The control of electrons at the level of the elementary charge $e$ was demonstrated experimentally already in the 1980s. Ever since, the production of an electrical current $ef$, or its integer multiple, at a drive frequency $f$ has been a focus of research for metrological purposes. This review discusses the generic physical phenomena and technical constraints that influence single-electron charge transport and presents a broad variety of proposed realizations. Some of them have already proven experimentally to nearly fulfill the demanding needs, in terms of transfer errors and transfer rate, of quantum metrology of electrical quantities, whereas some others are currently “just” wild ideas, still often potentially competitive if technical constraints can be lifted. The important issues of readout of single-electron events and potential error correction schemes based on them are also discussed. Finally, an account is given of the status of single-electron current sources in the bigger framework of electric quantum standards and of the future international SI system of units, and applications and uses of single-electron devices outside the metrological context are briefly discussed.

DOI: 10.1103/RevModPhys.85.1421
PACS numbers: 05.30.–d, 05.40.–a, 73.23.Hk

CONTENTS
I. Introduction 1422
II. Principles of Manipulating Single Electrons 1422
   A. Charge quantization on mesoscopic conductors 1422
   B. Sequential single-electron tunneling 1425
   C. Cotunneling, Andreev reflection, and other higher-order processes 1428
   D. Coulomb blockade of Cooper-pair tunneling 1430
E. Single-electron tunneling in semiconductor structures 1431
F. Influence of environment on tunneling 1433
G. Heating of single-electron devices 1434
III. Realizations 1436
   A. Normal-metal devices 1436
   B. Hybrid superconducting–normal-metal devices 1438
      1. Operating principles 1438
      2. Higher-order processes 1439
      3. Quasiparticle thermalization 1440
   C. Quantum-dot-based single-electron pumps and turnstiles 1442
      1. Introduction to quantum dots as electron pumps 1442

*jukka.pekola@aalto.fi
†On leave from Lebedev Physical Institute, Moscow 119991, Russia.
I. INTRODUCTION

The future definition of the ampere is foreseen to be based on manipulating the elementary charge $e$. Its most direct realization would be the transport of a known number of electrons. Over the past quarter of a century, we have witnessed progress toward ever better control of individual electrons. Since single-electron tunneling is by now a well-established subject, several reviews of its different aspects exist in the literature (Averin and Likharev, 1991; Averin and Nazarov, 1992a; Sohn, Kouwenhoven, and Schön, 1997; van der Wiel et al., 2002; Durrani, 2009).

Several milestones have been achieved in the progress toward a single-electron current source since the initial proposals of the single-charge oscillations (Averin and Likharev, 1985) and of the metrological triangle in the mid-1980s (Likharev and Zorin, 1985). The single-electron amper is based on transporting an electron with charge $e$, or rather a known number $N$ of electrons $Ne$ in each operation of a control parameter that is cyclically repeated at frequency $f$, so that the output dc current is ideally equal to $Ne f$. The needs of precision metrology generally state that this operation has to be performed at a relative error level not larger than $10^{-8}$ and at the same time the current level needs to be several hundreds of picoamperes (Feltin and Piquemal, 2009). Just a few years after the initial theoretical proposal of controlled single-electron tunneling (Averin and Likharev, 1986), the first metallic (Geerligs et al., 1990; Pothier, 1991; Pothier et al., 1992) and semiconducting (Kouwenhoven et al., 1991a) single-electron turnstiles and pumps demonstrated currents $I = Ne f$ with an error of a few percent, still orders of magnitude away from what is needed. As often in precision metrology, the pursuit of higher accuracy has been a pacemaker for understanding new physics, since the errors that need to be suppressed are often a result of interesting physical phenomena. For instance, quantum multielectron processes and nonequilibrium phenomena have been intensively studied in order to improve the performance of single-electron sources. In five years, the accuracy of single-electron pumps was remarkably improved by another 5 to 6 orders of magnitude (Keller et al., 1996) by effectively suppressing the so-called cotunneling current, but at the expense of significantly increased complexity of the device and reduced overall magnitude of the output current (a few picoamperes) of the pump. Alternative ideas were to be found. At the same time, single-electron conveyors in semiconducting channels using surface-acoustic wave (SAW) driving yielded promising results, in particular, in terms of significantly increased current level (Shilton, Talyanskii et al., 1996). Yet likely due to overheating effects in the channel, it may turn out to be difficult to suppress thermal errors to the desired level using this technique.

Interestingly there was a decade of reduced progress in the field, until in the 2000s several new proposals and implementations were put forward. The most promising of these devices are undeniably the sources based on a quantum dot (QD) (Blumenthal et al., 2007), with a single-parameter ac control (Kaestner, Kashcheyevs, Hein et al., 2008), and a superconductor-insulator–normal metal–insulator-superconductor (SINIS) turnstile (Pekola et al., 2008), which is a basic single-electron transistor with superconducting leads and normal-metal island. These simple devices promise high accuracy and a possibility to run many of them in parallel (Maisi et al., 2009). At around the same time, other promising ideas came out, for example, a quantum-phase-slip (QPS) based superconducting current standard (Mooij and Nazarov, 2006). Quantum phase slips provide the mechanism for the existence of the Coulomb-blockade (CB) effects in superconducting wires without tunnel barriers (Astaev et al., 2012) and could potentially lead to current standards producing larger currents. Currently we are definitely witnessing a period of intense activity in the field in a well-founded atmosphere of optimism.

II. PRINCIPLES OF MANIPULATING SINGLE ELECTRONS

A. Charge quantization on mesoscopic conductors

We begin by summarizing the essential concepts of single-electron device physics, with the emphasis on the topics needed for the subsequent discussion of the quantized current sources. We focus mostly on metallic devices since those have
an elaborated theory based on first principles. A brief discussion of how and to what extent the main concepts can be adapted to semiconductor structures is given in Sec. II.E.

As is well known from the elementary treatments of the Bohr model in quantum mechanics, the electrostatic energy of an electron in the hydrogen atom is roughly equal to the kinetic energy of its confinement in the atomic orbitals. The fact that the characteristic energy separation of levels in the confinement energy spectrum decreases much more rapidly than the electrostatic energy with the size of the confining region ensures then that in mesoscopic conductors which are large on the atomic scale, the electrostatic energy of individual electrons can be large even in the regime where the separation of the individual energy levels associated with quantum confinement of electrons is negligible.

As a characteristic estimate, the electrostatic energy of charge $e$ of one electron on a micrometer-size conductor is on the order of a milli-electron-volt, or 10 K in temperature units, and is many orders of magnitude larger than the energy separation $\delta E$ of electron confinement levels in the same conductor, which should be about 1 neV, well below all practical temperatures. As a result, at low but easily reachable temperatures in the kelvin and subkelvin range, the properties of mesoscopic conducting islands are dominated by the electrostatic energy of individual electrons, while small $\delta E$ provides one of the conditions that makes it possible to use macroscopic capacitances to quantitatively describe electrostatics of these conductors even in this “single-electron” regime. The charging energy $U$ of a system of such conductors can be expressed then as usual in terms of the numbers $n_j$ of excess electrons charging each conductor and the capacitance matrix $C$ [see, e.g., Landau and Lifshitz (1980a)]:

$$U(n_j) = \frac{e^2}{2} \sum_{i,j} [C^{-1}]_{ij} n_i n_j, \quad (1)$$

where the sum runs over all conductors in the structure.

The electrostatic energy (1) creates energy gaps separating different charge configurations $\{n_j\}$ which provide the possibility to distinguish and manipulate these charge configurations. Historically, one of the first observations of distinct individual electron charges occurred in Millikan’s experiments on motion of charged micrometer-scale droplets of oil, which produced the evidence that “all electrical charges, however produced, are exact multiples of one definite, elementary, electrical charge” (Millikan, 1911). In those experiments, the oil droplets were, however, charged randomly by an uncontrollable process of absorption of ions which exist normally in air. By contrast, in mesoscopic conductors, the charge states $n_j$ can be changed in a controllable way. Besides the charging energy (1), such a process of controlled manipulation of individual charges in mesoscopic conductors requires two additional elements. First are the tunnel junctions formed between the nearest-neighbor electrodes of the structure which enable the electron transfer between these electrodes, and the second is the possibility to control the electrostatic energy gaps by continuous variation of charges on the junctions (Averin and Likharev, 1986). The simplest way of varying the charges on the tunnel junctions continuously is by placing the electrodes in external electrical fields (Büttiker, 1987) that create continuously varying potential differences between the electrodes of the structure. Externally controlled gate voltages produced in this way can be used then to transfer individual electrons in the system of mesoscopic conductors.

A simple model of the sources of continuously varying external voltages is obtained by taking some of the electrodes of the structure described by the energy (1) to have very large self-capacitance and carry large charge, so that the tunneling of a few electrons does not affect the potentials created by them. For instance, the most basic single-electron structure, the single-electron box (SEB) (Lafarge et al., 1991), can be simplified to two electrodes, one main island carrying the charge $e n$, and the electrode with the charge $e (N - n)$ creating the gate voltage $V_g$ (see Fig. 1). Quantitatively, the structure in Fig. 1 is characterized by the capacitance matrix

$$C = \begin{pmatrix} C_0 & -C_m \\ -C_m & C_{eN} \end{pmatrix} \quad (2)$$

where $C_m > 0$. In the limit $N, C_0 \to \infty$, with $eN/C_0 = V_g$, $C_0$ and $C_{eN}$ have the meaning of the total capacitances of the gate electrode and the island, respectively, and the energy (1) of the charges shown in Fig. 1 reduces for the capacitance matrix (2) to

$$U = U_0 + Ec n^2 - e^2 n n_g/C_{eN}. \quad (3)$$

In this equation, $U_0$ is the $n$-independent energy of creating the source of the gate voltage, $U_0 = e^2 N^2/2C_0$ in this case, $E_C = e^2/2C_{eN}$ is the charging energy of one electron on the main electrode of the box, and $e n_g = C_{eN} V_g$ is the charge induced on this electrode by the gate voltage $V_g$ through the gate capacitance $C_g = C_{eN} - C_m$. As one can see from Eq. (3), the gate voltage $V_g$ indeed controls the energy gaps separating the different charge states $n$ of the main island and therefore makes it possible to manipulate individual electron transitions changing the island charge $en$.

Figure 2(a) shows a scanning electron micrograph of a realistic box structure, in which, in contrast to the schematic diagram of Fig. 1, one pays attention to satisfying several quantitative requirements on the box parameters. First, the capacitance $C_{eN}$ needs to be sufficiently small to have significant charging energy $E_C$, while the gate capacitance $C_g$...
remains not very small in comparison to \( C_S \), to be able to manipulate the charge \( en \) more easily and also to measure it. To satisfy this requirement, the box in Fig. 2(a) is composed of two relatively large similar-size islands with very small overlap between them. Its equivalent electric circuit is shown in Fig. 2(b). The charging energy of the box islands is determined by the fact that an isolated conductor can contain only an integer total number of electrons, with the charging energy producing energy gaps that differ by one electron transition, \( \delta \) \( n_g \) rounded to the nearest integer. This dependence of \( n \) on \( n_g \) means that one electron is added or removed from the box island, changing \( n \) by \( \pm 1 \), whenever \( n_g \) passes through a degeneracy point, i.e., \( n_g = 1/2 \) modulo an integer, at which point the charging energies (3) of the two charge states that differ by one electron transition, \( \delta n_g = 1 \), are equal. If the gate voltage increases monotonically, the dependence \( n(n_g) \) has the shape of the “Coulomb staircase” (Lafarge et al., 1991), with each step of the staircase corresponding to the addition of one electron with gate-voltage increase by \( \delta n_g = 1 \). If the gate voltage oscillates in time around the degeneracy point \( n_g = 1/2 \), as in Fig. 2(c), with an appropriate amplitude (\( \delta n_g \sim 1 \)), it induces back-and-forth electron transitions between the two charge states separated by one electron charge, which can be seen in Fig. 2(c) as the two-level telegraph signal of the detector measuring the box charge. Thus, Fig. 2(c) gives a practical example of manipulation of an individual electron transition in the SEB.

