirs should act like black bodies with respect to electrons, so that electrons entering the reservoir are completely absorbed, and then injected back into the device according to the distribution function of the reservoir. After implementing these boundary conditions, stable solutions of the transients were obtained for a resonant tunnelling device.

Recently, a simplified quantum kinetic equation was proposed for the study of dissipative transport in steady state (Datta 1989b). This equation is formulated in terms of the electron density per unit energy $n(r; E)$, which is proportional to the Green function integrated over k and t. The averaging over t is made possible by the restriction to steady state. On the other hand, the averaging over k is made possible by assuming a special (point-like) form for the inelastic scatterers. The resulting simplification in the kinetic equation makes it possible to obtain numerical solutions for realistic structures that can be compared with experiment. Such comparisons may

illuminate the relevant physics of dissipative quantum transport, and may help identify new phenomena arising from spatially correlated inelastic scattering processes.

4.4. Space-charge effects

An important point that often tends to be overlooked is the role of space-charge effects in transport phenomena (Landauer 1987). Device engineers are usually quite aware of the fact that the drift-diffusion equation needs to be solved *self-consistently* with the Poisson equation. Any transport equation that replaces the drift-diffusion equation for ultrasmall devices will need to be solved likewise.

Equilibrium band-bending can have a major quantitative (if not qualitative!) impact on transport. Consider, for instance, the simple tunnel barrier shown in figure 20. Neglecting any space-charge effects, the conduction band profile is drawn as a rectangular potential barrier (figure 20(*a*)), which merely reflects the presence of a wide bandgap material. Depending upon the internal doping profile, the correct equilibrium potential can have a number of different forms, as illustrated in figures 20(*b*) and (*c*). We belabour this point for good reason: even in this simple example, the impact of band-bending can be significant in terms of the tunnelling current, which is ***exponentially*** sensitive to the barrier height. In an arbitrary mesoscopic sample, the complex interplay between interference effects and the electron density and the electrostatic potential make it much more difficult to foresee what the effect of band-bending would be. But clearly such effects cannot be neglected *a priori*. A strict account of the equilibrium band-bending must be taken as a first step in the analysis of any device. In the limit of linear response, the *corrections* to this equilibrium potential due to the applied bias can be neglected.

For large-signal response, on the other hand, one needs the correct potential distribution under an applied bias. Self-consistent solutions of the Schrödinger and Poisson equations applied to resonant tunnelling diodes have shown that the bias is

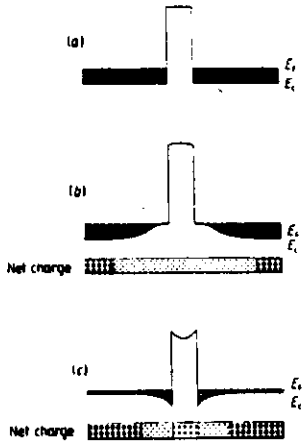


Figure 20. Equilibrium band-bending can significantly alter the potential profile of a device, as seen for a simple tunnelling barrier (*a*) neglecting space-charge effects, (*b*) with a lightly doped interior and (*c*) with a heavily doped barrier.

dropped non-uniformly across the device; much of the bias is absorbed by a depletion region on one side of the device, thereby scaling down the effective bias across the double-barrier region (Ohnishi *et al* 1986, Cahay *et al* 1987, Brennan 1987). Thus, higher applied biases are required in a self-consistent analysis to reach the NDR region. Another effect of space charge is to shift the resonant levels in the quantum well. Depending upon the occupation of these levels, two stable solutions of the electrostatic potential can be found, leading to a hysteresis in the I–V characteristic as the applied bias is ramped up and down. This hysteresis has been observed both theoretically (Kluksdahl *et al* 1989, Mains *et al* 1989) and experimentally (Zaslavsky *et al* 1989, Alves *et al* 1989), as discussed earlier in section 3.1.2. Note that these effects rely on changes in the space-charge distribution beyond the equilibrium band-bending discussed previously.

