

# Coulomb-Blockade Oscillations in Semiconductor Nanostructures

H. van Houten, C. W. J. Beenakker, and A. A. M. Staring

*Philips Research Laboratories, 5600 JA Eindhoven, The Netherlands*

Published in *Single Charge Tunneling*, edited by H. Grabert and M. H. Devoret, NATO ASI Series B294 (Plenum, New York, 1992).

## Contents

<b>I. Introduction</b>	1
A. Preface	1
B. Basic properties of semiconductor nanostructures	3
<b>II. Theory of Coulomb-blockade oscillations</b>	5
A. Periodicity of the oscillations	5
B. Amplitude and lineshape	7
<b>III. Experiments on Coulomb-blockade oscillations</b>	11
A. Quantum dots	11
B. Disordered quantum wires	12
C. Relation to earlier work on disordered quantum wires	14
<b>IV. Quantum Hall effect regime</b>	15
A. The Aharonov-Bohm effect in a quantum dot	15
B. Coulomb blockade of the Aharonov-Bohm effect	19
C. Experiments on quantum dots	20
D. Experiments on disordered quantum wires	22
<b>Acknowledgments</b>	23
<b>A. Conductance of a quantum dot coupled to two electron reservoirs</b>	24
<b>References</b>	27

## I. INTRODUCTION

### A. Preface

Coulomb-blockade oscillations of the conductance are a manifestation of single-electron tunneling through a system of two tunnel junctions in series (see Fig. 1).<sup>1,2,3,4,5</sup> The conductance oscillations occur as the voltage on a nearby gate electrode is varied. The number  $N$  of conduction electrons on an island (or dot) between two tunnel barriers is an integer, so that the charge  $Q = -Ne$  on the island can only change by discrete amounts  $e$ . In contrast, the electrostatic potential difference of island and leads changes continuously as the electrostatic potential  $\phi_{\text{ext}}$  due to the gate is varied. This gives rise to a net charge imbalance  $C\phi_{\text{ext}} - Ne$  between the island and the leads, which oscillates in a saw-tooth pattern with gate voltage ( $C$  is the mutual capacitance of island and leads). Tunneling is blocked at low temperatures, except near the degeneracy points of the saw-tooth, where the charge imbalance jumps from  $+e/2$  to  $-e/2$ . At these points the Coulomb blockade of tunneling is lifted and the conductance exhibits a peak. In metals these “Coulomb-blockade oscillations” are essentially a

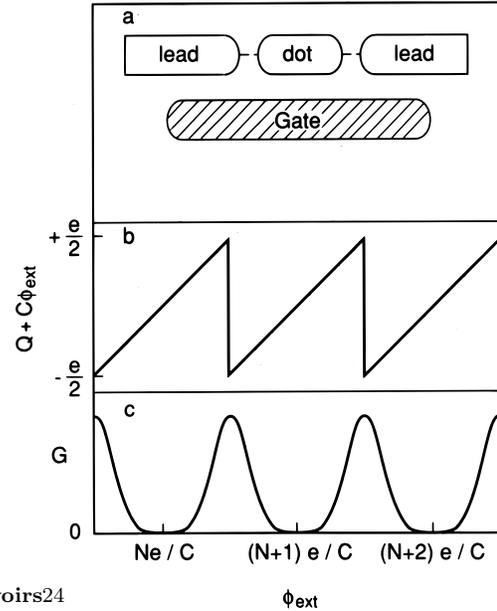


FIG. 1 (a) Schematic illustration of a confined region (dot) which is weakly coupled by tunnel barriers to two leads. (b) Because the charge  $Q = -Ne$  on the dot can only change by multiples of the elementary charge  $e$ , a charge imbalance  $Q + C\phi_{\text{ext}}$  arises between the dot and the leads. This charge imbalance oscillates in a saw-tooth pattern as the electrostatic potential  $\phi_{\text{ext}}$  is varied ( $\phi_{\text{ext}}$  is proportional to the gate voltage). (c) Tunneling is possible only near the charge-degeneracy points of the saw-tooth, so that the conductance  $G$  exhibits oscillations. These are the “Coulomb-blockade oscillations”.

classical phenomenon.<sup>6,7</sup> Because the energy level separation  $\Delta E$  in the island is much smaller than the thermal energy  $k_B T$ , the energy spectrum may be treated as a continuum. Furthermore, provided that the tunnel resistance is large compared to the resistance quantum  $h/e^2$ , the number  $N$  of electrons on the island may be treated as a sharply defined classical variable.

Coulomb-blockade oscillations can now also be studied in semiconductor nanostructures, which have a discrete energy spectrum. Semiconductor nanostructures are fabricated by lateral confinement of the two-dimensional electron gas (2DEG) in Si-inversion layers, or in GaAs-AlGaAs heterostructures. At low temperatures, the conduction electrons in these systems move over large distances (many  $\mu\text{m}$ ) without being scattered inelastically,

so that phase coherence is maintained. Residual elastic scattering by impurities or off the electrostatically defined sample boundaries does not destroy this phase coherence. The Fermi wavelength  $\lambda_F \sim 50$  nm in these systems is comparable to the size of the smallest structures that can now be made using electron-beam lithography. This has led to the discovery of a variety of quantum size effects in the ballistic transport regime. These effects may be adequately understood without considering electron-electron interactions.<sup>8</sup>

The first type of semiconductor nanostructure found to exhibit Coulomb-blockade oscillations is a narrow disordered wire, defined by a split-gate technique.<sup>9,10,11,12,13,14</sup> As shown in Fig. 2a, such a quantum wire may break up into disconnected segments if it is close to pinch-off. Conduction at low temperatures proceeds by tunneling through the barriers delimiting a segment, which plays the role of the central island in Fig. 1. The dominant oscillations in a wire typically have a well-defined periodicity, indicating that a single segment limits the conductance. Nevertheless, the presence of additional segments may give rise to multiple periodicities and to beating effects.

The second type of nanostructure exhibiting Coulomb-blockade oscillations is a small artificially confined region in a 2DEG (a quantum dot), connected by tunnel barriers either to narrow leads (Fig. 2b),<sup>15,16</sup> or to wide electron reservoirs (Fig. 2c).<sup>17</sup> The distinction between these two types of nanostructures is not fundamental, since a segment of a quantum wire delimited by two particularly strong scattering centers can be seen as a naturally formed quantum dot. Both types of structure are of interest:

Whereas artificially defined quantum dots are more suited to a study of the effect under relatively well-controlled conditions, the significance of the phenomenon of *periodic* conductance oscillations in *disordered* quantum wires lies in its bearing on the general problem of transport in disordered systems. It contradicts the presumed ubiquity of random conductance fluctuations in mesoscopic systems, and directly demonstrates the predominant role of electrostatic interactions in a disordered conductor.<sup>18</sup>

In a typical experiment, the segment of the wire, or the quantum dot, contains  $N \sim 100$  electrons, with an average energy level separation  $\Delta E \sim 0.2$  meV. At temperatures below a few Kelvin, the level spacing  $\Delta E$  exceeds the thermal energy  $k_B T$ , so that transport through the quantum dot proceeds by resonant tunneling. Resonant tunneling can by itself also lead to conductance oscillations as a function of gate voltage or Fermi energy. The interplay of resonant tunneling and the Coulomb blockade occurs when  $\Delta E$  and the charging energy  $e^2/C$  are of comparable magnitude (which is the case experimentally, where  $e^2/C \sim 1$  meV). This chapter reviews our current understanding of this interplay in semiconductor nanostructures. After a brief introduction to the properties of a 2DEG (based on Ref.<sup>8</sup>) we present in Sec. II a

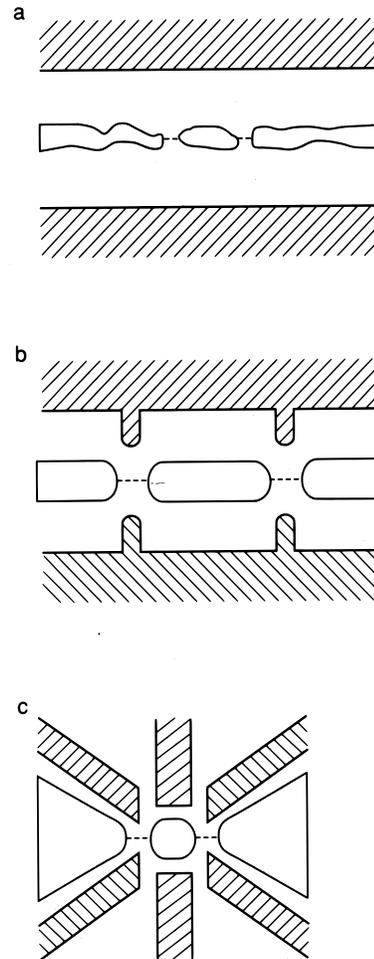


FIG. 2 Schematic top-view of three semiconductor nanostructures exhibiting Coulomb-blockade oscillations. Hatched regions denote gates, electron gas regions are shaded. Dashed lines indicate tunneling paths. (a) Disordered quantum wire with a single conductance limiting segment. (b) Quantum dot in a narrow channel. (c) Quantum dot between wide regions with separate sets of gates to modulate the tunnel barriers, and to vary the external potential of the dot.

discussion of the key results of a linear response theory for Coulomb-blockade oscillations in a quantum dot.<sup>19,20</sup> In Sec. III we review experimental results on quantum dots<sup>15,16,17</sup> and disordered quantum wires<sup>9,10,11,12,13,14</sup> in the absence of a magnetic field, and discuss to what extent they are now understood.

Kastner and collaborators<sup>9,10,15,21</sup> originally suggested that the conductance oscillations which they observed were due to the formation of a charge density wave or “Wigner crystal”. They inferred from a model due to Larkin and Lee,<sup>22</sup> and Lee and Rice,<sup>23</sup> that the conductance would be thermally activated because of the pinning of the charge density wave by impurities in the narrow channel. The activation energy would be determined by the most strongly pinned segment in the channel, and periodic oscillations in the conductance as a function of

gate voltage or electron density would reflect the condition that an integer number of electrons is contained between the two impurities delimiting that specific segment. A Wigner crystal is a manifestation of long-range order neglected in the theory of Coulomb-blockade oscillations. In a quantum wire with weak disorder (no tunnel barriers), a Wigner crystal may well be an appropriate description of the ground state.<sup>24</sup> The point of view adopted in this chapter, following Ref.<sup>25</sup>, is that the Coulomb blockade model is adequate for the present experiments in systems with artificial or natural tunnel barriers. We limit ourselves to a discussion of that model, and refer the reader to Ref.<sup>11</sup> for an exposition of the alternative point of view of Kastner and collaborators.

The Coulomb blockade and Wigner crystal models have in common that electron-electron interactions play a central role. In contrast, some authors have argued that resonant tunneling of non-interacting electrons can by itself explain the observed conductance oscillations.<sup>26,27</sup> We stress that one cannot discriminate between these two models on the basis of the periodicity of the oscillations. Conductance oscillations due to resonant tunneling through non-degenerate levels as well as Coulomb-blockade oscillations both have a periodicity corresponding to the addition of a single electron to the confined region. Other considerations (notably the absence of spin-splitting of the peaks in a magnetic field, and the large activation energy — by far exceeding  $\Delta E$ ) are necessary to demonstrate the inadequacy of a model based on resonant tunneling of non-interacting electrons.

Semiconductor nanostructures offer the additional intriguing possibility to study single-electron tunneling in the quantum Hall effect regime. This is the subject of Sec. IV. In this regime of a strong magnetic field, the one-electron states are extended along equipotential contours.<sup>8</sup> The contours of subsequent states within the same Landau level enclose one extra flux quantum  $h/e$ . States at the Fermi level are edge states circulating along the circumference of the quantum dot. If charging effects are negligible oscillations in the conductance of the dot are observed as a function of gate voltage or magnetic field, due to resonant tunneling through circulating edge states.<sup>28</sup> This is a manifestation of the Aharonov-Bohm effect, normally associated with magnetoconductance oscillations in a ring, rather than a dot. Circulating edge states, however, make the dot behave effectively as a ring<sup>29</sup> — at least for non-interacting electrons. As we will discuss, the single-electron charging energy can cause a “Coulomb blockade” of the Aharonov-Bohm effect in a quantum dot.<sup>30,31</sup> The magnetoconductance oscillations are suppressed when  $e^2/C$  becomes comparable to the Landau level separation  $\hbar\omega_c$  (with  $\omega_c = eB/m$ ). However, the periodic oscillations as a function of gate voltage remain. This difference illustrates how in the presence of charging effects magnetic and electrostatic fields play fundamentally different roles,<sup>12</sup> in contrast to the equivalent roles played in the diffusive or ballistic transport

regimes.\* An additional topic covered in Sec. IV is the effect of a magnetic field on the amplitude and position of the oscillations, from which detailed information can be obtained on the one-electron energy spectrum of the quantum dot.<sup>32</sup>

In this chapter we consider the Coulomb-blockade oscillations in zero magnetic field and in the *integer* quantum Hall effect regime. The generalization to the *fractional* quantum Hall effect is still an open problem, at least experimentally. Some theoretical considerations have been given,<sup>33</sup> but will not be considered here. We limit ourselves to the linear response regime, and do not discuss the non-linear current-voltage characteristics.<sup>34,35</sup> In metallic tunnel junctions with very different tunnel rates through the two barriers one finds steps in the current as a function of source-drain voltage.<sup>1,2</sup> This “Coulomb staircase” has recently also been observed in a quantum dot.<sup>36</sup> A third limitation is to stationary transport phenomena, so that we do not consider the effects of radio-frequency modulation of the source-drain or gate voltages. A new development in metals is the realization of a “turnstile clocking” of the current through an array of junctions at a value  $ef$ , with  $f$  the frequency of the modulation of the voltage on a gate.<sup>37,38</sup> These effects very recently also been observed in a quantum dot.<sup>36</sup> Concerning the types of sample, we limit ourselves to quantum dots and wires defined by a split-gate in a two-dimensional electron gas. Quantum dots may also be defined by etching a pillar out of a quantum well.<sup>39,40</sup> Such “vertical” structures have the advantage over the planar structures considered here that the thickness and height of the potential barriers separating the quantum dot from the leads can be tailored to a great precision during the epitaxial growth. A disadvantage is that it is more difficult to change the carrier density in the dot by means of a gate electrode.<sup>41</sup> In the planar structures based on a 2DEG not only the electron density, but also the geometry can be varied continuously using gates.

## B. Basic properties of semiconductor nanostructures

Electrons in a two-dimensional electron gas (2DEG) are constrained to move in a plane, due to a strong electrostatic confinement at the interface between two semiconductor layers (in the case of a GaAs-AlGaAs heterostructure), or at the interface between a semiconductor and an insulator (in the case of a Si-inversion layer, where the insulator is SiO<sub>2</sub>). The areal density  $n_s$  may be varied continuously by changing the voltage on a

---

\* Examples of this equivalence are the fluctuations in the conductance as a function of gate voltage or magnetic field due to quantum interference, and the sequence of quantized conductance plateaux (at integer multiples of  $e^2/h$ ) as a result of magnetic or electrostatic depopulation of one-dimensional subbands.<sup>8</sup>

gate electrode deposited on the top semiconductor layer (in which case isolation is provided automatically by a Schottky barrier) or on the insulator. The gate voltage is defined with respect to an ohmic contact to the 2DEG. The density under a gate electrode of large area changes linearly with the electrostatic potential of the gate  $\phi_{\text{gate}}$ , according to the plate capacitor formula

$$\delta n_s = \frac{\epsilon}{ed} \delta \phi_{\text{gate}}, \quad (1.1)$$

where  $\epsilon$  is the dielectric constant of the material of thickness  $d$  between gate and 2DEG. For GaAs  $\epsilon = 13\epsilon_0$ , whereas SiO<sub>2</sub> has  $\epsilon = 3.9\epsilon_0$ .

A unique feature of a 2DEG is that it can be given any desired shape using lithographic techniques. The shape is defined by etching a pattern (resulting in a permanent removal of the electron gas), or by electrostatic depletion using a patterned gate electrode (which is reversible). A local (partial) depletion of the 2DEG under a gate is associated with a local increase of the electrostatic potential, relative to the undepleted region. At the boundaries of the gate a potential step is thus induced in the 2DEG. The potential step is smooth, because of the large depletion length (of the order of 100 nm for a step height of 10 meV). This large depletion length is at the basis of the split-gate technique, used to define narrow channels of variable width with smooth boundaries.

The energy of non-interacting conduction electrons in an unbounded 2DEG is given by

$$E(k) = \frac{\hbar^2 k^2}{2m}, \quad (1.2)$$

as a function of momentum  $\hbar k$ . The effective mass  $m$  is considerably smaller than the free electron mass  $m_e$  as a result of interactions with the lattice potential (for GaAs  $m = 0.067 m_e$ , for Si  $m = 0.19 m_e$ , both for the (100) crystal plane). The density of states  $\rho_{2D}(E) \equiv dn(E)/dE$  is the derivative of the number of electronic states  $n(E)$  (per unit surface area) with energy smaller than  $E$ . In  $k$ -space, these states fill a circle of area  $A = 2\pi mE/\hbar^2$  [according to Eq. (1.2)], containing a number  $g_s g_v A/(2\pi)^2$  of states. The factors  $g_s$  and  $g_v$  account for the spin and valley-degeneracy, respectively (in GaAs  $g_v = 1$ , in Si  $g_v = 2$ ;  $g_s = 2$  in zero magnetic field). One thus finds  $n(E) = g_s g_v mE/2\pi\hbar^2$ , so that the density of states per unit area,

$$\rho_{2D} = g_s g_v \frac{m}{2\pi\hbar^2}, \quad (1.3)$$

is *independent* of the energy. In equilibrium, the states are occupied according to the Fermi-Dirac distribution function

$$f(E - E_F) = \left[ 1 + \exp\left(\frac{E - E_F}{k_B T}\right) \right]^{-1}. \quad (1.4)$$

At low temperatures  $k_B T \ll E_F$ , the Fermi energy (or chemical potential)  $E_F$  of a 2DEG is thus directly proportional to its sheet density  $n_s$ , according to

$$E_F = n_s / \rho_{2D}. \quad (1.5)$$

The Fermi wave number  $k_F \equiv (2mE_F/\hbar^2)^{1/2}$  is related to the density by  $k_F = (4\pi n_s/g_s g_v)^{1/2}$ . Typically,  $E_F \sim 10$  meV, so that the Fermi wavelength  $\lambda_F \equiv 2\pi/k_F \sim 50$  nm.

If the 2DEG is confined laterally to a narrow channel, then Eq. (2) only represents the kinetic energy from the free motion (with momentum  $\hbar k$ ) *parallel* to the channel axis. Because of the lateral confinement, the conduction band is split itself into a series of one-dimensional (1D) subbands, with band bottoms at  $E_n$ ,  $n = 1, 2, \dots$ . The total energy  $E_n(k)$  of an electron in the  $n$ -th 1D subband is given by

$$E_n(k) = E_n + \frac{\hbar^2 k^2}{2m}, \quad (1.6)$$

in zero magnetic field. Two frequently used potentials to model analytically the lateral confinement are the square well potential (of width  $W$ ), and the parabolic potential well (described by  $V(x) = \frac{1}{2}m\omega_0^2 x^2$ ). The confinement levels are given by  $E_n = (n\pi\hbar)^2/2mW^2$ , and  $E_n = (n - \frac{1}{2})\hbar\omega_0$ , respectively.

Transport through a very short quantum wire (of length  $L \sim 100$  nm, much shorter than the mean free path) is perfectly ballistic. When such a short and narrow wire forms a constriction between two wide electron gas reservoirs, one speaks of a quantum point contact.<sup>42</sup> The conductance  $G$  of a quantum point contact is quantized in units of  $2e^2/h$ .<sup>42,44</sup> This effect requires a unit transmission probability for all of the occupied 1D subbands in the point contact, each of which then contributes  $2e^2/h$  to the conductance (for  $g_s g_v = 2$ ). Potential fluctuations due to the random distribution of ionized donors have so far precluded any observation of the conductance quantization in longer quantum wires (even if they are considerably shorter than the mean free path in wide 2DEG regions). Quantum wires are extremely sensitive to disorder, since the effective scattering cross-section, being of the order of the Fermi wavelength, is comparable to the width of the wire. Indeed, calculations demonstrate<sup>45</sup> that a quantum wire close to pinch-off breaks up into a number of isolated segments. The Coulomb-blockade oscillations in a quantum wire discussed in Sec. III are associated with tunneling through the barriers separating these segments (see Fig. 2a).

A quantum dot is formed in a 2DEG if the electrons are confined in all three directions. The energy spectrum of a quantum dot is fully discrete. Transport through the discrete states in a quantum dot can be studied if tunnel barriers are defined at its perimeter. The quantum dots discussed in Sec. III are connected by quantum point contacts to their surroundings (see Figs. 2b and 2c). The quantum point contacts are operated close to pinch-off ( $G < 2e^2/h$ ), where they behave as tunnel barriers of adjustable height and width. The shape of such barriers differs greatly from that encountered in metallic tunnel junctions: the barrier height typically exceeds the Fermi energy by only a few meV, and the thickness of the barrier at  $E_F$  is large, on the order of 50 nm. This may

lead to a strong energy dependence of the tunnel rates, not encountered in metals.