One of the most interesting dynamic manifestations of the manipulation of individual electrons in a system of mesoscopic conductors is the possibility to arrange the system dynamics in such a way that electrons are transferred through it one by one, in a correlated fashion. This can be achieved, for instance, if the gate voltage \( V_g \) of the SEB grows in time at a constant rate such that effectively a constant dc “displacement” current \( I = en_g \) is injected in the box junction. The same dynamic would be obtained if real dc current \( f \) flows into a mesoscopic tunnel junction. In this case, correlated successive transfer of electrons one by one through the junction gives rise to “single-electron tunneling” oscillations (Averin and Likharev, 1986; Bylander, Duty, and Delsing, 2005) of voltage on the junction, \( dU/\partial (en) = e(n - n_g)/C_S \), with frequency \( f \) related to the current by the fundamental equation

\[
I = ef. \tag{4}
\]

More complex structures than a SEB or an individual tunnel junction, such as single-electron turnstile (Geerlings et al., 1990) and pump (Pothier et al., 1992; Keller et al., 1996) discussed below, make it possible to “invert” this relation and transfer one electron per period of the applied gate-voltage oscillation with frequency \( f \). The above discussion of the manipulation of individual electrons in the SEB shows that the charge states \( n \), while controlled by the gate voltage \( V_g \), remain the same in a range of variation of \( V_g \). Physically, such “quantization of charge” results from the fact that an isolated conductor can contain only an integer total number of electrons, with the charging energy producing energy gaps separating different electron number states. Charge quantization enables one to make the accuracy of manipulation of

![Figure 2](color online). Practical SEB. (a) Scanning electron micrograph of a realistic box structure, (b) its equivalent electric circuit, and (c) single-electron transitions in the box illustrating the “charge quantization”: a time-dependent gate voltage \( V_g(t) \) (sinusoidal curve) of an appropriate amplitude drives individual electron transitions changing the box state between the two discrete charge configurations, the electron on the left or on the right island. These two charge states are detected via the detector shown in the upper right corner of (a), whose two-level output current is synchronous with the oscillating \( V_g(t) \). Adapted from Saira et al., 2010 and Saira, Yoon et al., 2012.
individual electron charges in structures such as SEBs very high, in principle approaching the metrological level. Potentially metrological accuracy also extends to the transport in turnstiles and pumps, making the current sources based on single-electron tunneling promising candidates for creation of the quantum standard of electrical current.

B. Sequential single-electron tunneling

One of the key elements in manipulating individual electrons in systems of mesoscopic conductors is a tunnel junction, which provides the means to transfer electrons along the system, thus creating the dc current $I$ through it. A tunnel junction (Giaever, 1960) is a system of two conductors separated by a layer of insulator that is sufficiently thin to allow electrons to tunnel between the conductors (see Fig. 3). For normal conductors, the current through the junction at small applied voltages depends linearly on the voltage and is characterized by the tunnel conductance $G_T = 1/R_T$. In single-electron devices, $G_T$ should satisfy two contrasting requirements. To increase the current $I$ driven through the structure, e.g., to increase the allowed range of frequencies $f$ for which Eq. (4) is satisfied accurately, one should maximize $G_T$. On the other hand, charge quantization on the electrodes of the structure requires that they are well isolated from each other, i.e., $G_T$ should be small. The latter condition can be formulated more quantitatively requiring that the characteristic charging energy $E_C$ of the localized charge states is well defined despite the finite lifetime of these states $\sim G_T/C$, where $C$ is the typical junction capacitance in the structure $E_C \gg \hbar G_T/C$. This condition can be expressed as $G_T \ll 1/R_K$, where $R_K = \hbar/e^2 \approx 25.8 \text{ k}\Omega$ is the characteristic “quantum” resistance. When this condition is satisfied, the localized charge states provide an appropriate starting point for the description of a single-electron structure, while electron tunneling can be treated as a perturbation. In what follows, we mostly concentrate on such a regime of “strong Coulomb blockade” which is necessary for implementation of precise transport of individual electrons as required for quantized current sources.

The majority of practical metallic structures employ tunnel junctions based on barriers formed by either thermal or plasma oxidation of aluminum. The main reason for this is the superior properties of the aluminum oxide layer, in terms of its uniformity and electrical and noise properties. A typical barrier structure is shown in Fig. 3 that includes a high-resolution transmission-electron-microscopy (TEM) image of a cross section of an aluminum-based junction with amorphous Al$_2$O$_3$ tunnel barrier. From the point of view of the Landauer-Büttiker formula for electric conductance of a mesoscopic conductor, the junction tunnel conductance can be expressed as $G_T = (2/\hbar) \sum T_j$, where the sum is taken over spin-degenerate electron transport channels propagating across the junction, and $T_j$ is the quantum mechanical transmission coefficient of the insulator barrier for electrons in the $j$th channel. The condition of the strong Coulomb blockade $G_T \ll 1/R_K$ implies that all individual transmission coefficients are small, $T_j \ll 1$. Although the transmission coefficients $T_j$ are sensitive to the atomic-scale structure of the junction, the fact that the aluminum oxide layer is relatively uniform on an intermediate space scale larger than the individual atoms (Greibe et al., 2011) allows transport properties to be estimated semiquantitatively from the “bulk” properties of the barrier.

Since the tunnel current depends exponentially on the barrier parameters, the measured electron tunneling rates in high-resistance junctions and over the large voltage range allow one to estimate parameters of the aluminum oxide barrier [see, e.g., Tan et al. (2008)]; they yield a barrier height $U \approx 2 \text{ eV}$ and effective electron mass $m_{\text{eff}} \approx 0.5m_e$ in terms of the free electron mass $m_e$. While the dimensions of the typical tunnel junctions need to be small [on the order of $\sim 100 \text{ nm}$, also cf. Fig. 2(a)] in order to make the junction capacitance sufficiently low, they are still quite large on the atomic scale. In this regime, discreteness of the spectrum of the transverse modes $j$ is negligible, and the tunnel conductance $G_T$ is proportional to the junction area $A$. For the value of specific junction resistance $A/G_T \sim 10 \text{ k}\Omega \mu\text{m}^2$ typical for the tunnel junctions, estimates using the barrier parameters and the simplest assumption of ballistic transport in the junction give for the barrier transparency $T \sim 10^{-6}$ corresponding to barrier thickness close to $2 \text{ nm}$ (cf. Fig. 3). A barrier with this thickness effectively transmits only the electrons impinging on it orthogonally. This “focusing” effect means that the tunnel conductance can be expressed in terms of one maximum value of the transmission coefficient $G_T = \mathcal{N}T/R_K$, where the effective number $\mathcal{N}$ of the transport channels in the junction is not determined directly by the density of states (DOS) in the electrodes, but depends also on the characteristic “traversal energy” $\epsilon_0$ of the barrier, which gives the energy scale on which the barrier transparency changes with energy: $\mathcal{N} \approx Am\epsilon_0/2\pi\hbar^2$. For the parameters of the aluminum oxide barrier mentioned, this gives for the area per transport channel $A/\mathcal{N} \approx 1 \text{ nm}^2$. As will be discussed in Sec. II.C, some of the higher-order transitions in the single-electron structures, e.g., Andreev reflection (AR), depend separately on the barrier transmission coefficients $T_j$ and on the number $\mathcal{N}$ of the transport modes. In contrast to this, the lowest-order electron tunneling depends only on the total junction conductance $G_T$.

The most straightforward approach to the description of tunneling in the single-electron structures in the
strong Coulomb-blockade regime is based on the tunnel Hamiltonian method (Cohen, Falicov, and Phillips, 1962), in which the junction is modeled with the following Hamiltonian:

\[ H = H_1 + H_2 + H_T, \quad H_T = \sum_{k,p} T_{kp} c_k^\dagger c_p + \text{H.c.}. \]  

(5)

Here \( H_1 \) are the Hamiltonians of the junction electrodes, \( H_T \) is responsible for tunneling, with \( c_k \) and \( c_p \) denoting the electron destruction operators in the two electrodes, respectively, and \( T_{kp} \) are the tunneling amplitudes. In a typical metallic mesoscopic conductor, when discreteness of single-electron transport. In this case, one can treat the normal metal, is constant on a small energy scale of interest form a continuum with some density of states which in a particular the single-particle electron states is negligible, these states

\[ \Gamma(E) = \frac{G_T}{e^2} \int d\epsilon f(\epsilon)[1 - f(\epsilon + E)] \nu_j(\epsilon) \nu_2(\epsilon + E). \]  

(6)

In this expression, \( f(\epsilon) \) is the equilibrium Fermi distribution function, and \( \nu_j(\epsilon) \) is the density of the single-particle states in the \( j \)th electrode of the junction, \( j = 1, 2 \), in units of the normal density of states \( \rho_j \), which together with the average of the squares of the tunneling amplitudes determine the tunnel conductance \( G_T = 4\pi e^2(|T_{kp}|^2)\rho_1\rho_2/\hbar \). Equation (6) assumes that the energy \( E \approx E_C \) is much smaller than all internal energies of the junction in the normal state, in particular, the traversal energy \( \epsilon_0 \), a condition very well satisfied for practical metallic structures in which \( E_C \sim 1 \text{ meV} \), while \( \epsilon_0 \sim 1 \text{ eV} \). Using the standard properties of the Fermi distribution functions, one can see directly that the rate (6) of tunneling between the two equilibrium electrodes satisfies the necessary detailed balance condition \( \Gamma_{1\rightarrow 2}(-E) = e^{-E/k_BT}\Gamma_{2\rightarrow 1}(E) \). If, in addition, the densities of states are symmetric with respect to the chemical potentials of the electrodes, the tunneling rate is also symmetric, \( \Gamma_{1\rightarrow 2}(E) = \Gamma_{2\rightarrow 1}(E) \), and the detailed balance condition simplifies to \( \Gamma(-E) = e^{-E/k_BT}\Gamma(E) \). The detailed balance condition makes it possible to express the tunneling rate (6) in terms of the current-voltage characteristic \( I(V) \) of the junction at fixed bias voltage \( V \):

\[ \Gamma(E) = I(E/e)/(1 - e^{-E/k_BT}). \]  

(7)

For normal metal–insulator–normal metal (NIN) junctions, when both electrodes are in the normal (N) states, \( \nu_j(\epsilon) = 1 \), Eq. (6) gives, in agreement with Eq. (7), for the tunneling rate

\[ \Gamma(E) = \frac{G_T}{e^2} \frac{E}{1 - e^{-E/k_BT}}. \]  

(8)

Tunneling of individual electrons with the rate (8) is an irreversible dissipative process which converts the electrostatic energy change \( E \) into internal energy of the electron gas inside the junction electrodes. In accordance with this understanding, at small temperatures \( T \), the rate (8) vanishes as \( e^{E/k_BT} \) for energetically unfavorable transitions with \( E < 0 \), when the energy for the transition is taken from the thermal fluctuations of the electron reservoirs. In the regime of allowed transitions \( E \rightarrow E_C > 0 \), the magnitude of the typical transition rate \( \Gamma \sim G_T/C \) for the realistic values of the parameters \( G_T \sim 1 \text{ M\Omega} \), \( C \sim 10^{-16} - 10^{-15} \text{ F} \) is quite high, in the gigahertz range.

In superconductor-insulator–normal metal (SIN) junctions, when one of the junction electrodes is a superconductor (S), the BCS density of states \( \nu_j(\epsilon) = |\epsilon|/(\epsilon^2 - \Delta^2)^{1/2} \) for \( |\epsilon| > \Delta \), and vanishing otherwise, implies that at temperatures well below the superconducting energy gap \( \Delta \), the tunneling rate (6) is strongly suppressed and can be reduced into the kilohertz and even hertz range. Indeed, evaluating the integral in Eq. (6) for the SIN junction assuming \( k_BT \), \( E \ll \Delta \), one gets

\[ \Gamma(E) = \frac{G_T}{e^2} \sqrt{2\pi\Delta k_BT} e^{-\Delta/k_BT} \frac{\sinh(E/k_BT)}{1 - e^{-E/k_BT}}. \]  

(9)

Figure 4 shows the tunneling rate (9) measured in an SIN junction in the configuration of a “hybrid” SEB [see Fig. 2(a)], in which one of the islands of the box is a superconductor (aluminum), the other one being normal metal (copper). The electrostatic energy change \( E \) in the case of the box follows from Eq. (3) as \( E = U(n = 0) - U(n = 1) = 2E_C(n_g - 1/2) \), i.e., is proportional to the deviation of the gate voltage of the box from the degeneracy point \( n_g = 1/2 \). The measurements can be described well by Eq. (9) with reasonable values of parameters including the superconducting energy gap \( \Delta \) of aluminum.