In a self-consistent solution, the Poisson equation accounts for electron–electron interactions only in the Hartree approximation. It is possible that exchange and correlation will also play a significant role in small structures as they do in atoms and molecules. For example, Bandara *et al* (1988) have found a shift in the photoexcitation wavelengths associated with transitions between quantum well sub-bands. This shift is well described by including corrections in the sub-band energies due to exchange and correlation. On the whole, however, exchange and correlation effects are largely unexplored. Coulomb blockade in ultrasmall tunnelling junctions (Likharev 1988) can possibly be viewed as an example of a many-body correlation effect. It remains to be seen whether future experiments will uncover new situations in which these effects play a significant role.

5. Concluding remarks

The important realisation that has emerged over the last few years is that electronic circuits with dimensions less than a phase-breaking length behave much like microwave or optical networks. Once we accept this basic notion, most of the novel phenomena recently discovered in mesoscopic structures are easily understood. The lasting value of these discoveries thus lies in teaching us to look at electronic transport from a new viewpoint. This viewpoint naturally raises the possibility of electronic analogues of microwave and optical devices. Of course, device concepts based on controlling the phase of the wavefunction can only become practicable when the technology has advanced far enough to eliminate random variations in the phase (e.g. conductance fluctuations). Moreover, the fermionic nature of electrons leads to an important difference with light. A single-moded optical fibre can, in principle, carry any amount of power per unit frequency range. But a single-moded quantum wire can only carry 80 nA of current per meV, due to the exclusion principle. The operation of such quantum devices is usually limited to low bias voltages and low temperatures, to reduce the energy spread of electrons involved in transport. In addition to energy averaging at large biases, electron heating leads to increased electron–electron scattering, thereby destroying interference effects.

As the microelectronics industry continues its drive towards smaller device dimensions, it seems inevitable that quantum interference effects will play a more significant role in device operation, intentionally or otherwise. It seems likely that quantum devices that rely on such effects will find useful niches in a variety of specialised applications (Capasso 1989a, b). But the question of whether quantum devices will

ever make it 'big' by impacting the integrated circuits industry is a controversial one that only the future can answer. Leading researchers in this field have provided both optimistic (Bate 1988) and pessimistic (Landauer 1989a) outlooks.

To date, most quantum device concepts have been based on analogies with linear optics. However, a totally different class of quantum devices could emerge from analogies with non-linear optics. One can view space-charge effects as a large source of non-linearity inherent in electronic transport, that could be turned to an advantage. Non-linear optics, for instance, is based on the dependence of the dielectric constant on the light intensity, which is a second-order effect. The corresponding phenomenon for electrons is the dependence of the potential on the electron density which is a *first-order* effect. It may be possible to design novel switching devices based on an interplay between quantum effects and space-charge effects. One example is the possibility of engineering a space-charge induced transition in a lateral superlattice (Ferry 1981). Another example is the Coulomb blockade observed in ultrasmall tunnelling junctions (Likharev 1988). Such effects are expected to become increasingly significant as we go to smaller structures.

The role of space charge in electron waveguide transport has so far been neglected. This may be qualitatively different from the role that space charge plays in large-area devices where the electrostatic problem in essentially one dimensional. By contrast in waveguides, one has in general a three-dimensional electrostatic problem, and the effect of space charge may depend on the presence or absence of neighbouring ground planes. Clever design may make it possible to build non-linear electron waveguide networks with novel switching properties. Clearly major breakthroughs are needed before such exotic devices become practicable. But it is likely that quantum devices will eventually find uses that cannot currently be foreseen.

Acknowledgments

The authors are grateful to the National Science Foundation, the Semiconductor Research Corporation and the Office of Naval Research for their support.

