

INTRODUCTION TO MESOSCOPIC ELECTRON TRANSPORT

LEO P. KOUWENHOVEN

*Dept. of Applied Physics and DIMES
Delft University of Technology, 2600 GA Delft, The Netherlands*

GERD SCHÖN

*Institut für Theoretische Festkörperphysik
Universität Karlsruhe, 76128 Karlsruhe, Germany*

AND

LYDIA L. SOHN

*Dept. of Physics, Princeton University
Princeton, NJ 08544 - 0708, USA*

In this introductory chapter several basic concepts, relevant for mesoscopic electron transport, will be described. The aim is to provide a basis for several of the following Chapters of this volume. We, therefore, describe in the first Section various aspects of electron quantum transport in two-dimensional electron gases. This includes an estimate of typical material parameters as well as comments on the fabrication. We then describe the quantization of the conductance in point contacts and the edge state picture of the Quantum Hall effect. In the second Section we describe the theory of single-electron tunneling in systems with strong charging and Coulomb-blockade effects. Here we restrict ourselves to the simplest case where low order perturbation theory is sufficient. We first discuss metallic low-capacitance junction systems and then indicate the relevant extensions when dealing with transport through quantum dots with discrete levels. Many further, equally important aspects of mesoscopic transport can not be covered here. Examples are interference and weak localization effects, level statistics or the many body description of solids. Fortunately, some of those will be covered in the more specialized Chapter of this volume.

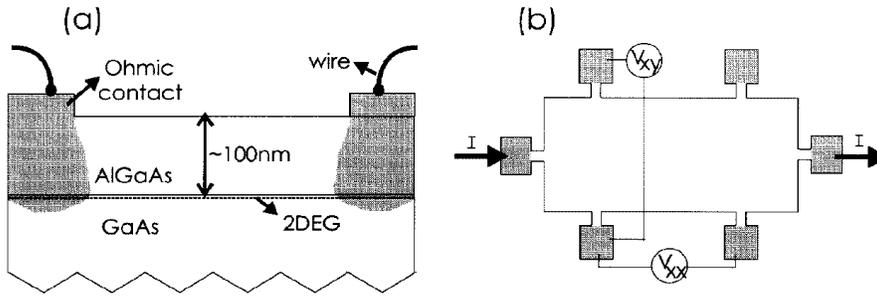


Figure 1. (a) Schematic cross section of a GaAs/AlGaAs heterostructure. The 2DEG is located at the interface between the GaAs substrate and the AlGaAs top layer. (b) Hall bar with six Ohmic contacts (shaded squares).

1. Electron Transport

1.1. 2DEG SYSTEMS AND THE QUANTUM HALL EFFECT

In this Section we introduce basic properties of a two-dimensional electron gas (2DEG) in a GaAs/AlGaAs heterostructure (see Fig. 1a). On the GaAs substrate a layer of typically 100 nm AlGaAs is grown. Somewhere halfway in the AlGaAs layer there is a thin layer where the Ga atoms are replaced by Si donor atoms. With a proper amount of Si one finds that at low temperature the only mobile electrons are located at the GaAs/AlGaAs interface. These free electrons are attracted by the GaAs since they can lower their energy in this smaller band gap material. They are also held as close as possible to their ionized Si^+ donors and thus they form a thin conducting layer near the GaAs/AlGaAs interface (for a review on growth of GaAs heterostructures see Ref. [1]). Since GaAs and AlGaAs can form a nearly perfect interface on the atomic scale and since the Si donors are spatially separated, the electrons experience very little scattering. Typical mean free paths are $10 \mu\text{m}$ and the record is close to $100 \mu\text{m}$. Before we discuss ballistic mesoscopic devices we first review a few important electron transport properties of 2DEG's. In Table 1.1 we have summarized a number of useful relations and some typical values for GaAs/AlGaAs heterostructures.

For transport experiments one first defines a so-called Hall-bar of typically $0.1\text{mm} \times 1\text{mm}$. At the edge of the Hall-bar one then locally evaporates a number of metal squares of roughly $50\mu\text{m} \times 50\mu\text{m}$ in size. By heating the whole sample this metal diffuses into the semiconductor where at some point it makes electrical contact to the 2DEG. Good contacts are characterized by linear current-voltage traces and are therefore called Ohmic contacts. Fig. 1b shows a Hall-bar with 6 Ohmic contacts in a configuration that allows measurements of the Hall resistance R_{xy} and the longitudinal

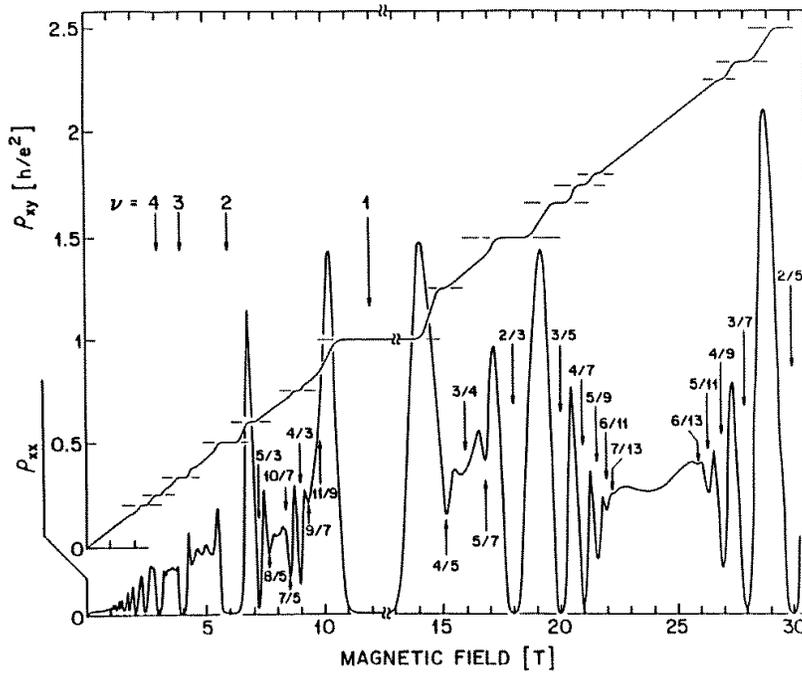


Figure 2. Measurement of the Hall resistance R_{xy} and the longitudinal resistance R_{xx} at 0.1 K. The mobility is $\mu = 1.5 \cdot 10^6 \text{ cm}^2/\text{Vs}$, the temperature is $T = 80\text{mK}$. (Figure provided by R. L. Willett.)

resistance R_{xx} . The sample is current biased and resulting voltages can be measured as a function of magnetic field.

Fig. 2 shows a remarkable measurement of R_{xy} and R_{xx} as a function of magnetic field measured at a temperature of 80mK. The Hall resistance shows plateaus where the longitudinal resistance has minima of virtually zero resistance. The striking aspect of this data is that the plateaus are precisely quantized at integer and fractional multiples of h/e^2 . The integer plateaus are known as the integer quantum Hall effect (QHE), and correspondingly, the fractional plateaus are referred to as the fractional QHE. The precision of the quantization of the former is so accurate that it now forms the international standard of resistance. An important quantity in the quantum Hall regime is the filling factor $\nu = hn_s/eB$. It is equal to the number of electrons divided by the number of flux quanta $\Phi_0 = h/e$. For filling factor $\nu = 1$ the system is in the center of the first plateau at h/e^2 , for $\nu = 2$ in the center of the second plateau at $h/2e^2$, and so forth. The filling factor is, for instance, convenient for determining the electron density. The oscillations in the longitudinal resistance, which have minima

at the same magnetic fields where the Hall resistance shows plateaus, are called Shubnikov-de Haas oscillations. The temperature dependence of the oscillation minima is an accurate determination of the mobility and mean free path of the 2DEG.

Table 1.1. Useful relations and system parameters (including spin degeneracy). We have chosen typical values for μ_e , and n_s , from which the other values are deduced.

– electron mobility	$\mu_e = 10^6 \text{ cm}^2/\text{Vs}$ (typical value)
– scattering time	$\tau = m^* \mu_e / e = 38 \text{ ps}$
– effective mass	$m^* = 0.067 m_0$
– electron density	$n_s = 2.8 \cdot 10^{15} \text{ m}^{-2}$ (typical value)
– Fermi energy	$E_F = \pi \hbar^2 n_s / m^* = 10 \text{ meV}$
– Fermi velocity	$v_F = (2E_F / m^*)^{1/2} = 2.3 \cdot 10^5 \text{ m/s}$
– Fermi wavelength	$\lambda_F = (2\pi / n_s)^{1/2} = 47 \text{ nm}$
– elastic mean free path	$l = v_F \tau = 8.7 \mu\text{m}$
– cyclotron radius at E_F	$r_c = m^* v_F / eB = 88 \text{ nm}$ at $B = 1 \text{ T}$
– angular cyclotron frequency	$\omega_c = eB / m^* = 2.6 \cdot 10^{12} \text{ rad/s}$
– magnetic length	$l_B = (\hbar / eB)^{1/2} = 8.1 \text{ nm}$ at $B = 10 \text{ T}$

1.2. E-BEAM FABRICATION OF A SUBMICRON SEMICONDUCTOR DEVICE

We briefly outline a standard procedure for fabricating small devices in a 2DEG. We start from a 2DEG confined in a GaAs/AlGaAs heterostructure. To laterally confine the electrons one must define a pattern on top of the heterostructure. As an example, we describe e-beam lithography, which is a technique also used in the fabrication of chips. Depending on the desired pattern, one can choose between many variations in the lithography process steps. One procedure is shown schematically in Fig. 3. An organic resist film ($\approx 100 \text{ nm}$ thick) is spun onto the substrate. Exposing the resist with an electron beam results in a molecular-mass difference between the exposed and unexposed parts (see Fig. 3a). An appropriate developer removes only the exposed resist, resulting in the mask pattern shown in Fig. 3b. Evaporated material now sticks only at the substrate where the resist has been removed (see Fig. 3c). The mask itself can be removed by dissolving the remaining resist (lift-off), leaving a small pattern on top of the substrate (see Fig. 3d). The minimum resolution of such a pattern with present day electron-beam lithography facilities is about 20 nm.

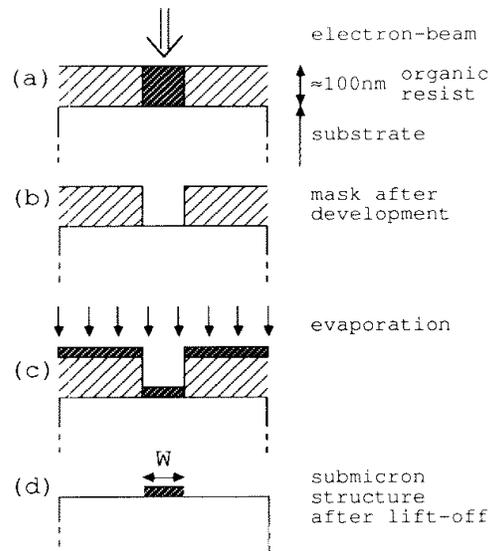


Figure 3. Outline of the electron-beam lithography procedure for fabricating submicron structures.

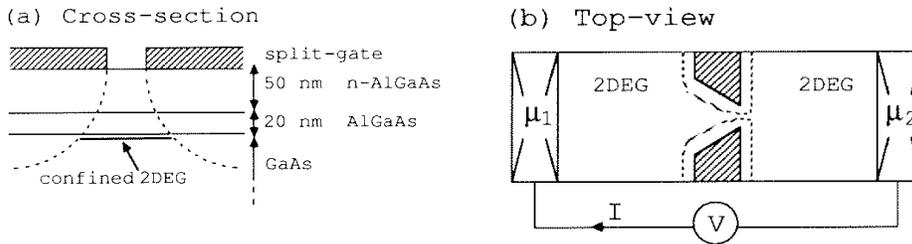


Figure 4. (a) Cross-section of a GaAs/AlGaAs heterostructure with typical layer thicknesses. A negative voltage V_G applied to the metal split-gate confines the electrons laterally in the 2DEG.

(b) Top-view of a QPC. The dotted line indicates the depletion region in the 2DEG, which is tuned by V_G . The two wide 2DEG regions act as reservoirs, emitting electrons through the QPC with energies up to their electrochemical potentials μ_1 and μ_2 . A voltage difference $V = (\mu_1 - \mu_2)/e$ results in a net current I through the QPC.

From this point there are basically two ways to transfer the pattern to the 2DEG. One way is to use the pattern as an etch mask. Etching removes the portion of the 2DEG not protected by the pattern. The boundaries of the etched pattern cause a depletion region, such that the conducting width in the 2DEG is unknown and often much smaller than the defined width [2].

Another, more flexible way to transfer the pattern to the 2DEG is to use it as a gate [3, 4]. Applying a negative voltage to the gate depletes the electron gas beneath it, thereby confining electron motion to the ungated region. For a split-gate geometry, shown in Fig. 4, this results in a narrow conducting channel. The advantage of the split-gate technique is that the conducting width of the point contact in the 2DEG can be tuned from the defined lithographic width of the pattern to zero, by making the gate voltage more negative. Transport between the two wide 2DEG regions in Fig. 4b occurs only through the point contact and can be studied as a function of the width by changing the gate voltage. The width W of the constriction can be made comparable to the Fermi wavelength, so this device is called a quantum point contact (QPC). The actual induced potential in the 2DEG is unknown, but self-consistent calculations [5] indicate that it has a saddle-shape (see Fig. 5a). In the constriction, electrons are confined in the lateral x -direction and slowed down by the presence of a potential barrier in the y -direction. Making the gate voltage more negative, simultaneously reduces the width and increases the barrier height. For zero-width or a barrier which is higher than the Fermi energy E_F of the 2DEG, the QPC is pinched-off and electron transport between the wide 2DEG regions is impossible. Fig. 5b shows a scanning electron micrograph of a double point contact device.