## II. THEORY OF COULOMB-BLOCKADE OSCILLATIONS

Part of the interest in quantum dots derives from the fact that their electronic structure mimicks that of an isolated atom — with the fascinating possibility to attach wires to this “atom” and study transport through its discrete electronic states. In this section we address this problem from a theoretical point of view, following Ref.<sup>19</sup>.

### A. Periodicity of the oscillations

We consider a quantum dot, which is weakly coupled by tunnel barriers to two electron reservoirs. A current  $I$  can be passed through the dot by applying a voltage difference  $V$  between the reservoirs. The linear response conductance  $G$  of the quantum dot is defined as  $G \equiv I/V$ , in the limit  $V \rightarrow 0$ . Since transport through a quantum dot proceeds by tunneling through its discrete electronic states, it will be clear that for small  $V$  a net current can flow only for certain values of the gate voltage (if  $\Delta E \gg k_B T$ ). In the absence of charging effects, a conductance peak due to resonant tunneling occurs when the Fermi energy  $E_F$  in the reservoirs lines up with one of the energy levels in the dot. This condition is modified by the charging energy. To determine the location of the conductance peaks as a function of gate voltage requires only consideration of the equilibrium properties of the system,<sup>19,30</sup> as we now discuss.

The probability  $P(N)$  to find  $N$  electrons in the quantum dot in equilibrium with the reservoirs is given by the grand canonical distribution function

$$P(N) = \text{constant} \times \exp\left(-\frac{1}{k_B T}[F(N) - NE_F]\right), \quad (2.1)$$

where  $F(N)$  is the free energy of the dot and  $T$  the temperature. The reservoir Fermi energy  $E_F$  is measured relative to the conduction band bottom in the reservoirs. In general,  $P(N)$  at  $T = 0$  is non-zero for a *single* value of  $N$  only (namely the integer which minimizes the thermodynamic potential  $\Omega(N) \equiv F(N) - NE_F$ ). In that case,  $G \rightarrow 0$  in the limit  $T \rightarrow 0$ . As pointed out by Glazman and Shekhter,<sup>5</sup> a non-zero  $G$  is possible only if  $P(N)$  and  $P(N + 1)$  are both non-zero for some  $N$ . Then a small applied voltage is sufficient to induce a current through the dot, via intermediate states  $N \rightarrow N + 1 \rightarrow N \rightarrow N + 1 \rightarrow \dots$ . To have  $P(N)$  and  $P(N + 1)$  both non-zero at  $T = 0$  requires that both  $N$  and  $N + 1$  minimize  $\Omega$ . A necessary condition is  $\Omega(N + 1) = \Omega(N)$ , or

$$F(N + 1) - F(N) = E_F. \quad (2.2)$$

This condition is also sufficient, unless  $\Omega$  has more than one minimum (which is usually not the case).

Equation (2.2) expresses the equality of the electrochemical potential of dots and leads. The usefulness of this result is that it maps the problem of determining the location of the conductance peaks onto the more familiar problem of calculating the electrochemical potential  $F(N + 1) - F(N)$  of the quantum dot, i.e. the energy cost associated with the addition of a single electron to the dot. This opens the way, in principle, to a study of exchange and correlation effects on the conductance oscillations in a quantum dot (e.g. along the lines of work by Bryant<sup>46</sup> and by Maksym and Chakraborty<sup>47</sup>).

At  $T = 0$  the free energy  $F(N)$  equals the ground state energy of the dot, for which we take the simplified form  $U(N) + \sum_{p=1}^N E_p$ . Here  $U(N)$  is the charging energy, and  $E_p$  ( $p = 1, 2, \dots$ ) are single-electron energy levels in ascending order. The term  $U(N)$  accounts for the charge imbalance between dot and reservoirs. The sum over energy levels accounts for the internal degrees of freedom of the quantum dot, evaluated in a mean-field approximation (cf. Ref.<sup>48</sup>). Each level contains either one or zero electrons. Spin degeneracy, if present, can be included by counting each level twice, and other degeneracies can be included similarly. The energy levels  $E_p$  depend on gate voltage and magnetic field, but are assumed to be independent of  $N$ , at least for the relevant range of values of  $N$ . We conclude from Eq. (2.2) that a peak in the low-temperature conductance occurs whenever

$$E_N + U(N) - U(N - 1) = E_F, \quad (2.3)$$

for some integer  $N$  (we have relabeled  $N$  by  $N - 1$ ).

We adopt the simple approximation of the orthodox model<sup>4</sup> of taking the charging energy into account macroscopically. We write  $U(N) = \int_0^{-Ne} \phi(Q') dQ'$ , where

$$\phi(Q) = Q/C + \phi_{\text{ext}} \quad (2.4)$$

is the potential difference between dot and reservoir, including also a contribution  $\phi_{\text{ext}}$  from external charges (in particular those on a nearby gate electrode). The capacitance  $C$  is assumed to be independent of  $N$  (at least over some interval). The charging energy then takes the form

$$U(N) = (Ne)^2/2C - Ne\phi_{\text{ext}}. \quad (2.5)$$

To make connection with some of the literature<sup>3,49</sup> we mention that  $Q_{\text{ext}} \equiv C\phi_{\text{ext}}$  plays the role of an “externally induced charge” on the dot, which can be varied continuously by means of an external gate voltage (in contrast to  $Q$  which is restricted to integer multiples of  $e$ ). In terms of  $Q_{\text{ext}}$  one can write

$$U(N) = (Ne - Q_{\text{ext}})^2/2C + \text{constant},$$

which is equivalent to Eq. (2.5). We emphasize that  $Q_{\text{ext}}$  is an externally controlled variable, via the gate voltage,

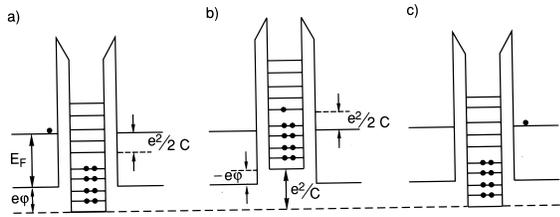


FIG. 3 Single-electron tunneling through a quantum dot, under the conditions of Eq. (2.6), for the case that the charging energy is comparable to the level spacing. An infinitesimally small voltage difference is assumed between the left and right reservoirs. (From Beenakker et al.<sup>31</sup>)

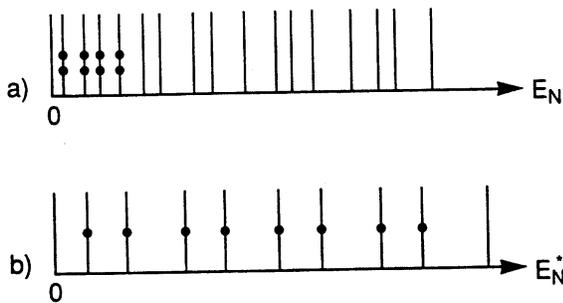


FIG. 4 Diagram of the bare energy levels (a) and the renormalized energy levels (b) in a quantum dot for the case  $e^2/C \approx 2\langle\Delta E\rangle$ . The renormalized level spacing is much more regular than the average bare level spacing  $\langle\Delta E\rangle$ . Note that the spin degeneracy of the bare levels is lifted by the charging energy. (From Staring et al.<sup>12</sup>)

regardless of the relative magnitude of the various capacitances in the system.

Substitution of Eq. (2.5) into Eq. (2.3) gives

$$E_N^* \equiv E_N + (N - \frac{1}{2}) \frac{e^2}{C} = E_F + e\phi_{\text{ext}} \quad (2.6)$$

as the condition for a conductance peak. The left-hand-side of Eq. (2.6) defines a renormalized energy level  $E_N^*$ . The renormalized level spacing  $\Delta E^* = \Delta E + e^2/C$  is enhanced above the bare level spacing by the charging energy. In the limit  $e^2/C\Delta E \rightarrow 0$ , Eq. (2.6) is the usual condition for resonant tunneling. In the limit  $e^2/C\Delta E \rightarrow \infty$ , Eq. (2.6) describes the periodicity of the classical Coulomb-blockade oscillations in the conductance versus electron density.<sup>3,4,5,6,7</sup>

In Fig. 3 we have illustrated the tunneling of an electron through the dot under the conditions of Eq. (2.6). In panel (a) one has  $E_N + e^2/2C = E_F + e\phi(N - 1)$ , with  $N$  referring to the lowest unoccupied level in the dot. In panel (b) an electron has tunneled into the dot. One now has  $E_N - e^2/2C = E_F + e\phi(N)$ , with  $N$  referring to the highest occupied level. The potential difference  $\phi$  between dot and reservoir has decreased by  $e/C$  (becoming negative), because of the added electron. Finally,

in panel (c) the added electron tunnels out of the dot, resetting the potentials to the initial state of panel (a).

Let us now determine the periodicity of the oscillations. Theoretically, it is convenient to consider the case of a variation of the Fermi energy of the reservoirs at constant  $\phi_{\text{ext}}$ . The periodicity  $\Delta E_F$  follows from Eq. (2.6),

$$\Delta E_F = \Delta E^* \equiv \Delta E + \frac{e^2}{C}. \quad (2.7)$$

In the absence of charging effects,  $\Delta E_F$  is determined by the irregular spacing  $\Delta E$  of the single-electron levels in the quantum dot. The charging energy  $e^2/C$  regulates the spacing, once  $e^2/C \gtrsim \Delta E$ . This is illustrated in Fig. 4, for the case that there is no valley degeneracy. The spin degeneracy of the levels is lifted by the charging energy. In a plot of  $G$  versus  $E_F$  this leads to a doublet structure of the oscillations, with a spacing alternating between  $e^2/C$  and  $\Delta E + e^2/C$ .

Experimentally, one studies the Coulomb-blockade oscillations as a function of gate voltage. To determine the periodicity in that case, we first need to know how  $E_F$  and the set of energy levels  $E_p$  depend on  $\phi_{\text{ext}}$ . In a 2DEG, the external charges are supplied by ionized donors and by a gate electrode (with an electrostatic potential difference  $\phi_{\text{gate}}$  between gate and 2DEG reservoir). One has

$$\phi_{\text{ext}} = \phi_{\text{donors}} + \alpha\phi_{\text{gate}}, \quad (2.8)$$

where  $\alpha$  (as well as  $C$ ) is a rational function of the capacitance matrix elements of the system. The value of  $\alpha$  depends on the geometry. Here we consider only the geometry of Figs. 2a,b in detail, for which it is reasonable to assume that the electron gas densities in the dot and in the leads increase, on average, equally fast with  $\phi_{\text{gate}}$ . For equidistant energy levels in the dot we may then assume that  $E_F - E_N$  has the same value at each conductance peak. The period of the oscillations now follows from Eqs. (2.6) and (2.8),

$$\Delta\phi_{\text{gate}} = \frac{e}{\alpha C}. \quad (2.9)$$

To clarify the meaning of the parameters  $C$  and  $\alpha$ , we represent the system of dot, gates and leads in Figs. 2a,b by the equivalent circuit of Fig. 5. The mutual capacitance of gates and leads does not enter our problem explicitly, since it is much larger than the mutual capacitances of gate and dot ( $C_{\text{gate}}$ ) and dot and leads ( $C_{\text{dot}}$ ). The capacitance  $C$  determining the charging energy  $e^2/C$  is formed by  $C_{\text{gate}}$  and  $C_{\text{dot}}$  in parallel,

$$C = C_{\text{gate}} + C_{\text{dot}}. \quad (2.10)$$

The period of the oscillations corresponds in our approximation of equidistant energy levels ( $E_F - E_N = \text{constant}$ ) to the increment by  $e$  of the charge on the dot with no change in the voltage across  $C_{\text{dot}}$ . This implies  $\Delta\phi_{\text{gate}} = e/C_{\text{gate}}$ , or

$$\alpha = C_{\text{gate}}/(C_{\text{gate}} + C_{\text{dot}}). \quad (2.11)$$

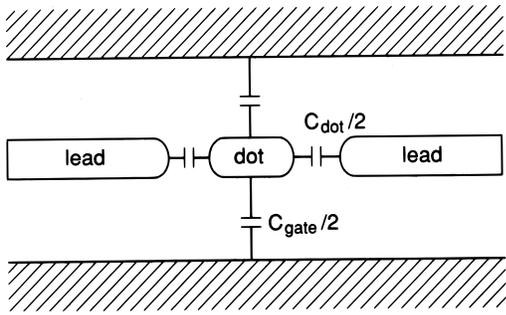


FIG. 5 Equivalent circuit of quantum dot and split gate. The mutual capacitance of leads and gate is much larger than that of the dot and the split gate ( $C_{\text{gate}}$ ), or the dot and the leads ( $C_{\text{dot}}$ ), and can be neglected.

Thus, in terms of the electrostatic potential difference between gate and 2DEG reservoirs, the period of the conductance oscillations is  $\Delta\phi_{\text{gate}} = e/C_{\text{gate}}$ . Note that this result applies regardless of the relative magnitudes of the bare level spacing  $\Delta E$  and the charging energy  $e^2/C$ .

In an experiment the gate voltage is the *electrochemical* potential difference  $V_{\text{gate}}$  between gate and leads, i.e. the difference in Fermi level, whereas so far we have discussed the period of the oscillations in terms of the *electrostatic* potential difference  $\phi_{\text{gate}}$ , i.e. the difference in conduction band bottoms. In one period, the change in Fermi energy in the dot and leads (measured with respect to their local conduction band bottom) is approximately equal to  $\Delta E$ . The change in Fermi energy in the (metal) gate is negligible, because the density of states in a metal is much larger than in a 2DEG. We thus find that the oscillation period  $\Delta V_{\text{gate}}$  in the geometry of Figs. 2a,b is

$$\Delta V_{\text{gate}} = \frac{\Delta E}{e} + \Delta\phi_{\text{gate}} = \frac{\Delta E}{e} + \frac{e}{C_{\text{gate}}}. \quad (2.12)$$

Note that  $C_{\text{dot}}$  does not affect the periodicity. In many of the present experiments  $\Delta E$  is a factor of 10 below  $e^2/C_{\text{gate}}$ , so that the differences between  $\Delta\phi_{\text{gate}}$  and  $\Delta V_{\text{gate}}$  are less than 10 %. Even in such a case, these differences are quite important, since their study yields direct information on the energy spectrum of the quantum dot.

In the case of a two-fold spin-degeneracy, the level separation  $E_{p+1} - E_p$  in a dot of area  $A$  alternates between 0 and  $\Delta E \sim 2\pi\hbar^2/mA$  [cf. Eq. (1.3)]. As mentioned above, this leads to a doublet structure of the oscillations as a function of  $E_F$ . To determine the peak spacing as a function of gate voltage we approximate the change in  $E_F$  with  $\phi_{\text{gate}}$  by  $\partial E_F/\partial\phi_{\text{gate}} \sim \Delta EC_{\text{gate}}/2e$ . We then obtain from Eqs. (2.6), (2.8), (2.10), and (2.11) that the spacing alternates between two values:

$$\Delta\phi_{\text{gate}}^{(1)} = \left(\frac{e}{C_{\text{gate}}}\right) \frac{e^2/C}{\Delta E/2 + e^2/C}, \quad (2.13)$$

$$\Delta\phi_{\text{gate}}^{(2)} = \left(\frac{e}{C_{\text{gate}}}\right) \frac{\Delta E + e^2/C}{\Delta E/2 + e^2/C}. \quad (2.14)$$

The average spacing equals  $e/C_{\text{gate}}$ , in agreement with Eq. (2.9) [derived for non-degenerate equidistant levels]. To obtain  $\Delta V_{\text{gate}}$  one has to add  $\Delta E/2e$  to the factor  $e/C_{\text{gate}}$  between brackets in Eqs. (2.13) and (2.14). If the charging energy dominates ( $e^2/C \gg \Delta E$ ), one has equal spacing  $\Delta\phi_{\text{gate}}^{(1)} = \Delta\phi_{\text{gate}}^{(2)} = e/C_{\text{gate}}$ , as for non-degenerate levels. In the opposite limit  $\Delta E \gg e^2/C$ , one finds instead  $\Delta\phi_{\text{gate}}^{(1)} = 0$ , and  $\Delta\phi_{\text{gate}}^{(2)} = 2e/C_{\text{gate}}$ . Thus, the period is effectively doubled, corresponding to the addition of *two* electrons to the dot, instead of one. This is characteristic for resonant tunneling of non-interacting electrons through two-fold spin-degenerate energy levels. An external magnetic field will resolve the spin-degeneracy, leading to a splitting of the conductance peaks which increases with the field.

## B. Amplitude and lineshape

Equation (2.6) is sufficient to determine the periodicity of the conductance oscillations, but gives no information on their amplitude and width, which requires the solution of a kinetic equation. For the linear response conductance in the resonant tunneling regime an analytical solution has been derived by Beenakker,<sup>19</sup> which generalizes earlier results by Kulik and Shekhter<sup>7</sup> in the classical regime. Equivalent results have been obtained independently by Meir, Wingreen, and Lee.<sup>20</sup> Related work on the non-linear current-voltage characteristics has been performed by Averin, Korotkov, and Likharev,<sup>34</sup> and by Groshev.<sup>35</sup> In this sub-section we summarize the main results of Ref.<sup>19</sup>, along with the underlying assumptions.

A continuum of states is assumed in the reservoirs, which are occupied according to the Fermi-Dirac distribution (1.4). The tunnel rate from level  $p$  to the left and right reservoirs is denoted by  $\Gamma_p^l$  and  $\Gamma_p^r$ , respectively. We assume that  $k_B T \gg h(\Gamma^l + \Gamma^r)$  (for all levels participating in the conduction), so that the finite width  $h\Gamma = h(\Gamma^l + \Gamma^r)$  of the transmission resonance through the quantum dot can be disregarded. This assumption allows us to characterize the state of the quantum dot by a set of occupation numbers, one for each energy level. (As we will discuss, in the classical regime  $k_B T \gg \Delta E$  the condition  $\Delta E \gg h\Gamma$  takes over from the condition  $k_B T \gg h\Gamma$  appropriate for the resonant tunneling regime.) We assume here that inelastic scattering takes place exclusively in the reservoirs — not in the quantum dot. (The effects of inelastic scattering in the dot for  $k_B T \gg h\Gamma$  are discussed in Ref.<sup>19</sup>.)

The equilibrium distribution function of electrons among the energy levels is given by the Gibbs distribution in the grand canonical ensemble:

$$P_{\text{eq}}(\{n_i\}) = \frac{1}{Z} \exp \left[ -\frac{1}{k_B T} \left( \sum_{i=1}^{\infty} E_i n_i + U(N) - N E_F \right) \right], \quad (2.15)$$

where  $\{n_i\} \equiv \{n_1, n_2, \dots\}$  denotes a specific set of occu-

pation numbers of the energy levels in the quantum dot. (The numbers  $n_i$  can take on only the values 0 and 1.) The number of electrons in the dot is  $N \equiv \sum_i n_i$ , and  $Z$  is the partition function,

$$Z = \sum_{\{n_i\}} \exp \left[ -\frac{1}{k_B T} \left( \sum_{i=1}^{\infty} E_i n_i + U(N) - N E_F \right) \right]. \quad (2.16)$$

The joint probability  $P_{\text{eq}}(N, n_p = 1)$  that the quantum dot contains  $N$  electrons *and* that level  $p$  is occupied is

$$P_{\text{eq}}(N, n_p = 1) = \sum_{\{n_i\}} P_{\text{eq}}(\{n_i\}) \delta_{N, \sum_i n_i} \delta_{n_p, 1}. \quad (2.17)$$

In terms of this probability distribution, the conductance is given by

$$G = \frac{e^2}{k_B T} \sum_{p=1}^{\infty} \sum_{N=1}^{\infty} \frac{\Gamma_p^l \Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} P_{\text{eq}}(N, n_p = 1) \times [1 - f(E_p + U(N) - U(N-1) - E_F)]. \quad (2.18)$$

This particular product of distribution functions expresses the fact that tunneling of an electron from an initial state  $p$  in the dot to a final state in the reservoir requires an occupied initial state and empty final state. Equation (2.18) was derived in Ref.<sup>19</sup> by solving the kinetic equation in linear response. This derivation is presented in the appendix. The same formula has been obtained independently by Meir, Wingreen, and Lee,<sup>20</sup> by solving an Anderson model in the limit  $k_B T \gg \hbar \Gamma$ .