References

Aharanov Y and Bohm D 1959 *Phys. Rev.* **115** 48–91
Al'tshuler B L and Khmel'nitskii D E 1985 *Pis'ma Zh. Eksp. Teor. Fiz.* **42** 291–3 (Engl. transl. *JETP Lett.* **42** 359–63)
Al'tshuler B L and Lee P A 1988 *Phys. Today* **41** 36–44
Alves E S, Eaves L, Henini M, Hughes O H, Leadbeater M L, Sheard F W and Toombs G A 1988 *Electron. Lett.* **24** 1190–1
Ando T, Fowler A B and Stern F 1982 *Rev. Mod. Phys.* **54** 437–672
Avishai Y and Band Y B 1989a *Phys. Rev.* B **40** 3429–32
—— 1989b *Phys. Rev. Lett.* **62** 2527–30
Azbel M Ya 1983 *Phys. Rev.* B **28** 4106–25
Bagwell P F and Orlando T P 1989a *Phys. Rev.* B **40** 1456–64
—— 1989b *Phys. Rev.* B **40** 3735–48
Bandara K M S V, Coon D D, O B, Lin Y F and Francombe M H 1988 *Appl. Phys. Lett.* **53** 1931–3
Baranger H U, Jalabert R A and Stone A D 1990 *Bull. Am. Phys. Soc.* **35** 492
Baranger H U and Stone A D 1989a *Science and Engineering of 1- and 0-Dimensional Semiconductors* ed S P Beaumont and C M Sotomayer-Torres (Nes York: Plenum)
—— 1989b *Phys. Rev. Lett.* **63** 414–17

—— 1989c Preprint (submitted to *Phys. Rev.* B)
Baranger H U, Stone A D and DiVincenzo D P 1988 *Phys. Rev.* B **37** 6521–4
Barker J R 1982 *Handbook on Semiconductors* vol 1 ed W Paul (Amsterdam: North-Holland) pp 617–59
Bate R T 1977 *Bull. Am. Phys. Soc.* **22** 407
—— 1988 *Sci. Am.* March 96–100
Beenakker C W J and Van Houten H 1989 *Phys. Rev. Lett.* **63** 1857–60
Beenakker C W J, van Houten H and van Wees B J 1989 *Advances in Solid State Physics* vol 29 ed U Rössler (Braunschweig: Pergamon/Vieweg)
Beltram F, Capasso F, Luryi S, Chu S-N G, Cho A Y and Sivco D L 1988 *Appl. Phys. Lett.* **53** 219–21
Benoit A, Umbach C P, Laibowitz R B and Webb R A 1987 *Phys. Rev. Lett.* **58** 2343–6
Benoit A D, Washburn S, Umbach P, Laibotitz and Webb R A 1986 *Phys. Rev. Lett.* **57** 1765–8
Beresford R, Luo L F and Wang W I 1989 *Appl. Phys. Lett.* **54** 1899–901
Berggren K-F, Roos G and van Houten H 1988 *Phys. Rev.* B **37** 10118–24
Bergmann G 1982 *Solid State Commun.* **42** 815–17
—— 1984 *Phys. Rep.* **101** 1–58
Bernstein G and Ferry D K 1987 *Z. Phys.* B **67** 449–52
Berthold K, Levi A F J, Walker J and Malik R J 1989 *Appl. Phys. Lett.* **54** 813–5
Brennan K F 1987 *J. Appl. Phys.* **62** 2392–400
Broekaart T P E, Lee W and Fonstad C G 1988 *Appl. Phys. Lett.* **53** 1545–7
Brown R J, Kelly M J, Pepper M, Ahmed H, Hasko D G, Peacock D C, Frost J E F, Ritchie D A and Jones G A C 1989 *J. Phys. Condens. Matter* **1** 6285–90
Bruinsma R and Bak P 1986 *Phys. Rev. Lett.* **56** 420–3
Büttiker M 1986a *Phys. Rev. Lett.* **57** 1761–4
—— 1986b *Phys. Rev.* B 3020–6
—— 1987 *Phys. Rev.* B 4123–6
—— 1988a *IBM J. Res. Devel.* **32** 317–34
—— 1988b *Phys. Rev.* B **38** 9375–89
—— 1988c *IBM J. Res. Devel.* **32** 63–75
Büttiker M, Imry Y, Landauer R and Pinhas S 1985 *Phys. Rev.* B **31** 6207–15
Bychkov Y A and Rashba E I 1984 *J. Phys. C: Solid State Phys.* **17** 6039–45
Cahay M, McLennan M J and Datta S 1988 *Phys. Rev.* B **37** 10125–36
Cahay M, McLennan M, Datta S and Lundstrom M S 1987 *Appl. Phys. Lett.* **50** 612–14
Caldeira A O and Leggett A J 1983 *Ann. Phys., NY* **149** 374–456
Capasso F 1989a *IEEE Trans. Electron Dev.* **ED-36** 2065–82
Capasso F (ed) 1989b *Physics of Quantum Electron Devices* (*Springer Series in Electronics and Photonics* vol 28) (Berlin: Springer)
Capasso F, Mohammed K and Cho A Y 1986 *IEEE J. Quantum Electron.* **QE-22** 1853–69
Caroli C, Combescot R, Nozieres P and Saint-James D 1972 *J. Phys. C: Solid State Phys.* **5** 21–42
Chang, L. L, Esaki L and Tsu R 1974 *Appl. Phys. Lett.* **24** 593–5
Cheeks T L, Roukes M L, Scherer A and Craighead H G 1988 *Appl. Phys. Lett.* **53** 1964–6
Chou, S Y, Allee D R, Pease R F W and Harris J S Jr 1989 *Appl. Phys. Lett.* **55** 176–8
Chow D H, McGill T C, Sou I K, Faurier J P and Nieh C W 1988 *Appl. Phys. Lett.* **52** 54–6
D'Amato J L and Pastawski H M 1990 (to appear in *Phys. Rev.* B)
Das B, Miller D C, Datta S, Reifenberger R, Hong W P, Bhattacharya P K, Singh J and Jaffe M 1989 *Phys. Rev.* B **39** 1411–4
Datta S 1989a *Superlattices and Microstructures* **6** 83–93
—— 1989b *Phys. Rev.* B **40** 5830–3
Datta S and Das B 1990 *Appl. Phys. Lett.* **56** 665–7
Datta S and McLennan M J 1989 *Purdue University Technical Report* **TR-EE 89-12** 47–54
Datta S, Melloch M R, Bandyopadhyay S and Lundstrom M S 1986 *Appl. Phys. Lett.* **48** 487–9
Datta S, Melloch M R, Bandyopadhyay S, Noren R, Vaziri M, Miller M and Reifenberger R 1985 *Phys. Rev. Lett.* **55** 2344–7
Demel T, Heitmann D, Grambow P and Ploog K 1988 *Appl. Phys. Lett.* **53** 2176–8
deVegvar P G N, Timp G, Mankiewich P M, Behringer R and Cunningham J 1989 *Phys. Rev.* B **40** 3491–4
deVegvar P G N, Timp G, Mankiewich P M, Cunningham J E, Behringer R and Howard R E 1988 *Phys. Rev.* B **38** 4326–8
DiVencenzo D P and Kane C L 1988 *Phys. Rev.* B **38** 3006–15
Dohler G, Tsu R and Esaki L 1975 *Solid State Commun.* **17** 317–20