Since the tunneling transitions described quantitatively by the rates (6)–(9) are inherently random stochastic processes,
dynamics of the structures in the strong Coulomb-blockade regime and electron transport properties including the dc current $I$, current noise, or even full statistical distribution of the transferred charge, can be obtained from the time evolution of the probabilities $p(n_j)$ of various charge configurations $\{n_j\}$ governed by the standard rate equation for the balance of the probability fluxes. The most basic single-electron system that allows for the flow of dc current through it and gives an example of such an equation is the single-electron transistor (SET) (Averin and Likharev, 1986; Fulton and Dolan, 1987; Likharev, 1987). The transistor can be viewed as a generalization of the SEB and consists of a mesoscopic conducting island connected by two tunnel junctions to the bulk electrodes that provide the transport voltage $V$ across it. The island is also coupled capacitively to the source of the gate voltage $V_g$ which controls the flow of current $I$ through the transistor between the two electrodes. An equivalent circuit of the transistor is shown in Fig. 5, and an example of its geometric structure can be seen in the upper right corner of Fig. 2(a), where it is used to measure the charge state of the SEB. The charge configuration of the transistor is characterized simply by the number $n$ of extra electrons on its central island, and accordingly, the rate equation describing its dynamics is

$$\dot{p}(n) = \sum_{j,z}[p(n \pm 1)\Gamma_j^{(z)}(n \pm 1) - p(n)\Gamma_j^{(z)}(n)],$$

(10)

where $p(n)$ is the probability distribution of the charge $en$ on the central island of the transistor, and the rates $\Gamma_j^{(z)}(n)$ describe the tunneling processes in junction $j$ with the tunnel conductance $G_j$ out of the state $n$ in the direction that increases (+) or decreases (−) by 1. The rates are given by Eq. (8) or (9), or their generalizations, depending on the nature of the transistor electrodes. They depend on the indices of $\Gamma$’s in Eq. (10) through the change $E$ of the charging energy $U$ of the transistor, which is a function of all these indices. The transistor energy $U$ consists of two parts, one that coincides with the charging energy (3) of the SEB in which $C_\Sigma = C_1 + C_2 + C_g$, and the other $U_V$ that is created by the transport voltage $V$:

$$U_V = -enV - enV(C_2 + C_g/2)/C_\Sigma.$$

(11)

Here $N$ is the number of electrons that have been transferred through the transistor. Both the dc current $I$ through the transistor (Averin and Likharev, 1991) and the current noise (Korotkov, 1994) can be calculated starting from Eq. (10).

The main physical property of the transistor transport characteristics is that they depend periodically on the gate voltage, in particular, $I(n_g + 1) = I(n_g)$. This dependence of the transistor current on the charge $en_g$ induced on its central island makes the SET a charge detector, with subelectron sensitivity approaching $(10^{-5} - 10^{-6}) e/Hz^{1/2}$ (Zimmerli et al., 1992; Krupenin, 1998; Roschier et al., 2001). As a result, the SET is the most standard charge detector for measurements of, e.g., individual electron dynamics in other single-electron structures [cf. Fig. 2(a)].

The hybrid SINIS of normal metal–insulator–superconductor–insulator–normal metal (NISIN) transistors have an additional important feature that distinguishes them from the SETs with normal electrodes. They provide the possibility to realize the regime of the quantized current [Eq. (4)], when driven by an ac gate voltage $V_g(t)$ of frequency $f$ (Pekola et al., 2008). This property of the hybrid SETs is one of the main topics of this review and is discussed in detail below.

The basic expression (6) for the tunneling rates assumes that the electrodes of the tunnel junction are in equilibrium at temperature $T$, with the implied assumption that this temperature coincides with fixed temperature of the whole sample. Since each electron tunneling event deposits an amount of heat $\Delta U$ into the electron system of the electrodes, this condition requires that the relaxation processes in the electrodes are sufficiently effective to maintain the equilibrium. The relaxation rates decrease rapidly with decreasing temperature, e.g., proportionally to $T^5$ for electron-phonon relaxation in an ordinary metal; see, e.g., Giazotto et al. (2006). This makes the relaxation insufficient and causes the overheating effects to appear at some low temperature, in practice around 0.1 K. Therefore, the overheating sets a lower limit to the effective temperature of the transitions, in this way limiting the accuracy of control over the individual electron transport.

One more assumption underlying Eq. (9) for the tunneling rate in SIN junctions is that the electron distribution function is given by the Fermi function $f(e)$. As known from statistical mechanics, even in equilibrium, this requires that the total effective number of particles that participate in forming this distribution is large. In normal-metal islands, this requirement is satisfied at temperatures much larger than the single-particle level spacing, $T \gg \delta E/k_B$, as is the case for practically all metallic tunnel junctions. In contrast to this, in superconducting islands, this condition can be violated at temperatures below the superconducting energy gap, $T \ll \Delta/k_B$, when the total number of the quasiparticle excitations in the electrode is no longer large. The temperature scale of the onset of this “individual quasiparticle” regime can be estimated from $\int_\Delta^\infty d\epsilon f(\epsilon)\nu(\epsilon)/\delta E \sim 1$. The main qualitative feature of this regime is the sensitivity of the electron transport properties of a superconducting island to the parity of the total number of electrons on it (Averin and Nazarov, 1992b; Tuominen et al., 1992). In particular, the charge tunneling rate (9) in the SIN junction should be modified in this case into the rates of tunneling of individual quasiparticles. For $T \gg \delta E/k_B$, these rates are still determined by the same
average \( \langle |T_{kp}|^2 \rangle \) over many single-particle states of the squares of the tunneling amplitudes (5) which gives the tunnel conductance \( G_T \), and therefore can be expressed through \( G_T \). In the regime of “strong” parity effects, when \( T \ll T^* = \Delta/k_B \ln N_{eff} \), where \( N_{eff} = (2\pi k_B T \Delta)^{1/2}/\delta E \) is the effective number of states for the quasiparticle excitations (Tuominen et al., 1992), an ideal BCS superconductor should reach the state with no quasiparticles, if the total number \( N_0 \) of electrons in the superconductor is even, and precisely one unpaired quasiparticle if \( N_0 \) is odd. Although many nonequilibrium processes in realistic superconductors lead to the creation of a finite density of quasiparticle excitations which do not “freeze out” at low temperatures [see, e.g., de Visser et al. (2011)], one can realize the situation with the number of quasiparticles controlled as in an ideal BCS superconductor, as, e.g., in Tuominen et al. (1992), Lafarge et al. (1993), and Saiara, Kemppinen et al. (2012). In this regime, the rates of sequential charge tunneling between the normal-metal electrode and a superconducting island depend on the parity of the total number \( N_0 \) of electrons in the island (Schön and Zaikin, 1994; Maisi et al., 2012). For \( T \ll T^* \), the tunneling rates both to and from the island are dominated by the one quasiparticle that exists on the island for odd \( N_0 \). When this quasiparticle is equilibrated to the edge of the quasiparticle spectrum at energy \( \Delta \), the rates of tunneling to and from the island (i.e., increasing and decreasing the charge \( eN \) of the island) coincide, and for \( |E| \ll \Delta \) are independent of the electrostatic energy change \( E \):

\[
\Gamma_{odd} = \frac{G_T \delta E}{4e^2}. \tag{12}
\]

For even \( N_0 \), when there are no quasiparticles on the island, tunneling necessarily involves the process of creation of a quasiparticle, making the tunneling rates dependent on the energy change \( E \):

\[
\Gamma_{even}(E) = \Gamma_{odd} N_{eff} e^{-(\Delta^2 - E^2)/k_B T}. \tag{13}
\]

In the hybrid superconductor and normal-metal structures, these tunneling rates determine the electron transport properties through a rate equation similar to Eq. (10).

C. Cotunneling, Andreev reflection, and other higher-order processes

The sequential tunneling discussed previously represents only the first nonvanishing order of the perturbation theory in the tunnel Hamiltonian \( H_T \) (5). In the strong Coulomb-blockade regime \( G_T \ll 1/R_K \), this approximation provides an excellent starting point for the description of electron transport, accounting quantitatively for the main observed properties of these structures. However, a more detailed picture of the transport should also include the tunneling processes of higher order in \( H_T \), which involve transfers of more than one electron in one or several tunnel junctions. Although for \( G_T \ll 1/R_K \) the rates of these more complex multistep electron “cotunneling” processes are small in comparison with the rates of the single-step sequential electron tunneling, they are frequently important either because they provide the only energetically allowed transport mechanism or because they limit the accuracy of control of the basic sequential single-electron transitions. The simplest example of the cotunneling is the current leakage in the SET in the CB regime (Averin and Odintsov, 1989; Geerligs, Averin, and Mooij, 1990), where the bias voltage \( V \) is smaller than the CB threshold and any single-step electron transfer that changes the charge \( en \) on the transistor island by \( \pm e \) (see Fig. 5) would increase the charging energy (3) and is suppressed. In this regime, only the two-step cotunneling process that consists of electron transfers in both junctions of the transistor in the same direction gains the bias energy (11). It achieves this by changing the number \( N \) of electrons transported through the transistor by 1 without changing the charge \( en \) on the island. Qualitatively, this process represents a quantum tunneling through the energy barrier created by the charging energy. Because of the discrete nature of charge transfer in each step of the cotunneling, its rate is not suppressed exponentially as for the usual quantum tunneling, and is smaller only by a factor \( G_T R_K \ll 1 \) than the rate of sequential tunneling processes.

In a hybrid SIN junction, in addition to the charging energy, the superconducting energy gap \( \Delta \) provides an extra energy barrier to tunneling of individual electrons, suppressing the sequential tunneling rate (9) at low temperatures \( T \ll \Delta/k_B \). The gap \( \Delta \) exists only for individual electrons, while pairs of electrons with zero total energy and momentum can enter a superconductor as a Cooper pair, in the process called Andreev reflection (Andreev, 1964). In tunnel junctions, AR can be described similarly to the cotunneling, as a perturbative two-step tunneling process, in which the transfer of the first electron is virtual and only the second electron transfer makes the process energetically favorable and real. Quantitatively, the rates of such multistep transitions can be determined through their higher-order transition amplitudes constructed according to the standard rules of perturbation theory [see, e.g., Landau and Lifshitz (1980b)]. For instance, in the simplest example of a two-step AR process in a hybrid single-electron box, the elementary amplitude \( A(\epsilon_k, \epsilon_l) \) of the process that takes two electrons in the normal electrode with energies \( \epsilon_k \) and \( \epsilon_l \) and transfers them into the superconductor as a Cooper pair can be written as

\[
A(\epsilon_k, \epsilon_l) = \sum_p u_p v_p T_{kp} T_{lp} \left( \frac{1}{\Omega_p + \epsilon_l - \epsilon_k} + \frac{1}{\Omega_p + \epsilon_i - \epsilon_l} \right). \tag{14}
\]

The two-step process goes through an intermediate state obtained as a result of the first step of the process. The intermediate states differ by the order of transfer of the two electrons and by the single-particle state of energy \( \epsilon_p \) in the superconductor in which the virtual quasiparticle with excitation energy \( \Omega_p = (\Delta^2 + \epsilon_p^2)^{1/2} \) is created. In addition to \( \Omega_p \), the energy of the intermediate state includes the charging energy barrier \( E_i \) to the transfer of one electron from the normal electrode to the superconductor. The standard BCS factors \( v_p = [(1 - \epsilon_p/\Omega_p)/2]^{1/2} \) and \( u_p = [(1 + \epsilon_p/\Omega_p)/2]^{1/2} \) enter Eq. (14) because \( u_p \) is the amplitude of state \( p \) being empty in the BCS ground state, thus allowing the first electron transfer, while \( u_p \) is the overlap of the doubly occupied orbital state \( p \) with the BCS ground state, which gives the amplitude of return to the ground state after the second electron transfer. Since no
trace of the intermediate states is left in the final state obtained after the whole AR process is complete, they should be summed over coherently, at the level of the amplitude \( A(\epsilon_k, \epsilon_l) \). By contrast, the initial states \( \epsilon_k \) and \( \epsilon_l \) of the electrons in the transition are left empty in the final state and can be used to distinguish between different transition processes. This means that they should be summed over incoherently, in the expression for the tunneling rate \( \Gamma_{AR} \). At small temperatures \( k_B T \ll \Delta \), one can neglect thermal excitations in the superconductor, obtaining the total AR tunneling rate as

\[
\Gamma_{AR} = \frac{2\pi}{\hbar} \sum_{k,l} |A(\epsilon_k, \epsilon_l)|^2 f(\epsilon_k) f(\epsilon_l) \delta(\epsilon_k + \epsilon_l + E),
\]

(15)

where \( E \) is the electrostatic energy change due to the complete AR tunneling process. The sum over all states \( \rho \) in the superconductor in Eq. (14) implies that the contribution of the individual quasiparticles [which is important in the parity-dependent transition rates (12) and (13)] is negligible in the amplitude \( A \), and individual quasiparticles affect \( \Gamma_{AR} \) only through the change of the charging conditions for tunneling.

