9
Single-Electron Tunneling

The charge stored on a capacitor is not quantized: it consists of polarization charges generated by displacing the electron gas with respect to the positive lattice ions and can take arbitrary magnitudes. The charge transfer across a tunnel junction, however, is quantized in units of the electron charge (*single-electron tunneling*), and may be suppressed due to the Coulomb interaction (*Coulomb blockade*). These simple facts lay the foundation for a new type of electronic device called single-electron tunneling (SET) devices. Coulomb blockade was first suggested back in 1951 by Gorter [123], who explained earlier experiments [164]. It remained largely unnoticed until, almost 40 years later, Fulton and Dolan built a transistor based on single-electron tunneling [109]. After introducing the concept of Coulomb blockade in Section 9.1, we will discuss basic single-electron circuits, in particular the double barrier and the single-electron transistor, in Section 9.2. Some examples and applications are given in Section 9.3.

9.1
The principle of Coulomb blockade

Consider a tunnel junction biased by a voltage V. The equivalent circuit of a tunnel junction consists of a "leaky" capacitor, i.e. a resistor R in parallel with a capacitor C (Fig. 9.1). For charges $|q| < e/2$, an electron tunneling across the barrier would increase the energy stored in the capacitor. This effect is known as *Coulomb blockade* [191]. For $|q| > e/2$, the tunneling event reduces the electrostatic energy, and the differential conductance is given by $dI/dV = 1/R$. Experimentally, it is far from easy to observe Coulomb blockade at a single tunnel barrier, for two reasons.

First of all, in order to avoid thermally activated electron transfers, $e^2/(8C) \geq k_B\Theta$ is required.

Question 9.1: *A typical tunnel junction patterned by angle evaporation is formed by a thin oxide layer (thickness 5 nm, dielectric constant $\epsilon \approx 5$). Estimate the maxi-*
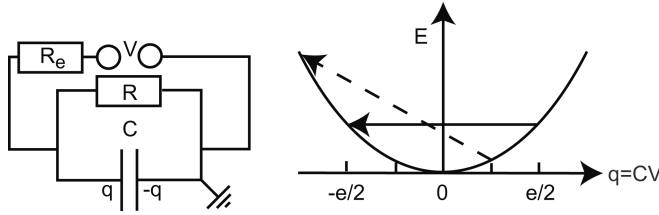
Fig. 9.1 Equivalent circuit and energy diagram of a single tunnel junction. The resistor R_e represents the low-frequency impedance of the environment.

mum area of the capacitor plates for Coulomb blockade to be observed at (a) 4.2 K and (b) 300 K.

Second, the resistance of the tunnel junction has to be "sufficiently large". We can speak of individual electrons tunneling through the barrier only if the tunnel events do not overlap, which means that the time between two successive events $\delta t \approx eR/V$ must be large compared to the duration τ of a tunnel event, which can be estimated as $\tau \approx \hbar/eV$ [178]. This leads to the condition $R \gg \hbar/e^2$. Furthermore, quantum fluctuations can destroy the Coulomb blockade as well. So far, we have neglected the fact that the tunnel junction is coupled to its environment, which is modeled by the resistance R_e in Fig. 9.1. More generally, the environment represents a frequency-dependent impedance, although here we restrict ourselves to very small frequencies, such that the impedance can be replaced by R_e.

In fact, our above line of arguing implicitly assumes the so-called *local rule*, which states that the tunneling rate across the junction is governed by the difference in electrostatic energy right before and right after the tunnel event. According to the *global rule*, on the other hand, the tunnel rate is determined by the electrostatic energy difference of the whole circuit. Since the environment inevitably includes some capacitances much larger than the capacitance of the tunnel junction, we may expect that, in this case, the Coulomb blockade vanishes.

The influence of the electromagnetic environment on the performance of tunnel junctions is discussed in detail in [125]. Here, we just give a simple argument. The local rule holds provided the tunnel junction is sufficiently decoupled from the environment. In the leads, quantum fluctuations of the charge take place. An estimate based on the Heisenberg uncertainty relation tells us what "sufficiently decoupled" actually means: for quantum fluctuations with a characteristic energy amplitude δE, the uncertainty relation $\delta E \, \delta t \geq \hbar/2$ holds. Coulomb blockade is only visible for energy fluctuations at the junction much smaller than $e^2/8C$, while the time scale is given by the time constant of the circuit: $\delta t \approx \tau = R_e C$.

Hence, Coulomb blockade can be observed on a single tunnel junction only if the resistance of the environment is of the order of the resistance quantum h/e^2 or higher. The influence of the environmental resistance on the Coulomb blockade has been calculated in [70] and is shown in Fig. 9.2. These considerations imply that it is not so easy to observe Coulomb blockade at a single tunnel junction. Since the environment has to be sufficiently decoupled, the resistance of the leads has to be larger than h/e^2. This generates Joule heating, which in turn makes it difficult to keep the electron temperature below $e^2/2Ck_B$. Nevertheless, Coulomb blockade has been observed in single tunnel junctions biased via wires of sufficiently high resistance (Fig. 9.3).
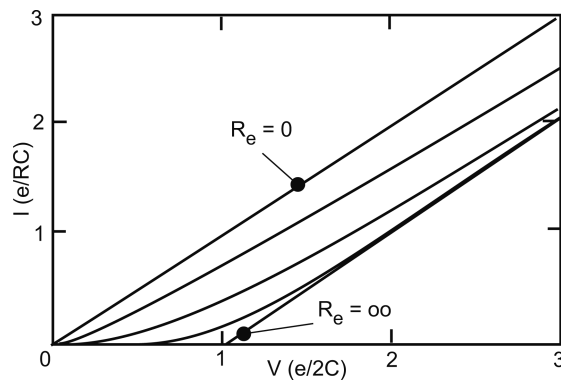

Fig. 9.2 Evolution of the $I-V$ characteristic of a single tunnel junction as the resistance of the environment R_e is increased. For $R_e > h/2e^2$, the Coulomb gap becomes clearly visible. The traces are shown for $R_e/R = 0, 0.1, 1, 10$, and ∞. After [70].
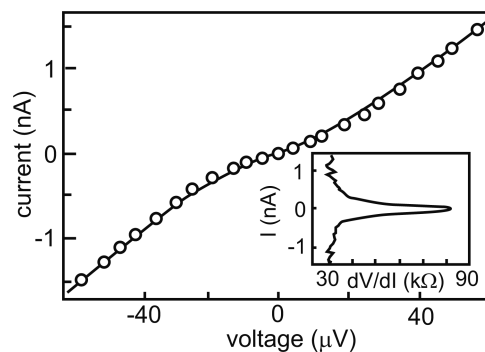

Fig. 9.3 The $I-V$ characteristic of Al–Al$_2$O$_3$–Al tunnel barriers, fabricated by angle evaporation. In order to suppress quantum fluctuations, the cross section of the Al wires is only $10\,\text{nm} \times 10\,\text{nm}$. The superconductive state has been destroyed by applying a magnetic field. After [57].

The limitations imposed by the need to decouple the environment from the tunnel junction can be relaxed by using two tunnel junctions in series (Fig. 9.4), since here quantum fluctuations at the island in between the junctions are strongly suppressed [125]. The number of electrons at the enclosed island can change only by tunneling across one of the barriers, an event essentially free of dissipation. The energy relaxation will take place somewhere in the leads, far away from the island. The resistance of relevance for the suppression of the quantum fluctuations is now that of a tunnel barrier, while the capacitance corresponds to the total capacitance of the island to its environment. Therefore, quantum fluctuations at the island can be suppressed easily without running into heating problems.
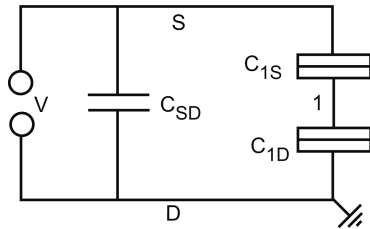

Fig. 9.4 A double barrier structure attached to source (S) and drain (D). C_{SD} denotes a residual capacitance between the two leads.

Question 9.2: *The self-capacitance of a metallic grain is sometimes estimated by $C_{self} = V/q$, where V denotes the potential of the grain and q the charge transferred onto it from infinity (at zero potential). For a sphere, C_{self} equals $4\pi\epsilon\epsilon_0 r$, whereas, for a circular disk, $C_{self} = 8\epsilon\epsilon_0 r$ (r denotes the radius of the island). Estimate C_{self} and the charging energy for some reasonable grain radii.*

9.2
Basic single-electron tunneling circuits

Before we discuss single-electron tunneling in the double barrier system, it is useful to have a look at the problem from a more general point of view, which is then used to analyze specific examples including the circuit of Fig. 9.4.