Doniach S and Sondheimer E H 1974 *Green's Functions for Solid State Physicists* (New York: Benjamin-Cummings)

Duke C B 1969 *Tunneling in Solids* (New York: Academic) (and references therein)

Eranen S and Sinkkonen J 1987 *Phys. Rev.* B **35** 2222–7

Esaki L 1958 *Phys. Rev.* **109** 603–4

—— 1974 *Rev. Mod. Phys.* **46** 237–44

—— 1986 *IEEE J. Quantum Electron.* QE-**22** 1611–24

Esaki L and Tsu R 1970 *IBM J. Res. Devel.* **14** 61–5

Ferry D K 1981 *Phys. Status Solidi* b **106** 63–71

Feynman R P and Vernon 1963 *Ann. Phys. NY* **24** 118

Fisher D S and Lee P A 1981 *Phys. Rev.* B **23** 6851–4

Ford C J B, Thornton T J, Newbury R, Pepper M, Ahmed H, Peacock D C, Ritchie D A, Frost J E F and Jones G A C 1988 *Phys. Rev.* B **38** 8518–21

—— 1989 *Appl. Phys. Lett.* **54** 21–3

Fowler A B 1985 *US Patent* no 4550330

—— 1988 *Workshop on Quantum Devices* (*Atlanta*)