The result of the summation over different single-particle states in Eqs. (14) and (15) depends on the detailed structure of the SIN junction. For instance, the quadratic dependence only to the total tunnel conductance \( G_T \) is limited by the angular dependence of the charging conditions for tunneling. The summing over different channels can be done directly and gives the effective number \( \mathcal{N} \) of the channels, as discussed in Sec. II.B, is limited by the angular dependence of the barrier transmission probabilities [see, e.g., Averin and Bardas (1995)]. In the ballistic approximation, Eqs. (14) and (15) give for the AR tunneling rate (Hekking et al., 1993; Averin and Pekola, 2008)

\[
\Gamma_{AR} = \frac{\hbar G^2_{12} \Delta^2}{16 \pi e^4 \mathcal{N}} \int d\epsilon f(-E/2)f(E/2) \left[ \sum \left| a(\pm \epsilon - E_l - E/2) \right|^2 \right],
\]

(16)

where

\[
a(\epsilon) = (\epsilon^2 - \Delta^2)^{-1/2} \ln \left[ \frac{\Delta - \epsilon + (\epsilon^2 - \Delta^2)^{1/2}}{\Delta - \epsilon - (\epsilon^2 - \Delta^2)^{1/2}} \right].
\]

Equation (16) is well defined if the relevant energies in the amplitude \( a(\epsilon) \) do not approach the edge of the superconducting energy gap \( \epsilon = \Delta \), which gives a logarithmically divergent contribution to \( \Gamma_{AR} \). This singularity can be smeared by many mechanisms, e.g., the nonuniformity of the gap \( \Delta \) or finite transmission probability of the barrier. In the single-electron tunneling regime, one of the main broadening mechanisms should be the lifetime of the intermediate charge state in the AR process and can be accounted for by replacing in Eq. (14) the energy \( E_i \) with \( E_i - i\gamma/2 \), where \( \gamma \) is the rate of sequential lowest-order tunneling out of the intermediate charge configuration.

Experimentally, individual AR processes can be observed directly in the time domain in the hybrid SEB (Maisi et al., 2011). This observation allows one to extract the rates of AR tunneling shown in Fig. 6 as a function of the normalized gate voltage \( n_g \) which determines the energies \( E_i \) and \( E \) of the transition. Figure 6 also shows the theoretical fit based on Eq. (16). One can see that Eq. (16) describes very well the shape of the curves. The fit requires, however, a considerably smaller (roughly by a factor 15) effective number \( \mathcal{N} \) of the transport channels to describe stronger AR tunneling processes. In practice, the fact that the magnitude of AR tunneling rates is larger by roughly a factor of 10 than the theoretical expectation for a given tunnel conductance \( G_T \) is a usual feature of the tunnel junctions [see also, e.g., Pothier et al. (1994) and Greibe et al. (2011)], and in principle can be qualitatively accounted for by the variation of the barrier thickness over the junction area. Unfortunately, there is so far no quantitative experimental or theoretical evidence that the barrier nonuniformity is indeed the reason for the discrepancy between the magnitude of the lowest-order and AR tunneling.

In the structures without superconducting electrodes, multistep electron transitions, in contrast to the AR processes, involve electron transfers in different directions and/or across different tunnel junctions, since a transfer of the two electrons in the same junction and the same direction cannot make the process energetically favorable in the absence of the pairing gap \( \Delta \). In the simplest example of the normal metal–insulator–normal metal–insulator–normal metal (NININ) SET, the two-step cotunneling process in the CB regime
discussed qualitatively at the beginning of this section consists of two-electron transfers across the two junctions of the transistor. Quantitatively, the rate of this process is dominated by the inelastic contribution \( \Gamma_{\text{in}} \), in which the single-particle states \( \epsilon_p \) and \( \epsilon_q \) of electrons in the central island of the transistor involved in the transfers are different (Averin and Odintsov, 1989). As a result, the occupation factors of these states are changed, i.e., electron-hole excitations are created after the process is completed, a fact that implies that contributions to the tunneling rates from different \( \epsilon_p \) and \( \epsilon_q \) should be summed over incoherently. The elementary amplitude \( A \) of this process consists then only of a sum over the two possibilities, one in which an electron is first transferred onto the island increasing the charging energy of the intermediate state by \( E^{(+)} \), and the other, in which an electron tunnels first from the island still increasing the charging energy but by a different amount \( E^{(-)} \):

\[
A = T_{k\ell} T_{q\ell} \left( \frac{1}{\epsilon_p + E^{(+)} - \epsilon_k} + \frac{1}{\epsilon_l + E^{(-)} - \epsilon_q} \right).
\] (17)

The total rate \( \Gamma_{\text{in}} \) is given then by the sum of \( |A|^2 \) over all single-particle states involved with the appropriate equilibrium occupation factors and can be expressed directly through the junction conductances as

\[
\Gamma_{\text{in}}(E) = \frac{hG_1G_2}{2\pi e^2} \int d\epsilon_p d\epsilon_q d\epsilon_{jf} f(\epsilon_j) [1 - f(\epsilon_p)]
\times \left[ 1 - f(\epsilon_j) \right] \delta(E - \epsilon_p - \epsilon_k - \epsilon_j - \epsilon_q)
\times \left[ \frac{1}{\epsilon_p + E^{(+)} - \epsilon_k} + \frac{1}{\epsilon_l + E^{(-)} - \epsilon_q} \right]^2,
\] (18)

where \( E = eV \) is the energy gain due to the transfer of electron charge \( e \) through both junctions of the transistor (see Fig. 5). Equation (18) shows explicitly that the second-order electron cotunneling that involves one virtual intermediate stage is indeed smaller than the rate (6) of sequential single-electron tunneling roughly by a factor \( R_K G_T \ll 1 \). The derivation above also makes it clear that the rate of the multistep electron transitions that go through a virtual intermediate stages with larger \( n \) would be suppressed much more strongly by a factor \( (R_K G_T)^n \).

If the energy gain \( E \) and thermal energy \( k_B T \) are smaller than the charging energy barriers \( E^{(\pm)} \), Eq. (18) for the inelastic cotunneling rate can be simplified to

\[
\Gamma_{\text{in}}(E) = \frac{hG_1G_2}{12\pi e^2} \left( \frac{1}{E^{(+)}_j} + \frac{1}{E^{(-)}_j} \right)^2 E^{2} \left[ 1 - e^{-2\pi k_B T} \right]^2.
\] (19)

This equation shows that, as a result of creation of excitations in the process of inelastic cotunneling, its rate decreases rapidly with decreasing \( E \) and \( T \). At very low energies, the process of cotunneling in the NININ transistor will be dominated by the elastic contribution, in which an electron is added to and removed from the same single-particle state of the transistor island, without creating excitations on the island. Because of the restriction on the involved single-particle states, the rate of such elastic contribution contains an additional factor on the order of \( \delta E/E_C \) (Averin and Nazarov, 1990) and can win over \( \Gamma_{\text{in}} \) only at very low temperatures, practically negligible for the structures based on the micrometer-scale metallic islands considered in this review.

The approach to multistep electron transitions in the single-electron structures illustrated in this section with the examples of Andreev reflection and electron cotunneling can be directly extended to other higher-order tunneling processes, e.g., cotunneling of a Cooper pair and an electron (Averin and Pekola, 2008), which together with Andreev reflection and electron cotunneling limit in general the accuracy of control over sequential single-electron transitions.

D. Coulomb blockade of Cooper-pair tunneling

In contrast to the tunneling processes considered previously, which involve electrons in the normal-metal electrodes, tunneling of Cooper pairs in a junction between two superconductors is intrinsically a dissipationless process (Josephson, 1962). As such, it should not be characterized by a tunneling rate but a tunneling amplitude. A quantitative form of the corresponding term in the junction Hamiltonian can be written most directly at low energies \( k_B T, E_C \ll \Delta \), when the quasiparticles cannot be excited in the superconducting electrodes of the junction, and the tunneling of the Cooper pair is well separated from the tunneling of individual electrons. In this regime, a superconductor can be thought of as a Bose-Einstein condensate of a “mesoscopically” large number of Cooper pairs which all occupy one quantum state. Transfer of one pair between two such condensates in the electrodes of a tunnel junction does not have any non-negligible effects on the condensates apart from changing the charge \( Q = 2e n \) on the junction capacitance by \( \pm 2e \). Therefore, the part of the Hamiltonian describing the tunneling of Cooper pairs should contain the terms accounting for the changes of the charge \( Q \). Using the standard notation \(-E_j/2\) for the amplitude of Cooper-pair tunneling and including the charging energy (3), one obtains the Hamiltonian of a superconductor-insulator-superconductor (SIS) tunnel junction or, equivalently, Cooper-pair box in the following form (Averin, Zorin, and Likharev, 1985; Büttiker, 1987):

\[
H = 4E_C(n - n_j)^2 - \frac{E_j}{2} \sum \{n(n \pm 1)\}.
\] (20)

Here \( n \) is the number of Cooper pairs charging the total junction capacitance, and \( n_j \) is the continuous (e.g., gate-voltage-induced) charge on this capacitance normalized now to the Cooper-pair charge \( 2e \). Similarly to the sequential tunneling rates, the Cooper-pair tunneling amplitude in the Hamiltonian (20) is a macroscopic parameter which receives contributions from all Cooper pairs in the condensate and can be expressed directly through the tunnel conductance \( G_T \) of the junction \( E_j = \pi G_T \Delta/2e \) (Ambegaokar and Baratoff, 1963), in agreement with the simple fact that the amplitude of the two-electron tunneling should have the same dependence on the barrier transparency as the rate of tunneling of one electron. In the situation of the junction (20) realized with the actual Bose-Einstein condensates of atoms, such a “Bose-Josephson junction” can contain a relatively small total number of particles, and the tunnel amplitude varies then with the difference \( n \) of the number of particles in the two
condensates; see, e.g., Fölling et al. (2007), Averin et al. (2008), and Cheinet et al. (2008).

The dependence on the ground state of the Hamiltonian (20) on the induced charge \( n_g \) allows for qualitatively similar control of the individual Cooper pairs as for individual electrons in the normal-state SEB discussed in Sec. II.B. If \( E_I \ll E_C \), precisely one Cooper pair is transferred through the junction, changing \( n \) by \( \pm 1 \), whenever \( n_g \) passes adiabatically through a degeneracy point \( n_g = 1/2 \) modulo an integer. This leads to the same staircase-like dependence \( n(n_g) \) as in the normal case, but with each step corresponding to the transfer of one more Cooper pair with the increase of \( n_g \) by 1. The main new element of the superconducting situation is that the SIS junction is intrinsically a coherent quantum system without dissipation, and if extrinsic sources of decoherence can be made sufficiently weak, should exhibit reversible dynamics of a simple quantum system. For instance, close to the degeneracy point \( n_g = n + 1/2 \) the two charge states with the same electrostatic energy, \( n \) and \( n + 1 \), are coupled by coherent quantum mechanical tunneling of a Cooper pair, and the junction behaves as a very basic quantum two-state system (Bouchiat et al., 1998; Nakamura, Pashkin, and Tsai, 1999). Such two-state dynamics and general coherent quantum dynamics of the Hamiltonian (20) serve as the basis for the development of superconducting quantum information devices; for reviews, see, e.g., Averin (2000) and Makhlín, Schön, and Shnirman (2001).

Superconducting junctions also exhibit the dynamics similar to the single-electron tunneling oscillations. If the induced charge \( n_g \) grows in time at a constant rate, so that effectively a dc displacement current \( I = 2e\dot{n}_g \) is injected into the junction, Cooper pairs are transferred through it in a correlated manner, one by one, giving rise to the “Bloch” oscillations (Averin, Zorin, and Likharev, 1985) of voltage across the junction, with frequency \( f \) related to the current \( I \):

\[
I = 2ef. \tag{21}
\]

The Hamiltonian (20) and its extensions to multijunction systems can be used to design time-dependent periodic dynamics with frequency \( f \) which transfer precisely one Cooper pair per period and therefore produce a dc current quantized according to Eq. (21). Although the system dynamics employed for such Cooper-pair pumping can be of different kinds [see, e.g., Hoehne et al. (2012)], the most typical is the adiabatic dynamics (Geerligs et al., 1991), in which the pumped charge is related (Pekola et al., 1999; Aunola and Toppari, 2003; Möttönen, Vartiainen, and Pekola, 2008) to Berry’s phase or (Faoro, Siewert, and Fazio, 2003) its non-Abelian extensions.