We will now discuss some limiting cases of the general result (2.18). We first consider the conductance of the individual barriers and the quantum dot in the high temperature limit  $k_B T \gg e^2/C, \Delta E$  where neither the discreteness of the energy levels nor the charging energy are important. The conductance then does not exhibit oscillations as a function of gate voltage. The high temperature limit is of interest for comparison with the low temperature results, and because its measurement allows a straightforward estimate of the tunnel rates through the barriers. The conductance of the quantum dot in the high temperature limit is simply that of the two tunnel barriers in series

$$G = \frac{G^l G^r}{G^l + G^r}, \quad \text{if } \Delta E, e^2/C \ll k_B T \ll E_F. \quad (2.19)$$

The conductances  $G^l, G^r$  of the left and right tunnel barriers are given by the thermally averaged Landauer formula

$$G^{l,r} = -\frac{e^2}{h} \int_0^{\infty} dE T^{l,r}(E) \frac{df}{dE}. \quad (2.20)$$

The transmission probability of a barrier  $T(E)$  equals the tunnel rate  $\Gamma(E)$  divided by the attempt frequency  $\nu(E) = 1/\hbar \rho(E)$ ,

$$T^{l,r}(E) = \hbar \Gamma^{l,r}(E) \rho(E). \quad (2.21)$$

If the height of the tunnel barriers is large, the energy dependence of the tunnel rates and of the density of states  $\rho$  in the dot can be ignored (as long as  $k_B T \ll E_F$ ). The conductance of each barrier from Eq. (2.20) then becomes

$$G^{l,r} = (e^2/h) T^{l,r} = e^2 \Gamma^{l,r} \rho \quad (2.22)$$

(where  $T, \Gamma$ , and  $\rho$  are evaluated at  $E_F$ ), and the conductance of the quantum dot from Eq. (2.19) is

$$G = e^2 \rho \frac{\Gamma^l \Gamma^r}{\Gamma^l + \Gamma^r} = \frac{e^2}{h} \frac{T^l T^r}{T^l + T^r} \equiv G_{\infty}, \quad \text{if } \Delta E, e^2/C \ll k_B T \ll E_F. \quad (2.23)$$

The conductance  $G_{\infty}$  in the high temperature limit depends only on the barrier height and width (which determine  $T$ ), not on the area of the quantum dot (which determines  $\rho$  and  $\Gamma$ , but cancels in the expression for  $G_{\infty}$ ).

The validity of the present theory is restricted to the case of negligible quantum fluctuations in the charge on the dot.<sup>4</sup> Since charge leaks out of the dot at a rate  $\Gamma^l + \Gamma^r$ , the energy levels are sharply defined only if the resulting uncertainty in energy  $\hbar(\Gamma^l + \Gamma^r) \ll \Delta E$ . In view of Eq. (2.21), with  $\rho \sim 1/\Delta E$ , this requires  $T^{l,r} \ll 1$ , or  $G^{l,r} \ll e^2/h$ . In the resonant tunneling regime of comparable  $\Delta E$  and  $k_B T$ , this criterion is equivalent to the criterion  $\hbar \Gamma \ll k_B T$  mentioned earlier. In the classical regime  $\Delta E \ll k_B T$ , the criterion  $\hbar \Gamma \ll \Delta E$  dominates. The general criterion  $\hbar \Gamma \ll \Delta E, k_B T$  implies that the conductance of the quantum dot  $G \ll e^2/h$ .

As we lower the temperature, such that  $k_B T < e^2/C$ , the Coulomb-blockade oscillations become observable. This is shown in Fig. 6. The classical regime  $\Delta E \ll k_B T$  was first studied by Kulik and Shekhter.<sup>6,7</sup> In this regime a continuum of energy levels in the confined central region participates in the conduction. If  $\Delta E \ll k_B T \ll e^2/C$ , only the terms with  $N = N_{\text{min}}$  contribute to the sum in Eq. (2.18), where  $N_{\text{min}}$  minimizes the absolute value of  $\Delta(N) = U(N) - U(N-1) + \bar{\mu} - E_F$ . [Here  $\bar{\mu}$  is the equilibrium chemical potential of the dot, measured relative to the bottom of the potential well.] We define  $\Delta_{\text{min}} \equiv \Delta(N_{\text{min}})$ . For energy-independent tunnel rates and density of states  $\rho \equiv 1/\Delta E$ , one obtains a line shape of individual conductance peaks given by

$$G/G_{\text{max}} = \frac{\Delta_{\text{min}}/k_B T}{\sinh(\Delta_{\text{min}}/k_B T)} \approx \cosh^{-2} \left( \frac{\Delta_{\text{min}}}{2.5 k_B T} \right), \quad (2.24)$$

$$G_{\text{max}} = \frac{e^2}{2\Delta E} \frac{\Gamma^l \Gamma^r}{\Gamma^l + \Gamma^r}. \quad (2.25)$$

The second equality in Eq. (2.24) is approximate, but holds to better than 1%. A plot of  $G/G_{\text{max}}$  versus  $\Delta_{\text{min}}$  is shown for an isolated peak in Fig. 7 (dashed curve).

Whereas the width of the peaks increases with  $T$  in the classical regime, the peak height (reached at  $\Delta_{\text{min}} = 0$ )

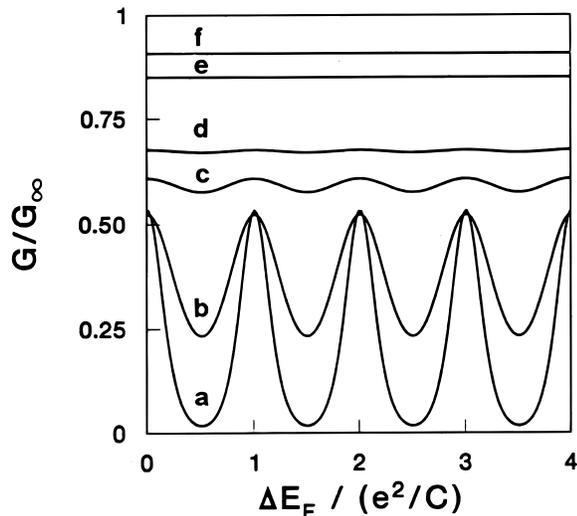


FIG. 6 Temperature dependence of the Coulomb-blockade oscillations as a function of Fermi energy in the classical regime  $k_B T \gg \Delta E$ . Curves are calculated from Eq. (2.18) with  $\Delta E = 0.01 e^2/C$ , for  $k_B T/(e^2/C) = 0.075$  (a), 0.15 (b), 0.3 (c), 0.4 (d), 1 (e), and 2 (f). Level-independent tunnel rates are assumed, as well as equidistant non-degenerate energy levels.

is temperature independent (compare traces (a) and (b) in Fig. 6). The reason is that the  $1/T$  temperature dependence associated with resonant tunneling through a particular energy level is canceled by the  $T$  dependence of the number  $k_B T/\Delta E$  of levels participating in the conduction. This cancellation holds only if the tunnel rates are energy independent within the interval  $k_B T$ . A temperature dependence of the conductance may result from a strong energy dependence of the tunnel rates. In such a case one has to use the general result (2.18). This is also required if peaks start to overlap for  $k_B T \sim e^2/C$ , or if the dot is nearly depleted ( $E_F \leq k_B T$ ). The latter regime does not play a role in metals, but is of importance in semiconductor nanostructures because of the much smaller  $E_F$ . The presence of only a small number  $E_F/\Delta E$  of electrons in a quantum dot leads also to a gate voltage dependence of the oscillations in the classical regime  $k_B T \gg \Delta E$ .

Despite the fact that the Coulomb blockade of tunneling is lifted at a maximum of a conductance peak, the peak height  $G_{\max}$  in the classical Coulomb-blockade regime  $\Delta E \ll k_B T \ll e^2/C$  is a factor of two smaller than the conductance  $G_\infty$  in the high temperature regime  $k_B T \gg e^2/C$  of negligible charging energy (in the case of energy-independent tunnel rates). The reason is a correlation between subsequent tunnel events, imposed by the charging energy. This correlation, expressed by the series of charge states  $Q = -N_{\min}e \rightarrow Q = -(N_{\min} - 1)e \rightarrow Q = -N_{\min}e \rightarrow \dots$ , implies that an electron can tunnel from a reservoir into the dot only half of the time (when  $Q = -(N_{\min} - 1)e$ ). The tunnel probability is therefore

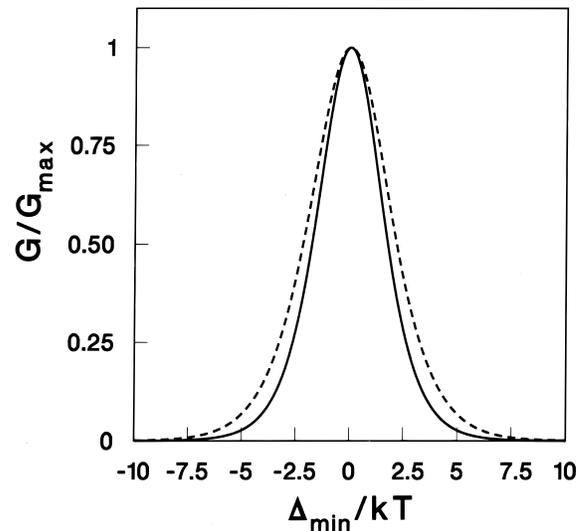


FIG. 7 Comparison of the lineshape of a thermally broadened conductance peak in the resonant tunneling regime  $h\Gamma \ll k_B T \ll \Delta E$  (solid curve) and in the classical regime  $\Delta E \ll k_B T \ll e^2/C$  (dashed curve). The conductance is normalized by the peak height  $G_{\max}$ , given by Eqs. (2.25) and (2.28) in the two regimes. The energy  $\Delta_{\min}$  is proportional to the Fermi energy in the reservoirs, cf. Eq. (2.26). (From Beenakker.<sup>19</sup>)

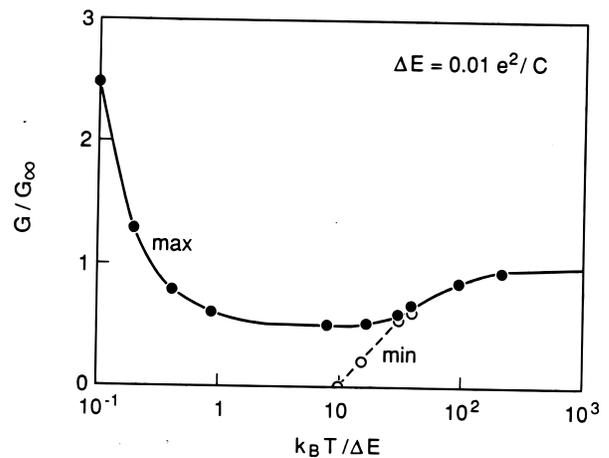


FIG. 8 Temperature dependence of the maxima (max) and the minima (min) of the Coulomb-blockade oscillations, in the regime  $h\Gamma \ll k_B T$ . The calculation, based on Eq. (2.18), was performed for the case of equidistant non-degenerate energy levels (at separation  $\Delta E = 0.01 e^2/C$ ), all with the same tunnel rates  $\Gamma^l$  and  $\Gamma^r$ .

reduced by a factor of two compared to the high temperature limit, where no such correlation exists.

The temperature dependence of the maxima of the Coulomb-blockade oscillations as obtained from Eq. (2.18) is plotted in Fig. 8. Also shown in Fig. 8 are the minima, which are seen to merge with the maxima as  $k_B T$  approaches  $e^2/C$ . In the resonant tunneling regime

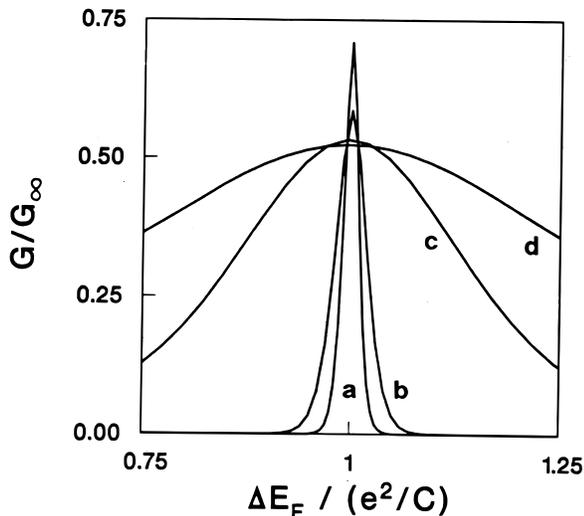


FIG. 9 Lineshape for various temperatures, showing the crossover from the resonant tunneling regime (a and b) where both the width and the peak height depend on  $T$ , to the classical regime (c and d) where only the width of the peak depends on  $T$ . Curves are calculated from Eq. (2.18) with  $\Delta E = 0.01 e^2/C$ , and for  $k_B T/\Delta E = 0.5$  (a), 1 (b), 7.5 (c), and 15 (d).

$k_B T \lesssim \Delta E$  the peak height increases as the temperature is reduced, due to the diminished thermal broadening of the resonance. The crossover from the classical to the quantum regime is shown in Fig. 9 [calculated directly from Eq. (2.18)].

In the case of well-separated energy scales in the resonant tunneling regime ( $h\Gamma \ll k_B T \ll \Delta E$ ), Eq. (2.18) can again be written in a simplified form. Now the single term with  $p = N = N_{\min}$  gives the dominant contribution to the sum over  $p$  and  $N$ . The integer  $N_{\min}$  minimizes the absolute value of

$$\Delta(N) = E_N + U(N) - U(N-1) - E_F. \quad (2.26)$$

We again denote  $\Delta_{\min} \equiv \Delta(N_{\min})$ . Equation (2.18) reduces to

$$\begin{aligned} G/G_{\max} &= 4k_B T f'(\Delta_{\min}) \\ &= \cosh^{-2} \left( \frac{\Delta_{\min}}{2k_B T} \right), \end{aligned} \quad (2.27)$$

$$G_{\max} = \frac{e^2}{4k_B T} \frac{\Gamma_{N_{\min}}^l \Gamma_{N_{\min}}^r}{\Gamma_{N_{\min}}^l + \Gamma_{N_{\min}}^r}. \quad (2.28)$$

As shown in Fig. 7, the lineshape in the resonant tunneling regime (full curve) is different from that in the classical regime (dashed curve), if they are compared at equal temperature. Equation (2.27) can be seen as the usual resonant tunneling formula for a thermally broadened resonance, generalized to include the effect of the charging energy on the resonance condition. Eqs. (2.27) and (2.28) hold regardless of the relative magnitude of

$\Delta E$  and  $e^2/C$ . As illustrated in Fig. 8, the peak height in the resonant tunneling regime increases monotonically as  $k_B T/\Delta E \rightarrow 0$ , as long as  $k_B T$  is larger than the resonance width  $h\Gamma$ .

No theory has been worked out for Coulomb-blockade oscillations in the regime  $k_B T \lesssim h\Gamma$  (although the theory of Meir et al.<sup>20</sup> is sufficiently general to be applicable in principle). For *non-interacting* electrons, the transmission probability has the Breit-Wigner form<sup>49,50,51</sup>

$$G_{\text{BW}} = \mathcal{G} \frac{e^2}{h} \frac{\Gamma^l \Gamma^r}{\Gamma^l + \Gamma^r} \frac{\Gamma}{(\epsilon/\hbar)^2 + (\Gamma/2)^2}. \quad (2.29)$$

Here  $\mathcal{G}$  is the degeneracy of the resonant level, and  $\epsilon$  is the energy separation of that level from the Fermi level in the reservoirs. In the presence of inelastic scattering with rate  $\Gamma_{\text{in}}$  one has to replace  $\Gamma$  by  $\Gamma + \Gamma_{\text{in}}$ .<sup>49,50,51</sup> This has the effect of reducing the conductance on resonance by a factor  $\Gamma/(\Gamma + \Gamma_{\text{in}})$ , and to increase the width of the peak by a factor  $(\Gamma + \Gamma_{\text{in}})/\Gamma$ . This is to be contrasted with the regime  $h\Gamma \ll k_B T \ll \Delta E$ , where inelastic scattering has no effect on the conductance. [This follows from the fact that the thermal average  $-\int G_{\text{BW}} f'(\epsilon) d\epsilon \approx \int G_{\text{BW}} d\epsilon/4k_B T$  is independent of  $\Gamma_{\text{in}}$ .] If inelastic scattering is negligible, and if the two tunnel barriers are equal, then the maximum conductance following from the Breit-Wigner formula is  $\mathcal{G}e^2/h$  — a result that may be interpreted as the fundamental contact conductance of a  $\mathcal{G}$ -fold degenerate state.<sup>50,52</sup> We surmise that the charging energy will lift the level degeneracy, so that the maximum peak height of Coulomb-blockade oscillations is  $G_{\max} = e^2/h$  for the case of equal tunnel barriers.

A few words on terminology, to make contact with the resonant tunneling literature.<sup>49,50</sup> The results discussed above pertain to the regime  $\Gamma \gg \Gamma_{\text{in}}$ , referred to as the “coherent resonant tunneling” regime. In the regime  $\Gamma \ll \Gamma_{\text{in}}$  it is known as “coherent sequential tunneling” (results for this regime are given in Ref.<sup>19</sup>). Phase coherence plays a role in both these regimes, by establishing the discrete energy spectrum in the quantum dot. The classical, or incoherent, regime is entered when  $k_B T$  or  $h\Gamma_{\text{in}}$  become greater than  $\Delta E$ . The discreteness of the energy spectrum can then be ignored.

We close this overview of theoretical results by a discussion of the activation energy of the minima of the conductance oscillations. It is shown in Ref.<sup>19</sup> that  $G_{\min}$  depends exponentially on the temperature,  $G_{\min} \propto \exp(-E_{\text{act}}/k_B T)$ , with activation energy

$$E_{\text{act}} = \frac{1}{2}(\Delta E + e^2/C) = \frac{1}{2}\Delta E^*. \quad (2.30)$$

This result holds for equal tunnel rates at two subsequent energy levels. The renormalized level spacing  $\Delta E^* \equiv \Delta E + e^2/C$ , which according to Eq. (2.7) determines the periodicity of the Coulomb-blockade oscillations as a function of Fermi energy, thus equals twice the activation energy of the conductance minima. The exponential decay of the conductance at the minima of

the Coulomb blockade oscillations results from the suppression of tunneling processes which conserve energy in the intermediate state in the quantum dot. Tunneling via a *virtual* intermediate state is not suppressed at low temperatures, and may modify the temperature dependence of the minima if  $\hbar\Gamma$  is not much smaller than  $k_B T$  and  $\Delta E$ .<sup>53,54</sup> For  $\hbar\Gamma \ll k_B T, \Delta E$  this co-tunneling or “macroscopic quantum tunneling of the charge” can be neglected.

### III. EXPERIMENTS ON COULOMB-BLOCKADE OSCILLATIONS

#### A. Quantum dots

Coulomb-blockade oscillations in the conductance of a quantum dot were first studied by Meirav, Kastner, and Wind.<sup>15</sup> The geometry of their device is shown in Fig. 2b. A split-gate electrode with a 300 nm wide slit is used to define a narrow channel. Small protrusions on each part of the split gate are used to define quantum point contacts in the narrow channel, 1  $\mu\text{m}$  apart. For sufficiently strong negative gate voltages the electron gas in the point contacts is depleted so that the channel is partitioned into a quantum dot, two tunnel barriers, and two leads. The width of the quantum dot is estimated to be 50 nm, whereas its length is about 1  $\mu\text{m}$ . The conductance of this device exhibits conductance peaks periodic in the gate voltage, at temperatures between 50 mK and 1 K (see Fig. 10a). Based on estimates of the gate capacitance, it was concluded that one electron was added to the quantum dot in each oscillation period. This conclusion was supported by experiments on devices with different values for the tunnel barrier separation.<sup>15</sup> Meirav et al. have also shown that the lineshape of an isolated peak could be fitted very well by a function of the form  $\cosh^{-2}(\gamma(V_{\text{gate}} - V_0)/2k_B T)$ . We note that, since the fit was done with  $\gamma$  and  $T$  as adaptable parameters, equally good agreement would have been obtained with the theoretical line shapes for the Coulomb-blockade oscillations in the classical or quantum regimes [Eqs. (2.25) and (2.28)].