Fowler A B, Wainer J J and Webb R A 1988 *IBM J. Res. Devel.* **32** 372–83

Frenkel J 1930 *Phys. Rev.* **36** 1604–18

Frensley W R 1985 *J. Vac. Sci. Technol.* B **3** 1261–6

—— 1986 *Phys. Rev. Lett.* **57** 2853–6

—— 1987 *Phys. Rev.* B **36** 1570–80

Gerhardts R R, Weiss D and von Klitzing K 1989 *Phys. Rev. Lett.* **62** 1173–6

Goldman V J, Tsui D C and Cunningham J E 1987 *Phys. Rev.* B **36** 7635–7

Hershfield S and Ambegaokar V 1988 *Phys. Rev.* B **38** 7909–12

Hirayama Y and Saku T 1989 *Appl. Phys. Lett.* **54** 2556–8

Hu P 1987 *Phys. Rev.* B **35** 4078–81

Iafrate G J and Krieger J B 1989 *Phys. Rev.* B **40** 6144–8

Imry Y 1986a Physics of mesoscopic systems *Directions in Condensed Matter Physics* ed G Grinstein and G Mazenko (Singapore: World Scientific) p 101

—— 1986b *Europhys. Lett.* **1** 249–56

—— 1989 *Nanostructure Physics and Fabrication* ed M A Reed and W P Kirk (San Diego: Academic) pp 379–88

Ismail K, Antoniadis D A and Smith H I 1989a *Appl. Phys. Lett.* **55** 589–91

Ismail K, Chu W, Antoniadis D A and Smith H I 1988 *Appl. Phys. Lett.* **52** 1071–3

Ismail K, Chu W, Yen A, Antoniadis D A and Smith H I 1989b *Appl. Phys. Lett.* **54** 460–2

Jain J K and Kivelson S A 1988 *Phys. Rev.* B **37** 4276–9

Jauho A P 1989 (to appear in *Solid State Electron.*)

Jauho A P and Ziep O 1989 *Phys. Scr.* T**25** 329–32

Jogai B and Wang K L 1985 *Appl. Phys. Lett.* **46** 167–8

Johnson M and Silsbee R H 1985 *Phys. Rev. Lett.* **55** 1790–3

Kadanoff L P and Baym G 1962 *Quantum Statistical Mechanics* (Reading, Mass: Benjamin)

Kane C L, Lee P A and DiVincenzo D P 1988 *Phys. Rev.* B **38** 2995–3005

Kane C L, Serota R A and Lee P A 1987 *Phys. Rev.* B **37** 6701–10

Kash K, Bhat R, Mahoney D D, Lin P S D, Scherer A, Worlock J M, Van der Gaag B P, Koza M and Grabbe P 1989 *Appl. Phys. Lett.* **55** 681–3

Keldysh L V 1965 *Sov. Phys.-JETP* **20** 1018

Kelly M J, Brown R J, Smith C G, Wharam D A, Pepper M, Ahmed H, Hasko D G, Peacock D C, Frost J E F, Newbury R, Ritchie D A and Jones G A C 1989 *Electron. Lett.* **25** 992–3

Kirchoefer S W, Newman H S and Comas J 1985 *Appl. Phys. Lett.* **46** 855–7

Kirczenow G 1988 *Phys. Rev.* B **38** 10958–61

Kluksdahl N C, Kriman A M, Ferry D K and Ringhofer C 1989 *Phys. Rev.* B **39** 7720–35

Kouwenhoven L P, van Wees B J, Harmans C J P M, Williamson J G, van Houten H, Beenakker C W J, Foxon C T and Harris J J 1989 *Phys. Rev.* B **39** 8040–3

Kubo R 1986 *Science* **233** 330–4

Landauer R 1957 *IBM J. Res. Devel.* **1** 223–31

—— 1970 *Phil. Mag.* **21** 863–7

—— 1987 *Z. Phys.* B **68** 217–28

—— 1988 *IBM J. Res. Devel.* **32** 306–16

—— 1989a *Phys. Today* **42**(10) 119–21

—— 1989b Electrons as guided waves in laboratory structures: strengths and problems *Preprint*

Landheer D, Liu H C, Buchanan M and Stoner R 1989 *Appl. Phys. Lett.* **54** 1784–6

Langreth D C 1976 *Linear and Nonlinear Electron Transport in Solids* ed J T Devreese and E Van Doren (New York: Plenum) p 2

Lee P A and Fisher D S 1981 *Phys. Rev. Lett.* **47** 882–5

Lee P A and Ramakrishnan T V 1985 *Rev. Mod. Phys.* **57** 287

Lee P A, Stone A D and Fukuyama H 1987 *Phys. Rev.* B **35** 1039–70

Leggett A J 1989 *Nanostructure Physics and Fabrication* ed M A Reed and W P Kirk (San Diego: Academic) pp 31–5