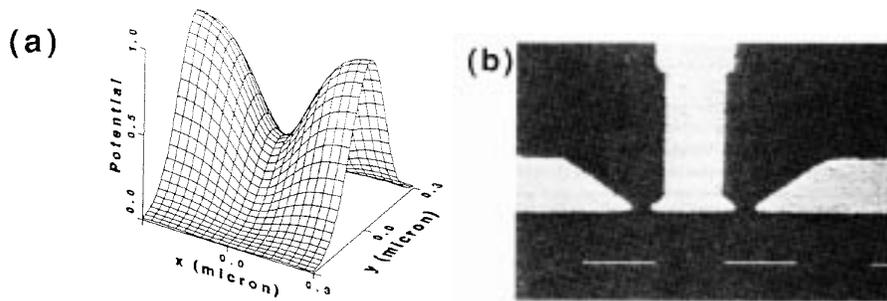


Figure 5. (a) Saddle-shaped potential induced in the 2DEG upon application of a negative gate voltage, resulting in lateral confinement in the x -direction and a potential barrier in the longitudinal y -direction (from Beenakker, and van Houten in Ref. [6]). (b) Scanning electron micrograph of a double-QPC device. The white areas are the Au gates, and the marker is $1 \mu\text{m}$ long. The QPCs are 250 nm wide and are separated by $1.5 \mu\text{m}$.

1.3. QUANTIZED CONDUCTANCE OF A POINT CONTACT

The resistance of a point contact in the classical ballistic regime is known as the Sharvin resistance [7]. The Sharvin resistance is entirely due to elastic backscattering at the geometrical narrowing of the ballistic point contact.

Dissipative processes, which bring the electron system into thermodynamic equilibrium, take place far away from the point contact (i.e. several times the inelastic mean free path). Therefore, the cause of the resistance in this system is spatially separated from its corresponding Joule heating. The classical conductance of a Sharvin point contact in a 2DEG is [8]

$$G_s = \frac{e^2}{\pi} \frac{dN^{2D}}{dE} v_F W . \quad (1)$$

The quantum mechanical 2D density of states per unit area, including a factor 2 for spin degeneracy, is $dN^{2D}/dE = m^*/\pi\hbar$, while $v_F = \hbar k_F/m^*$ is the Fermi velocity. We can rewrite the classical equation (1) to a semi-classical version which includes the conductance quantum $2e^2/h$

$$G_s = \frac{2e^2}{h} \frac{k_F W}{\pi} = \frac{2e^2}{h} \frac{2W}{\lambda_F} . \quad (2)$$

The Fermi wave-vector k_F or wavelength λ_F are related to the 2D electron density n_s by $k_F = 2\pi/\lambda_F = (2\pi n_s)^{1/2}$. The semi-classical form of the Sharvin conductance is continuous and linear in the width W . However, Eq. 2 suggests that we can expect deviations due to the wave nature of electrons whenever λ_F is of order W . We give a derivation of the quantum version of the Sharvin conductance below, but first discuss the experimental results.

The conductance of a point contact is measured by passing a current I through the sample and measuring the voltage V between the current source and drain (see Fig. 4b). Fig. 6 shows the conductance G in units of $2e^2/h = (12906 \Omega)^{-1}$ versus gate voltage V_G measured at $B = 0$. Assuming that the width varies linear with gate voltage, we indeed see that on average the conductance decreases linearly when the constriction is narrowed. However, around this classical dependence we see that G changes in quantized steps of $2e^2/h$. At $V_G = -2.2V$, the conductance becomes zero, corresponding to a pinched-off point contact. Fig. 6 further shows that on increasing the temperature the conductance quantization gradually disappears [9]. The conductance $G(V_G)$ is roughly linear at 4.2 K, in accordance with the classical dependence of Eq. 1. Although the classical result cannot explain the quantization, we note that the plateau values are obtained in Eq. 2 whenever the width W is an integer multiple of $\lambda_F/2$. Eq. 2 predicts that an increase in W of $\lambda_F/2$ (which is 21 nm in this sample) increases G by $2e^2/h$. In total 16 steps were observed between pinch-off and $V_G = -0.3V$ where the constriction is just formed in the 2DEG. The 16 steps give an estimate of the width W of about 340 nm, somewhat larger than the lithographic width of 250 nm, but consistent with the schematic depletion profile shown in Fig. 4(a). These considerations are reminiscent

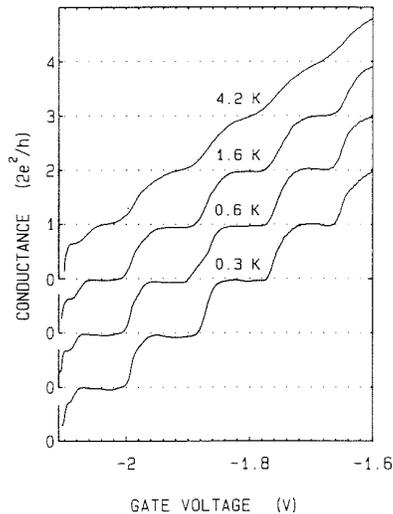


Figure 6. Conductance versus gate voltage at $B = 0$ and different temperatures. Increasing the temperature thermally averages the higher plateaus first (from Ref. [9]).

of the states of a particle-in-a-1D-box, which, as we show below, is the basic idea behind the conductance quantization.

We note that the conductance quantization is not as exact as the quantum Hall effect. First, a series resistance ($\approx 100 \Omega$) originating from the wide 2DEG regions has been subtracted to line up the plateaus at their quantized values [8]. Furthermore, the plateaus are not completely flat. This may be due to scattering at impurities in the vicinity of the QPC or, as we discuss below, the abruptness of the constriction.

We now discuss that the conductance quantization for transport through 1D subbands. If the potential which describes the transition from the wide 2DEG regions to the narrowest point in the QPC varies sufficiently smoothly (i.e. adiabatically), the potential variation in the x - and y -directions may be decoupled [10] (see Fig. 5a). The narrowest point forms the bottleneck of the QPC in the sense that it completely determines the transport properties. In this case we can calculate transport through a QPC from the Hamiltonian

$$H = \frac{p_x^2}{2m^*} + eV(x) + \frac{p_y^2}{2m^*}, \quad (3)$$

where $V(x)$ describe the transport at the bottleneck. For the confinement in the lateral x -direction, we follow Berggren et al. [11] and choose a parabolic confining potential $V(x) = 1/2m^*w_0^2x^2$. Self-consistent calculations of Laux et al. [5] have shown that small split-gate samples have a confinement close to such a parabola. The advantage of using a parabolic potential is that the

resulting Schrödinger equation can be written in the form of a harmonic oscillator having energy eigenvalues

$$E_n = \left(n - \frac{1}{2}\right)\hbar\omega_0 + \frac{\hbar^2 k_y^2}{2m^*} \quad (n = 1, 2, \dots) \quad (4)$$

which contains a free-electron kinetic energy dispersion in the longitudinal y -direction. In the lateral x -direction the energy states, indexed by $n = 1, 2, \dots$, are quantized and separated in energy by $\hbar\omega_0$.

Because the electron motion is free in one direction only, Eq. 4 describes 1D subbands. Fig. 7 shows the 1D subband dispersion versus longitudinal wave-vector k_y . The right-going electrons, with a velocity $\hbar v_n = dE_n/dk_y$, originate from the left 2DEG reservoir, which at zero temperature populates all the states up to its electrochemical potential μ_1 . Similarly, the left-going electron states are occupied up to μ_2 , the electrochemical potential of the right 2DEG reservoir (see Fig. 4b).

A voltage difference $V = (\mu_1 - \mu_2)/e$ between the two reservoirs results in a net current I , which is carried by the (uncompensated) electron states in the energy interval between μ_1 and μ_2 . Note that we define the Fermi energy as $E_F = \mu_1 = \mu_2$ when $V = 0$. The net current I at zero temperature is

$$I = e \sum_{n=1}^N \int_{\mu_2}^{\mu_1} dE \frac{1}{2} \frac{dN_n}{dE} v_n(E) T_n(E), \quad (5)$$

which includes the transmission probability of the n -th subband $T_n(E)$ to describe possible scattering events. Here N denotes the number of occupied subbands, i.e. the largest number for which $E_N(k_y = 0) < E_F$. The 1D spin-degenerate density of states is $dN_n/dE = 2/\pi(dE_n/dk_y)^{-1}$. The important aspect of 1D transport is the cancelation of the energy dependence in the product of velocity and density of states $(dN_n/dE)v_n = 4/h$. For small voltages ($eV \ll E_F$), one can take $T_n(E) = T_n(E_F)$. Substituting this in Eq. 5, one finds that the conductance $G = I/V = eI/(\mu_1 - \mu_2)$ is independent of energy

$$G = \frac{2e^2}{h} \sum_{n=1}^N T_n(E_F). \quad (6)$$

Eq. 6 is known as the 2-terminal Landauer formula [12]. If no backscattering takes place, so that $\sum_{n=1}^N T_n(E_F) = N$, Eq. 6 reduces to

$$G = \frac{2e^2}{h} N, \quad (7)$$

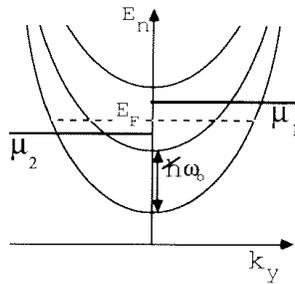


Figure 7. Energy E_n versus longitudinal wave-vector k_y from Eq. 4 at the bottleneck of a QPC assuming a parabolic confinement potential. The 1D subbands are separated by $\hbar\omega_0$. A net current results from the uncompensated occupied electron states in the interval between μ_1 and μ_2 , the electrochemical potentials of the two wide 2DEG reservoirs.

demonstrating that each occupied subband contributes $2e^2/h$ to the conductance. The subbands are called 1D current channels to emphasize that each channel carries the same amount of current.

In the experiment, a decreasing V_G increases the barrier in the QPC, and simultaneously increases the lateral confinement and consequently the energy splitting. Both effects increase the subband energies. As long as E_F is between two subband bottoms, N is constant and G is quantized. If a subband bottom moves through E_F , N changes by 1 and G by $2e^2/h$.

Several numerical calculations [13] have shown that Eq. 7 gives an accurate description of a QPC with the assumptions that impurity scattering is absent and that the potential variations are smooth. Sharp potential variations, possibly present at the entrance and exit of the QPC [14] or originating from impurities [15, 16] can give rise to backscattering and therefore destroy the quantization. The assumption of an adiabatic constriction in Eq. 3 is therefore no longer valid. At $T > 0$ the reservoirs inject electrons with a Fermi-Dirac distribution, which averages the conductance $G(T) = \int dE (-df/dE) G(E)$. Comparing this with the temperature dependence of the curves in Fig. 6, it is found that the subband separation gradually increases from about 1 meV at $V_G = -1V$ to 3 meV at $V_G = -2.1V$. This illustrates that thermal averaging has a stronger effect on the higher plateaus, as observed in Fig. 6.

1.4. DEPOPULATION OF 1D MAGNETO-ELECTRIC SUBBANDS

We now turn to the case of a QPC with an applied magnetic field in the z -direction. In the Hamiltonian of Eq. 3, the magnetic field B is included by substituting $(\mathbf{p} - e\mathbf{A})$ for the momentum \mathbf{p} . In the Landau gauge for the vector potential $\mathbf{A} = A_y = Bx$, the Schrödinger equation is once again

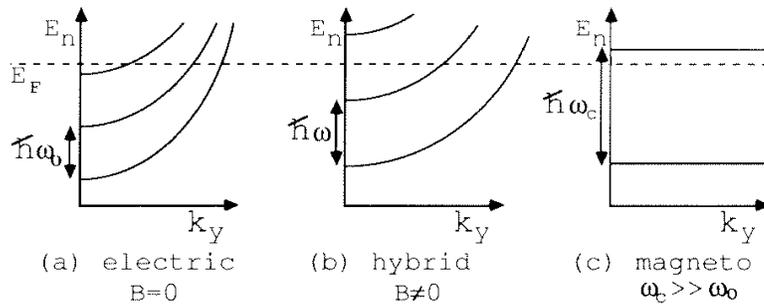


Figure 8. 1D subband dispersion for three values of the magnetic field, illustrating magnetic depopulation. The energy splitting is electric ($\hbar\omega_0$) in (a), and hybrid ($\hbar\omega$, with $\omega^2 = \omega_0^2 + \omega_c^2$) in (b). For large magnetic fields (c) the 1D subbands are Landau levels with an energy splitting of $\hbar\omega_c$.

that of a harmonic oscillator, but now with energy eigenvalues [11]

$$E_n = \left(n - \frac{1}{2}\right)\hbar\omega + \frac{\hbar^2 k_y^2}{2m_B}. \quad (8)$$

These eigenvalues describe hybrid magneto-electric subbands. With $\omega^2 = \omega_0^2 + \omega_c^2$, where the cyclotron frequency is $\omega_c = eB/m^*$, the energy separation is now a combination of the electrical confinement and the magnetic field. $m_B = m^*\omega^2/\omega_0^2$ is a magnetic field dependent effective mass yielding a smaller dispersion for larger magnetic fields. The influence of the magnetic field on the subband dispersion is shown schematically in Fig. 8. For $B = 0$ the subbands are determined by the electrical confinement only. A small magnetic field increases the subband splitting and reduces the dispersion. For large magnetic fields, the subbands have the magnetic energy separation and a vanishing dispersion. In this case, the subbands are the well-known Landau levels. It can be seen from Fig. 8 that on increasing the magnetic field, the number of occupied subbands decreases. This process is known as the depopulation of magneto-electric subbands.

One can show that the velocity and density of states also cancel in a magnetic field and that Eqs. (6) and (7) are still valid [17, 18]. From the above analysis it follows that a gradual transition exists between the quantized conductance $G = N2e^2/h$ at zero magnetic field (with N the number of occupied electric subbands in Fig. 8a) to the quantum Hall conductance $G_H = N_L 2e^2/h$ at a high magnetic field (with N_L the number of occupied Landau levels in Fig. 8c). We note that our particular choice of a parabolic confinement does not affect the general conclusions, such as the cancelation of velocity with the density of states, and the conductance quantization at zero and non-zero magnetic field.

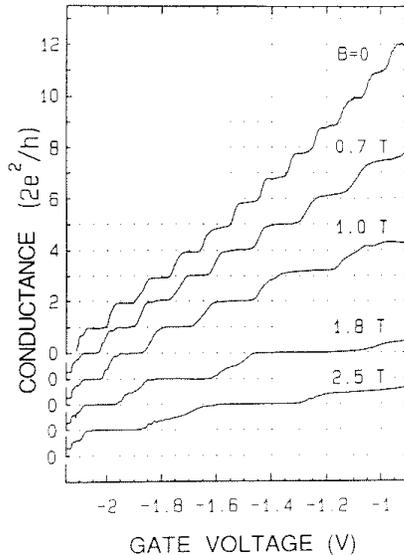


Figure 9. QPC conductance versus gate voltage at 0.6 K for several values of magnetic field. The increasing width of the plateaus demonstrate the increasing energy splitting in a magnetic field. The curves have been offset for clarity (from Ref. [18]).