E. Single-electron tunneling in semiconductor structures

One of the main features of metallic conductors used in the discussion of single-electron tunneling is the large density of free electrons in them, characterized quantitatively by the average electron-electron distance \( r \) that is not much larger than the Bohr’s radius \( a_0 \). In this regime, the electrostatic screening length \( \lambda \approx (ra_0)^{1/2} \) at low energies is also small, i.e., comparable to \( r \). This fact has several simplifying consequences for the discussion of single-electron tunneling. Most importantly, because of the strong screening, electrons are effectively noninteracting inside conductors at low energies relevant to the Coulomb-blockade transport. For normal metals, this makes it possible to describe the tunnel junction electrodes as reservoirs of noninteracting electrons—the model adopted above for the discussion of tunneling. (For superconducting electrodes, only superconducting pairing correlations are important.) Another consequence of a short, on the order of interatomic distance, screening length is that for all practical electrodes large on this scale the electron-electron interaction energy due to charging of the conductor as a whole is independent of the electron state inside the conductor and can be accurately described by macroscopic capacitances as was done in Eq. (1).

In the case of semiconductor single-electron structures based on quantum dots formed in two-dimensional conducting layers (see Sec. III.C for a brief discussion of the typical structures), the dot parameters, including carrier (usually, electron) concentration in the dot, can be controlled through external bias. Despite this control, and variability of the carrier concentration with the fabrication parameters, one can take \( n \approx 10^{12} \text{cm}^{-2} \) as a typical value of concentration, which corresponds to \( r \sim 10 \text{nm} \). Although this electron-electron distance is considerably larger than in a good metal, the Bohr radius \( a_0 = 4\pi\varepsilon_0\varepsilon_r\hbar^2/e^2m \) is also much larger in the semiconductors, e.g., gallium arsenide or silicon, used to fabricate quantum dots, because of the dielectric constant \( \varepsilon \sim 10 \) and effective mass \( m \) smaller than the free electron mass. This keeps the parameter \( r \approx r/a_0 \) which determines the strength of interaction effects in an electron gas [see, e.g., Mahan (1990)] in the same weak-interaction range \( r \sim 1 \) and makes it reasonable to describe a quantum dot in the same approximation as used previously for metallic islands: noninteracting electron gas inside the dot with the electron-electron interaction giving rise to the charging energy \( U(n) \) that depends only on the total number \( n \) of electrons in the dot and can be expressed through the constant dot capacitance \( C \) as in Eq. (3). Since quantum confinement of effectively noninteracting electrons inside the dot potential produces in addition an energy spectrum \( \epsilon_k \) of the single-particle states, the total dot Hamiltonian is then

\[
H = U(n) + \sum_k \epsilon_k c_k^\dagger c_k, \quad n = \sum_k c_k^\dagger c_k. \tag{22}
\]

Although this model of a quantum dot is the same as for the normal-metal islands, an important difference between the two situations is created by the difference in the absolute values of characteristic length scales \( r \) and \( a_0 \) which are much larger for quantum dots. Because of this, already relatively large dots with characteristic dimensions \( d \sim 100 \text{ nm} \) can contain a small total number of electrons, starting with \( n = 1 \), and have the single-particle level spacing \( \delta E \) comparable to the charging energy \( E_C \). This difference has two important consequences which make a quantitative description of single-electron transport in quantum dots in general more involved than in metallic structures. The larger level spacing \( \delta E \) reduces the number of the single-particle energy levels participating in the single-electron tunneling transitions through the dot making the nonequilibrium effects in energy distribution of electrons in the dot more prominent in the...
regime of nonlinear transport. The small number of levels involved in transport also creates statistical correlations between occupation factors of different levels even for electrons that are effectively noninteracting inside the dot, as in the Hamiltonian (22). Combined, all this means that in contrast to metallic islands, the dynamics of the total charge \( en \) on the dot is not decoupled from the electron dynamics in the dot and depends more strongly on the relaxation processes and the structure of the energy spectrum \( \varepsilon_k \), both of which are sensitive to the effect of disorder on the dot confining potential.

For weak tunneling, a quantitative description of single-electron transport through a quantum dot, in a structure similar to a single-electron transistor (see Fig. 3), is based on a kinetic equation similar to Eq. (10) in the metallic regime (Averin, Korotkov, and Likharev, 1991; Beenakker, 1991). As qualitatively discussed previously, in the case of a dot, this equation cannot be formulated directly in terms of the probability distribution \( p(n) \) of the charge on the dot, but requires the probability \( p(n; k_1, \ldots, k_n) \) that a given set of \( n \) single-particle states of the dot \( k_1, \ldots, k_n \) is occupied by electrons. Since we will not be dealing explicitly with spin-related phenomena, it is assumed for notational simplicity that \( k \) includes the spin index of the single-particle states. Also note that the order of indices of the occupied states in the argument of \( p(n; k_1, \ldots, k_n) \) is irrelevant.

Expressed through this probability, the kinetic equation reads

\[
\dot{p}(n; k_1, \ldots, k_n) = S_{\text{tun}} + S_{\text{rel}},
\]

where \( S_{\text{tun}} \) and \( S_{\text{rel}} \) are, respectively, the probability flows due to electron tunneling between the dot and external electrodes which changes the charge \( en \) of the dot by \( \pm e \), and electron transitions inside the dot (without changing \( n \)) due to electron-phonon or residual electron-electron interactions which lead to thermalization and energy relaxation of electrons in the dot. The terms \( S_{\text{tun}} \) are expressed through the rates of tunneling between state \( k \) in the dot and reservoir \( j \), which similarly to metallic tunnel junctions (5) can be written as

\[
\Gamma_{k,j} = 2\pi|T_{k,j}|^2\rho_j / h,
\]

where \( \rho_j \) denotes averaging over the states \( p \) in reservoir \( j \) which have density \( \rho_j \). In terms of these rates,

\[
S_{\text{tun}} = \sum_j \sum_{k \in [k]} \Gamma_{k,j} (p(n + 1; k_1, \ldots, k_n, k) - p(n; k_1, \ldots, k_n)) \times (1 - f(\varepsilon_k + E_{j,n+1})) - p(n; k_1, \ldots, k_n)
\]

\[
\times f(\varepsilon_k + E_{j,n+1}) + \sum_{k \in [k]} \Gamma_{k,j} (p(n - 1; k) - k)
\]

\[
\times f(\varepsilon_k + E_{j,n}) - p(n; k_1, \ldots, k_n)
\]

\[
\times (1 - f(\varepsilon_k + E_{j,n})),
\]

where \( E_{j,n} \) is the change of the energy \( U_{\text{tot}} = U(n) + U_V \) which consists of the charging energy \( U(n) \) [Eq. (3)] and the bias energy \( U_V \) [Eq. (11)], due to transfer of one electron into the \( j \)th electrode from the dot carrying charge \( en \).

The relaxation term \( S_{\text{rel}} \) in the kinetic equation can be written similarly. For instance, in the case of electron-phonon relaxation

\[
S_{\text{rel}} = \sum_k \sum_k \left\{ \gamma(\varepsilon_k - \varepsilon_j) p(n; \{k\} - k, l) - \gamma(\varepsilon_k - \varepsilon_j) p(n; k_1, \ldots, k_n) \right\},
\]

where \( \gamma(e)'s \) are the rates of the phonon-induced transitions between the electron states in the dot. These equations show that the general nonlinear single-electron transport through the dot depends quantitatively on its microscopic structure, in particular, energy relaxation rates. In general, sensitivity of the single-particle level structure of the dot to its geometric shape and the details of the confining potential [for a review, see, e.g., Reiman and Manninen (2002)] turns precise quantitative characteristics of the dot transport almost into the fingerprints of an individual quantum dot, even in the simplest situation of effectively noninteracting electrons in the dot.

In addition, at low electron densities, electron-electron correlations inside the dot can become important, leading to formation of a finite Wigner crystal in effectively both one-dimensional and two-dimensional dots; see, e.g., Häusler and Kramer (1993) and Filinov, Bonitz, and Lozovik (2001) and references therein. From the point of view of electron transport, the main characteristic feature of such a correlated electron state is additional energy dependence of the electron tunneling rates into the dot, with tunneling suppressed by correlations at low energies (Kane and Fisher, 1992; Averin and Nazarov, 1993; Matveev and Glazman, 1993).

Despite stronger influence of internal microscopic physics on the quantum-dot transport, the charging energy \( E_C \) associated with individual electrons still remains typically the dominant energy in comparison, e.g., to the level spacing \( \delta E \), in the case of quantum dots as well. Because of this, semiconductor quantum dots allow for qualitatively similar manipulation of individual electrons as do metallic structures. Moreover, semiconductor structures provide an additional flexibility in this respect, in that the islands and barriers defining the quantum dots can be tuned or even formed by applying external voltages to gate electrodes; see Sec. III.C for details. This is in contrast to metallic systems which are usually defined solely by the conducting and insulating regions of the fabricated structure. An example of the single-electron control in a GaAs/AlGaAs semiconductor structure, two quantum dots monitored with a quantum-point contact (QPC) operated as a charge detector (Küng et al., 2012), is shown in Fig. 7. In this system, the QPC detector distinguishes different charge states of the two dots and allows one to detect transitions of individual electrons between the dots and to or from the source and drain electrodes. The observed charge dynamics as seen, e.g., in Fig. 7(c), resembles that in the metallic SEB shown in Fig. 2. Such correlated single-electron transitions in semiconductor dots, combined with the possibility of the gate-voltage control of tunnel barriers, make it possible to pump electrons by direct periodic modulation of the two barriers of a dot (Kouwenhoven et al., 1991a).

Attempts to increase the magnitude of pumped current lead naturally to the situation when the barriers become nearly completely suppressed, and electrons can cross them not only by quantum tunneling through the barrier but also by classical motion over the barrier. Coulomb-blockade correlations among different charge states survive in this regime (Zimmerman et al., 2004) which should be, in particular,
relevant for the “dynamic” quantum dots formed and destroyed by rapidly changing gate voltages; see Blumenthal et al. (2007) and Fricke et al. (2013) and references therein. The process of electron transfer through the rapidly created or destroyed barriers can lead to a stochastic uncertainty in created charge state of the dot, which is described with a “decay-cascade” model (Kashcheyevs and Kaestner, 2010). Because of the uncertainty in microscopic dynamics underlying the electron transfer in quantum dots discussed above, the precise limits which the varying barriers impose on the accuracy of electron manipulation in dynamic quantum dots are still not fully understood (Zimmerman et al., 2004; Blumenthal et al., 2007; Fujiwara, Nishiguchi, and Ono, 2008; Kashcheyevs and Timoshenko, 2012; Lin and Zhang, 2012).

F. Influence of environment on tunneling

Tunneling in small junctions is influenced by the electromagnetic environment. The tunneling rates are modified by photon absorption or emission; Fig. 8 schematically depicts a process where the tunneling rate in a generic junction is enhanced by absorption of a photon from the environment.

The general theoretical framework of how this happens was put forward in seminal works by Devoret et al. (1990) and Girvin et al. (1990), and later expanded by Ingold and Nazarov (1992). The golden-rule-type tunneling rates discussed in the earlier sections get modified as

$$
\Gamma = \frac{1}{\hbar e^2 R_T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dE dE' \nu_i(E-\delta E) \nu_j(E'-\delta E') f_i(E-\delta E) f_j(E'-\delta E') \times [1-f_2(E')] \rho(E-E'),
$$

where $\nu_i(E), i = 1, 2$ are the normalized DOSs in the two electrodes, $f_i(E)$ are the corresponding energy distributions in the electrodes, and $\delta E$ is the energy cost in the tunneling event. The function $P(E)$ can be calculated as the following transformation using the phase-phase correlation function $J(t)$:

$$
P(E) = \frac{1}{2\pi \hbar} \int_{-\infty}^{\infty} \exp \left( J(t) + \frac{i}{\hbar} E t \right) dt.
$$

By modeling the environment by a frequency $\omega/2\pi$ dependent impedance $Z(\omega)$ in thermal equilibrium at temperature $T_{env}$, one obtains

$$
J(t) = \frac{1}{2} \int_{0}^{\infty} \frac{d\omega}{\omega} \text{Re}[Z(\omega)] \left\{ \coth \left( \frac{\hbar \omega}{2k_B T_{env}} \right) \left( \cos \omega t - 1 \right) - i \sin \omega t \right\},
$$

where $R_k = h/e^2$ is the resistance quantum.