Meirav et al. found that the temperature dependence of the peak width yielded an estimate for  $e^2/2C$  that was a factor of 3.5 lower than the value inferred from the periodicity. One way to possibly resolve this discrepancy is to note that the width of the peaks, as well as the activation energy, is determined by the charging energy  $e^2/2C$  with  $C = C_{\text{dot}} + C_{\text{gate}}$  [Eq. (2.10)]. This energy is smaller than the energy  $e^2/2C_{\text{gate}}$  obtained from a measurement of the periodicity  $\Delta V_{\text{gate}} \simeq e/C_{\text{gate}}$  [Eq. (2.12)]. Alternatively, a strong energy dependence of the tunnel rates may play a role.<sup>20</sup>

Meir, Wingreen, and Lee<sup>20</sup> modeled the experimental data shown in Fig. 10a by means of Eq. (2.18) (derived independently by these authors), using parameters consistent with experimental estimates ( $\Delta E = 0.1 \text{ meV}$ ,

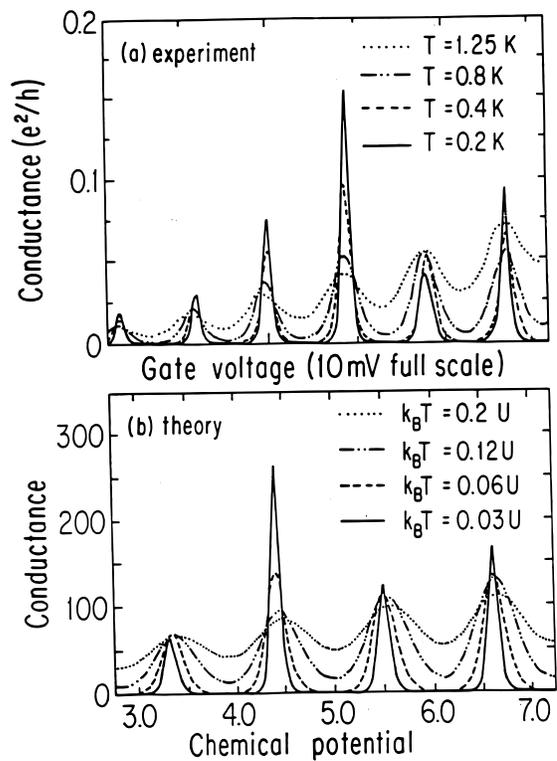


FIG. 10 (a) Measured conductance as a function of gate voltage in a quantum dot in the 2DEG of a GaAs-AlGaAs heterostructure, with a geometry as shown in Fig. 2b. (Experimental results obtained by U. Meirav, M. Kastner, and S. Wind, unpublished; U. Meirav, Ph.D. Thesis (M.I.T., 1990).) (b) Calculated conductance based on Eq. (2.18). The conductance is given in units  $\Gamma_1 C$ , and the chemical potential of the reservoirs is in units of  $e^2/C$ . The level spacing was taken to be  $\Delta E = 0.1 e^2/C$ . The tunnel rates of the levels increase in a geometric progression  $\Gamma_{p+1} = 1.5^p \Gamma_1$ , with  $\Gamma_4$  increased by an additional factor of 4 to simulate disorder. The temperature is quoted in units of  $e^2/C$ . (From Meir et al.<sup>20</sup>.)

$e^2/C = 1 \text{ meV}$ ). The results of their calculation are reproduced in Fig. 10b. The increasing height of successive peaks is due to an assumed increase in tunnel rates for successive levels ( $\Gamma_{p+1} = 1.5^p \Gamma_1$ ). Disorder is simulated by multiplying  $\Gamma_4$  by an additional factor of 4. No attempt was made to model the gate-voltage dependence of the experiment, and instead the chemical potential of the reservoirs was chosen as a variable in the calculations. Figs. 10a and 10b show a considerable similarity between experiment and theory. The second peak in the theoretical trace is the anomalously large  $\Gamma_4$  peak, which mimics the fourth peak in the experimental trace. In both theory and experiment a peak adjacent to the anomalously large peak shows a non-monotonic temperature dependence. This qualitative agreement, obtained with a consistent set of parameter values, supports the interpretation of the effect as Coulomb-blockade oscillations in the regime of a discrete energy spectrum.

It is possible that at the lowest experimental temper-

atures in the original experiment of Meirav et al.<sup>15</sup> the regime  $k_B T \lesssim h\Gamma$  of intrinsically broadened resonances is entered. An estimate of the average tunnel rates is most reliably obtained from the high-temperature limit, where the peaks begin to overlap. From Fig. 10a we estimate  $G_\infty \sim 0.1 e^2/h$ . For a symmetric quantum dot ( $\Gamma^l = \Gamma^r$ ) Eq. (2.23) with  $\rho \sim 1/\Delta E$  then implies  $h\Gamma \equiv h(\Gamma^l + \Gamma^r) \sim 0.4 \Delta E \sim 0.04 \text{ meV}$ . The condition  $k_B T \lesssim h\Gamma$  thus yields a crossover temperature of 500 mK. Meirav et al.<sup>15</sup> reported a saturation of the linear temperature dependence of the width of the peaks to a much weaker dependence for  $T \lesssim 500 \text{ mK}$ . It is thus possible that the approach of the intrinsically broadened regime  $k_B T \lesssim h\Gamma$  is at the origin of the saturated width at low temperatures (current heating of the electron gas<sup>15</sup> may also play a role). Unfortunately, as noted in Sec. II, a theory for the lineshape in this regime is not available.

We close the discussion of the experiments of Meirav et al. by noting that some of their samples showed additional periodicities in the conductance, presumably due to residual disorder. Thermal cycling of the sample (to room temperature) strongly affected the additional structure, without changing the dominant oscillations due to the quantum dot between the point contact barriers.

Williamson et al.<sup>17</sup> have studied the Coulomb-blockade oscillations using a quantum dot of the design shown in Fig. 2c. The device has three sets of gates to adjust the transmission probability of each tunnel barrier and the potential  $\phi_{\text{ext}}$  of the dot. (Because of the proximity of the gates the adjustments are not independent.) The tunnel barriers are formed by quantum point contacts close to pinch-off. A device with multiple gates in a lay-out similar to that of Fig. 2b was studied by Kouwenhoven et al.<sup>16</sup> From a measurement of the Coulomb-blockade oscillations for a series of values of the conductance of the individual quantum point contacts it has been found in both experiments that the oscillations disappear when the conductance of each point contact approaches the first quantized plateau, where  $G^{l,r} = 2e^2/h$ . It is not yet clear whether this is due to virtual tunneling processes, or to a crossover from tunneling to ballistic transport through the quantum point contacts. We note that this ambiguity does not arise in tunnel junctions between metals, where the area of the tunnel barrier is usually much larger than the Fermi wavelength squared, so that a barrier conductance larger than  $e^2/h$  can easily be realized *within* the tunneling regime. In semiconductors, tunnel barriers of large area can also be made — but it is likely that then  $e^2/C$  will become too small. A dynamical treatment is required in the case of low tunnel barriers, since the field across the barrier changes during the tunnel process.<sup>55</sup> Similar dynamic polarization effects are known to play a role in large-area semiconductor tunnel junctions, where they are related to image-force lowering of the barrier height.

## B. Disordered quantum wires

Scott-Thomas et al.<sup>9</sup> found strikingly regular conductance oscillations as a function of gate voltage (or electron gas density) in a narrow disordered channel in a Si inversion layer. The period of these oscillations differed from device to device, and did not correlate with the channel length. Based on estimates of the sample parameters, it was concluded that one period corresponds to the addition of a single electron to a conductance-limiting segment of the disordered quantum wire.

Two of us have proposed that the effect is the first manifestation of *Coulomb-blockade oscillations* in a semiconductor nanostructure.<sup>25</sup> To investigate this phenomenon further, Staring et al. have studied the periodic conductance oscillations in disordered quantum wires defined by a split gate in the 2DEG of a GaAs-AlGaAs heterostructure.<sup>12,13</sup> Other studies of the effect have been made by Field et al.<sup>11</sup> in a narrow channel in a 2D hole gas in Si, by Meirav et al.<sup>10</sup> in a narrow electron gas channel in an inverted GaAs-AlGaAs heterostructure, and by De Graaf et al.<sup>14</sup> in a very short split gate channel (or point contact) in a Si inversion layer. Here we will only discuss the results of Staring et al. in detail.

In a first set of samples,<sup>12</sup> a delta-doping layer of Be impurities was incorporated during growth, in order to create strongly repulsive scattering centers in the narrow channel. (Be is an acceptor in GaAs; some compensation was also present in the narrow Si inversion layers studied by Scott-Thomas et al.<sup>9</sup>) A second set of samples<sup>13</sup> did not contain Be impurities. The mean free path in the Be-doped samples in wide regions adjacent to the channel is  $0.7 \mu\text{m}$ . In the other samples it is  $4 \mu\text{m}$ . Close to pinch-off the channel will break up into a few segments separated by potential barriers formed by scattering centers. Model calculations have shown that statistical variations in the random positions of ionized donors in the AlGaAs are sufficient to create such a situation.<sup>45</sup> Indeed, both the samples with and without Be exhibited the Coulomb-blockade oscillations.

In Fig. 11a we reproduce representative traces of conductance versus gate voltage at various temperatures for a sample without Be.<sup>13</sup> Note the similarity to the results obtained for a single quantum dot shown in Fig. 10a. The oscillations generally disappear as the channel is widened away from pinch-off. No correlation was found between the periodicity of the oscillations and the channel length. At channel definition its width equals the lithographic width  $W_{\text{lith}} = 0.5 \mu\text{m}$ , and the sheet electron density  $n_s = 2.9 \times 10^{11} \text{ cm}^{-2}$ . As the width is reduced to  $0.1 \mu\text{m}$ , the density becomes smaller by about a factor of 2. (The estimate for  $W$  is based on typical lateral depletion widths of  $200 \text{ nm}/V$ ,<sup>8,45,46</sup> and that for  $n_s$  on an extrapolation of the periodicity of the Shubnikov-De Haas oscillations.) A  $3 \mu\text{m}$  long channel then contains some 450 electrons. Calculations for a split-gate channel<sup>56</sup> indicate that the number of electrons per unit length increases approximately linearly with gate voltage. The periodicity

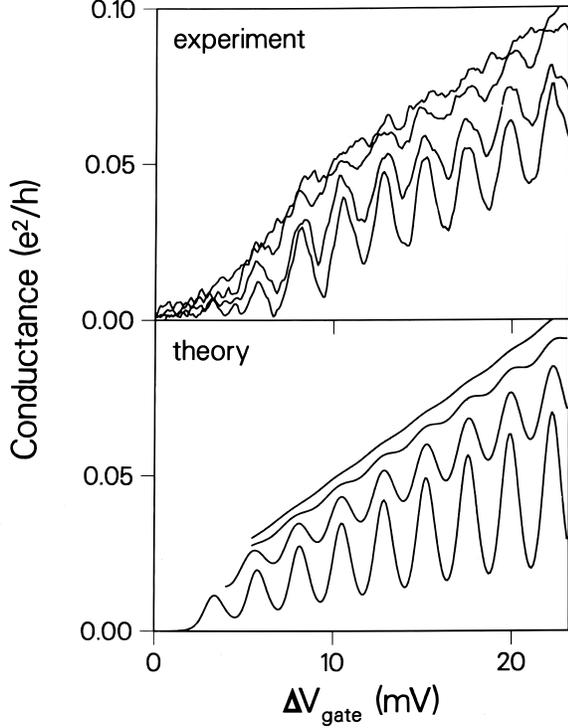


FIG. 11 (a) Measured conductance of an unintentionally disordered quantum wire in a GaAs-AlGaAs heterostructure, of a geometry as shown in Fig. 2a;  $T = 1.0, 1.6, 2.5,$  and  $3.2$  K (from bottom to top). (b) Model calculations based on Eq. (2.18), for  $\Delta E = 0.1$  meV,  $e^2/C = 0.6$  meV,  $\alpha = 0.27$ , and  $h\Gamma_p^{1,r} = 2.7 \times 10^{-2} p\Delta E$  ( $p$  labels spin-degenerate levels). (From Staring et al.<sup>13</sup>)

of the conductance oscillations as a function of gate voltage thus implies a periodicity as a function of density per unit length.

Our model for the Coulomb-blockade oscillations in a disordered quantum wire is essentially the same as that for a quantum dot, to the extent that a single segment limits the conductance. To calculate  $C_{\text{dot}}$  and  $C_{\text{gate}}$  is a rather complicated three-dimensional electrostatic problem, hampered further by the uncertain dimensions of the conductance limiting segment. Experimentally, the conductance peaks are spaced by  $\Delta V_{\text{gate}} \sim 2.4$  mV, so that from Eq. (2.12) we estimate  $C_{\text{gate}} \sim 0.7 \times 10^{-16}$  F. The length  $L$  of the quantum dot may be estimated from the gate voltage range  $\delta V_{\text{gate}} \sim 1$  V between channel definition and pinch-off:  $\delta V_{\text{gate}} \sim en_s W_{\text{lith}} L / C_{\text{gate}}$ , where  $n_s$  is the sheet density in the channel at definition. From the above estimate of  $C_{\text{gate}}$  and using  $\delta V_{\text{gate}} \sim 1$  V, we estimate  $L \sim 0.3 \mu\text{m}$ .<sup>†</sup> The width

<sup>†</sup> The estimated values for  $C_{\text{gate}}$  and  $L$  are consistent with what one would expect for the mutual capacitance of a length  $L$  of a

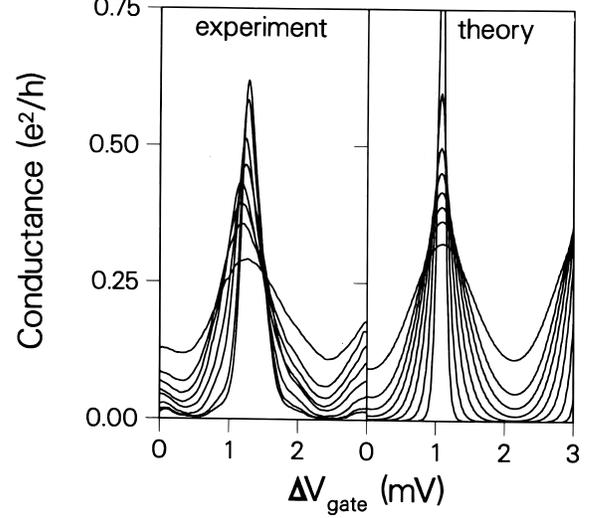


FIG. 12 Experimental and theoretical lineshapes of an isolated conductance peak in a Be-doped disordered quantum wire in a GaAs-AlGaAs heterostructure, at  $B = 6.7$  T, and  $T = 110, 190, 290, 380, 490, 590, 710,$  and  $950$  mK (from top to bottom). The theoretical curves have been calculated from Eq. (2.18), with  $\Delta E = 0.044$  meV (non-degenerate),  $e^2/C = 0.53$  meV,  $h\Gamma = 0.13$  meV, and  $\alpha = 0.27$ . (From Staring et al.<sup>13</sup>)

of the dot is estimated to be about  $W \sim 0.1 \mu\text{m}$  in the gate voltage range of interest. The level splitting in the segment is  $\Delta E \sim 2\pi\hbar^2/mLW \sim 0.2$  meV (for a 2-fold spin-degeneracy). Since each oscillation corresponds to the removal of a single electron from the dot, the maximum number of oscillations following from  $\Delta E$  and the Fermi energy  $E_F \sim 5$  meV at channel definition is given by  $2E_F/\Delta E \sim 50$ , consistent with the observations. From the fact that the oscillations are still observable at  $T = 1.5$  K, albeit with considerable thermal smearing, we deduce that in our experiments  $e^2/C + \Delta E \sim 1$  meV. Thus,  $C \sim 2.0 \times 10^{-16}$  F,  $C_{\text{dot}} = C - C_{\text{gate}} \sim 1.3 \times 10^{-16}$  F,<sup>‡</sup> and the parameter  $\alpha \equiv C_{\text{gate}}/C \sim 0.35$ . In Fig. 11 we compare a calculation based on Eq. (2.18) with the experiment, taking the two-fold spin-degeneracy of the energy levels into account<sup>13</sup> The tunnel rates were taken to increase by an equal amount  $0.027 \Delta E/h$  for each subsequent spin-degenerate level, at equal separation  $\Delta E = 0.1$  meV.

wire of diameter  $W$  running in the middle of a gap of width  $W_{\text{lith}}$  in a metallic plane (the thickness of the AlGaAs layer between the gate and the 2DEG is small compared to  $W_{\text{lith}}$ ):  $C_{\text{gate}} \sim 4\pi\epsilon L/2 \text{arccosh}(W_{\text{lith}}/W) \sim 0.9 \times 10^{-16}$  F (see Ref.<sup>57</sup>).

<sup>‡</sup> The mutual capacitance of dot and leads may be approximated by the self-capacitance of the dot,<sup>57</sup> which should be comparable to that of a two-dimensional circular disc of diameter  $L$ :  $C_{\text{dot}} \sim 4\epsilon L \sim 1.4 \times 10^{-16}$  F, consistent with the estimate given in the text.

The capacitances were fixed at  $e^2/C = 0.6$  meV and  $\alpha = 0.25$ . These values are consistent with the crude estimates given above. The Fermi energy was assumed to increase equally fast as the energy of the highest occupied level in the dot (cf. Sec. II.A). The temperature range shown in Fig. 11 is in the classical regime ( $k_B T > \Delta E$ ).

The resonant tunneling regime  $k_B T < \Delta E$  can be described qualitatively by Eq. (2.18), as shown in Fig. 12 for an isolated peak. The data was obtained for a different sample (with Be doping) in the presence of a magnetic field of 6.7 T. The parameter values used are  $\Delta E = 0.045$  meV,  $e^2/C = 0.53$  meV,  $h\Gamma = 0.13$  meV, and  $\alpha = 0.27$ . A fully quantitative theoretical description of the experimental lineshapes in Fig. 12 is not yet possible, because the experiment is in the regime of intrinsically broadened resonances,  $k_B T < h\Gamma$ , for which the theory has not been worked out.

The semi-quantitative agreement between theory and experiment in Figs. 11 and 12, for a consistent set of parameter values, and over a wide range of temperatures, supports our interpretation of the conductance oscillations as Coulomb-blockade oscillations in the regime of comparable level spacing and charging energies. Note that  $e^2/C_{\text{gate}} \sim 10 \Delta E$ , so that irregularly spaced energy levels would not easily be discernable in the gate voltage scans [cf. Eq. (2.12)]. Such irregularities might nevertheless play a role in causing peak height variations. Some of the data (not shown) exhibits beating patterns,<sup>12,13</sup> similar to those reported in Refs.<sup>9</sup> and<sup>11</sup>. These are probably due to the presence of multiple segments in the quantum wires.<sup>13</sup> Coulomb-blockade oscillations in arrays of tunnel junctions in the classical regime have been studied by several authors,<sup>58,59</sup>

As an alternative explanation of the conductance oscillations resonant tunneling of non-interacting electrons has been proposed.<sup>26,27</sup> There are several compelling arguments for rejecting this explanation (which apply to the experiments on a quantum dot as well as to those on disordered quantum wires). Firstly, for resonant tunneling the oscillations would be irregularly spaced, due to the non-uniform distribution of the bare energy levels [cf. Eq. (2.14)]. This is in contradiction with the experimental observations.<sup>11</sup> Secondly,<sup>12</sup> in the absence of charging effects the measured activation energy of the conductance minima would imply a level spacing  $\Delta E \sim 1$  meV. Since the Fermi energy  $E_F$  in a typical narrow channel is about 5 meV, such a large level spacing would restrict the possible total number of oscillations in a gate voltage scan to  $E_F/\Delta E \sim 5$ , considerably less than the number seen experimentally.<sup>9,12</sup> Thirdly, one would expect a spin-splitting of the oscillations by a strong magnetic field, which is not observed.<sup>11</sup> Finally, the facts that no oscillations are found as a function of *magnetic* field<sup>11,12</sup> and that the spin-splitting does not occur, all but rule out resonant tunneling of non-interacting electrons as an explanation of the oscillations as a function of gate voltage.

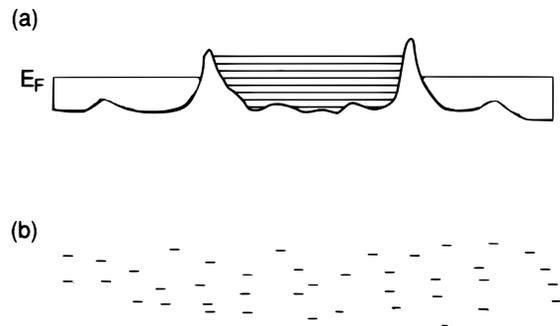


FIG. 13 (a) Coulomb-blockade oscillations occur in a disordered quantum wire as a result of the formation of a conductance limiting segment which contains many localized states. (b) Random conductance fluctuations due to variable range hopping between localized states (indicated by dashes) are found in the absence of such a segment.

### C. Relation to earlier work on disordered quantum wires

The disordered quantum wires discussed in this chapter exhibit *periodic* conductance oscillations as a function of gate voltage. The effect has been seen in electron and hole gases in Si<sup>9,11,14</sup> and in the electron gas in GaAs.<sup>10,12,13</sup> In contrast, previous work by Fowler et al.<sup>60</sup> and by Kwasnick et al.<sup>61</sup> on narrow inversion and accumulation layers in Si has produced sharp but *aperiodic* conductance peaks. How are these observations to be reconciled? We surmise that the explanation is to be found in the different strength and spatial scale of the potential fluctuations in the wire, as illustrated in Fig. 13.