Fig. 9 shows the conductance of a QPC versus gate voltage for several values of the magnetic field [18]. As can be seen, the quantization is preserved in a magnetic field. Above $B = 1$ T, spin-resolved plateaus develop at odd multiples of e^2/h . The depopulation can be seen from the fact that at a fixed gate voltage, the number of plateaus (or, equivalently, the number of occupied subbands), decreases with increasing magnetic field. From the measurements of Fig. 9, one can deduce subband splittings of about 1 meV at $V_G = -1$ V and 3 meV at $V_G = -2$ V, in agreement with the values obtained from the temperature dependence [9]. A third independent way to determine the subband splittings is by measuring the non-linear current-voltage characteristics of a QPC [19].

1.5. ELECTRON TRAJECTORIES IN A LOW MAGNETIC FIELD

In the previous Section, we discussed the influence of a magnetic field on the subband dispersion in momentum space. We now consider the electron motion in real space, which yields a simple physical picture of the QHE and associated effects. To elucidate the quantized electron motion in a high magnetic field, we first discuss classical electron trajectories in a low magnetic field.

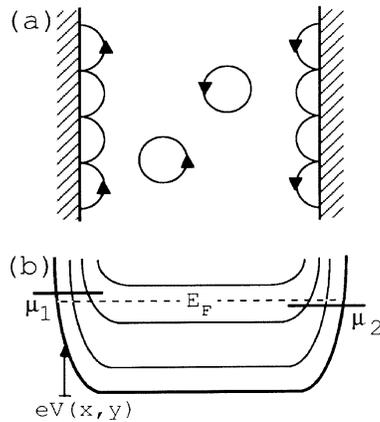


Figure 10. (a) Schematic classical electron trajectories in a magnetic field. (b) Corresponding quantum picture of the energy states of Eq. (10) along a cross-section of the 2DEG, illustrating the formation of edge channels at the boundary of the 2DEG (formed by the electrostatic potential energy $eV(x, y)$) where the Landau levels intersect the Fermi energy E_F .

In the absence of an electric field E , the balance of the Lorentz force $F_L = evB$ and the centripetal force $F = m^*v^2/r$ leads to a cyclotron motion of the electrons, with an angular frequency $\omega_c = eB/m^*$ and at the Fermi energy a cyclotron radius $r_c = m^*v_F/eB$ (see Fig. 10a). When the electric field $E = -\nabla V(x, y) \neq 0$, the electrons have a net drift velocity $v_D = E/B$. At the boundary of the sample, where E is large, the collisions at the edge result in skipping orbits. The electrons skip with the drift velocity along the edge of the sample. The direction of the velocity is opposite for the two opposite edges (see Fig. 10a).

The skipping orbit motion of electrons along a 2DEG boundary in a small magnetic field has been observed in an electron focusing experiment by van Houten et al. [20]. The geometry shown in Fig. 5b with two adjacent QPCs with a separation of $L = 3\mu\text{m}$ was used, where one QPC injects electrons into the 2DEG and the second QPC collects them. The injected electrons are focused by the magnetic field on the boundary between injector and collector at distances $2pr_c$ ($p = 1, 2, \dots$). Focusing into the collector occurs when $2pr_c = L$. In the experiment the collector voltage is measured as a function of magnetic field B . From the condition $2pr_c = L$ it follows that the largest number of electrons reaches the collector when $B_{\text{foc}} = 2pm^*v_F/eL$ which leads to periodic oscillations in the collector signal. This is shown in Fig. 11 together with a calculation. On average the voltage increases linearly with magnetic field, which is expected for the classical Hall resistance $V_{\text{Hall}}/I = B/en_s$. For curve (a), however, we see

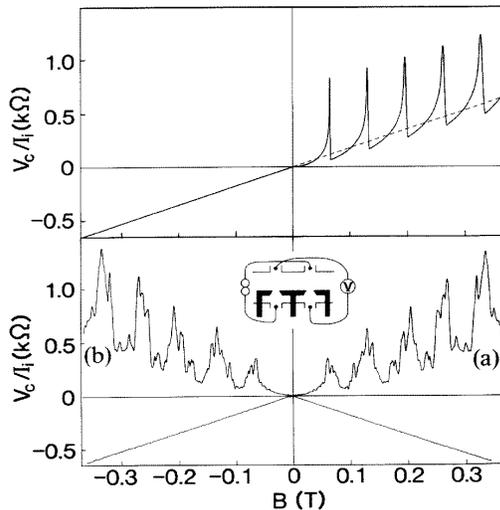


Figure 11. Bottom: Electron focusing ($T = 50$ mK, $L = 3.0\mu\text{m}$) in the configuration depicted in the inset. The two traces *a* and *b* are measured with interchanged current and voltage leads, and demonstrate the injector-collector reciprocity as well as the reproducibility of the fine structure. Top: Calculated classical focusing spectrum corresponding to the experimental trace *a* (50 nm wide point contacts were assumed). The dashed line is the extrapolation of the classical Hall resistance seen in reverse fields. (From Ref. [20].)

additional large oscillations with fine structure around the average Hall resistance for positive magnetic field. These oscillations occur at the expected focusing fields. For a negative magnetic field no focusing signal is observed since now the injected electrons are deflected away from the collector.

Curve (b) is taken with the current and voltage probes interchanged. The relation between (a) and (b) demonstrates Büttiker's reciprocity relation [21],

$$R_{ij,kl}(B) = R_{kl,ij}(-B). \quad (9)$$

It implies that upon interchange of the current probes i and j with the voltage probes k and l , one obtains the same resistance at the opposite magnetic field. This fundamental symmetry relation is clearly confirmed in Fig. 11, including the symmetry in the fine-structure. The latter has been explained by extending the classical focusing calculation of the top section to include quantum interference effects between trajectories being injected at different angles [20]. This focusing experiment demonstrates that the collisions at the boundary between injector and collector are highly specular, since diffusive boundary scattering would average the oscillations. In a subsequent experiment by Spector et al. [22] focusing signals were observed up to a distance between injector and detector as large as $64\mu\text{m}$.

1.6. THE QUANTUM HALL EFFECT ON MESOSCOPIC LENGTH SCALES

1.6.1. *Edge Channels*

In a high magnetic field the electron motion is quantized. The flux Φ enclosed by an electron in a cyclotron orbit equals an integer times the flux quantum $\Phi_0 = h/e$, and the quantized electron energies are:

$$E_n = \left(n - \frac{1}{2}\right)\hbar\omega_c + eV(x, y) \quad (10)$$

where $n = 1, 2, \dots$ is the spin-degenerate Landau level index, and we have ignored the Zeeman energy splitting $\pm g\mu_B B/2$. We assume that the electrostatic potential $V(x, y)$ is flat in the interior of the sample and rises at the boundary. Electrostatic variations due to impurities are ignored, because we are dealing with ballistic samples. Fig. 10b shows schematically the Landau energy levels of Eq. 10. The electron states at the left boundary are occupied up to μ_1 , the electrochemical potential of the current source, and at the right boundary up to μ_2 , the electrochemical potential of the current sink (see Fig. 4b). At the two sample boundaries, the electron states have opposite velocity directions, similar to the classical case of Fig. 10a.

The relevant electron states for linear transport are only those at the Fermi energy. As can be seen, these are located at the sample boundaries, where the Landau levels intersect the Fermi energy ($E_n = E_F$), and they extend in the direction perpendicular to the cross-section of Fig. 10b. The intersections are the current-carrying states, which are known as edge channels [17, 23, 24]. The net current I only results from the uncompensated states in the interval between μ_1 and μ_2 . The total current carried by the states below μ_2 is zero. The transport through edge channels is 1-dimensional [17]. Edge channels can therefore also be viewed as 1D current channels, each carrying a current $I_n = 2e/h(\mu_1 - \mu_2)$. With the Hall voltage $V_H = (\mu_1 - \mu_2)/e$ measured between the two sample boundaries, this directly gives the quantized Hall conductance $G_H = N_L I_n / V_H = N_L 2e^2/h$.

Up to now we have ignored all scattering processes. Büttiker [25] has pointed out that due to the spatial separation of the electron states with opposite velocity, backscattering requires scattering from one sample boundary to the other. Backscattering is therefore suppressed when the edge states between μ_1 and μ_2 are not connected by extended electron states. This is the case in Fig. 10b, where the Fermi energy is between two bulk Landau levels.

1.6.2. *Selective Probing of Edge Channels*

The above description of transport in the quantum Hall effect (QHE) regime, known as the Landauer-Büttiker formalism, is reviewed in Ref. [26]. This edge channel description gives an appealing physical picture of the

QHE. The question now arises about how to prove the existence of edge channels and whether they can really be viewed as independent current channels. We now discuss an experiment involving two adjacent QPCs which directly probes the transport through a particular edge channel. The theory for this experiment provides a simple illustration of Büttiker's multi-probe formalism which has been used very successfully to describe different kind of mesoscopic transport phenomena.

In Section 1.5 we discussed the electron focusing from one QPC to an adjacent second QPC by small magnetic fields ($B \leq 1\text{T}$). These fields are too small to quantize the electron motion, and the focusing can be explained in terms of classical cyclotron motion. We consider now the same geometry in the high field regime ($B \geq 1\text{T}$). Fig. 12a shows schematically two adjacent QPCs A and B defined in a 2DEG with an applied magnetic field such that two edge channels are occupied. The 2-terminal conductances G_A and G_B of the individual QPCs measure the number of transmitted channels and are quantized in multiples of $2e^2/h$ (see also Fig. 9). The Hall conductance G_H is normally thought to be independent of the characteristics of the current and voltage probes and to correspond directly to the number of occupied (spin-degenerate) Landau levels N_L in the 2DEG, $G_H = N_L 2e^2/h$. This is not true for the situation in Fig. 12a and we discuss here that for describing transport through mesoscopic conductors one has to include the properties of the measurement contacts.

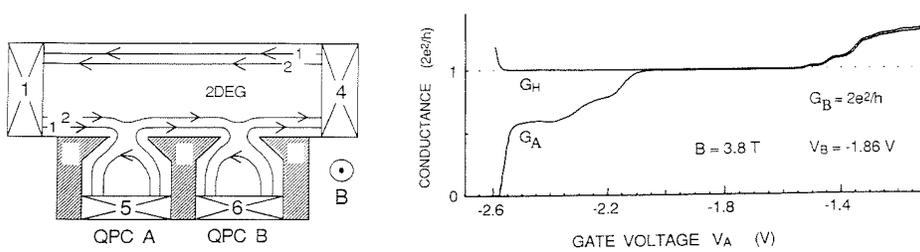


Figure 12. Left: Geometry used to observe an anomalous quantization of the Hall conductance, for which QPC A is used as current probe and QPC B as voltage probe. Right: Comparison between the Hall conductance G_H and the 2-terminal conductance of the current probe G_A . The voltage probe conductance is kept fixed at $G_B = 2e^2/h$. Although the number of occupied Landau levels in the 2DEG is unchanged, G_H follows the largest probe conductance (from Ref. [27])

We first assume that the edge channels shown in Fig. 12a are independent, i.e. no scattering events occur between different edge channels or, equivalently, electrons travel with conservation of quantum-subband-number. In this case the transport is adiabatic. The edge channels are oc-

cupied up to the electrochemical potential of the last Ohmic contact they have left. (We assume ideal Ohmic contacts, meaning that all incoming electrons are absorbed and all outgoing states are occupied up to an average electrochemical potential.) Consequently, in the region between the two QPCs, the two edge channels have an unequal population. Channel-2 is occupied up to μ_1 , while current probe 5 populates channel-1 up to μ_5 . This process is called selective population of edge channels. If the voltage probe B detects all edge channels like an ideal Ohmic contact, the regular quantum Hall conductance is measured. However, this is not the case in Fig. 12a where the voltage probe selectively detects only the first channel. The second channel is neither populated by the current probe nor detected by the voltage probe and, therefore, is not measured. The Hall conductance $G_{54,61}$ equals $2e^2/h$ instead of $4e^2/h$ which a regular Hall measurement would give. The current and voltage contacts do not measure all the 2DEG properties, but only those properties they "see", or couple to. In general one can show that in the absence of inter-edge channel scattering this Hall conductance is given by [27]

$$G_H = \max\{G_A, G_B\} \quad (11)$$

implying that G_H is completely determined by the characteristics of the probes and is independent of the number of occupied Landau levels in the 2DEG.

Fig. 12b compares the measured probe conductances G_A and G_B with the Hall conductance G_H . The magnetic field is kept fixed at 3.8 T, corresponding to $N_L = 2$ in the bulk 2DEG. The voltage on QPC B , defining the voltage probe, is also fixed such that only the first edge channel is transmitted, and therefore $G_B = 2e^2/h$. The voltage V_{gA} on QPC A , defining the current probe, is varied, resulting in a decreasing G_A . Again we note that a normal Hall measurement would give a constant conductance $G_H = 4e^2/h$. However, Fig. 12b shows a Hall conductance which virtually follows the largest probe conductance in agreement with Eq. 11 (for $V_{gA} > -1.5V$, $G_H = G_A > 2e^2/h$; for $-2.1V < V_{gA} < -1.5V$, $G_H = G_A = G_B = 2e^2/h$ corresponding to the edge channel flow of Fig. 12a; and for $V_{gA} < -2.1V$, $G_H = G_B = 2e^2/h$). The experiment demonstrates that on short distances of order μm the transport through edge channels is adiabatic, implying that they can be viewed as independent 1D current channels. Subsequent experiments have shown that a non-equilibrium population can persist up to larger distances (several tens of μm 's) [28]. In particular, it is found that the top most channel is virtually decoupled from the lower edge channels even over macroscopic distances of several times $100 \mu\text{m}$ [28, 29, 30, 31, 32]. These experiments have clarified the important role of measurement probes in the QHE regime.