Consider an arrangement of $(n + m)$ conductors embedded in some insulating environment. Each conductor i is at an electrostatic potential V_i, has a charge q_i stored on it, and has a capacitance C_{iD} to drain (ground).[1] Between

---

[1]) In publications, one frequently encounters an "antisymmetric bias condition", where a voltage of $V_S = +V/2$ is applied to the source, and the drain voltage is $V_D = -V/2$. The electrostatics is different in that case.

each pair of conductors i and j, there is a mutual capacitance C_{ij}. Some of these capacitances may belong to tunnel junctions, which allow electron transfers between the corresponding conductors. Furthermore, we assume that m conductors are connected to voltage sources, which we call *electrodes*, while the n remaining ones are *islands*.[2] For convenience, we enumerate the n islands from 1 to n, and the m electrodes from $n + 1$ to $n + m$.

The charges and potentials of the islands can be written in terms of an island charge vector $\vec{q}_I$ and potential vector $\vec{V}_I$, respectively. Similarly, charge and potential vectors can be written down for the electrodes, $\vec{q}_E$ and $\vec{V}_E$. The state of the system can be specified by the total charge vector $\vec{q} = (\vec{q}_I, \vec{q}_E)$. Equivalently, it can be characterized by the total potential vector defined as $\vec{V} = (\vec{V}_I, \vec{V}_E)$. Charge and potential vectors are related via the capacitance matrix $\underline{C}$:

$$\vec{q} = \underline{C}\vec{V} \tag{9.1}$$

We write $\underline{C}$ as

$$\underline{C} = \begin{pmatrix} \underline{C}_{II} & \underline{C}_{IE} \\ \underline{C}_{EI} & \underline{C}_{EE} \end{pmatrix} \tag{9.2}$$

The capacitance submatrices between type A and type B conductors (A, B can be electrodes or islands) are denoted by $\underline{C}_{AB}$. Note that the ground is *not* a conductor in terms of our definition, and that $\underline{C}$ is symmetric. The matrix elements of $\underline{C}$ are given by (see Appendix B)

$$(\underline{C})_{ij} = \begin{cases} -C_{ij} & j = 1, \ldots, n + m; \; j \neq i \\ C_{iD} + \sum\limits_{k=1; k \neq i}^{n+m} C_{ik} & j = i \end{cases}$$

The electrostatic energy[3] E is given by the energy stored at the islands, minus the work done by the voltage sources. Minimizing this energy gives us the ground state.

As we shall see, in single-electron circuits, usually the voltages applied to the electrodes are parametrically changed, and the initial island charge vector $\vec{q}_I$ given. As $\vec{V}_E$ is changed, the potential difference between two conductors connected by a tunnel junction may become sufficiently large for electrons to tunnel, resulting in a new charge configuration. Such charge rearrangements will take place as soon as the electrostatic energy of the new configuration is

2) The electrostatics of such systems in terms of the capacitance matrix is discussed in Appendix C.
3) The electrostatic energy is the free energy $E = U - \mu N$, where U is the total energy, μ is the electrochemical potential, and N is the number of electrons.

equal to, or smaller than, the energy of the original configuration. The charge transfer can be specified by the change of the charge vector $\Delta \vec{q} = \vec{q}_{\text{new}} - \vec{q}$. For a system initially in its ground state, we can find the parametric transition to a new ground state from the condition

$$\Delta E = E_{\text{new}} - E \leq 0 \tag{9.3}$$

It may look very cumbersome to calculate the energy differences of all the possible charge transfers and find its minimum. Usually, however, only very few electron transfers have to be considered.

In Eq. (9.3) ΔE is given by[4]

$$\Delta E[\vec{V}_{\text{E}}, \vec{q}_{\text{I}}, \Delta \vec{q}] = \Delta \vec{q}_{\text{I}} \underline{C}_{\text{II}}^{-1} [\vec{q}_{\text{I}} + \tfrac{1}{2} \Delta \vec{q}_{\text{I}} - \underline{C}_{\text{IE}} \vec{V}_{\text{E}}] + \Delta \vec{q}_{\text{E}} \vec{V}_{\text{E}} \tag{9.4}$$

This equation is an important relation, which can be used to analyze Coulomb blockade in all systems that can be characterized by a capacitance matrix. Note that it cannot be used to study Coulomb blockade at the single junction, since the crucial time scale involved there does not enter the formalism leading to Eq. (9.4). We are now ready to study the double barrier shown in Fig. 9.4.

9.2.1
Coulomb blockade at the double barrier

The system consists of one electrode (source S) and one island (1). In the following, islands will be labeled by arabic numbers and electrodes by capital letters. The capacitance matrix reads

$$\underline{C} = \begin{pmatrix} C_{11} & -C_{1S} \\ -C_{1S} & C_{SS} \end{pmatrix}$$

with $C_{11} = C_{1S} + C_{1D}$ and $C_{SS} = C_{1S} + C_{SD}$. The charge on the island is given by the number n of electrons tunneled onto it, plus an arbitrary background charge q_0, induced by the environment: $q = q_0 - ne$. Four different charge transfers are relevant. An electron can hop in both directions across C_{1S} or C_{1D}. For electron transfers across C_{1S}, we have $\vec{V} = (V_1, V)$, $\vec{q} = (q_0 - ne, q_S)$, and $\Delta \vec{q} = \pm e(-1, 1)$. Here "+ (−)" corresponds to a transfer of one electron from S to 1 (1 to S). Consequently, the energy difference reads, according to Eq. (9.4),

$$\Delta E[V, q_0 - ne, \pm e(-1, 1)] = \frac{e}{C_{11}} \left[\frac{e}{2} \pm (ne - q_0 + C_{1D}V) \right] \tag{9.5}$$

For tunnel events across C_{1D}, $\Delta \vec{q} = \pm e(-1, 0)$. Here, "+ (−)" corresponds to a transfer of one electron from D to 1 (1 to D). This gives

$$\Delta E[V, q_0 - ne, \pm e(-1, 0)] = \frac{e}{C_{11}} \left[\frac{e}{2} \pm (ne - q_0 - C_{1S}V) \right] \tag{9.6}$$

---

4) For a derivation of Eq. (9.4), see Appendix C.

Coulomb blockade is established only if all four energy differences are positive. This defines a voltage interval of vanishing current:

$$\text{Max}\left\{ \frac{1}{C_{1S}}[-q_0 + e(n - \tfrac{1}{2})], \ \frac{1}{C_{1D}}[q_0 - e(n + \tfrac{1}{2})] \right\}$$

$$< V < \text{Min}\left\{ \frac{1}{C_{1S}}[-q_0 + e(n + \tfrac{1}{2})], \ \frac{1}{C_{1D}}[q_0 - e(n - \tfrac{1}{2})] \right\} \qquad (9.7)$$

Let us study some special scenarios.

1. *No background charges.* The simplest situation is $n = 0$, no background charges ($q_0 = 0$), and identical junction capacitances $C_{1S} = C_{1D} = C_{11}/2$. Now Eq. (9.7) reads $-e/C_{11} \leq V \leq e/C_{11}$. For $V = 0$, we get

$$\Delta E[0, 0, e(\mp 1, \pm 1)] = \Delta E[0, 0, e(\pm 1, 0)] = e^2/(2C_{11})$$

All four charge transfer processes are suppressed (Fig. 9.5(a)). Applying a positive voltage $V = e/C_{11}$ to the source means that

$$\Delta E[V, 0, e(-1, 1)] = e^2/C_{11} > 0$$
$$\Delta E[V, 0, e(1, -1)] = 0 = \Delta E[V, 0, e(-1, 0)]$$

and

$$\Delta E[V, 0, e(1, 0)] = e^2/C_{11} > 0$$

At this voltage, an electron can either tunnel from drain to the island or from the island to source (Fig. 9.5(b)). Both processes have the same probability.