Coulomb-blockade oscillations require a small number of large potential spikes, so that a single segment limits the conductance (Fig. 13a). The random conductance fluctuations seen previously<sup>60,61</sup> are thought instead to be due to variable range hopping between a large number of localized states, distributed randomly along the length of the channel (Fig. 13b).<sup>62,63,64</sup> No segment containing a large number of states (localized within the same region) is present in the potential of Fig. 13b, in contrast to the situation shown in Fig. 13a. At large Fermi energy a transition eventually occurs to the diffusive transport regime in either type of wire. Both the regular Coulomb-blockade oscillations, and the random conductance peaks due to variable range hopping are then replaced by “universal” conductance fluctuations caused by quantum interference.<sup>65,66</sup>

Fowler et al.<sup>67</sup> have also studied the conductance of much shorter channels than in Ref.<sup>60</sup> (0.5  $\mu\text{m}$  long, and 1  $\mu\text{m}$  wide). In such channels they found well-isolated conductance peaks, which were temperature independent below 100 mK, and which were attributed to resonant tunneling. At very low temperatures a fine structure (some of it time-dependent) was observed. A numerical simulation<sup>68</sup> of the temporal fluctuations in the distribu-

tion of electrons among the available sites also showed fine structure if the time scale of the fluctuations is short compared to the measurement time, but large compared to the tunnel time. It is possible that a similar mechanism causes the fine structure on the Coulomb-blockade oscillations in a disordered quantum wire (cf. Fig. 11).

There have also been experimental studies of the effect of a strong magnetic field on variable range hopping<sup>69</sup> and on resonant tunneling through single impurity states.<sup>70</sup> We briefly discuss the work on resonant tunneling by Kopley et al.,<sup>70</sup> which is more closely related to the subject of this chapter. They observed large conductance peaks in a Si inversion layer under a split gate. Below the 200 nm wide slot in the gate the inversion layer is interrupted by a potential barrier. Pronounced conductance peaks were seen at 0.5 K as the gate voltage was varied in the region close to threshold. The peaks were attributed to resonant tunneling through single impurity states in the Si bandgap in the barrier region. The lineshape of an isolated peak could be fitted with the Breit-Wigner formula [Eq. (2.29)]. The amplitude of most peaks was substantially suppressed on applying a strong magnetic field. This was interpreted as a reduction of the tunnel rates because of a reduced overlap between the wavefunctions on the (asymmetrically placed) impurity and the reservoirs. The amplitude of one particular peak was found to be unaffected by the field, indicative of an impurity which is placed symmetrically in the barrier ( $\Gamma^r = \Gamma^l$ ). The width of that peak was reduced, consistent with a reduction of  $\Gamma$ . This study therefore exhibits many characteristic features of resonant tunneling through a single localized site. Yet, one would expect Coulomb interactions of two electrons on the site to be important, and indeed they might explain the absence of spin-splitting of the peaks in a strong magnetic field.<sup>70</sup> Theoretical work indicates that Coulomb interactions also modify the lineshape of a conductance peak.<sup>68,71</sup> The experimental evidence<sup>63,67,69,70</sup> is not conclusive, however.

## IV. QUANTUM HALL EFFECT REGIME

### A. The Aharonov-Bohm effect in a quantum dot

The Aharonov-Bohm effect is a quantum interference effect which results from the influence of the vector potential on the phase of the electron wavefunction. Aharonov and Bohm<sup>72</sup> originally considered the influence of the vector potential on electrons confined to a multiply-connected region (such as a ring), within which the magnetic field is zero. The ground state energy of the system is periodic in the enclosed flux with period  $h/e$ , as a consequence of gauge invariance. Coulomb repulsion does not affect this periodicity.

In the solid state, the Aharonov-Bohm effect manifests itself as a periodic oscillation in the conductance of a sample as a function of an applied magnetic field

**B.** A well-defined periodicity requires that the conducting paths through the sample enclose a constant area  $A$ , perpendicular to  $\mathbf{B}$ . The periodicity of the oscillations is then  $\Delta B = h/eA$ , plus possibly harmonics (e.g. at  $h/2eA$ ). The constant area may be imposed by confining the electrons electrostatically to a ring or to a cylindrical film.<sup>73,74</sup>

Entirely new mechanisms for the Aharonov-Bohm effect become operative in strong magnetic fields in the quantum Hall effect regime. These mechanisms do not require a ring geometry, but apply to singly-connected geometries such as a point contact<sup>75</sup> or a quantum dot.<sup>28,29</sup> As discussed below, these geometries behave as if they were multiply connected, because of circulating edge states. Resonant tunneling through these states leads to magnetoconductance oscillations with a fundamental periodicity  $\Delta B = h/eA$ , governed by the addition to the dot of a single quantum of magnetic flux  $h/e$ .

An essential difference with the original Aharonov-Bohm effect is that in these experiments the magnetic field extends into the conducting region of the sample. Since the periodicity is now no longer constrained by gauge invariance, this opens up the possibility, in principle, of an influence of Coulomb repulsion. We will discuss in the next subsection that the Aharonov-Bohm effect may indeed be *suppressed* by charging effects.<sup>30</sup> In this subsection we will first introduce the case of negligible charging effects in some detail.

If one applies a magnetic field  $\mathbf{B}$  to a metal, then the electrons move with constant velocity  $v_{\parallel}$  in a direction parallel to  $\mathbf{B}$ , and in a circular cyclotron orbit with tangential velocity  $v_{\perp}$  in a plane perpendicular to  $\mathbf{B}$ . The cyclotron frequency is  $\omega_c = eB/m$ , and the cyclotron radius is  $l_{\text{cycl}} = v_{\perp}/\omega_c$ . Quantization of the periodic cyclotron motion in a strong magnetic field leads to the formation of Landau levels

$$E_n(k_{\parallel}) = E_n + \frac{\hbar^2 k_{\parallel}^2}{2m}, \quad (4.1)$$

$$E_n = (n - \frac{1}{2})\hbar\omega_c, \quad (4.2)$$

labeled by the Landau level index  $n = 1, 2, \dots$ . In a field of 10 T (which is the strongest field that is routinely available), the Landau level separation  $\hbar\omega_c$  is about 1 meV (for  $m = m_e$ ). Consequently, in a metal the number of occupied Landau levels  $N_L \sim E_F/\hbar\omega_c$  is a large number, of order 1000. Even so, magnetic quantization effects are important at low temperatures, since  $\hbar\omega_c > k_B T$  for  $T < 10$  K. A familiar example is formed by the Shubnikov-De Haas oscillations in the magnetoresistance, which are caused by peaks in the density of states at the energies  $E_n$  which coincide with  $E_F$  for successive values of  $n$  as  $B$  is varied.

Because of the free motion along  $\mathbf{B}$ , the density of states in a metal does not vanish at energies between two Landau levels. Consequently, in metals magnetic quantum effects are relatively small. The situation is different in a 2DEG. Here the energy spectrum of the electrons becomes fully discrete in a strong perpendicular magnetic

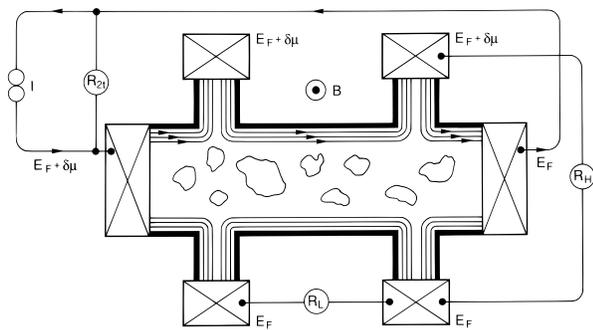


FIG. 14 Measurement configuration for the two-terminal resistance  $R_{2t}$ , the four-terminal Hall resistance  $R_H$ , and the longitudinal resistance  $R_L$ . The  $N_L$  edge channels at the Fermi level are indicated, arrows point in the direction of motion of edge channels filled by the source contact at chemical potential  $E_F + \delta\mu$ . The current  $N_L e \delta\mu / h$  is equipartitioned among the edge channels at the upper edge, corresponding to the case of local equilibrium. Localized states in the bulk do not contribute to the conductance. The resulting resistances are  $R_{2t} = R_H = h / N_L e^2$ ,  $R_L = 0$ . (From Beenakker and Van Houten.<sup>8</sup>)

field, since no free translational motion parallel to  $\mathbf{B}$  is possible. The vanishing of the density of states between Landau levels is at the origin of the pronounced magnetic quantum effects in a 2DEG. Well known is the integer quantum Hall effect, characterized by a vanishing longitudinal resistance  $R_L$  and a quantized Hall resistance  $R_H$  at values of  $h / N_L e^2$ . The distinction between a longitudinal and Hall resistance is topological (see Fig. 14): A four-terminal resistance measurement gives  $R_H$  if current and voltage contacts alternate along the boundary of the conductor, and  $R_L$  if that is not the case. There is no need to further characterize the contacts in the case of local equilibrium at the edge (in the opposite case the Hall resistance may take on anomalous values<sup>8</sup>). Frequently, the resistance of a sample is measured using only two contacts (which then act both as current and as voltage probes). In the quantum Hall effect regime, the two-terminal resistance  $R_{2t} = R_H + R_L = R_H$  is quantized at the same value as the Hall resistance.

The Fermi energy in a 2DEG is quite small (10 meV in conventional samples, 1 meV for samples with a very low density  $n_s \sim 10^{10} \text{cm}^{-2}$ ). Since, in addition, the effective mass is small, the extreme magnetic quantum limit  $N_L = 1$  is accessible. This is the realm of the fractional quantum Hall effect, studied in high-mobility samples at milli-Kelvin temperatures, and of the Wigner crystallization of the 2DEG. Both phenomena are due to electron-electron interactions in a strong magnetic field. This chapter is limited to the integer quantum Hall effect.

To the extent that broadening of the Landau levels by disorder can be neglected, the density of states (per unit area) in an unbounded 2DEG can be approximated by a

series of delta functions,

$$\rho(E) = g_s g_v \frac{eB}{h} \sum_{n=1}^{\infty} \delta(E - E_n). \quad (4.3)$$

The spin-degeneracy  $g_s$  is removed in strong magnetic fields as a result of the Zeeman splitting  $g \mu_B B$  of the Landau levels ( $\mu_B \equiv e \hbar / 2 m_e$  denotes the Bohr magneton; the Lande  $g$ -factor is a complicated function of the magnetic field in these systems<sup>76</sup>).

In the modern theory of the quantum Hall effect,<sup>77</sup> the longitudinal and Hall conductance (measured using two pairs of current contacts and voltage contacts) are expressed in terms of the transmission probabilities between the contacts for electronic states at the Fermi level. When  $E_F$  lies between two Landau levels, these states are *edge states* extended along the boundaries (Fig. 14). Edge states are the quantum mechanical analogue of *skipping orbits* of electrons undergoing repeated specular reflections at the boundary.<sup>8</sup> For a smooth confining potential  $V(\mathbf{r})$ , the edge states are extended along equipotentials of  $V$  at the guiding center energy  $E_G$ , defined by

$$E_G = E - (n - \frac{1}{2}) \hbar \omega_c, \quad (4.4)$$

for an electron with energy  $E$  in the  $n$ -th Landau level ( $n = 1, 2, \dots$ ). The confining potential should be sufficiently smooth that it does not induce transitions between different values of  $n$ . This requires that  $l_m V' \lesssim \hbar \omega_c$ , with  $l_m \equiv (\hbar / eB)^{1/2}$  the magnetic length (which plays the role of the wave length in the quantum Hall effect regime). Since the lowest Landau level has the largest guiding center energy, the corresponding edge state is located closest to the boundary of the sample, whereas the higher Landau levels are situated further towards its center.

In an open system, the single-electron levels with quantum number  $n$  form a 1D subband with subband bottom at  $E_n = (n - \frac{1}{2}) \hbar \omega_c$ . These 1D subbands are referred to as *edge channels*. Each of the  $N_L \sim E_F / \hbar \omega_c$  edge channels at the Fermi level contributes  $2e^2 / h$  to the Hall conductance if backscattering is suppressed. This happens whenever the Fermi level is located between two bulk Landau levels, so that the only states at  $E_F$  are those extended along the boundaries. Backscattering then requires transitions between edge states on *opposite* boundaries, which are usually far apart. In a very narrow channel, the Hall conductance may deviate from its quantized value  $N_L e^2 / h$  (and the longitudinal resistance may become non-zero) due to tunneling between opposite edges — a process that is strongly enhanced by disorder in the channel. The reason is that localized states at the Fermi energy may act as intermediate sites in a tunneling process from one edge to the other. We will come back to this point at the end of the section.

In a closed system, such as a quantum dot, the energy spectrum is fully discrete (for  $E_G$  less than the height  $E_B$  of the tunnel barriers which connect the dot to the leads).

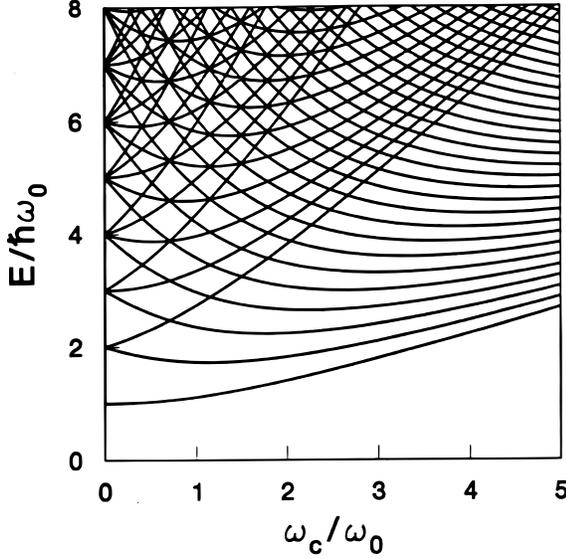


FIG. 15 Energy spectrum of a quantum dot with a harmonic confining potential as a function of magnetic field, according to Eq. (4.5). Spin-splitting is neglected.

An example which can be solved exactly is a quantum dot defined by a 2D harmonic oscillator potential  $V(r) = \frac{1}{2}m\omega_0^2 r^2$ . The energy spectrum is given by<sup>78,79</sup>

$$E_{nm} = \frac{1}{2}(n-m)\hbar\omega_c + \frac{1}{2}\hbar(\omega_c^2 + 4\omega_0^2)^{1/2}(n+m-1), \quad n, m = 1, 2, \dots \quad (4.5)$$

Each level has a two-fold spin-degeneracy, which is gradually lifted as  $B$  is increased. For simplicity, we do not take the spin degree of freedom into account. The energy spectrum (4.5) is plotted in Fig. 15. The asymptotes corresponding to the first few Landau levels are clearly visible.

In the limit  $\omega_0/\omega_c \rightarrow 0$  of a smooth potential and a fairly strong magnetic field, Eq. (4.5) reduces to

$$E_{nm} = \hbar\omega_c(n - \frac{1}{2} + (n+m-1)(\omega_0/\omega_c)^2), \quad (4.6)$$

which may also be written as

$$E_{nm} = (n - \frac{1}{2})\hbar\omega_c + V(R_{nm}), \quad B\pi R_{nm}^2 = (m + \gamma_n)\frac{h}{e}, \quad (4.7)$$

with  $\gamma_n = n - 1$ . Equation (4.7) is equivalent to the requirement that the equipotential of the edge state, of radius  $R_{nm}$ , encloses  $m + \gamma_n$  flux quanta. This geometrical requirement holds generally for smooth confining potentials, in view of the Bohr-Sommerfeld quantization rule

$$\frac{1}{h} \oint PdQ = m + \gamma_n. \quad (4.8)$$

The canonically conjugate variables  $P$  and  $Q$ , in the present case, are proportional to the guiding center co-

ordinates  $\mathbf{R} = (X, Y)$ , defined by

$$X = x - v_y/\omega_c, \quad (4.9)$$

$$Y = y + v_x/\omega_c, \quad (4.10)$$

in terms of the position  $\mathbf{r} = (x, y)$  and velocity  $\mathbf{v} = (v_x, v_y)$  of the electron. If one identifies  $Q \equiv X$ ,  $P \equiv eBY$ , one can verify the canonical commutation relation  $[Q, P] = i\hbar$  (using  $m\mathbf{v} = \mathbf{p} + e\mathbf{A}$ ,  $[x, p_x] = [y, p_y] = i\hbar$ ,  $[p_y, A_x] - [p_x, A_y] = i\hbar B$ ). The Bohr-Sommerfeld quantization rule thus becomes

$$\Phi = B \oint Y dX = \frac{h}{e}(m + \gamma_n), \quad (4.11)$$

which is the requirement that the flux  $\Phi$  enclosed by the guiding center drift is quantized in units of the flux quantum. To close the argument, we compute the guiding center drift  $\dot{\mathbf{R}} = B^{-2}\mathbf{E}(\mathbf{r}) \times \mathbf{B} \sim B^{-2}\mathbf{E}(\mathbf{R}) \times \mathbf{B}$ , in the approximation that the electric field  $\mathbf{E}$  does not vary strongly over the cyclotron radius  $|\mathbf{r} - \mathbf{R}|$ . In this case of a smoothly varying  $V$ , the motion of  $\mathbf{R}$  is along equipotentials at the guiding center energy  $E_G = E - (n - \frac{1}{2})\hbar\omega_c$ . The Bohr-Sommerfeld quantization rule can thus be written in the general form

$$E_{nm} = (n - \frac{1}{2})\hbar\omega_c + E_G(n, m), \quad (4.12)$$

where  $E_G(n, m)$  is the energy of the equipotential which encloses  $m + \gamma_n$  flux quanta. For the harmonic oscillator potential,  $\gamma_n = n - 1$ . For other smooth confining potentials  $\gamma_n$  may be different. (Knowledge of  $\gamma_n$  is not important if one only considers states within a single Landau level.)

Equation (4.12) does not hold for a hard-wall confining potential. An exact solution exists in this case for a circular disc<sup>80</sup> of radius  $R$ , defined by  $V(r) = 0$  for  $r < R$ , and  $V(r) = \infty$  for  $r > R$ . The case of a square disc was studied numerically by Sivan et al.<sup>29</sup> In Fig. 16a we show the energy spectrum as a function of  $B$  for the circular disc. (Fig. 16b is discussed in the following subsection.) The asymptotes correspond to the bulk Landau levels  $E_n = (n - \frac{1}{2})\hbar\omega_c$ . The first two Landau levels ( $n = 1, 2$ ) are visible in Fig. 16a. The states between the Landau levels are edge states, which extend along the perimeter of the disc. These circulating edge states make the geometry effectively doubly connected — in the sense that they enclose a well-defined amount of flux. Resonant tunneling through these states is the mechanism leading to the Aharonov-Bohm magnetoconductance oscillations in a quantum dot.

Three cases of interest are illustrated in Fig. 17. In a strong magnetic field, only edge states with  $n = 1$  corresponding to the first Landau level are occupied (Fig. 17a). As the field is reduced, also the second Landau level,  $n = 2$ , is occupied, as indicated in Fig. 17b. Tunneling through the quantum dot still occurs predominantly through the  $n = 1$  edge states, which have the largest tunnel probability through the barriers. If the height  $E_B$

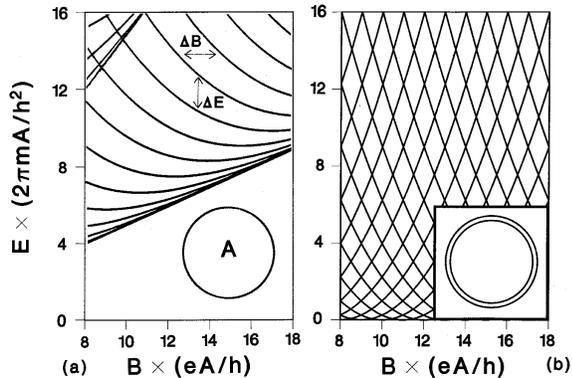


FIG. 16 Comparison of the energy levels in a disc and a ring. (a) Circular hard-wall disc (after Geerinckx et al.<sup>80</sup>). (b) Circular channel or ring of width  $W \ll l_m$  (after Büttiker et al.<sup>81</sup>). The levels in (b) are plotted relative to the energy of the bottom of the one-dimensional subband in the channel. The case  $W \gtrsim l_m$  is qualitatively the same as long as the area  $S$  of the annulus is much smaller than the area  $A$ . Spin-splitting is disregarded. (From Beenakker et al.<sup>30</sup>)

of the potential barriers is reduced, the  $n = 1$  edge states near the Fermi level may have  $E_G > E_B$ , so that they form an extended edge channel. The edge states with  $n > 1$  may still have  $E_G < E_B$ , and remain bound in the dot as before. As illustrated in Fig. 17c, resonant tunneling now occurs predominantly through the edge states belonging to the second Landau level.