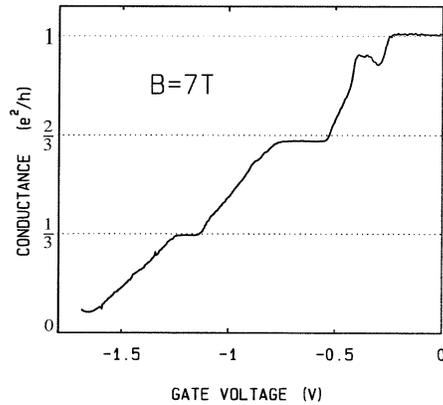


Figure 13. Two-terminal conductance of a quantum point contact at $B = 7\text{T}$. At zero gate voltage the conductance is $G = e^2/h$ implying that the 2DEG is spin-polarized at filling factor 1. On decreasing the gate voltage fractional plateaus are observed (from Ref. [37]).

1.6.3. *The Importance of Electron-Electron Interactions for the QHE on Mesoscopic Length Scales*

In our discussion we neglected Coulomb interactions between the electrons. For many experimental observations it is sufficient to give a description in terms of a non-interacting electron model. There are several exceptions, however. First of all, in very high mobility samples not only are plateaus observed at integer filling factors, but also at fractional filling factors (see Fig. 2). The theory for this fractional QHE needs to incorporate electron-electron interactions [33]. Recently, a new appealing formulation for the fractional QHE has been given in terms of composite particles [34]. Another recent development is the prediction [35] and observation [36] of so-called skyrmions.

Also, on mesoscopic length scales it has recently become appreciated that electron-electron interactions can be important under certain circumstances. First of all, the FQHE can also be observed at short ballistic length scales [37]. Fig. 13 shows the conductance of a QPC at 7 T. At zero gate voltage the conductance is equal to e^2/h showing that the filling factor of the bulk 2DEG is equal to one (i.e. spin resolved). On making the gate voltage more negative the conductance decreases. However, at fractional values $2/3$ and $1/3$ times e^2/h it shows quantized plateaus. The QPC used in this experiment shows quantized conductance at $B = 0$ implying that fractional plateaus can exist in a ballistic sub-micron device [37, 38]. Also, selective population and detection experiments of fractional edge channels have been performed [39, 40] in a similar way as in Section 1.6.2 for the case of integer edge channels. These experiments on the fractional QHE in

the ballistic regime can not be explained with the non-interacting electron models of Sections 1.6.1 and 1.6.2. Instead interacting models have been proposed by Beenakker [41], MacDonald [42] and Chang [43]. These models are the basis of recent models on self-consistent calculations of edge channels [44]. From these electrostatic calculations it follows that edge channels have a finite width. Regions where the Fermi energy is pinned in a Landau level are called a compressible liquid whereas regions where the Fermi energy is between Landau levels are called an incompressible liquid. Although, Coulomb interactions give edge channels a finite width the selective population experiments suggest that they remain 1D channels. In 1D Coulomb interactions are known to give rise to a new type of electron liquid. This so-called Luttinger liquid is discussed by Fisher and Glazman in this volume including predictions for correlated electron transport in fractional edge channels. Altogether, we can conclude this Section with the statement that the QHE is still a very interesting system for studying new electronic properties.

2. Single-electron tunneling

2.1. REVIEW

The electron transport in two-dimensional electron gases displays the quantum coherence of the electronic states, as described in the previous Section. Another important aspect of mesoscopic systems is the role of interactions. We mentioned already their importance in connection with the fractional QHE. Interaction effects also influence strongly electron tunneling through mesoscopic systems, and they lead to strong correlation effects. In this Section we will discuss situations where it is sufficient to approximate the Coulomb interaction by an effective capacitance model. This applies for metal junction systems where after a relaxation of the space distribution of the electrons the remaining Coulomb interaction is well described by the geometric capacitance of the junctions. The capacitance model also works remarkably well for the transport through small quantum dots. (See the Chapter of Kouwenhoven *et al.* for many facts on quantum dots.) The “charging effects” allow us to control single electron charges, which leads to a variety of single-electron effects, e. g. to the suppression of tunneling, a phenomenon known as the Coulomb blockade.

Single-electron effects have been studied for more than a decade, and a large number of papers, incl. several reviews have been devoted to the topic. Kulik and Shekhter [45] and Averin and Likharev [46] developed the perturbation theory of single-electron tunneling in metal junctions and discussed several consequences. Initial scepticism against the new concepts

was overcome when experiments proved to be successful. After early experiments by Fulton and Dolan [47], important breakthroughs were achieved in Delft by Mooij and Geerligs and in Saclay by Devoret, Estève, Urbina and further members of these groups [48, 49]. Their work is well summarized in the proceedings of an earlier NATO ASI *Single Charge Tunneling* [50].

In this Section we introduce the concepts and basic description of single-electron tunneling in systems with strong charging effects. For definiteness we first consider metallic electrodes with a large density of electron states. We study how the charging energy depends on the number of electrons and the transport and gate voltages applied to various parts of the system. The simplest model systems demonstrating these features are the “single-electron box” and the “single-electron transistor”. We then derive within perturbation theory the single-electron tunneling rates. In low capacitance systems it is crucial to account for the change in the charging energy associated with the tunneling process. A master-equation description accounts for the large-scale features of the current-voltage characteristic of the single-electron transistor. In the Coulomb-blockade regime, where single-electron tunneling is suppressed, higher-order processes such as coherent “cotunneling” of electrons through several junctions become observable. Finally we discuss single-electron tunneling through quantum dots, where the level quantization becomes observable as well.

Several extensions have been described in the literature. We mention only a few:

- The mesoscopic junction systems studied here are small such that charging effects and higher-order quantum processes play a role. On the other hand, they are large enough such that macroscopic current and voltage probes and sources can be coupled to the system. This makes the mesoscopic system susceptible to the influence of the electric circuit. A complete description has to include this circuit. The influence of the electrodynamic environment on single-electron tunneling has been reviewed in the article by Ingold and Nazarov in Ref. [50].
- In this introduction we study tunnel junctions with two normal-conducting electrodes (NN). If the system or part of it is superconducting the combination of single-electron tunneling, Cooper pair tunneling and Andreev reflection leads to further highly interesting effects [51]. For a review see the Chapter of Fazio and Schön in this volume.
- Many of the single-electron effects can be described within simple perturbation theory. A necessary requirement is that the resistance of the tunnel barriers is high compared to the quantum resistance $R_K = h/e^2 \approx 25.8 \text{ k}\Omega$. For tunnel junctions with lower resistance a more general formulation is needed [52, 53, 54]. A systematic description of tunneling in systems with strong charging effects is presented later in the Chapter of Schoeller.



Figure 14. a) An overlap junction with an oxide layer (hatched region), b) schematic diagram for a tunnel junction.

2.2. CHARGING ENERGY AND SINGLE-ELECTRON DEVICES

2.2.1. The Energy Scale

Modern lithographic techniques allow the fabrication of narrow metallic lines with widths down to approximately 20 nm, as well as tunnel junctions in overlap regions of such lines, as illustrated in Fig. 14. The structures are grown by evaporation of the metal, e. g. Aluminum, through masks onto the substrate. Tunnel junctions can be produced by shadow evaporation techniques, which involves two stages of evaporation at different angles. Between the two stages the first layer is allowed to oxidize. The junction is formed in the overlap region. They can be produced reliably with areas of $S = (100 \text{ nm})^2$ and below. The oxide layer is roughly $d = 10 \text{ \AA}$ thick, and the dielectric constant of the oxide is $\epsilon \approx 10$. Using the classical expression for the capacitance we arrive at $C = \epsilon S / (4\pi d) \approx 10^{-15} \text{ F}$.

The capacitance introduces an energy scale, the charging energy corresponding to a single-electron charge ($-e$),

$$E_C \equiv \frac{e^2}{2C}, \quad (12)$$

which characterizes all charging effects. It is of the order of $E_C \approx 10^{-4} eV$ if the capacitance is $C = 10^{-15} \text{ F}$, which corresponds to a temperature $E_C/k_B \approx 1 \text{ K}$. In a tunneling process the electrostatic energy changes by an amount of the order of magnitude of E_C . Hence we expect in the sub-Kelvin regime electron transport to be affected by charging effects. Similar properties have been observed in semiconductor nanostructures, for instance in quantum dots in 2-dimensional electron gases. The Coulomb energy in these systems again can be characterized by a capacitance which depends on the size of the dot and also may lie in the range of 10^{-15} F or less. Charging effects play a role in granular materials and ultimately even in molecular systems. Here the capacitance may be as low as 10^{-18} F , making single-electron tunneling observable even at room temperatures.

2.2.2. Single-Electron Box

We now analyze in more detail the charging energy of simple systems of tunnel junctions. It depends on the electron number in various parts of the particular system and the applied voltages. The first example is the

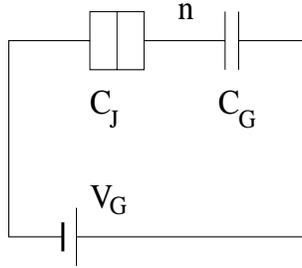


Figure 15. The single-electron box.

single-electron box, shown schematically in Fig. 15. It consists of a small metallic island, coupled via a tunnel junction with capacitance C_J to an electrode and via a capacitor C_G to a gate voltage source V_G . We choose the reference such that for $V_G = 0$ the lowest energy state of the system is charge-neutral, i. e. the electrons on the island compensate the charge of the ions; consequently there are $n = 0$ excess electrons on the island. If a gate voltage is applied the number of excess electrons on the island can change due to tunneling across the junction in *discrete* steps to $n = \pm 1, \pm 2, \dots$

While the total number of electrons on an island is integer, the charge is spatially distributed and in general shifted relative to the positive background. If a voltage is applied the surface charges on the capacitor plates, which are of equal magnitude but opposite sign on the two sides of each junction, are in general non-quantized. They are determined by the integer n and the non-quantized applied voltage. We obtain the charging energy from the following elementary arguments: the total excess charge of the box splits into two parts on the left and right capacitor plate $-ne = Q_L + Q_R$. The corresponding voltage drops add to the applied voltage $V_G = Q_L/C_J - Q_R/C_G$, and the charging energy is $Q_L^2/2C_J + Q_R^2/2C_G$. The relevant free energy is a Legendre transform of this energy, which also contains the work done by the voltage source $-V_G Q_R$. Elimination of Q_L and Q_R in favor of n and V_G yields, up to a contribution which does not depend on the variable n , the result

$$E_{\text{ch}}(n, Q_G) = \frac{(ne - Q_G)^2}{2C}. \quad (13)$$

Here $C = C_J + C_G$ is the total capacitance of the island. The effect of the voltage source is contained in the “gate charge” defined as $Q_G = C_G V_G$.

The charging energy E_{ch} is plotted in Fig. 16 as function of the gate charge for different excess electron numbers n . With increasing gate voltage, the electron number corresponding to the lowest energy state increases. It does so in discrete steps from n to $n + 1$ at the degeneracy points $Q_G/e =$

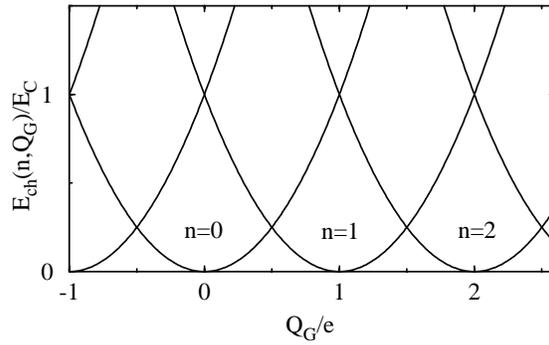


Figure 16. The charging energy of a single–electron box as a function of the gate voltage for different numbers n of electron charges on the island.

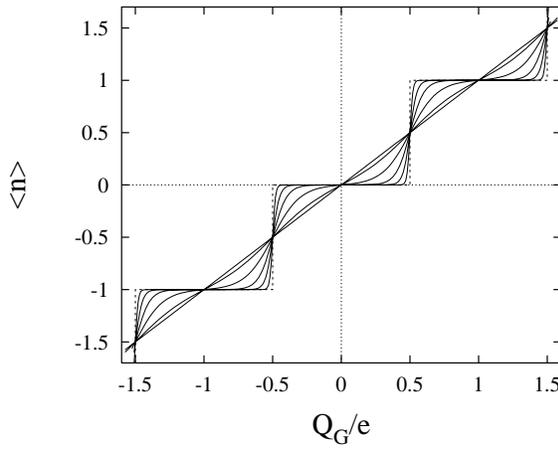


Figure 17. The average number of electron charges $\langle n \rangle$ on the island of a single–electron box as a function of the gate charge (voltage) for different temperatures $T/E_C = 0$ (dashed steps), 0.02, 0.05, 0.1, 0.2, 0.4, and 1 (nearly linear).

$n + 1/2$. Under the same conditions, the voltage of the island, $V_{\text{island}} = \partial E_{\text{ch}}/\partial Q_G$, displays a sawtooth dependence on the applied voltage.

At finite temperatures the steps and sawtooth dependence are washed-out, as follows from the classical statistical average

$$\langle n(Q_G) \rangle = \frac{1}{Z_{\text{ch}}} \sum_{n=-\infty}^{\infty} n e^{-E_{\text{ch}}(n, Q_G)/k_B T}, \quad (14)$$

where Z_{ch} is an obvious normalization. The result is displayed in Fig. 17 for different temperatures. The stepwise increase has been observed experimentally, e.g. by the Saclay group (see results in Ref. [50]). Their measurement

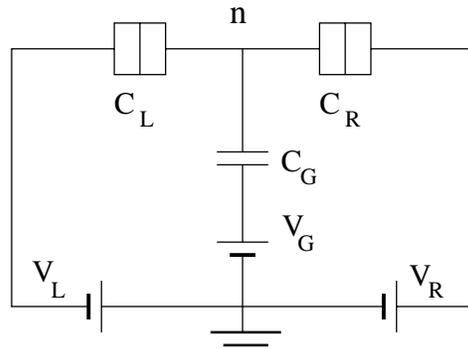


Figure 18. The SET transistor.

procedure will be discussed below. The experiments agree well with theoretical expectations. However, it is crucial to control heating and the noise from the measurement setup, which usually is at a temperature higher than that of the cryostat.