Often one can assume that the unintentional environment can be modeled as a wideband dissipative source in the form of an $RC$ circuit. For a purely resistive and capacitive environment

$$
\text{Re}[Z(\omega)] = R/[1 + (\omega RC)^2],
$$

where $R$ is the resistance of the environment and $C$ is the total capacitance including the junction capacitance and parallel shunt capacitors. This rather simple model has been successfully applied to explain several experimental observations; see, e.g., Martinis and Nahum (1993) and Hergenrother et al. (1995). For a system with intentionally enhanced capacitance, it could be used to account for experimental improvement of the characteristics of a normal metal–insulator-superconductor (NIS) junction and of a single-electron turnstile (Pekola et al., 2010). Further improvements were obtained by Saira et al. (2010) and Saira, Kemppinen (2012); see Sec. III.B. We show in Fig. 9 an NIS junction and its current-voltage characteristics under different experimental conditions.

Focusing on the single-electron sources, the environment has at least two effects to be considered. (i) The coupling of
the blackbody radiation of the hot surrounding environment can induce photon-assisted tunneling. (ii) The intentionally fabricated on-chip environment in the immediate vicinity of the single-electron circuit serves as a filter against external noise. Moreover, it can influence the tunneling rates in a way that improves the performance (Zorin et al., 2000; Lotkhov et al., 2001; Bubanja, 2011). Detailed discussion of the error processes in pumps, including those due to coupling to the environment, is given further in this review.

G. Heating of single-electron devices

Single-electron circuits operate optimally at low temperatures. The standard condition is that \( k_B T \ll E_C \), where \( E_C \) is the characteristic charging energy scale. Another condition in superconductor-based devices is that \( k_B T \ll \Delta \), where \( \Delta \) is the energy gap of the superconductor. Since thermal errors in synchronized transfer of electrons are typically proportional to \( e^{-E_C/k_BT} \), where \( E_{CR} \) is the characteristic energy (in the previous examples \( E_C \) or \( \Delta \)), it is obvious that the temperature needs to be more than an order of magnitude below \( E_{CR}/k_B \). At low temperatures the overheating becomes a critical issue (Giazotto et al., 2006). The energy relaxation between the electron system and the bath, typically formed by the phonons, becomes increasingly slow toward low temperatures. Moreover, the various heating rates are typically not scaling down similarly with decreasing temperature.

Heat is injected to the electron system, first and foremost, as Joule heating due to the current in a biased circuit. Other sources of heat include the application of dissipative gate voltages or magnetic flux injection, thermal radiation discussed in Sec. II.F, and shot-noise-induced dissipation by backaction from a charge or current detector. The steady-state temperature of the electron system is determined by the balance between the input powers and the heat currents via different relaxation channels. The injected energy relaxes to phonons via electron-phonon relaxation, to the leads by heat transport through the tunnel junctions, and by radiation to other dissipative elements in the cold circuit. We discuss these processes in more detail.

\textit{Joule heating and cooling:} In a biased circuit, the total Joule power is \( P = IV \), where \( V \) is the overall voltage and \( I \) is the current. This power can, however, be distributed very unevenly in the different parts of the circuit: in an extreme example, some parts may cool down whereas the others are heavily overheated. We now focus on dissipation in biased tunnel junctions. The basic example is a tunneling process in a junction between two conductors with essentially constant density of states, which is the case presented by normal metals. At the finite bias voltage \( V \) the tunneling electron leaves behind a holelike excitation and it creates an excited electron in the other electrode, i.e., both electrodes tend to heat up. Quantitatively we can write the expression of the power deposited in the, say, right electrode as

\[
P_R = \frac{1}{e^2 R_T} \int dE f_L(E - eV) - f_R(E). \tag{30}
\]

Here \( f_{L,R} \) refer to the energy distributions on the left (\( L \)) and right (\( R \)) sides of the junction, respectively. \( P_R = V^2/2R_T \) when \( f_L = f_R \), i.e., when the temperatures of the two sides are equal. By symmetry, or by direct calculation, we can verify that the same amount of power is deposited into the left electrode in this situation. Thus the total power dissipation equals \( P = P_L + P_R = V^2/R_T = IV \), as it should.

If one of the conductors is superconducting, the current-voltage characteristics are nonlinear and the power deposited into each electrode is given by

\[
P_{NS} = \frac{1}{e^2 R_T} \int dE f_{N,S}(E) f_N(E - eV) - f_S(E). \tag{31}
\]
Here $\tilde{E}_N = -(E - eV)$ and $\tilde{E}_S = E$, where $N$ and $S$ now refer to the normal and superconducting leads, respectively. The overall heating is again given by $IV$, but in this case, under bias conditions $eV = \Delta$, $P_N$ on the normal side can become negative [NIS cooling (Giazotto et al., 2006)] and $P_S$ on the superconductor side is always positive, i.e., it is heated up.

Finally, if both sides are superconducting, the current-voltage characteristics are highly nonlinear, but due to symmetry $P_L = P_R = IV/2 = P/2$.

Other heating sources: Overheating of a single-electron circuit can be caused by various other sources. ac gate voltages or ac magnetic fluxes can induce dissipative currents and heating due to dielectric losses, and single-electron electrometry or electrometry by a quantum-point-contact detector can cause effective heating due to the shot-noise backaction coupling to the single-electron circuit, just to mention a few possibilities.

Energy relaxation by conduction to leads: If a difference between the electronic temperatures $T_L$ and $T_R$ of the left and right leads exists, $\Delta T = T_L - T_R$, heat $P_{L-R}$ can flow electronically through the tunnel barrier. In the case of a normal-normal junction, we have

$$P_{L-R} = \frac{1}{e^2 R_T} \int dE \left[ f_L(E) - f_R(E) \right] = \frac{\pi^2 k_B^2}{6e^2 R_T} (T_L^2 - T_R^2),$$

(32)

where in the last step we assume that the junction is not biased. For a small temperature difference $\Delta T$ about the mean $T = (T_L + T_R)/2$ of the two temperatures, we can then write the thermal conductance $G_{th} = P_{L-R}/\Delta T$ of a NIN tunnel junction as

$$G_{th} = \frac{\pi^2 k_B^2 T}{3e^2 R_T},$$

(33)

which is the Wiedemann-Franz law for a conductor with resistance $R_T$. For either an NIS or SIS junction, heat conductance is exponentially small at low temperatures due to $\Delta$. Another mechanism for the heat flow is the diffusion in the leads. It is discussed, in particular, in superconducting leads in Sec. III.B.3.

Electron-phonon relaxation: Electron-phonon relaxation is one of the dominant and in many systems one of the best understood relaxation mechanisms. For a normal-metal conductor with a uniform temperature $T$ that differs from the bath phonon temperature $T_0$, one can write quite generally (Wellstood, Urbina, and Clarke, 1994)

$$P_{e-p} = \Sigma V (T^3 - T_0^3),$$

(34)

where $\Sigma$ is a material constant of the order of $10^9 \text{ W K}^{-5} \text{ m}^{-3}$ (Giazotto et al., 2006), and $V$ is the volume of the conductor. This equation holds amazingly well at subkelvin temperatures for various metals, irrespective of their dimensions. In single-electron devices, we typically consider dissipation in a small Coulomb-blockaded region, whose volume is small, and thus, according to Eq. (34), the coupling to the phonon bath is weak. Because of the small dimensions, one typically assumes a spatially uniform energy distribution on the conductor; moreover, the assumption of overheating with a well-defined electron temperature is also justified quite generally.

In some cases these assumptions are not necessarily valid. An important exception is given by superconductors where energy relaxation via phonon emission becomes extremely weak due to the energy gap. At low temperatures the relaxation is limited by the emission of $2\Delta$ phonons corresponding to the recombination of quasiparticles into Cooper pairs (Rothwarf and Taylor, 1967). In the past few years, several experiments have measured the relaxation rate in this context [see, e.g., Barends et al. (2008)], and the corresponding energy release rate was measured recently by Timofeev, Garcia et al. (2009). According to the latter measurement the recombination-related heat flux is strongly suppressed from that given in Eq. (34), being about 2 orders of magnitude weaker than in the normal state at the temperature $T = 0.3T_C$, where $T_C$ is the critical temperature of aluminum. At even lower temperatures the heat current is further suppressed, eventually exponentially as proportional to $e^{-\Delta/\hbar T}$. Besides recombination, the diffusive heat conduction is also strongly suppressed in a superconductor at $T \ll T_C$. This means that a superconductor is a poor material as a lead of a single-electron source, where nonequilibrium quasiparticles are injected at the rate $f$. The situation can be improved by inserting so-called quasiparticle traps into the circuit, discussed in Sec. III.B.3. Yet a fully superconducting Cooper-pair pump can be dissipationless ideally.

Heating and cooling by radiation: Coupling of a junction to the electromagnetic environment is associated with heat exchange. A hot environment can induce photon-assisted tunneling as discussed in Sec. II.F. The basic concept of radiative heat transport in an electric circuit has been known since the experiments of Johnson (1928) and Nyquist (1928) more than 80 years ago. Electromagnetic radiation on a chip has recently turned out to be an important channel of heat transport at low temperatures (Schmidt, Schoelkopf, and Cleland, 2004; Meschke, Guichard, and Pekola, 2006; Timofeev, Helle et al., 2009). If two resistors $R_1$ and $R_2$ at temperatures $T_1$ and $T_2$ are connected directly to each other in a loop, the heat exchange between them can be modeled by a Langevin-type circuit analysis as indicated in Fig. 10 by the voltage sources producing thermal noise. Assuming an idealized quantum limit, where the circuit transmits all frequencies

![FIG. 10. (a) Radiative heat flow is caused by the photons which carry energy between resistors $R_1$ and $R_2$ at temperatures $T_1$ and $T_2$, respectively. The heat transport can be modeled by having voltage fluctuations $\delta V_i$ as shown in (b). Here we have assumed total transmission. The assumption can be relaxed by adding a nonzero impedance to the loop.](image-url)
up to the thermal cutoff at \( \omega_{th} = k_B T_i / \hbar \), the net heat current between the two resistors is given by

\[
P_\gamma = \frac{R_1 R_2}{(R_1 + R_2)^2} \frac{\pi k_B^2}{12 \hbar} (T_1^2 - T_2^2).
\]

This is an interesting limit which applies for circuits on a chip where the stray capacitances and inductances are small enough such that the circuit low-pass cutoff frequency exceeds \( \omega_{th} \).

Equation (35) has an important limit for the maximum coupling with \( R_1 = R_2 \) and for small temperature differences \( |T_1 - T_2| \), namely, \( G_{th} = P_\gamma / (T_1 - T_2) = \pi k_B^2 T / 6 \hbar = G_Q \), the so-called quantum of thermal conductance (Pendry, 1983). Equation (35) can also be applied in the case where one of the resistors is replaced by an NIN tunnel junction with the corresponding resistance. Another important case is that of a hot resistor \( R \) at temperature \( T_{\text{env}} \), discussed in Sec. II.F. In this limit, with \( RC \) cut off as discussed in Sec. II.F, one finds that the heat absorption rate by a resistor or a normal tunnel junction (at \( T \ll T_{\text{env}} \)) is given by (Pekola and Hekking, 2007)

\[
P_\gamma = \frac{k_B T_{\text{env}}}{R T C}.
\]

III. REALIZATIONS

A. Normal-metal devices

Single-electron tunneling effects provide a means to transport electrons controllably one by one. In this respect the obvious choices are metallic single-electron circuits and semiconductor quantum dots. The metallic ones can be either in their normal or superconducting state or as hybrids of the two. The quantum dots, metallic hybrids, and superconducting circuits will be discussed in later sections. The first single-electron source was a metallic (nonsuperconducting) turnstile with four tunnel junctions and one active gate (Geerligs et al., 1990). The word “turnstile” refers to a device that is voltage biased at \( V_\text{th} \) between the external leads, but where the transport of electrons is impeded under idle conditions because of an energy gap. Under the active gate operation, electrons are transported synchronously one at a time. The finite voltage determines the direction of charge transport at the expense that the device is also dissipative. We discuss a more recent version of a turnstile in Sec. III.B.