Question 9.3: *Suppose that an electron has just tunneled from drain onto the island under these conditions. The system is in the state depicted in Fig. 9.5(b). Show that, now, an electron will tunnel from the island to source, and a current is established. Calculate the energy differences indicated in Fig. 9.5(c).*

The system thus oscillates between the situations depicted in Figs. 9.5(b) and (c). In each oscillation cycle, a single electron is transferred from drain to source. In addition, the tunnel events show a pair correlation. Shortly after an electron has tunneled from drain to the island, a tunneling process from the island to drain will take place, and vice versa.
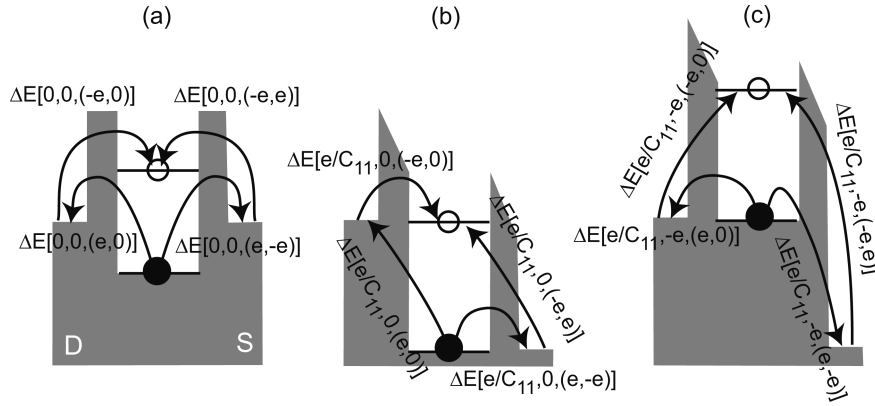
(a)   (b)   (c)

$\Delta E[0,0,(-e,0)]$ $\Delta E[0,0,(-e,e)]$

$\Delta E[e/C_{11},0,(-e,0)]$

$\Delta E[e/C_{11},-e,(e,0)]$

$\Delta E[0,0,(e,0)]$ $\Delta E[0,0,(e,-e)]$

$\Delta E[e/C_{11},0,(e,-e)]$

$\Delta E[e/C_{11},-e,(-e,0)]$

$\Delta E[e/C_{11},-e,(-e,e)]$

$\Delta E[e/C_{11},0,(e,0)]$

$\Delta E[e/C_{11},0,(-e,e)]$

$\Delta E[e/C_{11},-e,(e,e)]$

D   S

Fig. 9.5 Energy differences of the four electron transfers at the double barrier. Open circles denote empty states, while full circles correspond to occupied states. (a) No voltage is applied ($V = 0$), and Coulomb blockade is established. (b) $V = e/C_{11}$. Electrons can hop from drain onto the island, as well as from the island to source. (c) Differences in the electrostatic energy after an electron has, starting from the situation in (b), tunneled from drain onto the island.

2. *Effect of a background charge q_0.* Let us assume that $n = 0$, and $C_{1S} = C_{1D}$, which leads to the condition for Coulomb blockade

$$\text{Max}\left\{ \frac{2}{C_{11}}\left(-q_0 - \frac{e}{2} \right), \frac{2}{C_{11}}\left(q_0 - \frac{e}{2} \right) \right\}$$

$$< V < \text{Min}\left\{ \frac{2}{C_{11}}\left(-q_0 + \frac{e}{2} \right), \frac{2}{C_{11}}\left(q_0 + \frac{e}{2} \right) \right\}$$

This means that, by a non-zero q_0, the Coulomb gap can be reduced, but never be increased. In fact, for $q_0 = (j + \frac{1}{2})e$ with j being an integer, the Coulomb gap vanishes completely. Background charges can seriously hamper the observation of the Coulomb blockade, especially when they are time-dependent.

Question 9.4: *Draw the energy diagram corresponding to Figs. 9.5(a)–(c) for $q_0 = e/4$. Assume equal capacitances.*

Question 9.5: *Show that for $C_{1S} \neq C_{1D}$, the larger capacitance determines the Coulomb gap, which gets reduced compared to the Coulomb gap for identical junctions.*

Coulomb blockade in metallic islands has been known for a long time. As an example of the early indications, we take a look at an experiment of Giaever and Zeller [117]. The authors measured the current–voltage characteristic of a granular Sn film sandwiched between an oxide layer and metallic electrodes (Fig. 9.6). The average diameter of the Sn granules was 11 nm, such that single-electron tunneling is expected to play a role at low temperatures. The system contains an ensemble of double barriers in parallel. Therefore, we expect to observe a gap in the I–V characteristic around $V = 0$ that corresponds to the average single-electron charging energy. Leakage currents through the oxide in between the islands are quite small, since the conductance of tunnel barriers decreases exponentially with increasing barrier thickness. At zero magnetic field, both the Al electrodes as well as the Sn granules are in the superconductive state, and the superconductive energy gap strongly influences the transport measurements.[5] However, by applying a magnetic field, the superconductive state is destroyed and our previous model becomes applicable. The Coulomb gap manifests itself in an increased differential resistance around $V = 0$, compared to that observed at larger voltages.
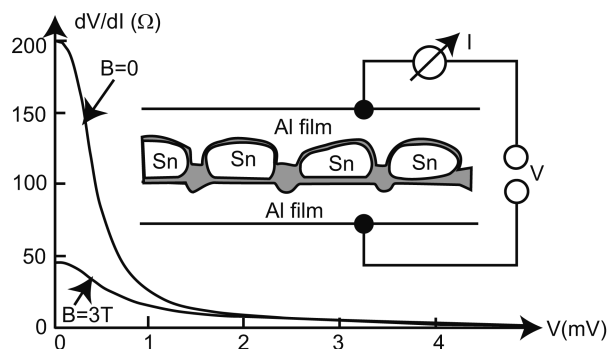

Fig. 9.6 The experiment of Giaever and Zeller. After [117]. A granular Sn film was embedded in an oxide layer and covered on both sides by Al, which acted as source and drain.

9.2.2

Current–voltage characteristics: The Coulomb staircase

Besides the Coulomb gap around $V = 0$, the Coulomb blockade generates under certain conditions a staircase-like structure in the current–voltage characteristic, known as a *Coulomb staircase*. In contrast to our earlier considerations concerning transport through mesoscopic structures, we study here a system of interacting electrons, and a charge transfer changes the electrostatic energy

5) Some information about the interplay of superconductivity and
single-electron tunneling can be gained from Paper P10.4.

as well. To include the interaction, we use the so-called transfer Hamiltonian model, which allows us to relate the change in energy ΔE due to a tunnel event with a tunnel rate $\Gamma(\Delta E)$. For the transmission coefficients calculated in earlier chapters, we always assumed that the energy is conserved. Here, however, the electrostatic energy changes as an electron tunnels, and the voltage sources do some work on the system.

Such situations can be conveniently dealt with by using Fermi's golden rule, which originates in time-dependent perturbation theory. The transfer Hamiltonian model starts from an impenetrable barrier, separating two electron gases. Tunneling is treated as a perturbation and is described by a perturbation Hamiltonian H_t, which is of no further interest to us here. The interested reader is referred to [90] for details. Applied to a tunnel barrier, Fermi's golden rule states that the transition rate for an electron in the initial state $|i\rangle$ to a final state $|f\rangle$ on the other side of the tunnel barrier is given by

$$\Gamma_{i \to f} = \frac{2\pi}{\hbar} |\langle i|H_t|f\rangle|^2 \, \delta(E_f - E_i - \Delta E) \tag{9.8}$$

Here, E_i and E_f denote the energies of the initial and final states with respect to the bottom of the conduction band, and the matrix element $\langle i|H_t|f\rangle$ describes the coupling of the left-hand side to the right-hand side of the tunnel barrier. This transition rate is just the transmission probability per unit time. In order to determine the total transition rate $\Gamma(\Delta E)$, we have to make the following considerations.

1. The tunneling rate at energy E will be proportional to the spectral electron density $n(e) = D_i(E)f(E)$. Here the index i denotes the side of the barrier that hosts state i, D_i is the relevant density of states, and $f(E)$ denotes the Fermi–Dirac distribution function.

2. Since we are dealing with fermions, the electrons can tunnel only into an empty state $|f\rangle$. The transfer rate for an electron in $|i\rangle$ will thus be proportional to $D_f(E + \Delta E)[1 - f(E + \Delta E)]$.

3. We have to integrate over all energies at which states with non-zero tunneling probability exist. These are all the states above the maximum of the conduction band bottoms on both sides $E_{cb,max}$.