In the quantum Hall effect regime scattering between edge channels can be neglected on length scales comparable to the diameter of the dot<sup>82</sup> (this is known as adiabatic transport<sup>8</sup>). The edge channels may then be treated as independent parallel conduction paths. The edge channels with  $E_G > E_B$  contribute  $e^2/h$  to the conductance. Resonant tunneling through the edge states with  $E_G < E_B$  gives an oscillating contribution to the conductance of the quantum dot as a function of magnetic field. The periodicity of the conductance oscillations can be deduced from the result (4.12) for the edge state energy spectrum. Resonant tunneling from the reservoir with Fermi energy  $E_F$  into an edge state in the quantum dot is possible when  $E_F = E_{nm}$  for certain quantum numbers  $n$  and  $m$ . For the edge states in the  $n$ -th Landau level the condition for resonant tunneling is that the equipotential at the guiding center energy  $E_G \equiv E_F - (n - \frac{1}{2})\hbar\omega_c$  should enclose  $m + \gamma_n$  flux quanta, for some integer  $m$ . Let  $A(B)$  denote the (magnetic field dependent) area of the equipotential at energy  $E_G$ . The  $m$ -th conductance peak occurs at a magnetic field  $B_m$  determined by  $B_m A(B_m) = (h/e)(m + \gamma_n)$ . The periodicity  $\Delta B \equiv B_{m+1} - B_m$  of the conductance oscillations from the  $n$ -th Landau level is obtained by expanding  $A(B)$  around  $B_m$ ,

$$\Delta B = \frac{\hbar}{e} [A(B_m) + B_m A'(B_m)]^{-1} \equiv \frac{\hbar}{e} \frac{1}{A_{\text{eff}}(B_m)}. \quad (4.13)$$

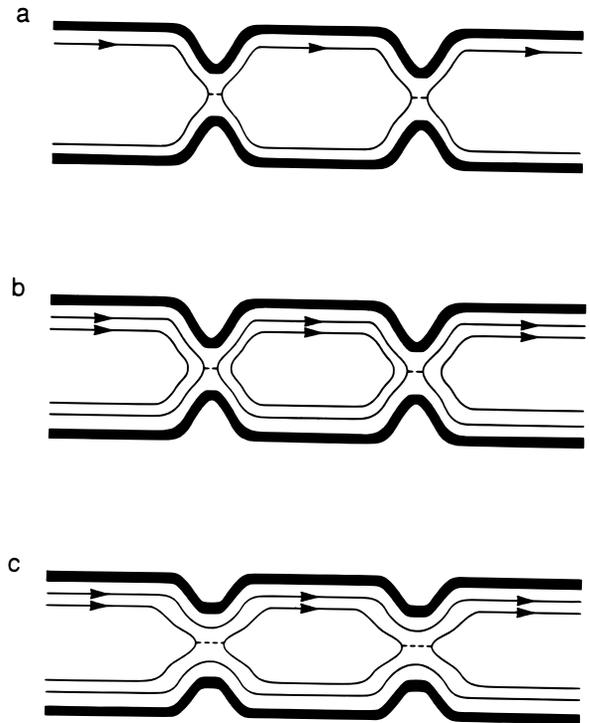


FIG. 17 Aharonov-Bohm magnetoconductance oscillations may occur due to resonant tunneling through circulating edge states. Tunneling paths are indicated by dashed lines. (a) Only the first Landau level is occupied. If the capacitance of the dot is sufficiently small, the Coulomb blockade suppresses the Aharonov-Bohm oscillations. (b) Two Landau levels are occupied. Resonant tunneling through the dot occurs predominantly through the first (outer) Landau level. The Aharonov-Bohm effect is not suppressed by the charging energy. (c) Two Landau levels are occupied, one of which is fully transmitted. Since the number of electrons in the dot is not discretized, no Coulomb blockade of the Aharonov-Bohm effect is expected.

The effective area  $A_{\text{eff}}(B)$  can differ substantially from the geometrical area  $A(B)$  in the case of a smooth confining potential.<sup>28</sup> The magnetoconductance oscillations are approximately periodic in  $B$  if the change in  $A_{\text{eff}}(B)$  in one period  $\Delta B$  is much smaller than the effective area itself. Since the change in  $A_{\text{eff}}$  is of order  $\hbar/eB$  per period, while  $A_{\text{eff}} \sim mh/eB$ , approximately periodic oscillations occur for  $m \gg 1$ . This is the Aharonov-Bohm effect in the quantum Hall regime, first observed by Van Wees et al.<sup>28</sup> Their experimental results (reproduced in Fig. 18) correspond to the situation of Fig. 17c with one (or more) fully transmitted edge channels.

We close this subsection by mentioning that resonant backscattering (or resonant reflection) can cause similar Aharonov-Bohm oscillations as those caused by resonant transmission. Resonant backscattering may occur via a localized state bound on a potential maximum, created artificially (for example in a ring) or created by the pres-

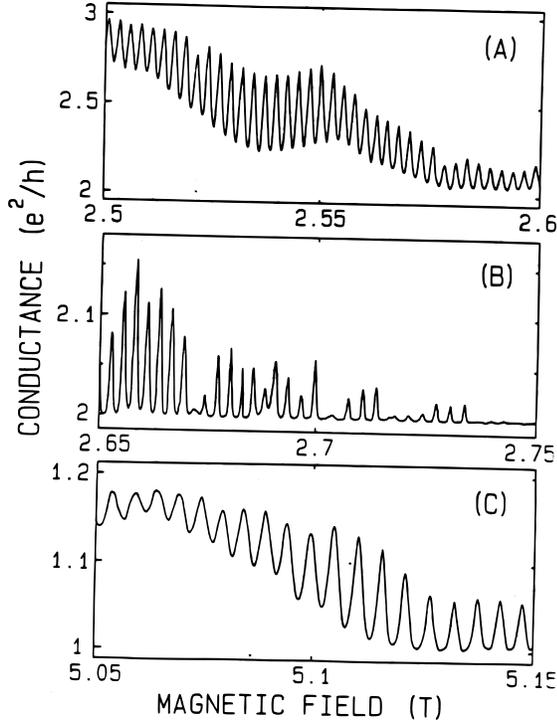


FIG. 18 Magnetoconductance of a quantum dot in the 2DEG of a GaAs-AlGaAs heterostructure of  $1.5 \mu\text{m}$  diameter, with point contacts at entrance and exit serving as tunnel barriers. The temperature is 30 mK. (a) and (b) Aharonov-Bohm magnetoconductance oscillations due to resonant tunneling through bound states belonging to the third (spin-split) edge channel. The first two (spin-split) Landau levels are fully transmitted (cf. Fig. 17c). (c) Resonant tunneling through bound states belonging to the second (spin-split) edge channel. The first (spin-split) edge channel is fully transmitted. (From Van Wees et al.<sup>28</sup>)

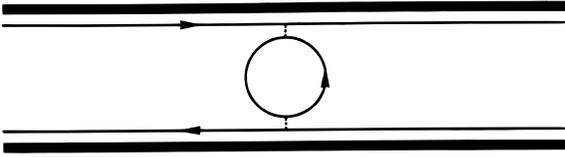


FIG. 19 A circulating edge state bound on a local potential maximum causes resonant backscattering, thereby providing an alternative mechanism for Aharonov-Bohm magnetoconductance oscillations. (From Beenakker and Van Houten.<sup>8</sup>)

ence of disorder.<sup>83</sup> The mechanism is illustrated in Fig. 19. Resonant backscattering leads to a periodic suppression of the conductance, in contrast to the periodic enhancement considered above.

## B. Coulomb blockade of the Aharonov-Bohm effect

Single-electron tunneling is governed by the transport of a single quantum of charge  $e$ . The Aharonov-Bohm effect is governed by the flux quantum  $h/e$ . The present subsection addresses the interplay of these two quanta of nature in the integer quantum Hall effect regime.

In the previous subsection we have discussed how resonant tunneling through circulating edge states can lead to magnetoconductance oscillations in a quantum dot with a well-defined periodicity  $\Delta B$ , similar to the usual Aharonov-Bohm effect in a ring. There is, however, an essential difference between the two geometries if only a single Landau level is occupied.<sup>30</sup> In each period  $\Delta B$  the number of states below a given energy increases by one in a dot — but stays constant in a ring. As a result, the Aharonov-Bohm oscillations in the magnetoconductance of a quantum dot are accompanied by an increase of the charge of the dot by one elementary charge per period. That is of no consequence if the Coulomb repulsion of the electrons can be neglected, but becomes important if the dot has a small capacitance  $C$  to the reservoirs, since then the electrostatic energy  $e^2/C$  associated with the incremental charging by single electrons has to be taken into account.

Following Ref.<sup>30</sup>, we analyze this problem by combining the results reviewed in the previous sections. We apply Eq. (2.6) to the energy spectrum shown in Fig. 16a. We consider here only the edge states from the lowest (spin-split) Landau level, so that the Aharonov-Bohm oscillations have a single periodicity. This corresponds to the strong-magnetic field limit. The magnetic field dependence of the edge states can be described approximately by a sequence of equidistant parallel lines,

$$E_p = \text{constant} - \frac{\Delta E}{\Delta B}(B - p\Delta B), \quad (4.14)$$

see Fig. 16a. For a circular quantum dot of radius  $R$  with a hard-wall confining potential, one can estimate<sup>29</sup>  $\Delta B \sim h/eA$  and  $\Delta E \sim \hbar\omega_c l_m/2R$ . For a smooth confining potential  $V(r)$  (with  $l_m V' \lesssim \hbar\omega_c$ ) one has instead the estimates  $\Delta B \sim (h/e)[A(B) + BA'(B)]^{-1} \sim (h/eA)[1 - \hbar\omega_c/RV'(R)]^{-1}$ ,<sup>28</sup> and  $\Delta E \sim h/\tau \sim l_m^2 V'(R)/R$ , where  $A(B)$  is the area enclosed by the equipotential of radius  $R$  at the guiding center energy  $V(R) = E - \frac{1}{2}\hbar\omega_c$  (cf. Eq. (4.4) for  $n = 1$ ). [The estimate for  $\Delta E$  results from the correspondence between the level spacing and the period  $\tau$  of the classical motion along the equipotential, with guiding-center-drift velocity  $V'(R)/eB$ .]

On substitution of Eq. (4.14) into Eq. (2.6), one finds the condition

$$N \left( \Delta E + \frac{e^2}{C} \right) = \frac{\Delta E}{\Delta B} B_N + E_F + \text{constant} \quad (4.15)$$

for the magnetic field value  $B_N$  of the  $N$ -th conductance peak. The  $B$ -dependence of the reservoir Fermi energy can be neglected in Eq. (4.15) in the case of a hard-wall confining potential (since  $dE_F/dB \approx \hbar\omega_c/B \ll$

$\Delta E/\Delta B$ ). The periodicity  $B^* \equiv B_{N+1} - B_N$  of the Aharonov-Bohm oscillations is thus given by

$$\Delta B^* = \Delta B \left( 1 + \frac{e^2}{C\Delta E} \right). \quad (4.16)$$

[In the case of a smooth confining potential, the term  $\Delta B$  in the enhancement factor of Eq. (4.16) should be replaced by the term  $\Delta B[1 + (\Delta B/\Delta E)(dE_F/dB)]^{-1} \sim h/eA$ , under the assumption that the Fermi energy in the reservoir is pinned to the lowest Landau level, i.e.  $E_F = \frac{1}{2}\hbar\omega_c$ .] We conclude from Eq. (4.16) that the charging energy enhances the spacing of two subsequent peaks in  $G$  versus  $B$  by a factor  $1 + e^2/C\Delta E$ . The periodicity of the magnetoconductance oscillations is lost if  $\Delta B^*$  becomes so large that the linear approximation (4.14) for  $E_p(B)$  breaks down. Since Eq. (4.14) holds at most over an energy range of the Landau level separation  $\hbar\omega_c$ , this suppression of the Aharonov-Bohm effect occurs when  $(\Delta E/\Delta B)\Delta B^* \gtrsim \hbar\omega_c$ , i.e. when  $e^2/C \gtrsim \hbar\omega_c$ .

The Aharonov-Bohm oscillations with bare periodicity  $\Delta B = h/eA$  are recovered if one makes a hole in the disc, which is sufficiently large that the area  $S$  of the conducting region is much smaller than the enclosed area  $A$ . The inner perimeter of the resulting ring supports a second set of edge states, which travel around the ring in opposite direction as the first set of edge states at the outer perimeter. We compare in Fig. 16 the energy spectrum for a disc<sup>80</sup> and a ring.<sup>81</sup> The two sets of clockwise and counter-clockwise propagating edge states in a ring are distinguished by the opposite sign of  $dE_p/dB$ , i.e. of the magnetic moment. Each set of edge states leads to oscillations in the magnetoconductance of a ring with the same period  $\Delta B$ , but shifted in phase (and in general with different amplitude, because the edge states at the inner perimeter have a smaller tunneling probability to the reservoir than those at the outer perimeter). The charging energy does not modify  $\Delta B$  in a ring, because

$$E_p(B) = E_p(B + \Delta B) \quad (\text{ring}).$$

In a disc, in contrast, one has according to Eq. (4.14),

$$E_p(B) = E_{p+1}(B + \Delta B) \quad (\text{disc}).$$

To illustrate the difference, we compare in Fig. 20 for disc and ring the renormalized energy levels  $E_p^*$  [defined in Eq. (2.6)]. The effect of the charging energy in a ring is to open an energy gap of magnitude  $e^2/C$  in  $E_p^*$ . This gap will not affect the conductance oscillations as a function of  $B$  (at constant or slowly varying  $E_F$ ). A controlled experimental demonstration of the influence of Coulomb repulsion on the AB effect may be obtained in a system which can be transformed from a disc into a ring. What we have in mind is a geometry such as shown in Fig. 21, which has an additional gate within the gates shaping the disc. By applying a negative voltage to this additional gate one depletes the central region of the quantum dot, thereby transforming it into a ring. In order

to estimate the mutual capacitance  $C$  between the undepleted quantum disc and the adjacent 2DEG reservoirs, we note that only a circular strip of width  $l_m$  and radius  $R$  along the circumference of the disc contributes to  $C$ . The central region of the dot is incompressible in the quantum Hall effect regime, and thus behaves as a dielectric as far as the electrostatics is concerned. The capacitance  $C$  contains contributions from the self-capacitance of this strip as well as from its capacitance to the gate. (We assume that the gate is electrically connected to the 2DEG reservoirs.) Both contributions are of order  $\epsilon R$ , with a numerical prefactor of order unity which depends only logarithmically on the width of the strip and the separation to the gate<sup>57</sup> ( $\epsilon$  is the dielectric constant). A dot radius of  $1 \mu\text{m}$  yields a charging energy  $e^2/C \simeq 1 \text{ meV}$  for  $\epsilon \simeq 10 \epsilon_0$ . This exceeds the level separation  $\Delta E \simeq \hbar\omega_c l_m / 2R \simeq 2 \times 10^{-5} \text{ eV}(\text{T}/B)$  at a field of a few T. A significant increase of the frequency of the AB oscillations should thus be observable on depletion of the central region of the dot, even for a relatively large radius of  $1 \mu\text{m}$ . To observe a full suppression of the AB effect in a sub-micron disc with  $e^2/C \gtrsim \hbar\omega_c$ , and its recovery on transformation to a ring, would be an ultimate test of the theory<sup>30</sup> reviewed here.

The difference between a ring and a disc disappears if more than a single Landau level is occupied in the disc. This occurs in the upper-left-hand corner in Fig. 16a. The energy spectrum in a disc now forms a mesh pattern which is essentially equivalent to that in a ring (Fig. 16b). There is no Coulomb-blockade of the Aharonov-Bohm effect in such a case,<sup>32</sup> as discussed below.

### C. Experiments on quantum dots

We propose that the observation in a quantum dot of Aharonov-Bohm magnetoconductance oscillations by Van Wees et al.<sup>82</sup> was made possible by the presence of one or more extended edge channels, as in Fig. 17c (all of the successful observations were, to our knowledge, made for  $G > e^2/h$ ). In the presence of extended states the charge on the dot varies continuously, so that the Coulomb blockade of the Aharonov-Bohm effect discussed above is not operative. A direct experimental test of this interpretation would be desirable. This could be done by repeating the experiment in different magnetic field regimes, both with and without the presence of an extended edge channel.

Even if the magnetoconductance oscillations are suppressed, it is still possible to observe Coulomb-blockade oscillations in the conductance as a function of gate voltage (at fixed magnetic field). Previous observations of conductance oscillations as a function of gate voltage which were not observed as a function of  $B$  have been attributed to the Aharonov-Bohm effect,<sup>84,85</sup> but might well have been Coulomb-blockade oscillations instead.

An extended edge channel is one way to remove the Coulomb blockade of the Aharonov-Bohm effect. A sec-

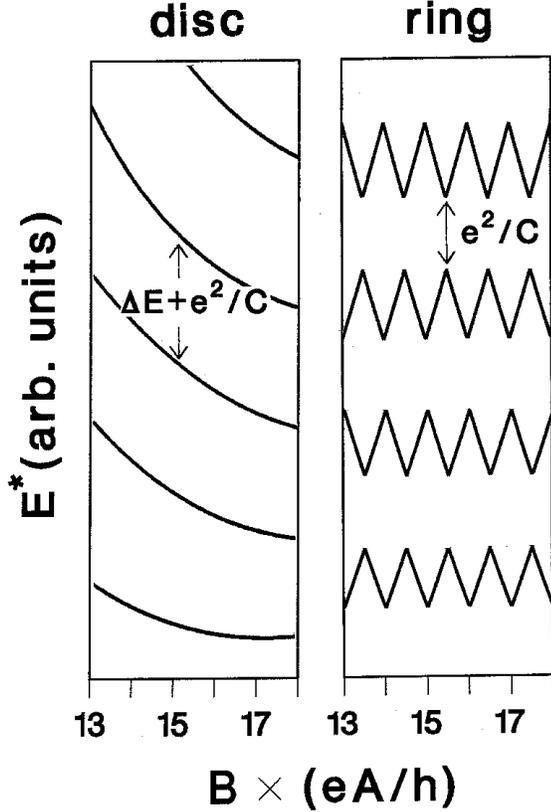


FIG. 20 Renormalized energy levels, defined by Eq. (2.6), corresponding to the bare energy levels shown in Fig. 16. (From Beenakker et al.<sup>30</sup>)

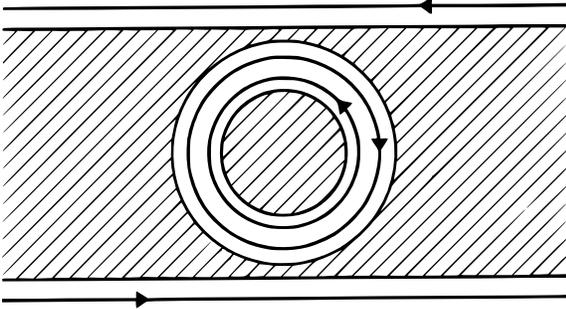


FIG. 21 Schematic layout of a semiconductor nanostructure proposed to demonstrate the Coulomb blockade of the Aharonov-Bohm effect in a quantum dot, and its recovery upon transformation of the device into a ring (by applying a negative voltage to the central gate). (From Beenakker et al.<sup>30</sup>)

ond circulating edge channel in the quantum dot is another way, exploited by McEuen et al.<sup>32</sup> They observed conductance oscillations both as a function of gate voltage and as a function of magnetic field in a quantum dot of the design shown in Fig. 2b. Their main experimental results are reproduced in Fig. 22. The trace of

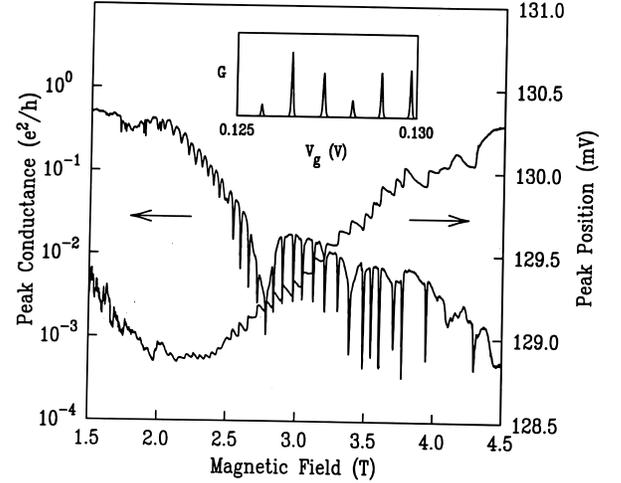


FIG. 22 Effect of a magnetic field on the height and position of a conductance peak in a quantum dot in a GaAs-AlGaAs heterostructure, of the design shown in Fig. 2b. The temperature is 100 mK. Inset: Coulomb-blockade oscillations as a function of gate voltage, for  $B = 3$  T. (From McEuen et al.<sup>32</sup>)

conductance versus gate voltage at  $B = 3$  T (Fig. 22, inset) exhibits the Coulomb-blockade oscillations, with an approximately constant periodicity. The main curves in Fig. 22 show that the height and position of a particular peak vary with  $B$  in a striking fashion. In the region between 2.5 and 3.5 T the peak height is periodically suppressed by as much as an order of magnitude, while the position of the peak oscillates synchronously around a slowly varying background. In this field regime two Landau levels are occupied in the dot, as in Fig. 17b, the lowest of which is spin-degenerate.