The most impressive results of the early days of single-electron sources were obtained by metallic multijunction pumps, operating in a nonsuperconducting state. A prototype of them, featuring the main principle, is the three-junction pump, with two islands and a gate to each of them; see Fig. 11(a). This kind of pump was successfully operated in 1991 by Pothier et al. (Pothier, 1991; Pothier et al., 1991, 1992). Figure 11(b) demonstrates the stability diagram of a three-junction pump, which is essentially the same as that of the more common double-island quantum-dot circuit. The two axes here are the two gate voltages \( n_{g1}, n_{g2} \) normalized by the voltage corresponding to charge displacement of one electron, i.e., \( n_{g1} = C_{g1} V_{g1} / e \), where \( C_{g1} \) is the gate capacitance of island \( i \). The stability diagram consists of lines separating different stable charge states on the islands, indicated by indices \( (n_1, n_2) \) in the figure. The important property of this stability diagram is the existence of the nodes where three different charge states become degenerate; these are the three states with the lowest energy. The pump is operated around such a node, setting the working point at this node by applying dc voltages to the two gates. We focus here on one such node that at \( n_{g1} = n_{g2} = 1/3 \). Now if the temporarily varying gate voltages with frequency \( f \) added to these dc gate biases are such that the cyclic trajectory encircles the node at \( n_{g1} = n_{g2} = 1/3 \) counterclockwise, one electron is transported through the pump from left to right. The simplest implementation of such a cyclic trajectory is a circle around the node, which is represented by two equal-amplitude (in \( n_{g1} \)) sinusoidal voltages applied to the two gates, phase shifted by 90°. We take point A as the starting point of the cycle. There the system is in the charge state \( (0,0) \). Upon crossing the first degeneracy line, the new stable charge state is \( (1,0) \), meaning that an electron has to tunnel from the left lead to island 1, while moving in this part of the stability diagram. Reversible pumping is achieved when \( f \) is so slow that the transition occurs right at the degeneracy line. If, however, the pumping frequency is too fast, the tunneling does not occur before meeting the next degeneracy line, and the pumping fails. Roughly speaking, the tunneling process is stochastic, where the decay time of the Poisson process is determined by the junction resistance (see Sec. II.B), and if the pumping frequency becomes comparable to the inverse decay time for tunneling, the desired event can be missed. In the successful cycle, on the contrary, the system next crosses the degeneracy line between charge states \( (1,0) \) and \( (0,1) \), and...
under the same conditions, the system transits to the new stable charge state by an electron tunneling from the left island to the right one. In the remaining part of the cycle, on crossing the last degeneracy line, an electron tunnels from the right island to the right lead, completing the cycle where charge $e$ (one electron) has been transported from the left lead to the right one. By cyclically repeating this path at frequency $f$, an average current $I = ef$ runs from right to left, and this current can be read, for instance, by a regular transimpedance amplifier.

One of the advantages of the three-junction pump over the early turnstiles is that the device can be operated, in principle, reversibly, since no external bias voltage is needed. It can pump even against moderate bias. Another difference between the pump and the turnstile above is that in the pump there are no unattended islands on which the charge would be poorly controlled. Yet early realizations using fully normal-metal conductors in both the turnstiles and pumps suffered from other error sources which made these devices relatively inaccurate, on the level of 1%, even at low operation frequencies. A fundamental error source in this case is cotunneling, discussed in Sec. II.C. To circumvent this problem, a pump with a longer array of junctions is desirable: the error rate due to cotunneling is effectively suppressed by increasing the number of junctions in the array.

Theoretical analysis of cotunneling in multijunction pumps in the form of $N$ junctions in series with nonsuperconducting electrodes was performed by Jensen and Martinis (1992) and Averin, Odintsov, and Vyshenskii (1993). Thermal cotunneling errors were analyzed with a focus on the cases $N = 4$ and 5. The conclusion of the analysis was that under realistic experimental conditions, the $N = 4$ pump fails to produce an accuracy better than about $10^{-5}$, insufficient for metrology, whereas $N = 5$ should be sufficiently good at low operation frequencies, as far as cotunneling is concerned. This is illustrated in Fig. 12, where a relative error of $10^{-8}$ was predicted for an $N = 5$ pump at the operation frequency of $f = 1.3$ MHz, assuming that the pump junctions have 

\[ R_T = 500 \text{k}\Omega \text{ and } C = 0.61 \text{fF}, \]

and that the working temperature is $T = 50$ mK. All these parameters are quite realistic. In subsequent experiments (Martinis, Nahum, and Jensen, 1994), an error rate of about 0.5 ppm was achieved, which is still orders of magnitude above the prediction based on cotunneling for their circuit parameters and experimental conditions. Next, focus was turned to an $N = 7$ pump where further improved results, 15 ppb, were obtained at pumping frequencies of about 10 MHz (Keller et al., 1996). This impressive result, depicted in Fig. 13, was proposed to present a capacitance standard based on electron counting (Keller et al., 1999) and it still today stands as the best achievement in this respect. However, one notes that the frequency at which such a multijunction pump can be operated is very low, resulting in currents that are too small for a metrological redefinition of the ampere.

An analysis of the pump accuracy in the framework of the orthodox theory including cotunneling was presented by Martinis, Nahum, and Jensen (1994) for the $N = 5$ pump, and by Kautz, Keller, and Martinis (1999) for the $N = 7$ pump. In both cases, the experimental error rates could be quantitatively explained by a theory in an intermediate temperature range, where the majority of the errors was due to thermally activated single-junction tunneling processes. The observed low-temperature saturation of the error rates was conjectured to arise from photon-assisted tunneling and cotunneling, considered in this context theoretically, e.g., by Martinis and Nahum (1993) and White and Wagner (1993). Kautz, Keller, and Martinis (2000) explained quantitatively the error rates observed in the earlier experiments performed on pumps with $N = 4$–7 junctions by including photon-assisted processes in the model. The dominating error mechanism in the experiments was found to be photon-assisted single-electron tunneling, with negligible contribution from cotunneling. Jehl et al. (2003) explained error processes in a voltage-biased $N = 7$ pump with the same model.

The rate at which photon-assisted events occur is determined by the spectral density of voltage fluctuations across the junction at frequencies $f_{ph} \sim \Delta E/h$, where $\Delta E$ is the

---

**FIG. 12.** Predicted relative cotunneling-induced error vs inverse temperature for multijunction pumps, with $N = 4$ (circles) and $N = 5$ (squares). The computer simulations (points) and the predictions of analytic results (lines) are shown. Parameters are $R_T = 20R_K$, $f = 4 \times 10^{-4}/R_T C$, and $CV/e = -0.15$. Adapted from Jensen and Martinis, 1992.

**FIG. 13.** The seven-junction pump. (Left) The schematic of the pump, with six islands, each with a gate. The electrons are pumped to and from the external island on the top, and the charge on the island is detected by a single-electron electrometer. (Middle) The voltage $V_p$ on the external island vs time when pumping $\pm e$ with a wait time of 4.5 s in between. (Right) The pumping error vs temperature of the measurement, demonstrating the 15 ppb accuracy at temperatures below 100 mK. Adapted from Keller et al., 1996.
increase in electrostatic energy for a particular tunneling process. To explain the observed leakage rates, the authors in the above studies used $S_1(f) = \alpha |f|^2$, with the fitting parameter $\alpha$ assuming values from $(5 \text{ nV})^2$ to $(50 \text{ nV})^2$. Kautz, Keller, and Martinis (2000) motivated the $f^{-1}$ frequency dependence of the power spectrum by the ubiquitous charge noise present in SET electrometers, typically observed at frequencies below 1 kHz. (Covington et al. 2000) applied a calibrated amplitude of microwave radiation to one terminal of the pump, and the resulting tunneling rates were shown to be described by the theory of photon-assisted tunneling. They suggested that the origin of the high-frequency photons responsible for error events in the pumping experiments is the presence of fluctuating nonequilibrium charges near the devices. In addition, we note that recent electron-trapping results reported by Kemppinen et al. (2011) for a two-junction SNS-type trap with a series resistor measured in an rf-tight sample stage seem to indicate a much smaller flux of harmful photons to the junctions than was observed by, e.g., Covington et al. (2000).

Another successful line of metallic single-electron pumps relies on a smaller number of junctions ($N = 3$ or 5) while employing a resistive on-chip environment to suppress harmful cotunneling and photon-assisted tunneling (Lotkhov et al., 2001; Camarota et al., 2012). Suppression of cotunneling by a high-impedance environment was first demonstrated by Zorin et al. (2000) through SET $I$-$V$ measurements, motivated by the earlier theoretical predictions (Golubev and Zaikin, 1992; Odintsov, Bubanja, and Schön, 1992).

B. Hybrid superconducting–normal-metal devices

1. Operating principles

The hybrid turnstile, originally proposed and demonstrated by Pekola et al. (2008), is based on a single-electron transistor where the tunnel junctions are formed between a superconductor and a normal metal; see Fig. 14, top left. In principle, it can be realized in either a SINIS or NISIN configuration (Averin and Pekola, 2008; Kemppinen, Kafanov et al., 2009). However, it has turned out for several reasons that the former one is the only potential choice of the two for accurate synchronized electron transport purposes (Averin and Pekola, 2008). One reason is that in the NISIN structure tunneling strongly heats the island due to Joule power and weak energy relaxation in the small superconducting island, whereas in the SINIS case the island is of normal metal, better thermalized to the bath, and under proper operation, it can be cooled, too (Kafanov et al., 2009). The NISIN turnstile may also suffer from unpredictable $1e$-$2e$ periodicity issues. Furthermore, a detailed analysis of the higher-order tunneling processes shows that cotunneling limits the fundamental accuracy of the NISIN turnstile, whereas uncertainties below $10^{-8}$ are predicted for the SINIS version (Averin and Pekola, 2008). Hence we focus on the SINIS turnstile here.

The stability diagram of a conventional single-electron transistor is composed of Coulomb diamonds on the gate voltage $V_g$-drain-source voltage $V_b$ plane; see Fig. 15. Gate voltages $V_g$ are again written in dimensionless form, normalized by the voltage corresponding to charge displacement of one electron $n_{e}$. In this case the adjacent diamonds touch each other at a single point at $V_b = 0$, implying that the charge state is not locked for all gate-voltage values. The operation of the SINIS turnstile, on the contrary, is based on the combined effect of the two gaps: the superconducting BCS gap expands the stability regions of the charge states and the neighboring regions overlap. The principle of operation of the turnstile is illustrated in Fig. 15. When the gate charge $n_e(t)$ alternates between two neighboring charge states, electrons are transported through the turnstile one by one. A nonzero voltage, which yields a preferred direction of tunneling, can be applied since the idle current is ideally zero in the range $\left|eV_f\right| < \Delta$ at any constant gate charge value. If the gate signal is extended to span $k + 1$ charge states, one obtains current plateaus with $k$ electrons pumped per cycle. However,
the first plateau around the symmetric (degeneracy of two neighboring charge states) dc position of the gate is optimal for metrology. Note that if a nonzero bias voltage is applied across a normal-state SET, a gate span between different charge states always passes a region where none of the states is stable and where the current can freely flow through the device [white region in Fig. 15(a)]. Hence the normal-state SET cannot act as a turnstile even in principle, except for an experimentally infeasible gate sequence where \( n_g \) jumps abruptly between its extreme values in the Coulomb-blockaded parts of the stability diagram.

Figure 14 presents data obtained from a basic turnstile operated under various conditions (Pekola et al., 2008). Several wide current plateaus with increasing gate amplitude \( A_g \) can be seen. The gate drive is expressed here as \( n_g(t) = n_{g0} + A_g w(t) \), where \( n_{g0} \) and \( A_g \) are the gate offset and drive amplitude, respectively. The gate wave form of unit amplitude is denoted by \( w(t) \). The optimal gate drive is symmetric with respect to the two charge states: therefore in later sections we assume that \( n_{g0} = 1/2 \). In these first experiments, the accuracy of a synchronized charge transport as \( I = N_{ef} \), with \( N \) the integer index of a plateau, could be verified within about 1%.