Therefore, the total transition rate is given by

$$\Gamma_{1 \to 2}(\Delta E) = \frac{2\pi}{\hbar} \int_{E_{cb,max}}^{\infty} |\langle i|H_t|f\rangle|^2 D_i(E) D_f(E - \Delta E)$$
$$\times f(E)[1 - f(E - \Delta E)] \, dE \tag{9.9}$$

Now, 1 and 2 denote the conductors that contain the initial and final states, respectively. For large energy barriers, we can safely assume that the matrix elements of H_t will be approximately independent of energy. Second, we assume that the density of states does not depend on energy, either since the electron gas is two-dimensional, or since the voltage drop is sufficiently small. Furthermore,

$$f(E)[1 - f(E - \Delta E)] = \frac{f(E) - f(E - \Delta E)}{1 - \exp(\Delta E/k_B\Theta)}$$

If we further consider only cases where the temperature is sufficiently low, we can approximate the Fermi functions by step functions, and obtain

$$\Gamma_{1\to 2}(\Delta E) = \frac{1}{Re^2} \frac{\Delta E}{1 - \exp(\Delta E/k_B\Theta)} \tag{9.10}$$

Here, the resistance R of the tunnel barrier has been defined as

$$R = \frac{\hbar}{2\pi e^2 |\langle i|H_t|f\rangle|^2 D^2} \tag{9.11}$$

(see Exercise E9.2). The current is then obtained from the difference of tunnel rates in both directions,

$$I = e[\Gamma_{1\to 2}(\Delta E_{1\to 2}) - \Gamma_{2\to 1}(\Delta E_{2\to 1})]$$

Let us apply this result to the island of Fig. 9.4. For a steady state, the average charge at the island is constant, and the current from source to the island is given by

$$I(V) = e \sum_{n=-\infty}^{\infty} p(n)[\Gamma_{1\to S}(\Delta E_{1\to S}(n)) - \Gamma_{S\to 1}(\Delta E_{S\to 1}(n))] \tag{9.12}$$

Equivalently, $I(V)$ can be expressed in terms of the drain tunneling rates. Here, we denote the tunneling rate from 1 to source by $\Gamma_{1\to S}(\Delta E_{1\to S}(n))$, while the reverse process is denoted accordingly.

Of course, the energy differences now depend on the number of excess electrons n stored on the island. The probability of finding n electrons on the island is denoted by $p(n)$. We expect this function to be peaked around one number, which is given by the sample parameters and by V. The steady state condition furthermore requires that the probability for making a transition between two charge states (characterized by n) is zero. This means that the rate of electrons entering the island occupied by n electrons equals the rate of electrons leaving the island when occupied by $(n+1)$ electrons:

$$\begin{aligned}
&p(n)[\Gamma_{1\to S}(\Delta E_{1\to S}(n)) + \Gamma_{1\to D}(\Delta E_{1\to D}(n))] \\
&= p(n+1)[\Gamma_{S\to 1}(\Delta E_{S\to 1}(n+1)) + \Gamma_{D\to 1}(\Delta E_{D\to 1}(n+1))]
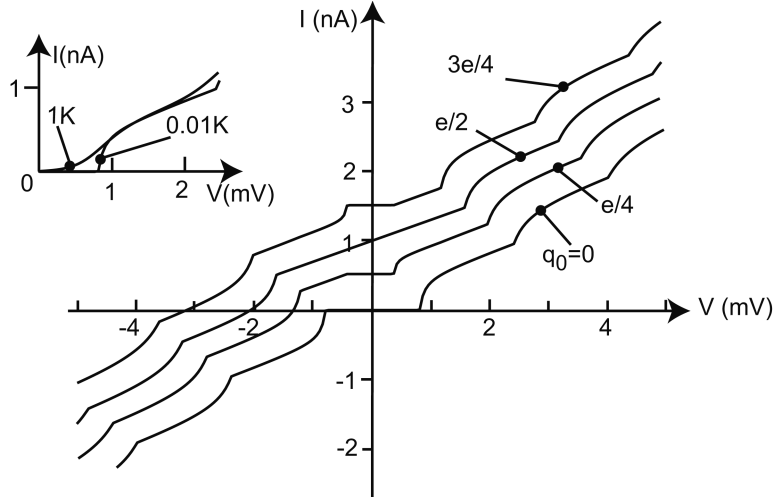\end{aligned} \tag{9.13}$$

Fig. 9.7 Coulomb staircase as calculated from Eq. (9.12), for different background charges q_0. The structure is periodic in q_0, with a period of one elementary charge. Typical sample parameters have been assumed, namely $C_{1S} = C_{1D} = 0.1\,\text{fF}$, $R_{1S} = 20\,\text{M}\Omega$, $R_{1D} = 1\,\text{M}\Omega$, at a temperature of $T = 10\,\text{mK}$. The inset shows the thermal smearing of the Coulomb gap (for $q_0 = 0$) as the temperature is increased to $1\,\text{K}$.

We are now ready to calculate the $I(V)$ characteristic. Equation (9.13), together with the normalization condition

$$\sum_{n=-\infty}^{\infty} p(n) = 1$$

allows us to obtain $p(n)$, which we insert in Eq. (9.12). This requires some numerics, which is considerably simplified by the fact that only a few occupation numbers have non-vanishing probabilities.

Fig. 9.7 shows staircases calculated from Eq. (9.12) for different background charges. The staircases are periodic in q_0 with a period of one elementary charge. Qualitatively, the staircase can be understood as follows: Suppose the tunnel rate across junction S is much larger than that across junction D, and the voltage applied is positive. The voltage now drops completely across junction D, i.e. $V_{1D} \approx V$. From Eq. (9.4), we calculate from $\Delta E[V, -ne, e(-1, 0)] = 0$ the threshold voltages $V(n_0)$ and $V(n_0 + 1)$, which differ by $\Delta V = e/C_{1S} \approx e/C_{11}$. If the voltage is increased by this amount, an additional electron can jump on the island via the drain junction. This increases the current (which is governed by $\Gamma_{1 \to D}$ and by $\Gamma_{D \to 1}$) by $\Delta I = e/R_{1D}C_{11}$ for sufficiently low temperatures, as can be seen by inserting

$$e\Delta V = \Delta E[V, -(n+1)e, e(-1, 0)] - \Delta E[V, -ne, e(-1, 0)]$$
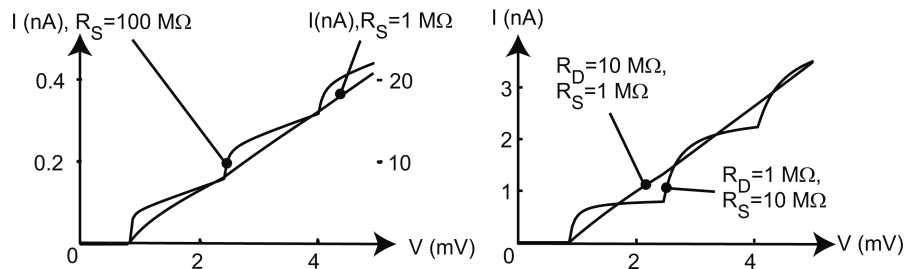
Fig. 9.8 Steps of the Coulomb staircase for various sample parameters, as calculated for $T = 10\,\text{mK}$. Left: $C_{1S} = C_{1D} = 0.1\,\text{fF}$, $R_{1D} = 1\,\text{M}\Omega$. For $R_{1S} = R_{1D}$, the steps are absent, while for $R_{1S} = 100R_{1D}$, they are quite pronounced. Right: Coulomb staircase of an island with two junctions of both different capacitances and different resistances, i.e. $C_{1S} = 0.1\,\text{fF}$, $C_{1D} = 1\,\text{aF}$.

in Eq. (9.12). The markedness of the staircase steps strongly depends on the sample parameters (Fig. 9.8). The steps become most pronounced if both the resistance and the capacitance of one junction are large compared to those of the second junction. Experimentally, however, this is hard to achieve, since small tunnel resistances tend to correspond to small capacitances as well. An analytical model for the Coulomb staircase in this limit is discussed in Paper P9.2.

Particularly beautiful Coulomb staircases have been observed in scanning tunneling experiments on clusters, where the experimental setup consists of a conducting granule or cluster, deposited on an insulating layer on top of a conducting substrate. The tip of a scanning tunneling microscope (STM) is positioned on top of the cluster (Fig. 9.9(a)) and the current is measured as a function of the voltage applied to the STM tip with respect to the substrate [6]. In such experiments, the resistance of one barrier is given by the distance between tip and granule, which can be changed over a wide range. Fig. 9.9(b) shows typical experimental data.