These observations have been explained by McEuen et al. in terms of the theory of Coulomb-blockade oscillations in the resonant tunneling regime. The one-electron energy spectrum in the range of two occupied Landau levels is shown in Fig. 23 (for the case of a parabolic confining potential, cf. Fig. 15). The experiment is performed at 100 mK, which is presumably in the resonant tunneling regime  $k_B T < \Delta E$ . Thus, only a single state participates in the conduction through the dot. As indicated in Fig. 23 (heavy line), this state belongs alternately to the first and the second Landau level (corresponding, respectively, to the falling and rising line segments of the sawtooth in Fig. 23). Thus the tunnel rate into this state is alternately large and small. The periodic suppression of the peak height in Fig. 22 reflects this difference in tunnel rates.

According to Eqs. (2.6) and (2.8), the gate voltage of the  $N$ -th peak shifts with  $B$  according to

$$\delta\phi_{\text{gate}} = \frac{1}{\alpha e} \frac{\partial(E_N - E_F)}{\partial B} \delta B, \quad (4.17)$$

with  $\alpha$  defined in Eq. (2.11). McEuen et al. determined  $\alpha$  from the temperature dependence of the peak with,

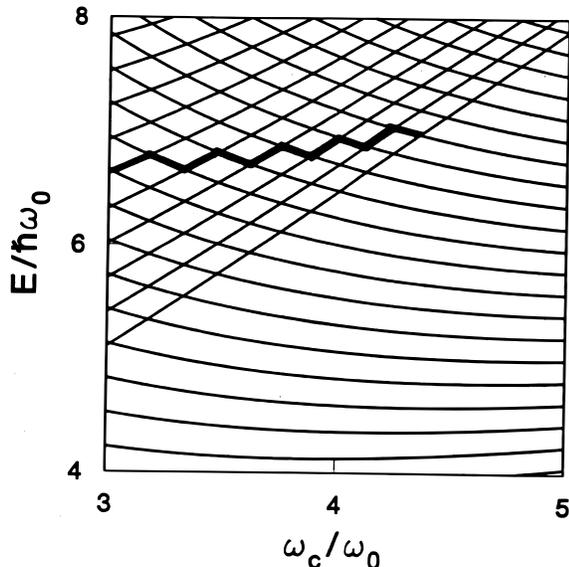


FIG. 23 Close-up of the energy spectrum of Fig. 15 (after McEuen et al.<sup>32</sup>). The heavy line indicates the energy of the highest occupied state for a fixed number (23) of electrons in the dot. In each period of the saw-tooth a single electron is transferred from the second Landau level (rising lines) to the first (falling lines).

and neglected the change in  $E_F$  with  $B$ , as well as the difference between the electrostatic potential  $\phi_{\text{gate}}$  and the measured electrochemical potential  $V_{\text{gate}}$ . The measured shift of the peak position with  $B$  (see Fig. 22) then directly yields the shift in energy  $E_N$ . In this way they were able to map out the one-electron spectrum of the dot (Fig. 24). (To arrive at the bare energy spectrum a constant charging energy  $e^2/C$  was subtracted for each consecutive level.) The similarity of Figs. 23 and 24b is quite convincing. An unexplained effect is the gap in the spectrum around 0.2 meV. Also, the level spacing in the first Landau level (the vertical separation between the falling lines in Fig. 24b) appears to be two times smaller than that in the second Landau level (rising lines). Although this might be related to spin-splitting,<sup>32</sup> we feel that it is more likely that the assumption of a magnetic-field independent  $E_F$  is not justified. If, as should be expected,  $E_F$  is pinned to the second Landau level, then a proper correction for the Fermi level shift with  $B$  would lead to a clock-wise rotation of the entire level spectrum in Fig. 24b around  $(B, E) = (0, 0)$ . The agreement with the theoretical spectrum would then improve.

Coulomb-blockade oscillations as a function of gate voltage in the quantum Hall effect regime were studied by Williamson et al.<sup>17</sup> in a quantum dot of the design shown in Fig. 2c. They found that the amplitude of the oscillations was strongly enhanced compared to zero field, whereas the period was not much affected. (A similar enhancement of the amplitude has been seen in disordered quantum wires, and possible explanations are

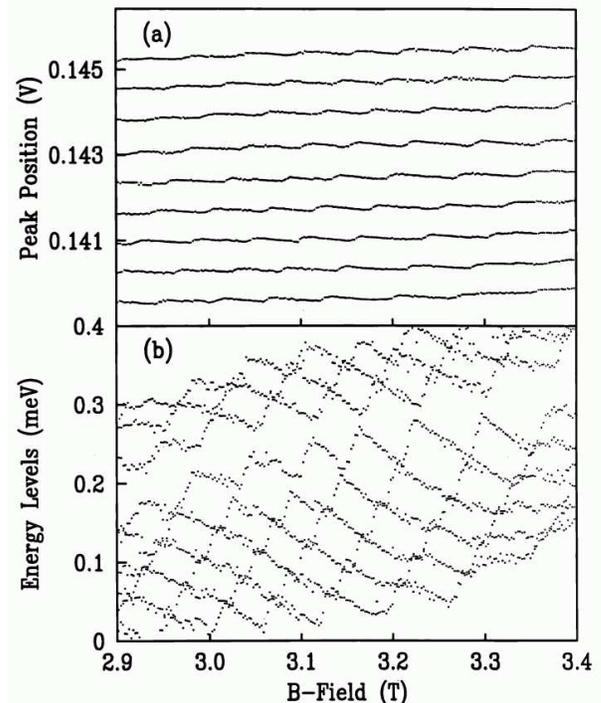


FIG. 24 (a) Peak position as a function of magnetic field for a series of consecutive Coulomb-blockade oscillations in a quantum dot with two occupied Landau levels. (b) Energy spectrum of the dot obtained from the data in (a) after subtraction of the charging energy. (From McEuen et al.<sup>32</sup>)

discussed below.) Representative traces of conductance versus gate voltage at zero field and for  $B = 3.75$  T are reproduced in Fig. 25. The oscillations in the presence of a field are quite spectacular, of amplitude comparable to  $e^2/h$ . These experiments are in the regime where the conductance of the individual barriers approaches  $e^2/h$  as well, and virtual tunneling processes may be important. Experimentally, the conductance minima are not exponentially suppressed (see Fig. 25), even though the temperature was low (100 mK). In addition, the conductance maxima in the zero-field trace exceed  $e^2/h$ . These observations are also indicative of virtual tunneling processes.<sup>53,54</sup> Finally, we would like to draw attention to the slow beating seen in the amplitude of the oscillations at zero field, which is suppressed at  $B = 3.75$  T. Instead, a weak doublet-like structure becomes visible, reminiscent of that reported by Staring et al.<sup>12</sup> for a disordered quantum wire in a strong magnetic field (see Fig. 26), discussed below. Further experimental and theoretical work is needed to understand these intriguing effects of a magnetic field.

#### D. Experiments on disordered quantum wires

The effect of a parallel and perpendicular magnetic field on the conductance oscillations in a narrow channel

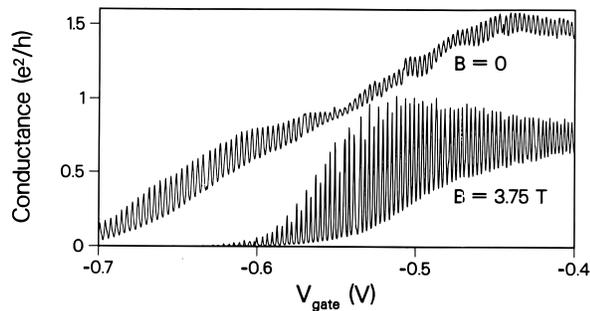


FIG. 25 Effect of a magnetic field on the conductance oscillations in a quantum dot in a GaAs-AlGaAs heterostructure, with a geometry as in Fig. 2c. The temperature is 50 mK. This is an effective two-terminal conductance (obtained from a four-terminal conductance measurement, with the voltage measured diagonally across the dot [Ref.<sup>8</sup>, page 183].) (From Williamson et al.<sup>17</sup>)

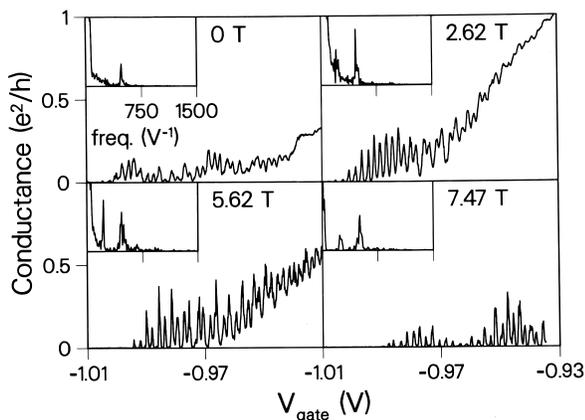


FIG. 26 Effect of a magnetic field on the Coulomb-blockade oscillations a disordered quantum wire (as in Fig. 11), at 50 mK. Insets: Fourier transforms of the data, with the vertical axes of the curves at 0 T and 7.47 T magnified by a factor 2.5, relative to the curves at 2.62 T and 5.62 T. (From Staring et al.<sup>12</sup>)

in a Si inversion layer has been studied by Field et al.<sup>11</sup> Staring et al.<sup>12,13</sup> investigated the effect of a perpendicular field on disordered quantum wires in the 2DEG of a GaAs-AlGaAs heterostructure. Some of the data is reproduced in Fig. 26. The Fourier transforms of the traces of conductance versus gate voltage (insets) demonstrate a  $B$ -independent dominant frequency of  $450 \text{ V}^{-1}$ . Curiously, as the magnetic field is increased a second peak in the Fourier transform emerges at about half the dominant frequency. This second peak corresponds to an amplitude modulation of the peaks, as is most clearly seen in the trace at 5.62 T where high and low peaks alternate in a doublet-like structure. This feature is characteristic of this particular sample. Other channels showed different secondary effects, such as a much more rapid oscillation

superposed on the conductance trace for certain values of the magnetic field.<sup>12</sup> It is likely that the presence of additional segments in the wire plays a role. The period  $\delta V_{\text{gate}} \sim 2.2 \text{ mV}$  of the dominant conductance oscillations is remarkably insensitive to a strong magnetic field. Spin-splitting of the peaks was not observed, even at the highest fields of 8 T. These qualitative observations agree with our interpretation of the effect as Coulomb-blockade oscillations. In Sec. III.B we have already had occasion to show that the temperature dependence of the lineshape of an isolated peak was well accounted for by Eq. (2.18), for a set of parameter values consistent with zero-field experiments.

The height of the conductance peaks is enhanced by a field of intermediate strength ( $2 \text{ T} < B < 6 \text{ T}$ ), followed by a decrease at stronger fields ( $B \sim 7.5 \text{ T}$ ). Also the width of the peaks is reduced in a strong magnetic field. The largest isolated peaks (found in a different sample<sup>13</sup>) approach a height of  $e^2/h$ , measured two-terminally. A similar enhancement of the amplitude of the Coulomb-blockade oscillations by a magnetic field was observed in a quantum dot<sup>17</sup> (see Fig. 25). One explanation is that the inelastic scattering rate is reduced by a magnetic field. In the low-temperature regime  $k_B T \lesssim h\Gamma$  this makes the peaks higher and narrower (cf. Sec. II.B). In a disordered quantum wire the magnetic suppression of backscattering provides another mechanism for an enhancement of the peak height because of the resulting reduction in series resistance.<sup>13</sup> Additionally, the modulation of the Fermi level in the quantum Hall effect regime may lead to a non-monotonic variation with  $B$  of the transmission probability  $T(E_F)$ , and thus presumably of the tunnel rates  $h\Gamma$ . The level degeneracy varies with  $B$ , becoming large when the Fermi energy coincides with a bulk Landau level in the dot. This may also give rise to variations in the peak height.<sup>34</sup> These are tentative explanations of the surprising magnetic field dependence of the amplitude of the Coulomb-blockade oscillations, which remains to be elucidated.

We close this subsection by noting that Staring et al.<sup>12</sup> also measured magnetoconductance traces at fixed gate voltage. In contrast to the gate voltage scans, these exhibited irregular structure only, with strong features corresponding to depopulation of Landau levels. The absence of regular oscillations constitutes the first experimental evidence for the predicted<sup>30</sup> Coulomb blockade of the Aharonov-Bohm effect.

## Acknowledgments

Valuable discussions with S. Colak, L. P. Kouwenhoven, N. C. van der Vaart, J. G. Williamson, and the support of J. Wolter and M. F. H. Schuurmans are gratefully acknowledged. Our experimental work has been made possible by C. T. Foxon who has grown the necessary samples by molecular beam epitaxy, and by C. E. Timmering who took care of the technology. We have

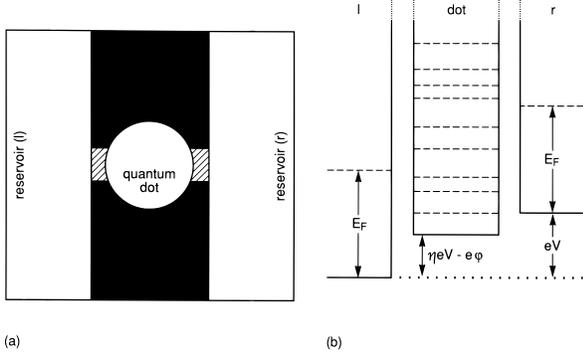


FIG. 27 (a) Schematic cross-section of the geometry studied in this appendix, consisting of a confined region (“quantum dot”) weakly coupled to two electron reservoirs via tunnel barriers (hatched). (b) Profile of the electrostatic potential energy (solid curve) along a line through the tunnel barriers. The Fermi levels in the left and right reservoirs, and the discrete energy levels in the quantum dot are indicated (dashed lines).

benefitted from interactions with the participants of the NATO ASI on Single Charge Tunneling. We thank our colleagues at M.I.T., Delft, and Philips for their permission to reproduce some of their results. This research was partly funded under the ESPRIT basic research action project 3133.

## APPENDIX A: Conductance of a quantum dot coupled to two electron reservoirs

Following the treatment by Beenakker,<sup>19</sup> we derive in this appendix Eq. (2.18) for the conductance of a confined region which is weakly coupled via tunnel barriers to two electron reservoirs. The confined region, or “quantum dot”, has single-electron energy levels at  $E_p$  ( $p = 1, 2, \dots$ ), labeled in ascending order and measured relative to the bottom of the potential well. Each level contains either one or zero electrons. Spin degeneracy can be included by counting each level twice, and other degeneracies can be included similarly. Each reservoir is taken to be in thermal equilibrium at temperature  $T$  and chemical potential  $E_F$ . A continuum of states is assumed in the reservoirs, occupied according to the Fermi-Dirac distribution

$$f(E - E_F) = \left[ 1 + \exp\left(\frac{E - E_F}{kT}\right) \right]^{-1}. \quad (\text{A1})$$

In Fig. 27 we show schematically a cross-section of the geometry, and the profile of the electrostatic potential energy along a line through the tunnel barriers.

A current  $I$  can be passed through the dot by applying a potential difference  $V$  between the two reservoirs. The tunnel rate from level  $p$  to the left and right reservoirs in Fig. 27 is denoted by  $\Gamma_p^l$  and  $\Gamma_p^r$ , respectively. We assume that both  $kT$  and  $\Delta E$  are  $\gg h(\Gamma^l + \Gamma^r)$  (for all levels

participating in the conduction), so that the finite width  $h\Gamma = h(\Gamma^l + \Gamma^r)$  of the transmission resonance through the quantum dot can be disregarded. This assumption allows us to characterize the state of the quantum dot by a set of occupation numbers, one for each energy level. (As discussed in Sec. II.B, the restriction  $kT, \Delta E \gg h\Gamma$  results in the conductance being much smaller than the quantum  $e^2/h$ .) We also assume conservation of energy in the tunnel process, thus neglecting contributions of higher order in  $\Gamma$  from tunneling via a *virtual* intermediate state in the quantum dot.<sup>53,53</sup> We finally assume that inelastic scattering takes place exclusively in the reservoirs — not in the quantum dot. The effect of inelastic scattering in the quantum dot is considered in Ref.<sup>19</sup>.

Energy conservation upon tunneling from an initial state  $p$  in the quantum dot (containing  $N$  electrons) to a final state in the left reservoir at energy  $E^{f,l}$  (in excess of the local electrostatic potential energy), requires that

$$E^{f,l}(N) = E_p + U(N) - U(N - 1) + \eta eV. \quad (\text{A2})$$

Here  $\eta$  is the fraction of the applied voltage  $V$  which drops over the left barrier. (As we will see, this parameter  $\eta$  drops out of the final expression for the conductance in linear response.) The energy conservation condition for tunneling from an initial state  $E^{i,l}$  in the left reservoir to a final state  $p$  in the quantum dot is

$$E^{i,l}(N) = E_p + U(N + 1) - U(N) + \eta eV, \quad (\text{A3})$$

where [as in Eq. (A2)]  $N$  is the number of electrons in the dot *before* the tunneling event. Similarly, for tunneling between the quantum dot and the right reservoir one has the conditions

$$E^{f,r}(N) = E_p + U(N) - U(N - 1) - (1 - \eta)eV \quad (\text{A4})$$

$$E^{i,r}(N) = E_p + U(N + 1) - U(N) - (1 - \eta)eV \quad (\text{A5})$$

where  $E^{i,r}$  and  $E^{f,r}$  are the energies of the initial and final states in the right reservoir.

The stationary current through the left barrier equals that through the right barrier, and is given by

$$I = -e \sum_{p=1}^{\infty} \sum_{\{n_i\}} \Gamma_p^l P(\{n_i\}) (\delta_{n_p,0} f(E^{i,l}(N) - E_F) - \delta_{n_p,1} [1 - f(E^{f,l}(N) - E_F)]) . \quad (\text{A6})$$

The second summation is over all realizations of occupation numbers  $\{n_1, n_2, \dots\} \equiv \{n_i\}$  of the energy levels in the quantum dot, each with stationary probability  $P(\{n_i\})$ . (The numbers  $n_i$  can take on only the values 0 and 1.) In equilibrium, this probability distribution is the Gibbs distribution in the grand canonical ensemble:

$$P_{\text{eq}}(\{n_i\}) = \frac{1}{Z} \exp \left[ -\frac{1}{kT} \left( \sum_{i=1}^{\infty} E_i n_i + U(N) - N E_F \right) \right], \quad (\text{A7})$$

where  $N \equiv \sum_i n_i$ , and  $Z$  is the partition function,

$$Z = \sum_{\{n_i\}} \exp \left[ -\frac{1}{kT} \left( \sum_{i=1}^{\infty} E_i n_i + U(N) - N E_F \right) \right]. \quad (\text{A8})$$

The non-equilibrium probability distribution  $P$  is a stationary solution of the kinetic equation

$$\begin{aligned} \frac{\partial}{\partial t} P(\{n_i\}) &= 0 \\ &= - \sum_p P(\{n_i\}) \delta_{n_p,0} (\Gamma_p^l f(E^{i,l}(N) - E_F) + \Gamma_p^r f(E^{i,r}(N) - E_F)) \\ &\quad - \sum_p P(\{n_i\}) \delta_{n_p,1} (\Gamma_p^l [1 - f(E^{f,l}(N) - E_F)] + \Gamma_p^r [1 - f(E^{f,r}(N) - E_F)]) \\ &\quad + \sum_p P(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots) \delta_{n_p,0} (\Gamma_p^l [1 - f(E^{f,l}(N+1) - E_F)] + \Gamma_p^r [1 - f(E^{f,r}(N+1) - E_F)]) \\ &\quad + \sum_p P(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots) \delta_{n_p,1} (\Gamma_p^l f(E^{i,l}(N-1) - E_F) + \Gamma_p^r f(E^{i,r}(N-1) - E_F)). \end{aligned} \quad (\text{A9})$$

The kinetic equation (A9) for the stationary distribution function is equivalent to the set of detailed balance equations (one for each  $p = 1, 2, \dots$ )

$$\begin{aligned} &P(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots) (\Gamma_p^l [1 - f(E^{f,l}(\tilde{N} + 1) - E_F)] + \Gamma_p^r [1 - f(E^{f,r}(\tilde{N} + 1) - E_F)]) \\ &= P(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots) (\Gamma_p^l f(E^{i,l}(\tilde{N}) - E_F) + \Gamma_p^r f(E^{i,r}(\tilde{N}) - E_F)), \end{aligned} \quad (\text{A10})$$

with the notation  $\tilde{N} \equiv \sum_{i \neq p} n_i$ .

A similar set of equations formed the basis for the work of Averin, Korotkov, and Likharev on the Coulomb staircase in the non-linear  $I-V$  characteristic of a quantum dot.<sup>34</sup> To simplify the solution of the kinetic equation, they assumed that the charging energy  $e^2/C$  is much greater than the average level spacing  $\Delta E$ . In this chapter we restrict ourselves to the regime of linear response, appropriate for the Coulomb-blockade oscillations. Then the conductance can be calculated exactly and analytically.