2.2.3. Single-Electron Transistor

Another fundamental example is provided by the single-electron transistor shown in Fig. 18. Here an island is coupled via two tunnel junctions to a transport voltage source $V = V_L - V_R$ such that a current can flow. The island is, furthermore, coupled capacitively to a gate voltage V_G . The charging energy of the system depends again on the integer number of electrons n on the island and the continuous voltages. Some algebra along the lines outlined for the electron box produces again $E_{\text{ch}}(n, Q_G) = (ne - Q_G)^2/2C$. For the transistor $C = C_L + C_R + C_G$ is the total capacitance of the island, i.e. the sum of the two junction capacitances and the gate capacitance, and all three voltage sources define the gate charge

$$Q_G = C_G V_G + C_L V_L + C_R V_R . \quad (15)$$

In a tunneling process, increasing the island charge from n to $n + 1$, the charging energy changes by

$$E_{\text{ch}}(n + 1, Q_G) - E_{\text{ch}}(n, Q_G) = \left(n + \frac{1}{2} - \frac{Q_G}{e} \right) \frac{e^2}{C} . \quad (16)$$

The second energy differences are equally spaced and can be tuned by the gate voltage. The situation is illustrated in the energy scheme shown in Fig. 19. The differences in charging energy are plotted in the center. We further display the Fermi levels of the two leads which are shifted by the applied potentials $V_{L/R}$.

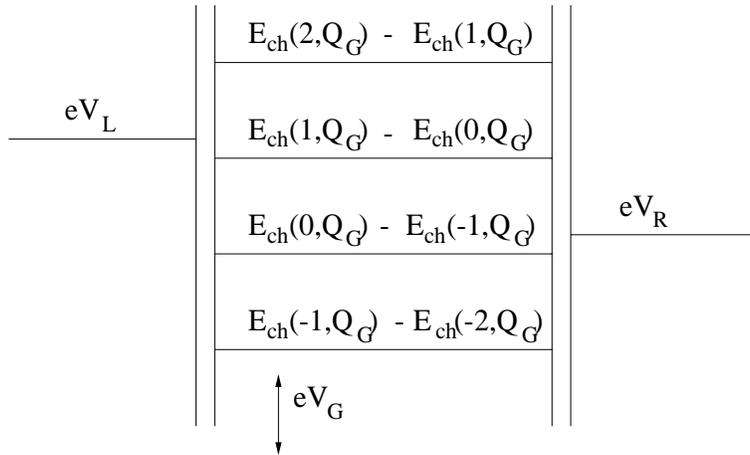


Figure 19. In the island the energy differences corresponding to the addition or removal of an electron charge are shown. They can be shifted by the gate voltage V_G . The electrochemical potentials of the leads are shifted relative to each other by the transport voltage $V = V_L - V_R$.

For definiteness we assume that the energy of the electrons in the left lead is higher than that in the right lead. Then at low temperature tunneling from the left lead to the island (transition from n to $n + 1$) is possible if the electrochemical potential in the left lead eV_L is high enough to compensate for the increase in charging energy of the island

$$eV_L > E_{\text{ch}}(n + 1, Q_G) - E_{\text{ch}}(n, Q_G) . \quad (17)$$

Similarly tunneling from the island (transition from $n + 1$ to n) to the right lead is possible at low temperature only if

$$E_{\text{ch}}(n + 1, Q_G) - E_{\text{ch}}(n, Q_G) > eV_R . \quad (18)$$

Both conditions have to be satisfied simultaneously in order for a current to flow through the transistor. It is obvious from the figure that at low transport voltages, depending on the gate voltage V_G we may be either in a Coulomb blockade regime or have a finite current. By varying the gate voltage we produce the Coulomb oscillations, i.e. the e -periodic dependence of the conductance on Q_G .

Additional devices can be examined (for a review see Estève's article in Ref. [50]): (i) The electron trap is similar to the electron box except that it contains at least two junctions in series. In contrast to the box the trap has metastable charge states. (ii) Two traps can be combined to build the electron turnstile, which can serve as a current source [48]. A suitable ac-gate voltage with frequency f allows the controlled transfer of

a single–electron per cycle. Hence the current is $I = ef$. (iii) Finally, in single–electron pumps a current is driven by two phase-shifted ac-voltages applied to different islands [49]. In this case a current $I = ef$ is transported even at vanishing transport voltages. The turnstile and pump can serve as a current standards, if one manages to minimize the effect of missed cycles, of thermal fluctuations, and of quantum fluctuations. This requires low frequencies, low temperatures, and a design (many junctions) which minimizes higher order quantum tunneling processes.

Many properties of single–electron systems can be understood by considering only the energy of different charge configurations. However, a detailed understanding of the I - V characteristics requires knowledge of the tunneling rates of the electrons, which will be next topic.

2.3. TUNNELING RATES AND I - V CHARACTERISTICS

In this Subsection we introduce the Hamiltonian of the SET transistor. Using simple golden-rule arguments we derive the rate for the transfer of a single electron charge across the tunnel barriers. It depends crucially on the change in the charging energy. The transition rates enter a master equation, from which we obtain the current-voltage characteristic. If a tunneling process would increase the charging energy it is suppressed at low temperature. This phenomenon is called “Coulomb blockade”. This “orthodox theory” was developed in Refs. [45, 46].

2.3.1. *The Single–electron Tunneling Rate*

For definiteness we consider a SET transistor, shown in Fig. 18, which consists of a metallic island coupled via tunneling barriers to two leads and capacitively to an external gate voltage. Its Hamiltonian is

$$H = H_L + H_I + H_R + H_{\text{ch}} + H_t . \quad (19)$$

Here, $H_L = \sum_{k,\sigma} \epsilon_k c_{k,\sigma}^\dagger c_{k,\sigma}$ describes the noninteracting electrons with wave vector k in the left lead, with similar expressions for the island H_I (with states denoted by q) and the right lead H_R . We allow that the leads have different electrochemical potentials. The Coulomb interaction H_{ch} is assumed to depend only on the total charge on the island, as discussed above,

$$H_{\text{ch}} = \frac{(\hat{n}e - Q_G)^2}{2C} . \quad (20)$$

The number operator of excess electrons on the island is given by $\hat{n} = \sum_{q,\sigma} c_{q,\sigma}^\dagger c_{q,\sigma} - N_+$, where the number of positively charged ions of the island has been subtracted. Charge transfer processes are described by the

standard tunneling Hamiltonian, for instance tunneling in the left junction between the states k and q by

$$H_{t,L} = \sum_{k,q,\sigma} T_{k,q} c_{k,\sigma}^\dagger c_{q,\sigma} + \text{h.c.} . \quad (21)$$

We determine the transition rates by golden-rule arguments. The tunneling rate of an electron in the left junction is

$$\gamma_L^+(n) = \frac{1}{e^2 R_{t,L}} \int_{-\infty}^{\infty} d\epsilon_k \int_{-\infty}^{\infty} d\epsilon_q f_L(\epsilon_k) [1 - f_I(\epsilon_q)] \delta(\delta E_{\text{ch}} + \epsilon_q - \epsilon_k) . \quad (22)$$

It describes the tunneling from one of the many states k in the left lead to one of the available states q in the island. In this process the electron number is increased from n to $n + 1$. The crucial point is that the energy, which is conserved as expressed by the δ -function, contains the energies of the electron states $\epsilon_{k/q}$, but also the *charging energy*. The latter depends on the electron number and the applied voltages V_G and $V_{L/R}$. In the process considered it changes by

$$\delta E_{\text{ch}} = E_{\text{ch}}(n + 1, Q_G) - E_{\text{ch}}(n, Q_G) - eV_L . \quad (23)$$

We further introduced the tunnel conductance of the left junction

$$\frac{1}{R_{t,L}} = \frac{4\pi e^2}{\hbar} N_I(0) \Omega_I N_L(0) \Omega_L |T|^2 . \quad (24)$$

It depends on the tunnel matrix elements $T_{k,q}$, which at this stage can be considered as constants, as well as the densities of states at the Fermi level, $N_{I/L}(0)$, and the volumes, $\Omega_{I/L}$, of the island and lead. Equivalent expressions apply for the reverse process $\gamma_L^-(n + 1)$, decreasing the island charge from $n + 1$ to n , and for the tunneling processes in the other barriers.

In equilibrium the distribution functions $f_{I/L}$ are Fermi functions, and the integrals over the electron states in Eq. (22) can be performed explicitly. The resulting ‘‘single-electron tunneling’’ (SET) rate is [46]

$$\gamma_L^+(n) = \frac{1}{e^2 R_{t,L}} \frac{\delta E_{\text{ch}}}{\exp[\delta E_{\text{ch}}/k_B T] - 1} . \quad (25)$$

At low temperatures, $k_B T \ll |\delta E_{\text{ch}}|$, if the charging energy would increase in a tunneling process, the tunneling is suppressed, $\gamma \rightarrow 0$. This phenomenon is called ‘‘Coulomb blockade’’ of electron tunneling. If charging energy is gained the rate is

$$\gamma_L^+(n) = \frac{1}{e^2 R_{t,L}} |\delta E_{\text{ch}}| \quad \text{for } \delta E_{\text{ch}} \leq 0, T \rightarrow 0 . \quad (26)$$

At finite temperatures all processes are allowed. The forward and backward rates satisfy detailed balance, $\gamma_L^+(n)/\gamma_L^-(n+1) = \exp[-\delta E_{\text{ch}}/k_B T]$.

A familiar limit of what is described above is a single voltage-biased tunnel junction, where δE_{ch} is replaced by $-eV$, independent of n . In this case (25) yields a linear current-voltage relation, $I_t = e[\gamma^+ - \gamma^-] = V/R_t$. We can also reverse the argument. The two requirements — (i) a linear characteristic in the voltage-biased case and (ii) detailed balance — uniquely determine the expression for the rate to be of the form (25).

2.3.2. Master Equation for Sequential Tunneling

Given the electron tunneling rates we can set up a master equation for the probability $P(n, t)$ to find the island in a state with n electrons. The probability changes by tunneling in the left and right junctions. Hence

$$\begin{aligned} \frac{d}{dt} P(n, t) = & - [\gamma_L^+(n) + \gamma_L^-(n) + \gamma_R^+(n) + \gamma_R^-(n)] P(n, t) \\ & + [\gamma_L^+(n-1) + \gamma_R^+(n-1)] P(n-1, t) \\ & + [\gamma_L^-(n+1) + \gamma_R^-(n+1)] P(n+1, t). \end{aligned} \quad (27)$$

The rates and probabilities also determine the current. In the left junction the current is

$$I_L(t) = -e \sum_n [\gamma_L^+(n) - \gamma_L^-(n)] P(n, t). \quad (28)$$

In most cases we apply dc-voltages and are interested in the dc-current. In this case we need only the stationary solution of the master equation, and the currents in the left and right junctions are equal $I = I_L = I_R$.

As an example we consider a junction with symmetric bias $V_L = -V_R = V/2$. At low temperatures and low transport voltages (except at symmetry points) only two different charge states – and those transitions which connect both – have an appreciable probability. For instance, if $ne < Q_G < (n+1)e$ we need to consider only $P(n)$ and $P(n+1)$ and the two transitions $\gamma_L^+(n)$ and $\gamma_R^+(n)$ increasing the island charge from n to $n+1$ electrons, as well as the two reverse transitions $\gamma_L^-(n+1)$ and $\gamma_R^-(n+1)$. The energy changes determining the rates $\gamma_L^+(n)$ and $\gamma_L^-(n+1)$ are

$$\delta E_{\text{ch}} = \pm \left[\left(n + \frac{1}{2} - \frac{Q_G}{e} \right) \frac{e^2}{C} - \frac{eV}{2} \right], \quad (29)$$

respectively, while for the transitions in the right junction eV is replaced by $-eV$. In the 2-state limit the stationary probability and current become

$$P(n) = \frac{\gamma_L^-(n+1) + \gamma_R^-(n+1)}{\gamma_\Sigma} \quad ; \quad P(n+1) = 1 - P(n)$$

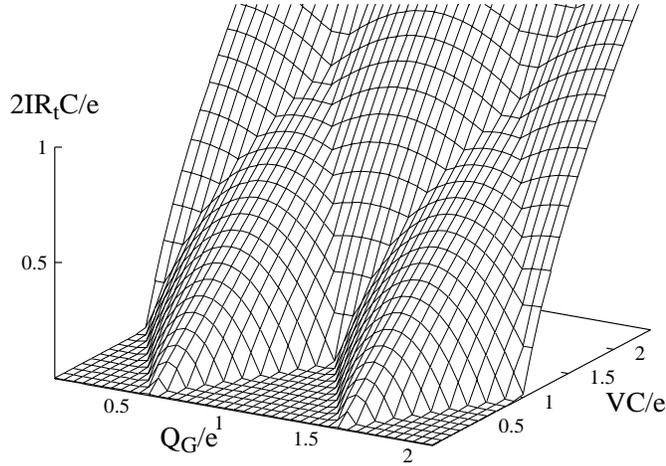


Figure 20. The current of a symmetric transistor is shown as a function of gate and transport voltage. At low temperatures and low transport voltages $VC/e < 1$ only two charge states play a role, and the Coulomb oscillations are clearly demonstrated. At larger transport voltages more charge states are involved.

$$I = -e \frac{\gamma_L^+(n)\gamma_R^-(n+1) - \gamma_R^+(n)\gamma_L^-(n+1)}{\gamma_\Sigma}, \quad (30)$$

where $\gamma_\Sigma = \gamma_L^+(n) + \gamma_R^+(n) + \gamma_L^-(n+1) + \gamma_R^-(n+1)$.

This expression is readily analyzed by inspection of (29). At low temperatures the tunneling process in the left junction from n to $n+1$, with rate $\gamma_L^+(n)$, is allowed when $Q_G - (n+1/2)e \geq -VC/2$. On the other hand, the transition which carries on the charge to the right electrode with rate $\gamma_R^-(n+1)$ is allowed when $Q_G - (n+1/2)e \leq VC/2$. Both coexist in a window of width CV around $Q_G = (n+1/2)e$. The other two processes are not allowed simultaneously, in fact they are suppressed in the window just mentioned. Therefore, at low temperature the current is

$$I = \frac{1}{4R_t} \left[V - \frac{4e}{C^2V} \left(\frac{Q_G}{e} - n - \frac{1}{2} \right)^2 \right] \quad \text{for} \quad -\frac{VC}{2e} \leq \frac{Q_G}{e} - n - \frac{1}{2} \leq \frac{VC}{2e}, \quad (31)$$

while it vanishes outside the window. For simplicity we have assumed in (31) that the two junctions have the same tunneling resistance $R_t = R_{t,L} = R_{t,R}$.