9.2.3
The SET transistor

In 1987, Fulton and Dolan [109] published a seminal experiment: By angle evaporation, a small metallic island was patterned, coupled to source and drain via tunnel barriers with cross sections in the range of $50\,\text{nm} \times 50\,\text{nm}$. In addition, a third electrode (the *gate electrode*) was defined such that the gate–island resistance approaches infinity, and thus couples to the island only capacitively. In this way, the effective background charge and thus the width of the Coulomb gap can be tuned continuously with the gate voltage, and, for sufficiently small source–drain voltages, the current flowing from source
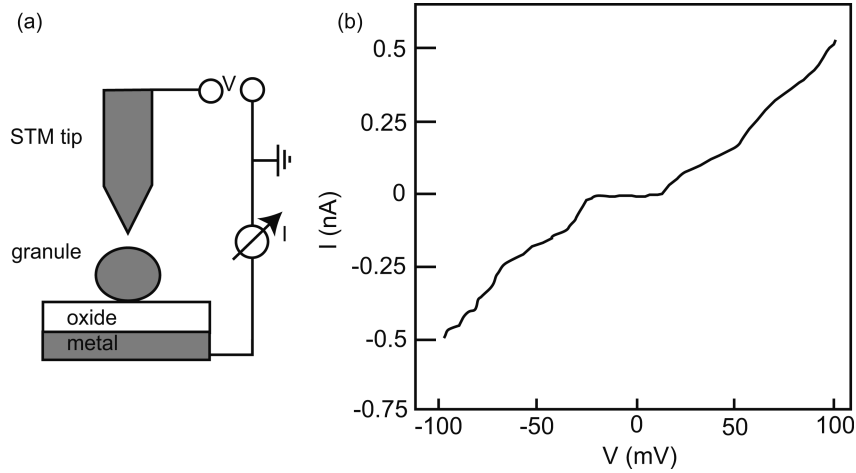
(a)

STM tip

granule

oxide
metal

(b)

Fig. 9.9 (a) One experimental setup for measuring the Coulomb staircase. (b) Experimental data, a least squares fit of which gives the parameters $C_S = 2\,\text{aF}$, $C_D = 4.14\,\text{aF}$, $R_S = 34.9\,\text{M}\Omega$, $R_D = 132\,\text{M}\Omega$, and an offset charge of $0.12e$. Here, the granule was a small indium droplet on top of an oxidized conducting substrate. The temperature was $4.2\,\text{K}$. The measurement is adapted from [6].
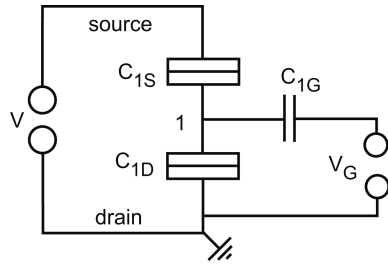
source

C_{1S} C_{1G}

V

1

C_{1D} V_G

drain

Fig. 9.10 Schematic diagram of a SET transistor.

to drain can be controlled. The system constitutes a transistor based on the Coulomb blockade and is known as a single-electron tunneling (SET) transistor. Its equivalent circuit is shown in Fig. 9.10.

For simplicity, let us assume that the background charge vanishes for zero gate voltage. This is no restriction of generality, since additional background charges can always be compensated for by a gate voltage offset. The inverse capacitance matrix now reads $(\underline{C}^{-1})_{11} = 1/C_{11}$, and $(\underline{C}^{-1})_{ij} = 0$ otherwise. Furthermore, $C_{IE} = (-C_{1S}, -C_{1G})$. The electrode voltage vector is given by $\vec{V}_E = (V, V_G)$, while the island charge vector reads $\vec{q}_I = -ne$. The Coulomb gap is given by the onset of the same tunneling events as for the single island studied above. Now, however, the Coulomb gap depends upon the gate
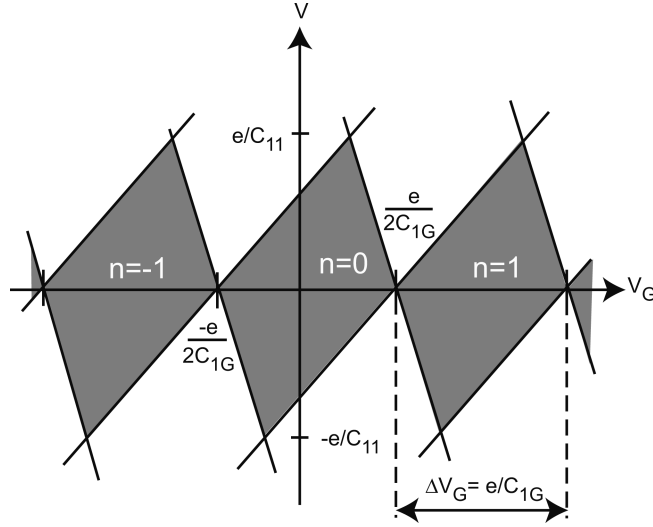
Fig. 9.11 Stability diagram of a single-electron transistor. Within the diamonds, Coulomb blockade is established, while outside, a current flows between source and drain. The slopes of the boundaries are given by $C_{1G}/(C_{11} - C_{1S})$, and by $-C_{1G}/C_{1S}$, respectively.

voltage. The corresponding energy differences are

$$[\Delta E[(V, V_G), -ne, \pm e(-1, 1)] = \frac{e}{C_{11}}[\tfrac{1}{2}e \pm (C_{11} - C_{1S})V \pm ne \mp C_{1G}V_G]$$

$$\Delta E[(V, V_G), -ne, \pm e(-1, 0)] = \frac{e}{C_{11}}[\tfrac{1}{2}e \mp C_{1S}V \pm ne \mp C_{1G}V_G]$$

Coulomb blockade is established if all four energy differences are positive. For each n, this condition defines a stable, diamond-shaped region in the (V_G, V) plane, with the four boundaries given by the onset conditions:

$$\Delta E[(V, V_G), -ne, \pm e(-1, 1)] = 0 \quad \implies$$

$$V(V_G, n) = \frac{C_{1G}}{C_{11} - C_{1S}}V_G - \frac{e(n \pm \tfrac{1}{2})}{C_{11} - C_{1S}}$$

$$\Delta E[(V, V_G), -ne, \pm e(-1, 0)] = 0 \quad \implies$$

$$V(V_G, n) = -\frac{C_{1G}}{C_S}V_G + \frac{e(n \pm \tfrac{1}{2})}{C_{1S}} \tag{9.14}$$

These stable regions are known as Coulomb diamonds, and line up along the V_G-axis (Fig. 9.11).

Fig. 9.12 shows a measurement of the stability diagram of a Al–Al$_2$O$_3$ single-electron transistor. The experimentally obtained shape of the Coulomb diamonds, as well as the current–voltage characteristic, agree very well with
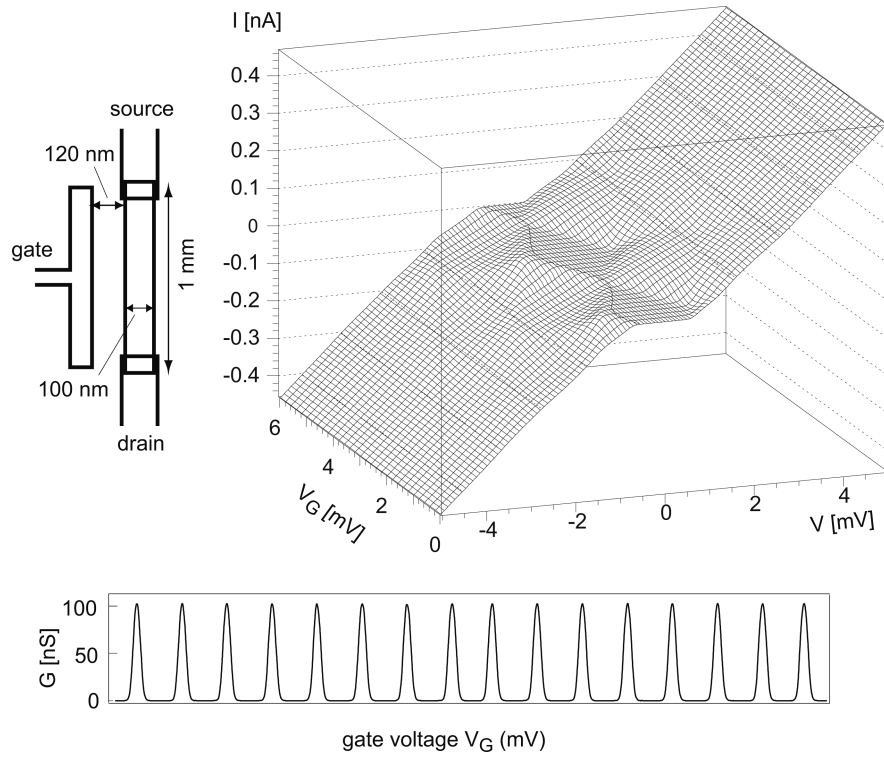
Fig. 9.12 Stability diagram of an Al–Al$_2$O$_3$ SET transistor (its dimensions are shown to the left), measured at a temperature of $30\,$mK. At the bottom, Coulomb blockade oscillations are shown for $V = 10\,\mu$V. Adapted from [111].