The (two-terminal) linear response conductance  $G$  of the quantum dot is defined as  $G = I/V$  in the limit  $V \rightarrow 0$ . To solve the linear response problem we substitute

$$P(\{n_i\}) \equiv P_{\text{eq}}(\{n_i\}) \left( 1 + \frac{eV}{kT} \Psi(\{n_i\}) \right) \quad (\text{A11})$$

into the detailed balance equation (A10), and linearize with respect to  $V$ . One finds

$$\begin{aligned} &P_{\text{eq}}(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots) (\Psi(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots) (\Gamma_p^l + \Gamma_p^r) [1 - f(\epsilon)] - [\Gamma_p^l \eta - \Gamma_p^r (1 - \eta)] kT f'(\epsilon)) \\ &= P_{\text{eq}}(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots) (\Psi(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots) (\Gamma_p^l + \Gamma_p^r) f(\epsilon) + [\Gamma_p^l \eta - \Gamma_p^r (1 - \eta)] kT f'(\epsilon)), \end{aligned} \quad (\text{A12})$$

where  $f'(\epsilon) \equiv df(\epsilon)/d\epsilon$ , and we have abbreviated  $\epsilon \equiv E_p + U(\tilde{N} + 1) - U(\tilde{N}) - E_F$ .

Equation (A12) can be simplified by making subsequently the substitutions

$$1 - f(\epsilon) = f(\epsilon) e^{\epsilon/kT}, \quad (\text{A13})$$

$$P_{\text{eq}}(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots) = P_{\text{eq}}(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots) e^{-\epsilon/kT}, \quad (\text{A14})$$

$$kT f'(\epsilon) (1 + e^{-\epsilon/kT}) = -f(\epsilon). \quad (\text{A15})$$

The factors  $P_{\text{eq}}$  and  $f$  cancel, and one is left with the simple equation

$$\Psi(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots) = \Psi(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots) + \frac{\Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} - \eta. \quad (\text{A16})$$

The solution is

$$\Psi(\{n_i\}) = \text{constant} + \sum_{i=1}^{\infty} n_i \left( \frac{\Gamma_i^r}{\Gamma_i^l + \Gamma_i^r} - \eta \right). \quad (\text{A17})$$

The constant first term in Eq. (A17) takes care of the normalization of  $P$  to first order in  $V$ , and need not be determined explicitly. Notice that the first order non-equilibrium correction  $\Psi$  to  $P_{\text{eq}}$  is *zero* if  $\eta = \Gamma_i^r/(\Gamma_i^l + \Gamma_i^r)$  for all  $i$ . This will happen in particular for two identical tunnel barriers (when  $\eta = \frac{1}{2}$ ,  $\Gamma_i^l = \Gamma_i^r$ ). Because of the symmetry of the system, the distribution function then contains only terms of *even* order in  $V$ .

Now we are ready to calculate the current  $I$  through the quantum dot to first order in  $V$ . Linearization of Eq. (A6), after substitution of Eq. (A11) for  $P$ , gives

$$\begin{aligned} I &= -e \frac{eV}{kT} \sum_p \sum_{\{n_i\}} \Gamma_p^l P_{\text{eq}}(\{n_i\}) (\delta_{n_p,0} \eta kT f'(\epsilon) + \delta_{n_p,1} \eta kT f'(\epsilon)) \\ &\quad + \Psi(\{n_i\}) \delta_{n_p,0} f(\epsilon) - \Psi(\{n_i\}) \delta_{n_p,1} [1 - f(\epsilon)] \\ &= \frac{e^2 V}{kT} \sum_p \sum_{\{n_i\}} \Gamma_p^l P_{\text{eq}}(\{n_i\}) \delta_{n_p,0} f(E_p + U(N+1) - U(N) - E_F) \\ &\quad \times [\eta + \Psi(n_1, \dots, n_{p-1}, 1, n_{p+1}, \dots) - \Psi(n_1, \dots, n_{p-1}, 0, n_{p+1}, \dots)] \\ &= \frac{e^2 V}{kT} \sum_p \sum_{\{n_i\}} \frac{\Gamma_p^l \Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} P_{\text{eq}}(\{n_i\}) \delta_{n_p,0} f(E_p + U(N+1) - U(N) - E_F). \end{aligned} \quad (\text{A18})$$

In the second equality we have again made use of the identities (A13)–(A15), and in the third equality we have substituted Eq. (A16). Notice that the parameter  $\eta$  has dropped out of the final expression for  $I$ .

We define the equilibrium probability distributions

$$P_{\text{eq}}(N) = \sum_{\{n_i\}} P_{\text{eq}}(\{n_i\}) \delta_{N, \sum_i n_i} = \frac{\exp(-\Omega(N)/kT)}{\sum_N \exp(-\Omega(N)/kT)}, \quad (\text{A19})$$

$$\begin{aligned} F_{\text{eq}}(E_p|N) &= \frac{1}{P_{\text{eq}}(N)} \sum_{\{n_i\}} P_{\text{eq}}(\{n_i\}) \delta_{n_p,1} \delta_{N, \sum_i n_i} \\ &= \exp(\mathcal{F}(N)/kT) \sum_{\{n_i\}} \exp\left(-\frac{1}{kT} \sum_{i=1}^{\infty} E_i n_i\right) \delta_{n_p,1} \delta_{N, \sum_i n_i}. \end{aligned} \quad (\text{A20})$$

Here  $\Omega(N)$  is the thermodynamic potential of the quantum dot, and  $\mathcal{F}(N)$  is the free energy of the internal degrees of freedom:

$$\Omega(N) = \mathcal{F}(N) + U(N) - N E_F, \quad (\text{A21})$$

$$\mathcal{F}(N) = -kT \ln \left[ \sum_{\{n_i\}} \exp\left(-\frac{1}{kT} \sum_{i=1}^{\infty} E_i n_i\right) \delta_{N, \sum_i n_i} \right]. \quad (\text{A22})$$

The function  $P_{\text{eq}}(N)$  is the probability that the quantum dot contains  $N$  electrons in equilibrium; The function  $F_{\text{eq}}(E_p|N)$  is the conditional probability in equilibrium that level  $p$  is occupied given that the quantum dot contains  $N$  electrons. In terms of these distribution functions, the conductance  $G = I/V$  resulting from Eq.

(A18) equals

$$\begin{aligned} G &= \frac{e^2}{kT} \sum_{p=1}^{\infty} \sum_{N=0}^{\infty} \frac{\Gamma_p^l \Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} P_{\text{eq}}(N) [1 - F_{\text{eq}}(E_p|N)] \\ &\quad \times f(E_p + U(N+1) - U(N) - E_F). \end{aligned} \quad (\text{A23})$$

In view of Eqs. (A13) and (A14), Eq. (A23) can equivalently be written in the form

$$\begin{aligned} G &= \frac{e^2}{kT} \sum_{p=1}^{\infty} \sum_{N=1}^{\infty} \frac{\Gamma_p^l \Gamma_p^r}{\Gamma_p^l + \Gamma_p^r} P_{\text{eq}}(N) F_{\text{eq}}(E_p|N) \\ &\quad \times [1 - f(E_p + U(N) - U(N-1) - E_F)]. \end{aligned} \quad (\text{A24})$$

Redefining  $P_{\text{eq}}(N) F_{\text{eq}}(E_p|N) = P_{\text{eq}}(N, n_p = 1)$  we find Eq. (2.18) as it appears in Sec. II.B.

## References

- [1] L. S. Kuz'min and K. K. Likharev, Pis'ma Zh. Eksp. Teor. Fiz. **45**, 389 (1987). [JETP Lett. **45**, 495 (1987)].
- [2] T. A. Fulton and G. J. Dolan, Phys. Rev. Lett. **59**, 109 (1987).
- [3] K. Mullen, E. Ben-Jacob, R. C. Jaklevic, and Z. Schuss, Phys. Rev. B **37**, 98 (1988); M. Amman, K. Mullen, and E. Ben-Jacob, J. Appl. Phys. **65**, 339 (1989).
- [4] K. K. Likharev, IBM J. Res. Dev. **32**, 144 (1988); D. V. Averin and K. K. Likharev, in: *Mesoscopic Phenomena in Solids*, ed. by B. L. Al'tshuler, P. A. Lee, and R. A. Webb, (Elsevier, Amsterdam, 1991). This is a comprehensive review of single-electron tunneling in metals.
- [5] L. I. Glazman and R. I. Shekhter, J. Phys. Condens. Matter **1**, 5811 (1989).
- [6] R. I. Shekhter, Zh. Eksp. Teor. Fiz. **63**, 1410 (1972) [Sov. Phys. JETP **36**, 747 (1973)].
- [7] I. O. Kulik and R. I. Shekhter, Zh. Eksp. Teor. Fiz. **68**, 623 (1975) [Sov. Phys. JETP **41**, 308 (1975)].
- [8] For a review of theoretical and experimental aspects of quantum transport in semiconductor nanostructures, see: C. W. J. Beenakker and H. van Houten, Solid State Physics **44**, 1 (1991); cond-mat/0412664
- [9] J. H. F. Scott-Thomas, S. B. Field, M. A. Kastner, H. I. Smith, and D. A. Antoniadis, Phys. Rev. Lett. **62**, 583 (1989).
- [10] U. Meirav, M. A. Kastner, M. Heiblum, and S. J. Wind, Phys. Rev. B **40**, 5871 (1989).
- [11] S. B. Field, M. A. Kastner, U. Meirav, J. H. F. Scott-Thomas, D. A. Antoniadis, H. I. Smith, and S. J. Wind, Phys. Rev. B **42**, 3523 (1990).
- [12] A. A. M. Staring, H. van Houten, C. W. J. Beenakker, and C. T. Foxon, in: *High Magnetic Fields in Semiconductor Physics III*, ed. by G. Landwehr, (Springer, Berlin, 1991).
- [13] A. A. M. Staring, H. van Houten, C. W. J. Beenakker, and C. T. Foxon, Phys. Rev. B, to be published.
- [14] C. de Graaf, J. Caro, S. Radelaar, V. Lauer, and K. Heyers, submitted to Phys. Rev. B.
- [15] U. Meirav, M. A. Kastner, and S. J. Wind, Phys. Rev. Lett. **65**, 771 (1990).
- [16] L. P. Kouwenhoven, N. C. van der Vaart, A. T. Johnson, C. J. P. M. Harmans, J. G. Williamson, A. A. M. Staring, and C. T. Foxon, *Festkörperprobleme/Advances in Solid State Physics* **31** (Vieweg, Braunschweig, 1991).
- [17] J. G. Williamson, A. A. M. Staring, H. van Houten, L. P. Kouwenhoven, and C. T. Foxon, to be published; A. A. M. Staring, J. G. Williamson, H. van Houten, C. W. J. Beenakker, L. P. Kouwenhoven, and C. T. Foxon, Physica B, to be published.
- [18] A. L. Efros and B. I. Shklovskii, in: *Electron-Electron Interactions in Disordered Systems*, ed. by A. L. Efros and M. Pollak, (North-Holland, Amsterdam, 1985).
- [19] C. W. J. Beenakker, Phys. Rev. B **44**, 1646 (1991).
- [20] Y. Meir, N. S. Wingreen, and P. A. Lee, Phys. Rev. Lett. **66**, 3048 (1991).
- [21] M. A. Kastner, S. B. Field, U. Meirav, J. H. F. Scott-Thomas, D. A. Antoniadis, and H. I. Smith, Phys. Rev. Lett. **63**, 1894 (1989).
- [22] A. I. Larkin and P. A. Lee, Phys. Rev. B **17**, 1596 (1978).
- [23] P. A. Lee and T. M. Rice, Phys. Rev. B **19**, 3970 (1979).
- [24] D. V. Averin and K. K. Likharev, unpublished.
- [25] H. van Houten and C. W. J. Beenakker, Phys. Rev. Lett. **63**, 1893 (1989).
- [26] F. M. de Aguiar and D. A. Wharam, Phys. Rev. **43**, 15 (1991).
- [27] J. Mašek and B. Kramer, submitted to Europhys. Lett.
- [28] B. J. van Wees, L. P. Kouwenhoven, C. J. P. M. Harmans, J. G. Williamson, C. E. Timmering, M. E. I. Broekaart, C. T. Foxon, and J. J. Harris, Phys. Rev. Lett. **62**, 2523 (1989).
- [29] U. Sivan and Y. Imry, Phys. Rev. Lett. **61**, 1001 (1988); U. Sivan, Y. Imry, and C. Hartzstein, Phys. Rev. B **39**, 1242 (1989).
- [30] C. W. J. Beenakker, H. van Houten, and A. A. M. Staring, Phys. Rev. B **44**, 657 (1991).
- [31] C. W. J. Beenakker, H. van Houten, and A. A. M. Staring, in: *Granular Nano-electronics*, ed. by D. K. Ferry, R. Barker, and C. Jacoboni, (Plenum, New York, 1991).
- [32] P. L. McEuen, E. B. Foxman, U. Meirav, M. A. Kastner, Y. Meir, N. S. Wingreen, and S. J. Wind, Phys. Rev. Lett. **66**, 1926 (1991).
- [33] P. A. Lee, Phys. Rev. Lett. **65**, 2206 (1990).
- [34] D. V. Averin and A. N. Korotkov, Zh. Eksp. Teor. Fiz. **97**, 1661 (1990) [Sov. Phys. JETP **70**, 937 (1990)]; A. N. Korotkov, D. V. Averin, and K. K. Likharev, Physica B **165** & **166**, 927 (1990); D. V. Averin, A. N. Korotkov, and K. K. Likharev, submitted to Phys. Rev. B.
- [35] A. Groshev, Phys. Rev. B **42**, 5895 (1990); A. Groshev, T. Ivanov, and V. Valtchinov, Phys. Rev. Lett. **66**, 1082 (1991).
- [36] L. P. Kouwenhoven, A. T. Johnson, N. C. van der Vaart, C. J. P. M. Harmans, and C. T. Foxon, unpublished.
- [37] L. J. Geerligs, V. F. Anderegg, P. A. M. Holweg, J. E. Mooij, H. Pothier, D. Estève, C. Urbina, and M. H. Devoret, Phys. Rev. Lett. **64**, 2691 (1990).
- [38] H. Pothier, P. Lafarge, C. Urbina, D. Estève, and M. H. Devoret, Europhys. Lett. **17**, 249 (1992).
- [39] M. A. Reed, J. N. Randall, R. J. Aggarwal, R. J. Matyi, T. M. Moore, and A. E. Wetsel, Phys. Rev. Lett. **60**, 535 (1988).
- [40] B. Su, V. J. Goldman, and J. E. Cunningham, preprint.
- [41] W. B. Kinard, M. H. Weichold, G. F. Spencer, and W. P. Kirk, in: *Nanostructure Physics and Fabrication*, ed. by M. A. Reed and W. P. Kirk, (Academic, New York, 1989).
- [42] H. van Houten, C. W. J. Beenakker, and B. J. van Wees, in: *Nanostructured Systems*, ed. by M. A. Reed, (a volume of Semiconductors and Semimetals, Academic, New York, 1991).
- [43] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon, Phys. Rev. Lett. **60**, 848 (1988).
- [44] D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. C **21**, L209 (1988).
- [45] J. A. Nixon and J. H. Davies, Phys. Rev. B **41**, 7929 (1990).
- [46] G. W. Bryant, Phys. Rev. Lett. **59**, 1140 (1987).
- [47] P. A. Maksym and T. Chakraborty, Phys. Rev. Lett. **65**, 108 (1990).
- [48] A. Kumar, S. E. Laux, and F. Stern, Phys. Rev. B **42**, 5166 (1990).
- [49] M. Büttiker, Phys. Rev. B **33**, 3020 (1986).
- [50] M. Büttiker, IBM J. Res. Dev. **32**, 63 (1988).

- [51] A. D. Stone and P. A. Lee, Phys. Rev. Lett. **54**, 1196 (1985).
- [52] R. Landauer, in: *Localization, Interaction, and Transport Phenomena*, ed. by B. Kramer, G. Bergmann, and Y. Bruynseraede, (Springer, Berlin, 1985).
- [53] D. V. Averin and A. A. Odintsov, Phys. Lett. A **140**, 251 (1989); D. V. Averin and Yu. V. Nazarov, Phys. Rev. Lett. **65**, 2446 (1990).
- [54] L. I. Glazman and K. A. Matveev, Pis'ma Zh. Eksp. Teor. Fiz. **51**, 425 (1990) [JETP Lett. **51**, 484 (1990)].
- [55] Yu. V. Nazarov, Solid State Comm. **75**, 669 (1990); A. N. Korotkov and Yu. V. Nazarov, Physica B, to be published.
- [56] S. E. Laux, D. E. Frank, and F. Stern, Surf. Sci. **196**, 101 (1988).
- [57] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, New York, 1960).
- [58] K. K. Likharev, N. S. Bakhvalov, G. S. Kazacha, and S. I. Serdyukova, IEEE Trans. Magn. **25**, 1436 (1989).
- [59] M. Amman, E. Ben-Jacob, and K. Mullen, Phys. Lett. A **142**, 431 (1989).
- [60] A. B. Fowler, A. Hartstein, and R. A. Webb, Phys. Rev. Lett. **48**, 196 (1982); R. A. Webb, A. Hartstein, J. J. Wainer, and A. B. Fowler, Phys. Rev. Lett. **54**, 1577 (1985).
- [61] R. F. Kwasnick, M. A. Kastner, J. Melngailis, and P. A. Lee, Phys. Rev. Lett. **52**, 224 (1984); M. A. Kastner, R. F. Kwasnick, J. C. Licini, and D. J. Bishop, Phys. Rev. B **36**, 8015 (1987).
- [62] P. A. Lee, Phys. Rev. Lett. **53**, 2042 (1984).
- [63] R. K. Kalia, W. Xue, and P. A. Lee, Phys. Rev. Lett. **57**, 1615 (1986).
- [64] C. J. Adkins, in: *Hopping and Related Phenomena*, ed. by H. Fritsche and M. Pollak, (World Scientific, New Jersey, 1990).
- [65] B. L. Al'tshuler Pis'ma Zh. Eksp. Teor. Fiz. **41**, 530 (1985) [JETP Lett. **41**, 648 (1985)].
- [66] P. A. Lee and A. D. Stone, Phys. Rev. Lett. **55**, 1622 (1985).
- [67] A. B. Fowler, G. L. Timp, J. J. Wainer, and R. A. Webb, Phys. Rev. Lett. **57**, 138 (1986); A. B. Fowler, J. J. Wainer, and R. A. Webb, IBM J. Res. Dev. **32**, 372 (1988).
- [68] M. Green and M. Pollak, in: *Hopping and Related Phenomena*, ed. by H. Fritsche and M. Pollak, (World Scientific, New Jersey, 1990).
- [69] J. J. Wainer, A. B. Fowler, and R. A. Webb, Surf. Sci. **196**, 134 (1988).
- [70] T. E. Kopley, P. L. McEuen, and R. G. Wheeler, Phys. Rev. Lett. **61**, 1654 (1988).
- [71] T. K. Ng and P. A. Lee, Phys. Rev. Lett. **61**, 1768 (1988).
- [72] Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959); **123**, 1511 (1961).
- [73] S. Washburn and R. A. Webb, Adv. Phys. **35**, 375 (1986).
- [74] A. G. Aronov and Yu. V. Sharvin, Rev. Mod. Phys. **59**, 755 (1987).
- [75] P. H. M. van Loosdrecht, C. W. J. Beenakker, H. van Houten, J. G. Williamson, B. J. van Wees, J. E. Mooij, C. T. Foxon, and J. J. Harris, Phys. Rev. B **38**, 10162 (1988).
- [76] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
- [77] M. Büttiker, Phys. Rev. B **38**, 9375 (1988).
- [78] V. Fock, Z. Phys. **47**, 446 (1928).
- [79] C. G. Darwin, Proc. Camb. Phil. Soc. **27**, 86 (1930).
- [80] F. Geerinckx, F. M. Peeters, and J. T. Devreese, J. Appl. Phys. **68**, 3435 (1990).
- [81] M. Büttiker, Y. Imry, and R. Landauer, Phys. Lett. A **96**, 365 (1983).
- [82] B. J. van Wees, E. M. M. Willems, C. J. P. M. Harmans, C. W. J. Beenakker, H. van Houten, J. G. Williamson, C. T. Foxon, and J. J. Harris, Phys. Rev. Lett. **62**, 1181 (1989).
- [83] J. K. Jain, Phys. Rev. Lett. **60**, 2074 (1988).
- [84] R. J. Brown, C. G. Smith, M. Pepper, M. J. Kelly, R. Newbury, H. Ahmed, D. G. Hasko, J. E. F. Frost, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. Condens. Matter **1**, 6291 (1989).
- [85] C. G. Smith, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. C **21**, L893 (1989).