The current through a symmetric SET transistor is plotted as a function of transport and gate voltages in Fig. 20. For gate voltages such that Q_G/e is close to an integer, the current vanishes below a threshold bias voltage $V_{\text{th}}(Q_G = ne) = e/C$. This is a manifestation of the Coulomb blockade. At non-integer values of Q_G/e the threshold voltage is lower $V_{\text{th}}(Q_G) =$

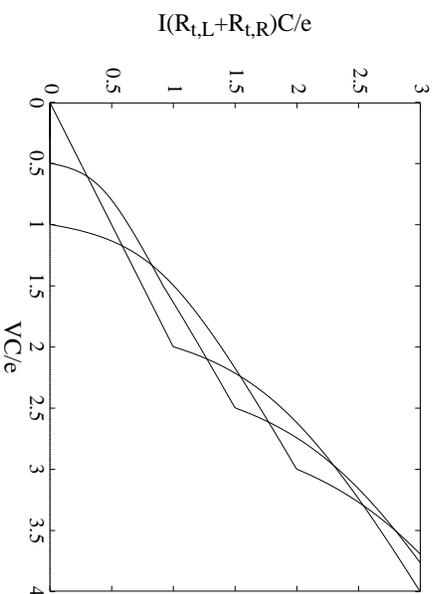


Figure 21. Coulomb staircase: The current of an asymmetric transistor with different tunneling resistances in the two junctions $R_{t,R} = 10R_{t,L}$ is shown as a function of the transport voltage for $Q_G = 0$ (pronounced Coulomb blockade, $Q_G/e = 0.25$ (intermediate), and $Q_G/e = 0.5$ (linear at low voltage)).

$\min_n \{2|Q_G - (n + 1/2)e/C\}$. One finds a series of evenly spaced peaks centered around half-integer values of $Q_G/e = n + 1/2$, each of parabolic shape as given by Eq. (31). These are called “Coulomb oscillations”.

The strong dependence of $I(Q_G, V)$ on the gate voltage makes the SET transistor a highly sensitive “electrometer”. Small changes of polarization charges by fractions of an electron charge influence a macroscopic measurement current. It has been used, for instance, to measure the charge in the electron box $\langle n(Q_G) \rangle$ in Fig. 15.

For larger transport voltages more charge states play a role even at low temperatures. For illustration we consider a junction with symmetric bias $V_L = -V_R = V/2$ and $Q_G = 0$, where the lowest energy state has $n = 0$ electrons in the island. At transport voltages exceeding a threshold $V_{th,0} = e/C$ tunneling sets in to the state with $n = 1$. Above this voltage the electrochemical potential in the left lead is sufficient to compensate the increase in charging energy of the island. Since this state with $n = 1$ is unstable against a tunneling process in the right junction, a current is transported through the system. At the same voltage tunneling processes involving the state with $n = -1$ are possible. At still higher voltages further charge states $|n| \geq 1$ play a role. This leads to a series of voltages $V_{th,n} = (2n+1)e/C$, each marking the threshold above which another pair of charge states becomes populated and a new channel for the conductance opens. The increase in conductance is limited, however, due to the normalization condition for the $P(n)$. Still for suitable parameters (differing conductances of the two junctions or different capacitances) the current increases in the

shape of a staircase. The phenomenon got accordingly the name “Coulomb staircase” [55]. The behavior is demonstrated in the plot of Fig. 21.

2.4. HIGHER ORDER TUNNELING PROCESSES

If sequential single-electron tunneling is suppressed by the Coulomb blockade, higher-order processes such as coherent “cotunneling” through several junctions becomes crucial (Averin and Nazarov in Ref. [50]). As a specific example we consider a SET transistor, biased such that the current in lowest-order perturbation theory vanishes (see Fig. 19). At low temperatures sequential tunneling is exponentially suppressed in this regime since the energy of a state with an excess charge on the island lies above the Fermi levels of the leads. On the other hand, if a transport voltage is applied, a higher-order tunneling process transferring an electron charge coherently through the *total* system is energetically allowed. In this case the state with an excess electron charge in the island exists only virtually. Standard second-order (or fourth, depending on the counting) perturbation theory yields the rate

$$\gamma_{i \rightarrow f} = \frac{2\pi}{\hbar} \left| \sum_{\psi} \frac{\langle i | H_t | \psi \rangle \langle \psi | H_t | f \rangle}{E_{\psi} - E_i} \right|^2 \delta(E_i - E_f). \quad (32)$$

The energy of the intermediate virtual state lies above the initial one, $E_{\psi} - E_i > 0$, but it enters only into the denominator rather than into the exponent of the sequential tunneling rate. Hence the higher-order rate is nonzero even at very low temperatures.

When analyzing the process we have to pay attention to the following:

- (i) There are actually two channels which add coherently. Either an electron tunnels first from the left lead onto the island, and then an electron tunnels from the island to the other lead. In this case the increase in charging energy of the intermediate state compared with the initial one is $\delta E_L = E_{\text{ch}}(n+1, Q_G) - E_{\text{ch}}(n, Q_G) - eV_L$. Or an electron tunnels first out of the island to the right lead, and another electron from the left lead replaces the charge. In this case the increase in energy of the intermediate state is $\delta E_R = E_{\text{ch}}(n-1, Q_G) + eV_R - E_{\text{ch}}(n, Q_G)$. Both amplitudes have to be added before the matrix element is squared.

- (ii) The leads contain a macroscopic number of electrons. Therefore, with overwhelming probability the outgoing electron will come from a different state than the one which the incoming electron occupies. Hence, after the process an electron-hole excitation is left in the island, which explains why it is called “inelastic” cotunneling. This scenario is visualized in Fig. 22.

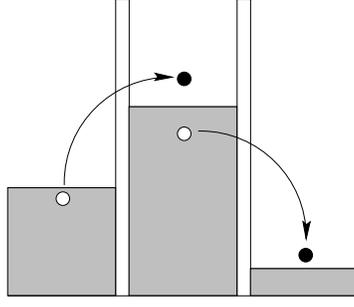


Figure 22. In an inelastic cotunneling process two electrons tunnel coherently, i.e. in a single quantum process, in the left and right junction. The result is the transfer of charge through the system even in the Coulomb blockade regime. A particle-hole excitation is left in the island.

Transitions involving different excitations are added incoherently. The resulting rate for inelastic cotunneling is

$$\begin{aligned} \gamma_{\text{cot}} &= \frac{\hbar}{2\pi e^4 R_{t,L} R_{t,R}} \\ &\times \int_{k \in L} d\epsilon_k \int_{q \in I} d\epsilon_q \int_{q' \in I} d\epsilon_{q'} \int_{k' \in R} d\epsilon_{k'} f(\epsilon_k) [1 - f(\epsilon_q)] f(\epsilon_{q'}) [1 - f(\epsilon_{k'})] \\ &\times \left[\frac{1}{\epsilon_q + \delta E_L - \epsilon_k} + \frac{1}{\epsilon_{k'} + \delta E_R - \epsilon_{q'}} \right]^2 \delta(eV + \epsilon_k - \epsilon_q + \epsilon_{q'} - \epsilon_{k'}). \quad (33) \end{aligned}$$

At $T = 0$ the integrals can be performed analytically. The result for in the Coulomb blockade regime ($eV \ll \delta E_L, \delta E_R$) is

$$\gamma_{\text{cot}} = \frac{\hbar}{12\pi e R_{t,L} R_{t,R}} \left(\frac{1}{\delta E_L} + \frac{1}{\delta E_R} \right)^2 V^3. \quad (34)$$

At finite temperatures forward and backward processes occur. They obey a detailed balance relation $\gamma_{\text{cot}}(-V) = \exp(-eV/k_B T) \gamma_{\text{cot}}(V)$. The current then is

$$I(V) \approx \frac{\hbar}{12\pi e^2 R_{t,L} R_{t,R}} \left(\frac{1}{\delta E_L} + \frac{1}{\delta E_R} \right)^2 [(eV)^2 + (2\pi k_B T)^2] V. \quad (35)$$

In the Coulomb blockade regime of a SET transistor the V^3 dependence of the cotunneling current has been observed. In systems with N junctions a corresponding N -th order process (or $2N$ -th order, depending on the counting) leads to a current $I \propto V^{2N-1}$. As an example we consider $N=4$ junctions with $C = 10^{-15}$ F and tunneling resistance R_t . In this

case (see D. Estève in Ref. [50]) $\gamma_{\text{cot}} = (2.5 \times 10^{-3}/\text{sec}) (V/\mu\text{V})^7 (\text{k}\Omega/R_t)^4$. These cotunneling processes limit the accuracy of the single-electronturnstile even under the most favorable situations, i.e. low T and low frequency, where thermally activated multi-electron transfer processes and missed cycles play little role.

The expression for the cotunneling rate presented above displays several important properties: (i) The expansion parameter is the dimensionless tunneling conductance R_K/R_t , where the quantum resistance R_K serves as reference. (ii) The approximate expression given diverges when the intermediate and initial or final states are degenerate. This divergence is removed by life-time broadening effects. The complete cotunneling theory, recently analyzed in Ref. [56], describes well the logarithmic temperature dependence observed in the experiments of the Saclay group on junctions with tunneling resistances comparable to the quantum resistance [57]. (iii) Higher order processes and eventually resonant tunneling processes are most essential near the points of degeneracy of the charging energy, $Q_G/e = n + 1/2$. The Chapter of Schoeller in this volume presents the theoretical framework to describe tunneling beyond perturbation theory.

There exists also the process of “elastic cotunneling” where *one* electron tunnels through the total system, leaving no excitations in the island. It is the dominant process in the Coulomb blockade regime of tunneling through a single-level quantum dot. However, in the metallic junction it is usually not important, since its rate is smaller by a prefactor $\propto 1/\Omega_I N_I(0)$ (i.e. inversely proportional to the number of states of the island) compared to the inelastic cotunneling rate. The exception is the range of very low voltages and temperatures $k_B T, eV \ll \sqrt{E_C/\Omega N(0)}$, since elastic cotunneling yields a current which is linear in the applied voltage.

Single-electron tunneling is also influenced by the response and the fluctuations of the electromagnetic circuit where it is embedded. This is particularly important in single junctions. In this case Coulomb blockade effects can only be observed if the junction is in series with a nearby large resistor of the order of the quantum resistance or larger. In systems of junctions the tunneling resistance of one junction usually provides the required decoupling of the remaining junctions and single-electron effects are observable. Here we do not have the space to present this theory. It is reviewed in the article by Ingold and Nazarov in Ref. [50].

2.5. SINGLE-ELECTRON TUNNELING THROUGH QUANTUM DOTS

In this Subsection we describe the influence of Coulomb blockade phenomena on single-electron tunneling through ultrasmall quantum dots. The important difference compared to the metallic case is the quantization of

the energy levels inside the quantum dot with typical separation δ . It can be resolved in transport experiments, when the level spacing exceeds the temperature $\delta > k_B T$. In this case one encounters the phenomenon of resonant tunneling through discrete levels.

Many of the experiments showing single-electron effects in quantum dots can be explained by lowest order perturbation theory in tunneling. As in the metallic case, the theory is based on a classical master equation with golden-rule rates which describe incoherent transport through the device. This means that the electrons tunnel sequentially, i.e. they lose their phase memory before the next tunneling process occurs. The “orthodox theory”, initially developed for metallic islands [45, 46], was later generalized to quantum dots with discrete spectra [58, 59, 60, 61]. In this Section we describe this golden-rule theory for the simplest systems. Quantum dots with exact many-body wave functions in the few electron limit have been studied in Ref. [62, 63, 64], while coupled quantum dots have been considered in Ref. [65, 66]. The effect of time-dependent fields have been described in Refs [67, 68] for metallic junctions and in Ref. [69] for the Coulomb blockade model, and in Refs. [70, 71, 72] for the metallic case in the presence of a heat bath. Various extensions, incl. experimental results and further references, will be reviewed in the Chapter of Kouwenhoven *et al.* later in this volume. A powerful theory which allows a consistent treatment of higher order tunneling processes and further generalizations will be presented in the Chapter by Schoeller.

The electron tunneling through a quantum dot is described by the Hamiltonian $H = H_{\text{res}} + H_{\text{D}} + H_{\text{t}}$, where

$$H_{\text{res}} = \sum_{r=L,R} \left[\sum_{k,\sigma} \epsilon_{kr} a_{k,\sigma r}^\dagger a_{k,\sigma r} + eV_r \hat{n}_r \right], \quad (36)$$

$$H_{\text{D}} = \sum_s E_s |s\rangle \langle s|, \quad (37)$$

$$H_{\text{t}} = \sum_{r=L,R} \sum_{k,l,\sigma} T_{kl}^r a_{k,\sigma r}^\dagger a_{l,\sigma \text{D}} + \text{h.c.}, \quad (38)$$

describe the reservoirs, the dot, and the tunneling, respectively.

The reservoirs are assumed to have noninteracting single-particle states labeled by the reservoir index r , wave vector k and spin σ . The voltage of reservoir r is denoted by V_r , and \hat{n}_r denotes the total particle number.

The eigenstates of the isolated dot are denoted by $|s\rangle$ with energy E_s . For the Coulomb blockade model, the states $|s\rangle$ of the dot are specified by the set of all occupation numbers for the single particle states: $|s\rangle =$

$|\{n_{l,\sigma D}\}\rangle$. In this case, the dot energy is given by

$$E_s = \sum_{l,\sigma} \epsilon_{lD} \hat{n}_{l,\sigma D} + E_C (n - n_G)^2, \quad (39)$$

where \hat{n} is the particle number operator of the dot. The charging energy $E_C = e^2/2C$ and the gate charge $Q_G = -en_G = \sum_{i=L,R,G} C_i V_i$ coincide with the expressions (13) and (15) introduced in the metallic case. The general notation in terms of the many-body wave functions $|s\rangle$ is introduced here to include cases where the states of the dot cannot be described by single particle states, see e.g. Refs. [73, 62, 63, 74].

The tunneling part describes charge transfer processes between the reservoirs and the dot. The tunneling matrix elements are conveniently combined in the spectral function

$$\gamma_{r,ll'}(E) = \frac{2\pi}{\hbar} \sum_k T_{kl}^{r*} T_{kl'}^r \delta(E - \epsilon_{kr}), \quad (40)$$

which depends on the energy and the single-particle states l, l' involved. A typical value of the spectral function defines an energy scale γ , which characterizes the intrinsic broadening of the single-particle states of the dot by quantum fluctuations from tunneling. If $\gamma \ll k_B T$, thermal fluctuations dominate over quantum fluctuations, and we can use golden-rule theory.