the model just developed. For $|V| < e/C_{11}$, the current oscillates strongly as a function of the gate voltage, an effect known as *Coulomb blockade oscillations*. Current peaks occur at $V_{\mathrm{G}} = (e/C_{1\mathrm{G}})(n + \frac{1}{2})$. In each gate voltage period $\Delta V_{\mathrm{G}} = e/C_{1\mathrm{G}}$, n changes by one. It is important to point out that these oscillations have nothing to do with resonant tunneling. Neither did we assume phase coherence, nor does the nearest-neighbor spacing of the energy levels have to be larger than $k_{\mathrm{B}}\Theta$! In fact, for the system shown in Fig. 9.12, the level spacing is well below $1\,\mu$eV. We shall see in the following chapter on quantum dots how single-electron tunneling coexists with size quantization. The weak structures outside the diamonds correspond to Coulomb staircases for each gate voltage, telling us that the two tunnel barriers are not identical.

The line shape of the Coulomb blockade resonances in the limit of negligible source–drain voltage has been derived in [184] and in [23]. The typical experimental situation is characterized by $h\Gamma \ll \Delta \ll k_{\mathrm{B}}\Theta \ll E_{\mathrm{C}}$. This is known as the *metallic regime*. Here, Γ denotes the coupling of the island to the

leads, while Δ is the spacing of the discrete (kinetic) energy levels of the island. Coulomb blockade is well pronounced in this regime, but many energy levels carry current. The line shape of the conductance resonances is given to a good approximation by

$$G(E) = \frac{e^2 D_{\text{island}}}{2} \frac{\Gamma^S \Gamma^D}{\Gamma^S + \Gamma^D} \cosh^{-2} \left(\frac{E - E_{\text{max}}}{2.5 k_B \Theta} \right) \tag{9.15}$$

Here, D_{island} is the density of states in the dot, $\Gamma^{S,D}$ denote the couplings of the dot to source and drain, while E_{max} is the energy at the peak amplitude. Note that the gate voltage can be transformed into an energy via $\delta E = eC_{1G}/C_{11}\delta V_G$. Increasing the temperature thus broadens the resonances, but does not change the peak conductance. Since the conductance of an individual energy level of the island scales as $1/\Theta$ (see Exercise E8.4), and the number of contributing states is proportional to Θ, the total temperature dependence of the peak conductance just cancels [23].

It is important to realize that Coulomb oscillations do not measure the density of states of the island, but the addition spectrum. The density of states tells us how many electrons can be in the system at a particular energy, for a *fixed* number of electrons. The addition spectrum, on the other hand, tells us at which energies electrons can be *added to* the system. If the system is interacting, these two quantities are different, a fact that is clearly demonstrated here. Besides being a somewhat unconventional transistor with an oscillatory transconductance dI/dV_G, this device is extremely sensitive to charges in the vicinity of the island and can thus be used as an electrometer, as used, for example, to study the electrochemical potential in semiconductor heterostructures [161, 321]. Particularly appealing is the integration of a SET transistor in the tip of a scanning probe microscope, which results in an electrometer of both high spatial and charge resolution [131, 340]. The charge resolution is ultimately limited by shot noise; a sensitivity of 10^{-4} electrons has been demonstrated experimentally in [348].

Question 9.6: *Estimate the charge resolution δq achievable with the single-electron transistor of Fig. 9.12. Assume the operation point is in the wing of a Coulomb blockade resonance, and assume a current resolution of 10 fA.*

In transistor operation, its advantage is low power consumption, since, for switching, the charge needs to be changed by only a small fraction of e. Schemes for a digital logic based on single-electron tunneling have been developed, and experimental implementations are being investigated [7, 178]. One problem is to reduce the island size sufficiently in order to operate the devices at room temperature. To date, there are several reports on SET tran-
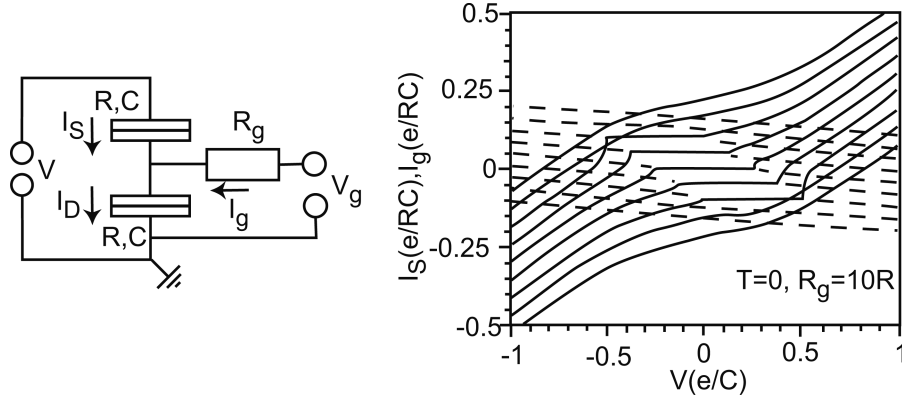
Fig. 9.13 Current–voltage characteristics of a resistively coupled single-electron transistor. Shown is both the source current (solid lines) and the gate current (dashed lines) for V_G varying from $-e/2C$ to $e/2C$ in steps of $e/8C$. The traces are offset vertically for clarity. (Adapted from [179].)

sistors operating at room temperature (see e.g. [275]), but production of such devices is by no means standard. In addition, the switching is strongly disturbed by fluctuating background charges, although a charge stability of 0.01 elementary charges over weeks has been demonstrated in silicon-based SET transistors [349]. Furthermore, the voltage gain in such transistors is limited.

These limitations can be overcome, in principle, by using resistively coupled single-electron transistors. The circuit is shown in Fig. 9.13: the gate couples to the island via a gate resistance $R_G \gg h/(2e^2)$. In describing this device, Eq. (9.10) has to be modified, since charge can also flow from the gate onto the island:

$$p(q)\left[\Gamma_{S\to 1}(\Delta E(q)) + \Gamma_{D\to 1}(\Delta E(q)) + \frac{1}{R_G C_{11}}\frac{\partial}{\partial q}(q - V_G C_{11} + V C_{1D})\right]$$
$$= p(q+e)[\Gamma_{1\to S}(\Delta E(q+e)) + \Gamma_{1\to D}(\Delta E(q+e))] \qquad (9.16)$$

Now $p(q)$ is the probability density of finding the total charge q on the island. The corresponding current–voltage characteristics are shown in Fig. 9.13.

In this device, the gate voltage keeps the island potential fixed at long time scales ($t \gg 1/R_G C_{11}$). If, however, V is sufficiently large and an electron can tunnel from S into the island, the gate response is too slow to prevent an additional voltage buildup at the drain junction, and the electron is able to tunnel to drain. If $|V_G| > e/C_{1D}$, a gate current starts to flow, and the island is open. Therefore, there is only one Coulomb diamond, centered around $(V, V_G) = (0, 0)$. The transconductance is no longer oscillatory in V_G, and the device is much less sensitive to fluctuating background charges. Fabricating such a transistor, however, hits some experimental difficulties that have yet to

be overcome: the heating problem is similar to that in a single tunnel junction, and the stray capacitance between gate and island should be negligible. In addition, this Coulomb diamond is much more sensitive to thermal smearing and noise than those in "conventional" SET transistors [179].

9.3
SET circuits with many islands: The single-electron pump

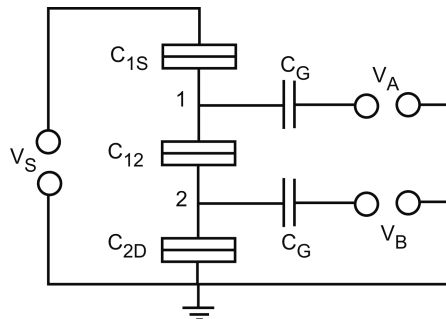As an example of a more complex SET circuit, we study the system of two islands in series, also known as a *single-electron pump* (Fig. 9.14).