Lenstra D and Smokers R T M 1988 *Phys. Rev.* B **38** 6452–60

Likharev K K 1988 *IBM J. Res. Devel.* **32** 144–58

Lommer G, Malcher F and Rössler V 1988 *Phys. Rev. Lett.* **60** 728–31

Maekawa S, Isawa Y and Ebisawa H 1987 *J. Phys. Soc. Japan* **56** 25–8

Mahan G D 1987 *Phys. Rep.* **145** 251–318

Mains R K, Sun J P and Haddad G I 1989 *Appl. Phys. Lett.* **55** 371–3

Mason B A and Hess K 1989 *Phys. Rev.* B **39** 5051–69

Mendez E E, Calleja E, Goncalves da Silva C E T, Chang L L and Wang W I 1986 *Phys. Rev.* B **33** 7368–70

Miller D C, Lake R K, Datta S, Lundstrom M S, Melloch M R and Reifenberger R 1989 *Nanostructure Physics and Fabrication* ed M A Reed and W P Kirk (San Diego: Academic) pp 165–74

Muttalib K A, Pichard J-L and Stone A D 1987 *Phys. Rev. Lett.* **59** 2475–8

Ohnishi H, Inata T, Muto S, Yokoyama N and Shibatomi A 1986 *Appl. Phys. Lett.* **49** 1248–50

Peeters F M and Vasilopoulos P 1989 *Appl. Phys. Lett.* **55** 1106–8

Prange R E and Girvin S M (eds) 1987 *The Quantum Hall Effect* (Berlin: Springer)

Rammer J and Smith H 1986 *Rev. Mod. Phys.* **58** 323

Ravaioli U, Osman M A, Pötz W, Kluksdahl N and Ferry D K 1985 *Physica* **134B** 36–40

Reed M A, Frensley W R, Matyi R J, Randall J N and Seabaugh A C 1989 *Appl. Phys. Lett.* **54** 1034–6

Reed M A, Randall J N, Aggarwal R J, Matyi R J, Moore T M and Wetsel A E 1988 *Phys. Rev. Lett.* **60** 535–7

Roukes M L, Scherer A, Allen S J, Craighead H G, Ruihen R M, Beebe E D and Harbison J P 1987 *Phys. Rev. Lett.* **59** 3011–4

Sakaki H, Wagatsuma K, Hamasaki J and Saito S 1976 *Thin Solid Films* **36** 497

Serota R A, Feng S, Kane C, and Lee P A 1987 *Phys. Rev.* B **36** 5031–4

Skocpol W J, Mankiewich P M, Howard R E, Jackel L D and Tennant D M 1986 *Phys. Rev. Lett.* **56** 2865–8

Skocpol W J, Mankiewich P M, Howard R E, Jackel L D, Tennant D M and Stone A D 1987 *Phys. Rev. Lett.* **58** 2347–50

Smith C G, Pepper M, Newbury R, Ahmed H, Hasko D G, Peacock D C, Frost J E F, Ritchie D A, Jones G A C and Hill G 1989 *J. Phys.: Condens. Matter* **1** 6763–70

Söderström J R, Chow D H and McGill T C 1989 *Appl. Phys. Lett.* **55** 1348–50

Sollner T C L G, Brown E R and Le H Q 1988 *Lincoln Lab. J.* **1** 89–106

Sollner T C L G, Goodhue W D, Tannenwald P E, Parker C D and Peck D D 1983 *Appl. Phys. Lett.* **43** 588–90

Sols F, Macucci M, Ravaioli V and Hess K 1989 *Appl. Phys. Lett.* **54** 350–2

Stern A, Aharanov Y and Imry Y 1989 (preprint)

Stone A D 1985 *Phys. Rev. Lett.* **54** 2692–5

Stone A D and Lee P A 1985 *Phys. Rev. Lett.* **54** 1196–9

Stone A D and Szafer A 1988 *IBM J. Res. Devel.* **32** 384–413

Szafer A and Stone A D 1989 *Phys. Rev. Lett.* **62** 300–3

Taylor R P, Leadbeater M L, Whittington G P, Main P C, Eaves L, Beaumont S P, McIntyre I, Thoms S and Wilkinson C D W 1988 *Surf. Sci.* **196** 52–8