The reservoirs are treated as large systems in equilibrium described by the grand canonical density matrix $\rho_{\text{res}}^{\text{eq}}$. The tunneling rate for a transition of the dot from state $|s'\rangle$ to $|s\rangle$ when $p = \pm 1$ charges have been added from reservoir r to the dot is given by the golden-rule expression

$$\gamma_{r,ss'}^p = \frac{2\pi}{\hbar} \sum_{\substack{\chi\chi' \\ n_r(\chi') = n_r(\chi) + p}} \rho_{\text{res}}^{\text{eq}}(\chi') |\langle \chi s | H_t | \chi' s' \rangle|^2 \delta(E_s - E_{s'} + E_\chi - E_{\chi'} - p\mu_r). \quad (41)$$

Here χ denote the states of the reservoirs with particle numbers $n_r(\chi)$. The energy conserving δ -function includes the change $\mu_r = eV_r$ of the electrostatic energy, which is regarded here as the effective electrochemical potential of reservoir r . The change of the electrostatic energy of the dot is included in $E_s - E_{s'}$.

Inserting the tunneling Hamiltonian H_t from Eq. (38) into the golden-rule rate (41) yields

$$\gamma_{r,ss'}^p = \sum_{ll'} \gamma_{r,ll'}(E_s - E_{s'} - \mu_r) \langle s | a_{lD}^\dagger | s' \rangle \langle s' | a_{l'D} | s \rangle f_r^p(E_s - E_{s'}), \quad (42)$$

where $f_r^+(\omega) = f(\omega - \mu_r)$ is the Fermi distribution of reservoir r , while $f_r^- = 1 - f_r^+$

The rates can be used as an input for a master equation. Consequently, the stationary dc-probability distribution P_s for the dot and the stationary dc-tunneling current in reservoir r can be calculated from

$$0 = \sum_{s'} [\gamma_{ss'} P(s') - \gamma_{s's} P(s)], \quad (43)$$

$$I_r = e \sum_{ss'} [\gamma_{r,ss'}^+ P(s') - \gamma_{r,s's}^- P(s)], \quad (44)$$

with $\gamma_{ss'} = \sum_{rp} \gamma_{r,ss'}^p$. Current conservation $\sum_r I_r = 0$ follows from the property

$$\sum_r [\gamma_{r,ss'}^+ - \gamma_{r,ss'}^-] = \gamma_{ss'} (n_s - n_{s'}), \quad (45)$$

where n_s denotes the number of particles in the dot in state s .

The rates satisfy the detailed balance relation $\gamma_{r,s's}^- / \gamma_{r,ss'}^+ = \exp[\beta(E_s - E_{s'} - \mu_r)]$. As a consequence, the equilibrium solution of the master equation (43) is the grand canonical distribution

$$P^{\text{eq}}(s) = \frac{1}{Z} e^{-\beta(E_s - \mu n_s)}, \quad (46)$$

which applies when all electrochemical potentials are the same $\mu_r = \mu$. In this case the dc-current (44) is zero.

Using detailed balance we can write the tunneling rates as

$$\gamma_{r,ss'}^+ = f_r^+(E_s - E_{s'}) \gamma_{r,ss'}^\Sigma, \quad \gamma_{r,s's}^- = f_r^-(E_s - E_{s'}) \gamma_{r,ss'}^\Sigma, \quad (47)$$

where

$$\gamma_{r,ss'}^\Sigma = \gamma_{r,ss'}^+ + \gamma_{r,s's}^- \quad (48)$$

is the sum of tunneling ‘in’ and tunneling ‘out’ rates. As a consequence the current (44) can be expressed as

$$I_r = e \sum_{ss'} \gamma_{r,ss'}^\Sigma [f_r^+(E_s - E_{s'}) P(s') - f_r^-(E_s - E_{s'}) P(s)]. \quad (49)$$

A current can flow through the structure if both the tunneling ‘in’ and tunneling ‘out’ rate are non-zero. At low temperatures we consider the transition between two dot states $s_n \leftrightarrow s_{n+1}$, with n and $n+1$ particles in the dot, respectively. For tunneling ‘in’ from reservoir r we need $\Delta E = E_{s_{n+1}} - E_{s_n} < \mu_r$, and a similar relation for tunneling ‘out’ to reservoir r' . Both conditions are satisfied simultaneously if the excitation energy ΔE lies in the window of the effective potentials of the reservoirs, $\mu_{r'} < \Delta E < \mu_r$.

The energy- and state-dependence of the spectral function $\gamma_{r,ll'}(\omega)$ accounts for the energy-dependence of the density of states in the reservoirs,

and mesoscopic fluctuations of the conductance peak heights. Here we concentrate on the simplest case and set

$$\gamma_{r,l'l'}(\omega) \approx \delta_{l'l',r}, \quad (50)$$

In this case we find from (42) that the sum of the tunneling ‘in’ and ‘out’ rates defined in (48), differs from this scale

$$\gamma_{r,ss'}^\Sigma = \gamma_r \sum_l |\langle s | a_{lD}^\dagger | s' \rangle|^2 \quad (51)$$

only by a constant factor, since the Fermi functions have canceled out. Inserting this relation in (49), and using current conservation $\sum_r I_r = 0$ to eliminate the term which is independent of the Fermi functions, we find

$$I_r = e \sum_{r'ss'} \gamma_{r',r'} \sum_l |\langle s | a_{lD}^\dagger | s' \rangle|^2 [P(s) + P(s')] \times [f_r(E_s - E_{s'}) - f_{r'}(E_s - E_{s'})], \quad (52)$$

with $\gamma_r = \sum_{r'} \gamma_{r',r}$. This formula is frequently used in the literature. In linear response, it reduces to a form first derived by Beenakker [59]. In this case we set $\mu_r = \mu + e\delta V_r$, and find

$$I_r = \sum_{r'} G_{rr'} (\delta V_r - \delta V_{r'}), \quad (53)$$

where the conductance is

$$G_{rr'} = -e^2 \sum_{ss'} \gamma_{r',r'} \sum_l |\langle s | a_{lD}^\dagger | s' \rangle|^2 [P(s)^{\text{eq}} + P(s')^{\text{eq}}] f'(E_s - E_{s'} - \mu). \quad (54)$$

In the general case, many excitation energies $E_s - E_{s'}$ can lie between μ_r and $\mu_{r'}$ and are relevant for transport. However, only those transitions $s' \rightarrow s$ will occur for which the initial probability $P(s')$ is not too small. For temperatures and bias voltages smaller than the level spacing δ and the charging energy E_{ch} , only the ground states s_n^0 of the dot will have nonzero occupation probability. This means that only the energies

$$\Delta_n = E_{s_{n+1}^0} - E_{s_n^0} \quad (55)$$

are relevant. The transition from s_n^0 to an excited state s_{n+1}^* does not occur since, upon increasing the gate voltage, the transition $s_n^0 \rightarrow s_{n+1}^0$ occurs sooner and afterwards the dot is already in the $n+1$ -particle ground state. Thus, we obtain the same physical picture as shown in Fig. 19 with the

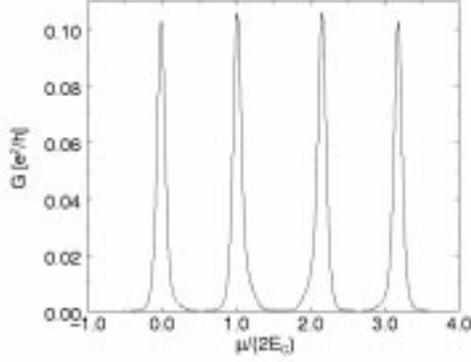


Figure 23. Linear conductance versus μ for two doubly degenerate levels with $\bar{\epsilon}_1 = 0$, $\bar{\epsilon}_2 = 25$, $T = 5$, $E_C = 75$, and $\epsilon^L = \epsilon^R = /2$. The distance between the second and third main resonance is larger due to the finite level spacing. All resonances involving excited states are hidden.

only difference that the distance Δ between adjacent excitation energies is no longer a constant.

This behavior is reflected in the formula (54) for the conductance matrix in linear response. Due to the derivative of the Fermi function, the conductance will be maximal when μ coincides with one of the excitation energies within a range set by the temperature. The energy difference $E_s - E_{s'}$ is varied experimentally by the gate voltage. Thus, the conductance shows a series of resonances with varying distance between the peaks and a line shape which is approximately given by the derivative of the Fermi distribution function. Between the resonances, transport is not possible and the system is in the Coulomb blockade regime. As an example, we show the Coulomb oscillations in Fig. 23 for the Coulomb blockade model with two spin-degenerate levels with energies $\epsilon_1 < \epsilon_2$. According to (39) the excitation energies, describing the energy changes of the dot when a particle is added in level l to a state with n particles, are given by

$$\Delta_{nl} = \bar{\epsilon}_{lD} + 2nE_C. \quad (56)$$

Here $\bar{\epsilon}_{lD} = \epsilon_{lD} + E_C(1 - n_G)$ describes the effective level position, which is tuned by the gate voltage. As a consequence, we observe four resonances corresponding to the excitation energies $\Delta_{01} = \bar{\epsilon}_1$, $\Delta_{11} = \bar{\epsilon}_1 + 2E_C$, $\Delta_{22} = \bar{\epsilon}_2 + 4E_C$, and $\Delta_{32} = \bar{\epsilon}_2 + 6E_C$. As explained above, all other excitation energies $\Delta_{02} = \bar{\epsilon}_2$, $\Delta_{12} = \bar{\epsilon}_2 + 2E_C$, $\Delta_{21} = \bar{\epsilon}_1 + 4E_C$, and $\Delta_{31} = \bar{\epsilon}_1 + 6E_C$ are hidden because they involve excited states.

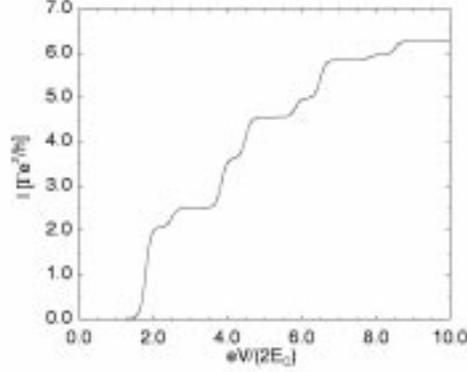


Figure 24. The dc current in nonlinear response versus $eV = e(V_L - V_R)$, with $V_L = -V_R = V/2$ and $C_G V_G/e = 1$ fixed, for two doubly degenerate levels with $\bar{\epsilon}_1 = 0$, $\bar{\epsilon}_2 = 50$. The other parameters are $T = 5$, $E_C = 75$, and $\Gamma^L = \Gamma^R = \Gamma/2$. All one-particle excitations of the dot are visible.

At finite bias voltage all excitations are in principle visible since the excited states acquire a finite occupation probability. This holds at least in the absence of certain selection rules arising from the matrix element $\langle s|a_{iD}^\dagger|s'\rangle$ in (52). For a constant density of states of the leads the I-V-characteristic shows steps each time a new excitation becomes relevant. This result is shown in Fig. 24 for the same example as before. Equivalently, the differential conductance dI/dV shows peaks as function of the bias voltage. As can be seen, all eight excitation energies mentioned before are visible.

The effects of strong correlations on the dot are not only reflected in the increase of the distance between adjacent resonances but also in the line shape of an individual peak as a function of the applied gate voltage. To show this explicitly, we consider the Coulomb blockade model for a single spin-degenerate state with energy ϵ on the dot. For $E_C \gg T$ we can restrict ourselves to the transition between an empty and a singly occupied dot, $n = 0, 1$. According to relation (56), the excitation energy for this transition is given by $\Delta = \bar{\epsilon}$. From the master equation and (52) we find

$$I_r = 2e \sum_{r'} \frac{\Gamma_r, r'}{\Gamma_r} \frac{1}{1 + \sum_{r''} \frac{\Gamma_{r''}}{\Gamma_r} f_{r''}(\Delta)} [f_r(\Delta) - f_{r'}(\Delta)], \quad (57)$$

$$G_{rr'}(\mu) = -2e^2 \frac{\Gamma_r, r'}{\Gamma_r} \frac{1}{1 + f(\Delta - \mu)} f'(\Delta - \mu). \quad (58)$$

The current contains an asymmetry factor $1/(1 + \sum_r \frac{\Gamma_r}{\Gamma_r} f_r(\Delta))$ which is absent either for a nondegenerate level or for the noninteracting case $E_C =$

0. This factor arises from correlations since double occupancy of the dot is forbidden. This gives rise to particle-hole asymmetry and, consequently, to an asymmetric line shape of the differential conductance as a function of Δ as shown in Fig. 25 for finite bias voltage. The maximal value of the conductance in linear response is given by

$$G_{rr'}^{max} = 4\pi \frac{e^2}{h} \frac{r, r'}{3T}. \quad (59)$$

For a nondegenerate level or for the noninteracting case with one degenerate level, the factor $2/3$ has to be replaced by $1/2$ or 1 , respectively. This can be easily understood. At the maximum of the conductance several states of the dot have the same probability. For large charging energy the doubly occupied state can be excluded, and we have two degenerate excitations which can be used for transport, and three possible states of the dot (the empty dot and two degenerate states with one electron). Each excitation contributes equally to the current but has to be multiplied with the probability $1/3$ of the initial state. This explains the factor $2/3$. For a nondegenerate level we have only one excitation and two states, resulting in a factor $1/2$. For a noninteracting model with one degenerate level we have four excitations (two for each transition $n = 0 \rightarrow n = 1$ and $n = 1 \rightarrow n = 2$) and four possible states, giving a factor 1 . The reduction of the current by Coulomb repulsion is obvious, since certain processes are blocked. In contrast to the noninteracting case, we have seen that the presence of degenerate states does *not* give rise to a pure multiplicative factor of the degeneracy. The reason is that Coulomb interaction induces a correlation between the levels. When one level is occupied, the other is not allowed to be occupied due to the strong on-site Coulomb repulsion.

Selection rules occur due to the matrix element $|\langle s|a_{ID}^\dagger|s'\rangle|^2$ in (52). Spin conservation allows only transitions where the total spin of the states s and s' differs by $\pm 1/2$. For the discussion of spin blockade effects and related negative differential conductances we refer to Refs. [62, 63, 74].