Fig. 9.14 Circuit of two islands in series. Each island can be tuned by a nearby gate electrode.

Via a tunnel junction, island 1 is coupled to source and island 2 to drain. The total capacitances C_{11} of both islands are assumed to be equal. Furthermore, we neglect several capacitance matrix elements (except those shown in Fig. 9.14) and assume that electrode A (B) couples only to island 1 (2), with equal capacitances. Nevertheless, V_B influences V_1 via the inter-island capacitance C_{12} and vice versa. We will not study the effect of a source–drain bias voltage. Rather, we are interested in the ground state of the system as a function of V_A and V_B. We assume that we can probe this state by applying a negligibly small source–drain voltage. Hence, we set $V_S = 0$. The island charge vector is given by $-e(n_1, n_2)$, and the electrode voltage vector by $(V_A, V_B, 0)$. The capacitance matrices of interest are

$$\underline{C}_{\mathrm{II}} = \begin{pmatrix} C_{11} & -C_{12} \\ -C_{12} & C_{22} \end{pmatrix}$$

$$\underline{C}_{\mathrm{IE}} = \begin{pmatrix} -C_G & 0 & -C_{1S} \\ 0 & -C_G & 0 \end{pmatrix}$$

with $C_{11} = C_{22} = C_{1S} + C_{12} + C_G = C_{2D} + C_{12} + C_G$. Six electron transfers are of importance.

1. An electron tunnels between source and 1, $\Delta \vec{q}_I = e(\pm 1, 0)$, $\Delta \vec{q}_E = e(0, 0, \mp 1)$. The onset of this transfer is determined by

$$\Delta E[\vec{V}_E, -e(n_1, n_2), \Delta \vec{q}] = 0 \quad \Longrightarrow$$
$$C_{11}[C_G V_A - (n_1 \mp \tfrac{1}{2})e] = -C_{12}[C_G V_B - n_2 e] \tag{9.17}$$

2. An electron is transferred between drain and 2, $\Delta \vec{q}_I = e(0, \pm 1)$, $\Delta \vec{q}_E = (0, 0, 0)$, which gives

$$C_{11}[C_G V_B - (n_2 \mp \tfrac{1}{2})e] = -C_{12}[C_G V_A - n_1 e] \tag{9.18}$$

3. Finally, electrons can be exchanged between 1 and 2, $\Delta \vec{q}_I = e(\pm 1, \mp 1)$, $\Delta \vec{q}_E = (0, 0, 0)$, leading to

$$V_A - \frac{e}{C_G}(n_1 \mp \tfrac{1}{2}) = V_B - \frac{e}{C_G}(n_2 \pm \tfrac{1}{2}) \tag{9.19}$$

These boundaries define regions of stable electron configurations in the (V_A, V_B) plane, each of which is characterized by the island charge vector that corresponds to the lowest energy. For $C_{12} \to 0$, islands 1 and 2 are no longer coupled. It becomes impossible to influence island 1 by V_B and vice versa. In this limit, the stability diagram consists of squares given by conditions 1 and 2. Condition 3 plays no role, since the corresponding lines just touch two corners of the square (Fig. 9.15(a)).

Question 9.7: *Investigate the stability diagram of the double island in the limit of connected islands.*
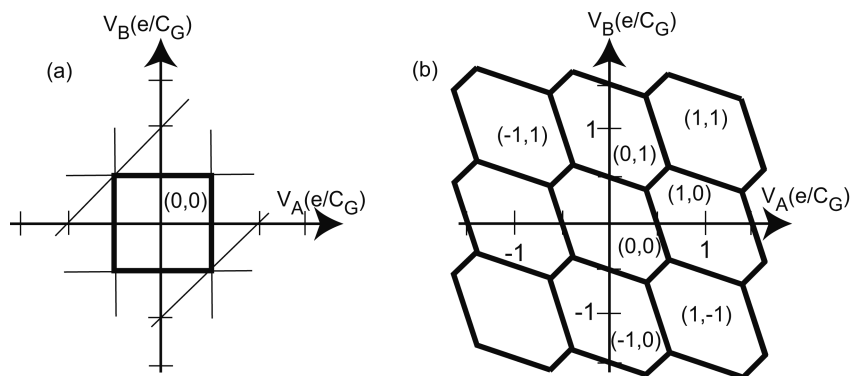

Fig. 9.15 Stability diagram of the two-island system of Fig. 9.14, (a) for completely decoupled islands and (b) for an inter-dot capacitance $C_{12} = C_G$.

The general situation is shown in Fig. 9.15(b): the boundaries (1) and (2) tilt for $C_{12} > 0$, and the stable regions develop a hexagonal shape. A current can pass from source to drain only if electrons can tunnel between the two islands as well as between island 1 (2) and source (drain). This degeneracy exists only at the corners of the elongated hexagons.

Question 9.8: *Study the effect of cross capacitances on the stability diagram. Consider equal capacitances between gate A (B) and island 2 (1), which are much smaller than C_G.*

The charge configuration of the double island system can be directly monitored by coupling a SET transistor to each island (Fig. 9.16). In this setup, the SET transistor labeled by 3 (4) serves as an electrometer to measure the charge on island 1 (2) [5]. In Fig. 9.16(a), the current through the double island is shown as a contour plot. As expected, current flows predominantly at the corners of the stable regions. Figs. 9.16(b) and (c) show the conductance of the electrometers 3 and 4, respectively, which is a measure of the charge on island 1 (2). The transition of the island charges is clearly visible as a sharp increase of the electrometer conductance along the direction that corresponds to changing the charge at the measured island. In Fig. 9.16(d), the difference signal of the two electrometers is shown, which emphasizes that, in each stable region, the charge configuration is really a different one.

In [245], it has been demonstrated for the first time that, with this device, electrons can be "pumped" by the gate voltages. The current can even be made to flow in the opposite direction of the source–drain bias voltage drop. In order to understand this experiment, we first consider the effect of a non-zero bias voltage: it shifts the boundaries of the stability diagram and generates triangular regions at the corners of the hexagons. Inside the triangles, Coulomb blockade becomes impossible. In order to operate the pump, the DC component of the gate voltages V_A and V_B are adjusted such that the device is located within one of these triangles (Fig. 9.17(a)).

Question 9.9: *Calculate the shifts of the boundaries given in Fig. 9.17(a).*

In addition, an AC voltage is applied to gates A and B, with a phase shift of (not necessarily exactly) $\pm\pi/2$. For sufficiently large AC amplitudes, the trajectory of the device state is a circle enclosing the triangle. Circling around the triangle labeled "P" in the positive direction corresponds to a sequence of states $(n_1, n_2) \rightarrow (n_1 + 1, n_2) \rightarrow (n_1, n_2 + 1) \rightarrow (n_1, n_2)$. This means that, for each round trip, one electron is transferred from source to drain, independent
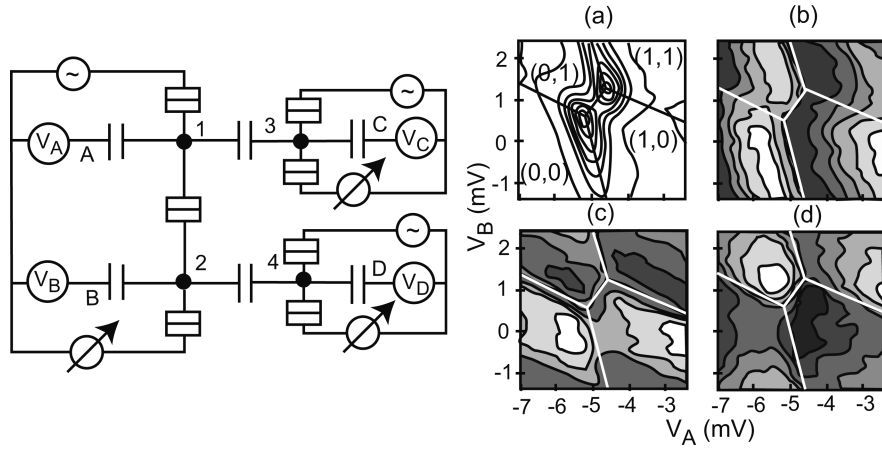
Fig. 9.16 Measurement of the stability diagram of the double island system. Left: Equivalent circuit of the double island system 1 and 2, with each island coupled to a SET transistor acting as an electrometer. Right: (a) conductance of the double island as a function of the gate voltages V_A and V_B in a contour diagram; (b, c) conductance of electrometer 3 (4), respectively; (d) difference signal of the two electrometers. (Adapted from [5].)
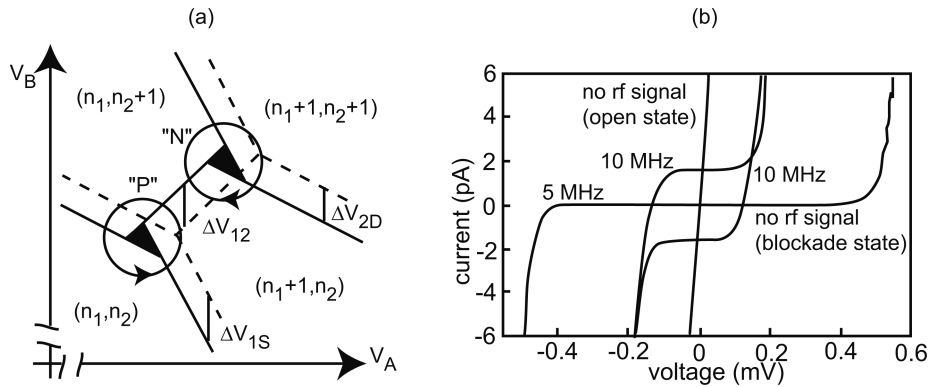