Thornber K 1978 *Solid State Electron.* **21** 259

Timp G, Baranger H U, deVegvar P, Cunningham J E, Howard R E, Behringer R and Mankiewich P M 1988 *Phys. Rev. Lett.* **60** 2081–4

Timp G, Behringer R, Sampere S, Cunningham J E and Howard R E 1989a *Nanostructure Physics and Fabrication* ed M A Reed and W P Kirk (San Diego: Academic) pp 331–45

Timp G, Chang A M, Cunningham J E, Chang T Y, Mankiewich P, Behringer R and Howard R E 1987a *Phys. Rev. Lett.* **58** 2814–7

Timp G, Chang A M, Mankiewich P, Behringer R, Cunningham J E, Chang T Y and Howard R E 1987b *Phys. Rev. Lett.* **59** 732–5

Timp G, Mankiewich P M, deVegvar P, Behringer R, Cunningham J E, Howard R E, Baranger H U and Jain J K 1989b *Phys. Rev.* B **39** 6227–30

Tsu R and Esaki L 1973 *Appl. Phys. Lett.* **22** 562–4

Tsubaki K, Tokura Y, Fukui T, Saito H and Susa N 1989 *Electron. Lett.* **25** 728

Umbach C P, Santhanam P, van Haesendonck C and Webb R A 1987 *Appl. Phys. Lett.* **50** 1289–91

van Houten H, Beenakker C W J, Broekaart M E I, Heigman M G J, van Wees B J, Mooij H E and Andre J B 1988 *Acta Electron.* **28** 27–38

van Houten H, van Wees B J, Heijman M G J and Andre J P 1986 *Appl. Phys. Lett.* **49** 1781–3

van Wees B J, Kouwenhoven L P, van Houten H, Beenakker C W J, Mooij J E, Foxon C T and Harris J J 1988b *Phys. Rev.* B **38** 3625–27

van Wees B J, van Houten H, Beenakker C W J, Williamson J G, Kouwenhoven L P, van der Marel D and Foxon C T 1988a *Phys. Rev. Lett.* **60** 848–50

van Wees B J, Willems E M M, Kouwenhoven L P, Harmans C J P M, Williamson J G, Foxon C T and Harris J J 1988c *Phys. Rev.* B **39** 8066–9

Vassell M O, Lee J and Lockwood H F 1983 *J. Appl. Phys.* **54** 5206–13

Warren A C, Antoniadis D A, Smith H I and Melngailis J 1985 *IEEE Electron Dev. Lett.* **EDL-6** 294–6

Washburn S 1988 *IBM J. Res. Devel.* **32** 335–46

Washburn S and Webb R A 1986 *Adv. Phys.* **35** 375–422.

Webb R A and Washburn S 1988 *Phys. Today* **41**(12) 47–53

Wharam D A, Pepper M, Ahmed H, Frost J E F, Hasko D G, Peacock D C, Ritchie D A and Jones G A C 1988b *J. Phys. C: Solid State Phys.* **21** L887–91

Wharam D A, Thornton T J, Newbury R, Pepper M, Ahmed H, Frost J E F, Hasko D G, Peacock D C, Ritchie D A and Jones G A C 1988a *J. Phys. C: Solid State Phys.* **21** L209–14

Whitaker J F, Mourou G A, Sollner T C L G and Goodhue W D 1988 *Appl. Phys. Lett.* **53** 385–7

Wigner E 1932 *Phys. Rev.* **40** 749–59

Wingreen N S, Jacobsen K W and Wilkins J W 1988 *Phys. Rev. Lett.* **61** 1396–9

Winkler R W, Kotthaus J P and Ploog K 1989 *Phys. Rev. Lett.* **62** 1177–80

Yamaguchi E 1985 *Phys. Rev.* B **32** 5280–8

Yamamoto M and Hohkawa K 1988 *Extended Abstracts of the 20th (1988 International) Conference on Solid State Devices and Materials* (Tokyo: Japan Society of Applied Physics) pp 495–8

Zaslavsky A, Goldman V J and Tsui D C 1988 *Appl. Phys. Lett.* **53** 1408–10