Conclusions

Many further fundamental concepts of mesoscopic electron transport could not be included in this Introduction, but will be covered in following Chapters. Interference and localization effects and extensions will be discussed in the Chapters of Stern and of Imry. The Chapter by Kouwenhoven, Markus, McEuen, Tarucha, Westervelt, and Wingreen will cover the wide field of electron transport through quantum dots. Eaves includes in his Chapter concepts related to chaos. De Jong and Beenakker review noise properties of electron transport. Also Büttiker and Christen's Chapter deals

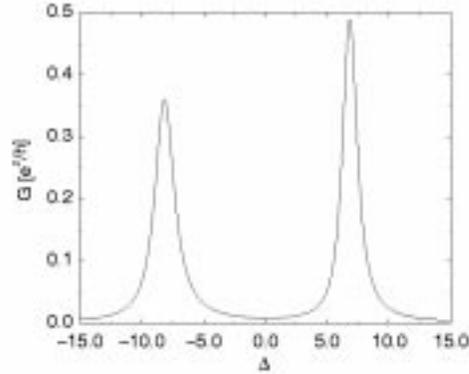


Figure 25. The differential conductance as a function of $\Delta = \bar{\epsilon}$ for a two-fold degenerate level with large charging energy E_C so that double occupancy can be neglected. $T = 0.25$, $\mu_L = -\mu_R = 15$, and $\Gamma^L = \Gamma^R = \Gamma/2$.

with extensions to time-dependent phenomena. A systematic discussion of tunneling beyond perturbation theory is presented in Schoeller's Chapter, while Fisher and Glazman discuss transport in 1D interacting systems.

Superconductivity adds further degrees of freedom to mesoscopic electron transport. The properties of normal metal – superconductor heterostructures are described in the Chapter of the Saclay group, while Fazio and Schön describe single-charge tunneling in superconducting junction systems and further review the theory of quantum transport in NS heterostructures. The very existence of superconductivity in ultrasmall particles is investigated in the article of Ralph et al., while van Wees and Takayanagi address transport through semiconductor – superconductor systems.

The field of scanning probe microscopy has advanced substantially in recent years. It is reviewed by Sohn et al. The transport through quantum point contacts still reveals new results as described in two Chapters by van Ruitenbeek and by Garcia et al.

We included in this Book also two peripheral Chapters. Yamamoto describes concepts of quantum optics to solid state physicists, while DiVincenzo reviews the novel field of quantum computing.

References

1. *Molecular Beam Epitaxy: Fundamentals and Current Status*, M. A. Herman and H. Sitter, Springer-Verlag, NY (1985); *The Technology and Physics of Molecular Beam Epitaxy*, ed. E. H. C. Parker, Plenum Press, NY (1985)
2. H. van Houten, B. J. van Wees, M. G. J. Heijman, and J. P. Andre, *Appl. Phys. Lett.* **49**, 1781 (1986).

3. T. J. Thornton, M. Pepper, H. Ahmed, D. Andrews, and G. J. Davies, *Phys. Rev. Lett.* **56**, 1198 (1986).
4. H. Z. Zheng, H. P. Wei, D. C. Tsui, and G. Weimann, *Phys. Rev. B* **34**, 5635 (1986).
5. S. E. Laux, D. J. Frank, and F. Stern, *Surf. Sci.* **196**, 101 (1988).
6. H. Ehrenreich and D. Turnbull, eds., *Solid State Physics* (Academic Press, New York, 1991).
7. Yu. V. Sharvin, *Zh. Eksp. Teor. Fiz.* **48**, 984 (1965) [*Sov. Phys. JETP* **21**, 655 (1965)].
8. 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).
9. B. J. van Wees, L. P. Kouwenhoven, E. M. M. Willems, C. J. P. M. Harmans, J. E. Mooij, H. van Houten, J. G. Williamson, and C. T. Foxon, *Phys. Rev. B* **43**, 12431 (1991).
10. L. I. Glazman, G. B. Lesovik, D. E. Khmel'nitskii, and R. I. Shekter, *JETP Lett* **48**, 238 (1988); L. I. Glazman, and A. V. Khaetskii, *J. Phys. : Condens. Matter* **1**, 5005 (1989).
11. K. F. Berggren, T. J. Thornton, D. J. Newson, and M. Pepper, *Phys. Rev. Lett.* **57**, 1769 (1986).
12. R. Landauer, *IBM J. Res. Dev.* **1**, 223 (1957); *Phys. Lett.* **85A**, 91 (1981); *J. Phys. Condens. Matter* **1**, 8099 (1989).
13. A. Szafer and A. D. Stone, *Phys. Rev. Lett.* **62**, 300 (1989); E. G. Haanappel and D. van der Marel, *Phys. Rev. B* **39**, 5484 (1989); G. Kirczenow, *Phys. Rev. B* **39**, 10452 (1989).
14. Several authors have obtained transmission resonances in the calculated conductance of narrow constrictions, see Ref. [13].
15. J. Nixon and J. Davies, *Phys. Rev. B* **41**, 7929 (1990).
16. J. G. Williamson, C. E. Timmering, C. J. P. M. Harmans, J. J. Harris, and C. T. Foxon, *Phys. Rev. B* **42**, 7675 (1990).
17. B. I. Halperin, *Phys. Rev. B* **25**, 2185 (1982).
18. B. J. van Wees, L. P. Kouwenhoven, H. van Houten, C. W. J. Beenakker, J. E. Mooij, C. T. Foxon, and J. J. Harris, *Phys. Rev. B* **38**, 3625 (1988).
19. L. P. Kouwenhoven, B. J. van Wees, C. J. P. M. Harmans, J. G. Williamson, H. van Houten, C. W. J. Beenakker, C. T. Foxon, and J. J. Harris, *Phys. Rev. B* **39**, 8040 (1989).
20. H. van Houten, B. J. van Wees, J. E. Mooij, C. W. J. Beenakker, J. G. Williamson, and C. T. Foxon, *Europhys. Lett.* **5**, 721 (1988); C. W. J. Beenakker, H. van Houten, and B. J. van Wees, *Europhys. Lett.* **7**, 359 (1988); H. van Houten et al., *Phys. Rev. B* **39**, 8556 (1989).
21. M. Büttiker, *Phys. Rev. Lett.* **57**, 1761, (1986).
22. J. Spector, H. L. Stormer, K. W. Baldwin, L. N. Pfeiffer, and K. W. West, *Surf. Sci.* **228**, 283 (1990).
23. P. Streda, J. Kucera, and A. H. MacDonald, *Phys. Rev. Lett.* **59**, 1973 (1987).
24. J. K. Jain and S. A. Kivelson, *Phys. Rev. Lett.* **60**, 1542 (1988).
25. M. Büttiker, *Phys. Rev. B* **38**, 9375 (1988).
26. M. Reed, ed., *Semiconductors and Semimetals* (Academic Press, New York, 1990).
27. 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).
28. G. Müller, D. Weiss, S. Koch, K. von Klitzing, H. Nickel, W. Schlapp, and R. Lösch, *Phys. Rev. B* **42**, 7633 (1990).
29. S. Komiyama, H. Hirai, S. Sasa, and S. Hiyamizu, *Phys. Rev. B* **40**, 12566 (1989).
30. B. J. van Wees, E. M. M. Willems, L. P. Kouwenhoven, C. J. P. M. Harmans, J. G. Williamson, C. T. Foxon, and J. J. Harris, *Phys. Rev. B* **39**, 8066 (1989).
31. B. W. Alphenaar, P. L. McEuen, R. G. Wheeler, and R. N. Sacks, *Phys. Rev. Lett.* **64**, 677 (1990).

32. P. L. McEuen, A. Szafer, C. A. Richter, B. W. Alphenaar, J. K. Jain, A. D. Stone, R. G. Wheeler, and R. N. Sacks, *Phys. Rev. Lett.* **64**, 2062 (1990).
33. For a review see T. Chakraborty and P. Pietiläinen, *The fractional quantum Hall effect*, Springer series in Solid-State Sciences 85 (Springer-Verlag 1988).
34. For a review see S. C. Zhang, *Int. J. of Mod. Phys. B*, **6**, 25 (1992).
35. S. L. Sondhi, A. Karlhede, S. A. Kivelson, and E. H. Rezayi, *Phys. Rev. B* **47**, 16419 (1993); and H. A. Fertig, L. Brey, R. Cote, and A. H. MacDonald, *Phys. Rev. B* **50**, 11018 (1994).
36. S. E. Barrett, G. Dabbagh, L. N. Pfeiffer, K. W. West, and R. Tycko, *Phys. Rev. Lett.* **74**, 5112 (1995).
37. L. P. Kouwenhoven, unpublished.
38. G. Timp, R. Behringer, J. E. Cunningham, R. E. Howard, *Phys. Rev. Lett.* **63**, 2268 (1989).
39. L. P. Kouwenhoven, B. J. van Wees, N. C. van der Vaart, C. J. P. M. Harmans C. E. Timmering, and C. T. Foxon, *Phys. Rev. Lett.* **64**, 685 (1990).
40. A. M. Chang and J. E. Cunningham, *Phys. Rev. Lett.* **69**, 2114 (1992).
41. C. W. J. Beenakker, *Phys. Rev. Lett.* **64**, 216 (1990).
42. A. H. MacDonald *Phys. Rev. Lett.* **64**, 220 (1990).
43. A. M. Chang, *Solid State Commun.* **74**, 871 (1990).
44. P. L. McEuen, E. B. Foxman, J. Kinaret, U. Meirav, M. A. Kastner, N. S. Wingreen, and S. J. Wind, *Phys. Rev. B* **45**, 11419 (1992); D. B. Chklovskii, B. I. Shklovskii, and L. I. Glazman, *Phys. Rev. B* **46**, 4026 (1992).
45. I. O. Kulik and R. I. Shekhter, *Zh. Eksp. Teor. Fiz.* **68**, 623 (1975) [*Sov. Phys. JETP* **41**, 308 (1975)].
46. D. V. Averin and K. K. Likharev, *J. Low Temp. Phys.* **62**, 345 (1986); and in *Mesoscopic Phenomena in Solids*, B. L. Altshuler, P. A. Lee and R. A. Webb, eds., p. 173 (Elsevier, Amsterdam, 1991).
47. T. A. Fulton and G. J. Dolan, *Phys. Rev. Lett.* **59**, 109 (1987).
48. L. J. Geerligs et al., *Phys. Rev. Lett.* **64**, 2691 (1990).
49. H. Pothier et al., *Physica B* **169**, 573 (1991), *Europhys. Lett.* **17**, 249 (1992)
50. *Single Charge Tunneling*, NATO ASI Series, Vol. B 294, eds. H. Grabert and M. H. Devoret, (New York, Plenum Press 1992).
51. *Mesoscopic Superconductivity*, Proceedings of the NATO ARW, edited by F. W. J. Hekking, G. Schön, and D. V. Averin, *Physica B* **203** (1994).
52. G. Schön and A. D. Zaikin, *Phys. Rep.* **198**, 237 (1990).
53. L. I. Glazman and K. A. Matveev, *Sov. Phys. JETP.* **71**, 1031 (1990); K. A. Matveev, *Sov. Phys. JETP* **72**, 892 (1991).
54. J. König, H. Schoeller, G. Schön, and R. Fazio, in *Quantum Dynamics of Submicron Structures*, NATO ASI Series B **291**, edited by H. Cerdeira et al. (Kluwer 1995), p. 221; J. König, H. Schoeller, and G. Schön, *Europhys. Letters* **31**, 31 (1995).
55. K. Mullen, E. Ben-Jacob, R. C. Jaklevic, and Z. Schuss, *Phys. Rev. B* **37**, 98 (1988).
56. J. König, H. Schoeller, and G. Schön, *Phys. Rev. Lett.* **78**, 4482 (1997).
57. P. Joyez, V. Bouchiat, D. Estève, C. Urbina, and M. H. Devoret, preprint.
58. D. V. Averin and A. N. Korotkov, *Zh. Eksp. Teor. Fiz.* **97**, 1661 (1990) [*Sov. Phys. JETP* **70**, 937 (1990)]; D. V. Averin, A. N. Korotkov, and K. K. Likharev, *Phys. Rev. B* **44**, 6199 (1991).
59. C. W. J. Beenakker, *Phys. Rev. B* **44**, 1646 (1991).
60. Y. Meir, N. S. Wingreen, and P. A. Lee, *Phys. Rev. Lett.* **66**, 3048 (1991).
61. H. van Houten, C. W. J. Beenakker, and A. A. M. Staring, in Ref. [50].
62. D. Weinmann, W. Häusler, W. Pfaff, B. Kramer, and U. Weiss, *Europhys. Lett.* **26**, 467 (1994); D. Weinmann et al *Phys. Rev. Lett.* **74**, 984 (1995).
63. K. Jauregui, W. Häusler, D. Weinmann, and B. Kramer, *Phys. Rev. B* **53** R1713 (1996).
64. D. Pfannkuche and S. E. Ulloa, *Phys. Rev. Lett.* **74**, 1194 (1995).
65. R. H. Blick, R. J. Haug, J. Weis, D. Pfannkuche, K. von Klitzing, and K. Eberl,

- Phys. Rev. B **53**, 7899 (1996).
66. G. Klimeck, G. Chen, and S. Datta, Phys. Rev. B **50**, 2316 (1994); Phys. Rev. B **50**, 8035 (1994).
 67. L.P. Kouwenhoven, S. Jauhar, K. McCormick, D. Dixon, P.L. McEuen, Yu.V. Nazarov, N.C. van der Vaart, and C.T. Foxon, Phys. Rev. B **50**, 2019 (1994).
 68. I. A. Devyatov and K. K. Likharev, Physica B **194-196**, 1341 (1994).
 69. C. Bruder and H. Schoeller, Phys. Rev. Lett. **72**, 1076 (1994).
 70. G. Ingold and Yu. V. Nazarov, in Ref. [50].
 71. A. A. Odintsov, G. Falci, and G. Schön, Phys. Rev. B **44**, 13089 (1991).
 72. K. Flensberg, S. M. Girvin, M. Jonson, D. R. Penn, and M. D. Stiles, Phys. Scripta T **42**, 189 (1992).
 73. D. Pfannkuche, V. Gudmundsson, and P. A. Maksym, Phys. Rev. B **47**, 2244 (1993).
 74. W. Häusler, Ann. Physik **5**, 401 (1996).