Fig. 9.17 (a) A non-zero bias voltage shifts the boundaries of the stability diagram in the V_B direction by

$$\Delta V_{1S} = -\frac{V_S}{C_G}\left(C + \frac{C^2}{C_{12}} - C_{12}\right)$$

$$\Delta V_{2D} = -\frac{V_S}{C_G}\frac{C_{12}^2}{C}$$

$$\Delta V_{12} = \frac{V_S}{C_G}C_{12}$$

respectively. As a result, triangular shaped regions are formed in which Coulomb blockade no longer exists. The circles denote the trajectories of the device as small AC voltages are applied to gates A and B. (b) Operation of the electron pump at different frequencies. The actual phase shift of the AC signal was $\pm 130°$. Also shown are the I–V characteristics in the center and at a corner of a stable region, without an AC voltage applied. After [193].

of the direction and magnitude of the bias voltage. The current plateaus of the single-electron pump are shown in Fig. 9.17(b) as a "P" point is encircled with two different frequencies in positive (phase shift $\pi/2$) and negative (phase shift $-\pi/2$) directions. Note that the current is independent of the sign of V_S within a window around $V_S = 0$.

Also shown is the current–voltage characteristic when no AC signal is applied. Here, the current plateaus are absent. Provided the trajectory encloses the triangle completely and the AC amplitude is sufficiently small, such that other electron transfers are impossible, the current is coupled to the frequency via

$$I = ef$$

Furthermore, for the system to follow the frequency, f has to be smaller than the inverse time constant $1/\tau$ of the device, given by roughly $\tau = R_{12}C_{12}$. Encircling type "N" points in the same direction, or switching the direction in type "P" points, respectively, reverses the sign of the current.

Frequencies are the most accurate quantities we have in physics (the "NIST-F1 standard" is currently the frequency standard in the US and has an accuracy of 10^{-15}). This raises the question whether the single-electron pump can be used as a current standard, with the current coupled to a frequency (at present, currents can be defined with a relative accuracy of 10^{-6} [203]). Here, the low current that can be pumped through a single-electron pump constitutes a problem. We may, however, rephrase this question and ask: How accurate is the number of electrons pumped? It turns out that the accuracy is dominated by multi-junction tunneling events, so-called cotunneling. Even with Coulomb blockade established, an electron may tunnel onto the island virtually. If this electron, or a different one, tunnels off the island across the second barrier, a real current results. Cotunneling can be suppressed by increasing the number of tunnel junctions [14, 15, 203]. Fig. 9.18 shows an example where the cotunneling has been suppressed by placing high on-chip resistors in series with the SET device [193].[6]

Keller and coworkers [172] used an electron pump (see Fig. 9.19) that consists of six islands in series to charge a capacitor with an accuracy of 10^{-8}, i.e. the uncertainty is one electron for 10^8 pumped electrons. By measuring the voltage drop V across the capacitor after pumping N electrons, the capacitance $C = Ne/V$ could be determined with a standard deviation of 3×10^{-7}.

---

6) The results shown in Figs. 9.16 and 9.17 have actually been obtained with a thin-film Cr resistor located at the entrance and exit of the electron pump (see [193]).
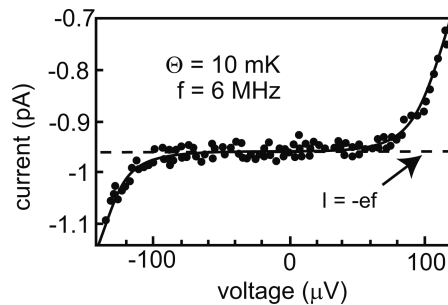
Fig. 9.18 Comparison between the observed current plateau of the single-electron pump (circles) and the current I expected from $I = ef$. Close to the center of the plateau, a relative error of 10^{-6} is found. Here, cotunneling has been suppressed by resistors in series with the single-electron pump. Adapted from [193].
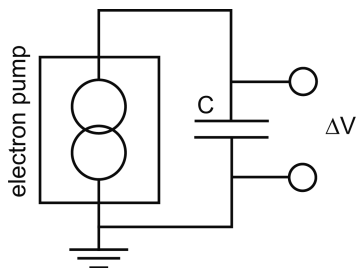

Fig. 9.19 Principle of the capacitance standard: the single-electron pump, consisting of several SET transistors in series, transfers a well defined number of electrons onto the plate of a capacitor, and the voltage drop is measured.

Papers and Exercises

P9.1 In [110], a single-electron transistor is used for detecting charge rearrangements in the substrate. How does this work?

P9.2 Hanna and Tinkham [140] developed an analytical model for the Coulomb staircase in the limit of strongly differing junction couplings. Work out their model and reconstruct the authors' "$I(V)$ phase diagram" in Fig. 1b of that paper.

P9.3 Geerligs et al. [113] demonstrated the operation of a *single-electron turnstile*, a slightly different concept for counting electrons than the single-electron pump. Explain the pumping mechanism of the single-electron turnstile.

P9.4 Superconductivity adds a new and exciting twist to single-electron tunneling. Work out the basic modifications due to superconductivity. A good starting point is Fitzgerald et al. [96].

E9.1 The *"single-electron tunneling box* consists of an island in between a tunnel barrier and a capacitor with infinite resistance (see Fig. 9.20). The tunnel resistance is sufficiently high to suppress quantum fluctuations. Calculate the number of excess electrons on the island as a function of the voltage.
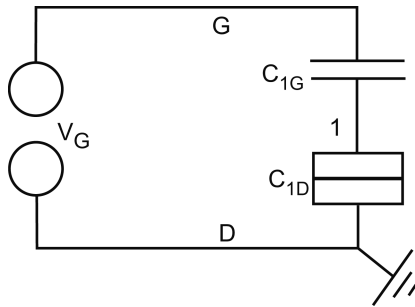

Fig. 9.20 Equivalent circuit of the SET box for Exercise E9.1.

E9.2 Calculate the current through a tunnel barrier in the absence of single-electron charging effects. Show that our definition of the resistance in Eq. (9.11) is reasonable for small voltages applied, since Ohm's law is obtained.

E9.3 Modify the double island system of Fig. 9.14 such that both source and drain couple to island 1 only. Island 2 "dangles" (see Fig. 9.21). In the limit of zero source–drain bias voltage, what does the phase diagram in the (V_A, V_B) plane look like? Discuss the relevance of direct electron transfers between island 2 and the source/drain contacts. Assume identical capacitances.
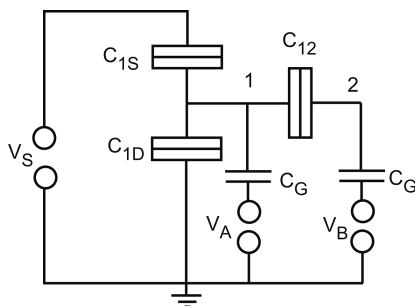

Fig. 9.21 Sketch of the double island system of Exercise E9.3.

Further Reading

A classic review article was written at the beginning of the "single-electron tunneling age" by Averin and Likharev [13]. A stimulating book containing collections of articles on various aspects of single-electron tunneling phenomena is [124]. Furthermore, [178] is an article entitled "Coulomb blockade and digital single electron devices", which focuses on the relevant aspects of a future single-electron logic.