A rough estimate for the optimal bias voltage \( V_b \) is obtained by considering the dominant thermal errors (Pekola et al., 2008). The probability of an electron tunneling against bias, i.e., “in the wrong direction” is given by \( \exp[-(eV_b/k_B T)] \). This error would lead to no net charge transferred during a pumping cycle, but it can be suppressed by increasing \( V_b \). On the other hand, increasing \( V_b \) increases the probability of transporting an extra electron in the forward direction. The magnitude of this kind of an error can be estimated as \( 1 - \exp[-(2\Delta - eV_b)/k_B T] \), since there is an energy cost given by the voltage distance from the conduction threshold at \( 2\Delta/e \). Combining these conditions, we obtain a trade-off \( eV_b = \Delta \) as the optimum bias voltage, where the thermal error probability is \( \exp(-\Delta/k_B T) \). The combined thermal error probability is \( \sim 10^{-7} \) at realistic temperatures of about 100 mK and with the BCS gap of aluminum \( \Delta/k_B = 2.5 \) K. The exact optimum of the bias close to the value given here depends on many other processes to be discussed. Experimentally, however, the choice \( eV_b = \Delta \) is a good starting point.

The optimal gate drive amplitude \( A_g \) lies somewhere between the threshold amplitudes for forward and backward tunneling which are, for the optimum bias voltage, \( A_{g,0} = \Delta/4E_c \) and \( A_{g,br} = 3\Delta/4E_c \), respectively. The subgap leakage is maximized at the degeneracy point \( n_{g0} = 1/2 \). In this respect, a square-wave signal is optimal. On the other hand, passing the threshold for forward tunneling too quickly tends to heat the island, whereas a sine signal can also cool it. Hence the optimal wave form is of some intermediate form.

The SINIS turnstile presents the choice of a single-electron source which is easy to manufacture and operate, and whose characteristics can be analyzed theoretically into great detail. It promises high accuracy as discussed in Sec. III.B.2. Its operation in a parallel configuration is straightforward thanks to the simple element of a single turnstile, and therefore it can yield higher currents than the other fixed-barrier single-electron sources presented in Sec. III.A. Thus it can be considered as a promising candidate in providing a realization of the amper.

### 2. Higher-order processes

As for the fully normal-metal pumps, the idealized picture of electron transport based on single-electron tunneling is disturbed by simultaneous tunneling of several electrons. Owing to the gap in the quasiparticle excitation spectrum of a BCS superconductor, elastic cotunneling takes place only when the bias voltage over the device exceeds \( 2\Delta/e \). The turnstile operation is achieved with voltages well below this threshold and hence cotunneling is suppressed, in contrast to purely normal-metal devices. As a general rule, any process that leaves behind an unpaired electron on a superconducting electrode incurs an energy penalty equal to \( \Delta \).

For hybrid structures, the lowest-order turnstile process where the energy cost of breaking a Cooper pair can be avoided is Andreev tunneling (Andreev, 1964), i.e., a complete Cooper-pair tunneling through a junction. Andreev tunneling has been studied thoroughly with single NIS junctions (Blonder, Tinkham, and Klapwijk, 1982; Eiles, Martinis, and Devoret, 1993; Lafarge et al., 1993; Heikkila and Nazarov, 1994; Pothier et al., 1994; Rajauria et al., 2008; Greibe et al., 2011; Maisi et al., 2011) as well as in so-called Cooper-pair splitters where the electrons of a Cooper pair tunnel to different normal-metal regions (Hofstetter, Csonka, and Nygrod, 2009; Herrmann et al., 2010; Wei and Chandrasekhar, 2010). In the case of a SINIS turnstile, Andreev tunneling manifests itself as two electrons being added to or removed from the island. Consecutively, increasing the charging energy of a device makes Andreev tunneling energetically unfavorable, suppressing it (Averin and Pekola, 2008; Maisi et al., 2011). The impact of Andreev tunneling on the accuracy of a turnstile has been directly observed on the pumped current (Aref et al., 2011). In Figs. 16(a) and 16(b), stability diamonds for single-electron and Andreev tunneling are shown for high-\( E_C \) and low-\( E_C \) devices, respectively. The pumping plateau of the high-\( E_C \) device, shown in Fig. 16(c), is free of Andreev tunneling whereas the low-\( E_C \) sample exhibits it as seen in Fig. 16(d).

For high-charging energy devices where Andreev tunneling is suppressed, the process limiting the accuracy of the SINIS turnstile is cotunneling of a Cooper pair and a single electron (Averin and Pekola, 2008). In this process, the island will be charged or discharged by a single electron while another electron effectively passes through the device. The net energy change is that of the corresponding single-electron process, plus the energy gained in transporting the Cooper pair from one electrode to another, which equals \( 2eV_b \) in the forward direction. Hence, the process cannot be made energetically unfavorable in a working turnstile. However, it can be suppressed relative to the first-order processes by making the junctions opaque enough. Ideally, to obtain an accuracy of \( 10^{-7} \), one needs to limit the speed of an aluminum-based turnstile to a few tens of pA (Averin and Pekola, 2008). This theoretically predicted maximum operation speed is expected to slow down by an additional factor of 3 due to nonuniformity of the tunnel barriers (Averin et al., 2011; Maisi et al., 2011). Thus 10 pA is expected to be the optimum yield per aluminum-based turnstile. In addition to the Cooper-pair electron cotunneling, the cotunneling of two Cooper pairs through the device increases the leakage current (Zaikin, 1994). In optimized devices discussed above, the Cooper-pair electron cotunneling is
density and is independent of the biasing at low energies $\Gamma_{qp} = n_{qp}/[2e^2R_F(D(E_F))]$. It should be compared to the rate at which we pump electrons. As discussed, we obtain roughly 10 pA from a turnstile free of higher-order tunneling errors at an accuracy of $10^{-7}$. The tunneling resistance of such a device is approximately $R_F = 1 \, \text{M}\Omega$. To ensure that the quasiparticle excitations do not cause errors on this level, we require the tunneling rate to satisfy $\Gamma_{qp} < 10^{-7} \times 10 \, \text{pA}/e$. With parameter values $D(E_F) = 1.45 \times 10^{47} \, \text{J}^{-1} \, \text{m}^{-3}$ and $\Delta = 200 \, \mu\text{eV}$, we need $n_{qp} < 0.04 \, \mu\text{m}^{-3}$. Such a level is demonstrated in an experiment without active driving of the system (Saira, Kemppinen et al., 2012) and is sensitive to the filtering and shielding of the sample. Also, the trapping of quasiparticles was shown to be important in this experiment.

Next we consider the relaxation of the quasiparticles. In turnstile operation, injection of hot quasiparticles through the tunnel junction drives the quasiparticle system of the superconductor actively out of equilibrium. We model the quasiparticle relaxation in the superconductor in terms of heat flow and obtain a diffusion equation for $n_{qp}$. Such an approach has been used to model several experiments (Ullom, Fisher, and Nahum, 1998; Rajauria, Courtois, and Pannetier, 2009; O’Neil et al., 2011; Peltonen et al., 2011; Knowles, Maisi, and Pekola, 2012). The heat flow of quasiparticles $J$ follows the equation $\nabla \cdot J = -p$, where $p$ is the power per unit volume removed from the quasiparticles. We use Fourier’s law of heat conduction $\mathbf{J} = -\kappa S \nabla T_{qp}$, where

$$\kappa S = \frac{6}{\pi^2} \frac{L_0 T_{qp}}{\rho_n} (\beta \Delta)^2 e^{-\beta \Delta}$$

is the heat conductivity of a superconductor (Bardeen, Rickayzen, and Tewordt, 1959). Here $L_0$ is the Lorenz number and $\rho_n$ is the resistivity in the normal state. By taking the derivatives only over strong exponential dependences and using Eq. (37), we obtain a diffusion equation

$$D \nabla^2 n_{qp} = p,$$

where the coefficient

$$D = \frac{\sqrt{2(k_B T_{qp} \Delta)^{1/2}}}{\sqrt{\pi e^2 \rho_n D(E_F)}}$$

is assumed to be constant. To write down the source term on the right side of Eq. (38), we consider the available mechanisms of heat conduction. Electron-phonon coupling is an inherent relaxation mechanism for quasiparticles inside a superconductor. However, it is so weak that the resulting decay length of $n_{qp}$ is usually on the millimeter scale (Martinis, Ansmann, and Aumentado, 2009; Rajauria, Courtois, and Pannetier, 2009; O’Neil et al., 2011). Typically, to enhance the relaxation, one uses so-called quasiparticle traps (Pekola et al., 2000; Rajauria, Courtois, and Pannetier, 2009; O’Neil et al., 2011), which are normal-metallic regions connected to the superconductor either directly or via an oxide layer. Once the hot quasiparticles enter the trap, the stronger electron-phonon relaxation in a normal metal removes their excess energy. A perfect quasiparticle trap forces the quasiparticle temperature at the interface to equal the electronic temperature of the normal metal. In the context of Eq. (38), this can be implemented as a boundary condition for $n_{qp}$. The boundary condition at the

FIG. 16 (color online). (a) Stability diamonds for single-electron tunneling (solid lines) and Andreev tunneling (dotted lines) for a sample with $E_C > \Delta$. (b) Stability diamonds for $E_C < \Delta$. (c) The first pumping plateau of the high-$E_C$ device as a function of the gate-voltage amplitude $\Delta$. The solid symbols show pumped current with $f = 10 \, \text{MHz}$ and three different bias voltages. Dotted lines are the simulated traces with the corresponding biasing. (d) The same data as in (c) but now for the low-$E_C$ device showing excess current due to Andreev tunneling. Adapted from Aref et al., 2011.
junction is obtained by setting the heat flow equal to the power injected by the quasiparticle current.

When the trap is connected via an oxide barrier, the heat is carried by quasiparticle tunneling. The orthodox theory result for the source term in such a configuration is

$$ p = \frac{2\pi r}{e^d} \int_{\Delta} dE E n_{\Delta}(E)(e^{-\beta E} - e^{-\beta_d E}) $$

$$ = \frac{\sigma_T}{e^d D(E_F) d} (n_{\text{qp}} - n_{\text{qgq}}), \quad (39) $$

which is obtained by setting the chemical potential difference of the trap and the superconductor to zero and assuming the diffusion to take place in two dimensions which is well justified for the thin films typically used in the samples. We also assumed $k_B T_{\text{qg}} \ll \Delta$. Here $\sigma_T$ is the electrical conductance per unit area of the trap, $d$ is the thickness of the superconducting film, $\beta = 1/k_B T_{\text{qg}}$, and $\beta_0 = 1/k_B T_0$, where $T_0$ is the temperature of the normal-metal electrons. We denote by $n_{\text{qgq}}$ the quasiparticle density of a fully thermalized superconductor, i.e., one where $T_{\text{qg}} = T_0$.

We consider some typical geometries of superconducting leads used in devices. First take a lead with a constant cross section as shown in Fig. 17(a). We assume that a heat flow $P_{\text{inj}}$ is injected at one end of the line, and that the other end is thermally anchored by a direct trap. For the lead itself, we assume a trap connected via an oxide barrier to be located on top. We can solve Eqs. (38) and (39) analytically in one dimension to obtain

$$ n_{\text{qp}}(x) = \frac{1}{D} \frac{e^{\sqrt{2}(x-x_0)}}{e^{\sqrt{2}x}} \frac{1}{(e^{2\sqrt{2}x} + 1) - 1} \frac{P_{\text{inj}}}{w d} + n_{\text{qgq}}. $$

Here

FIG. 17 (color online). Two typical geometries for a superconducting bias lead: (a) A lead having a constant cross section determined by the thickness $d$ and width $w$. The length of the line is $l$. (b) A sector-shaped lead characterized by an opening angle $\theta$, initial radius $r_0$, and final radius $r_f$. For the picture $\theta$ is set to 180°. The colored parts on top denote a quasiparticle trap connected via an oxide barrier. (c) Quasiparticle density $n_{\text{qp}}$ along a constant-cross-section line with various oxide trap transparencies $k$, and (d) along an opening line. In the plots, $n_{\text{qp}}$ is scaled by $n_{\text{qgq}} = D P_{\text{inj}}/A_p$, where the injection area $A_p$ equals $w d$ for (c) and $\theta r_0 d$ for (d). For the leads in (b) and (d), we also use the notation $x = (r - r_0)/l$ with $l = r_f - r_0$ and have used values $r_0 = 20 \text{ nm}$ and $r_f = 5 \mu \text{m}$.

$$ k = \frac{\sqrt{\pi} \rho_0 \sigma_T \Delta}{\sqrt{2}(k_B T_{\text{qg}} \Delta)^{1/2} d}. $$

and $x$ is the coordinate along the wire starting at the injection side ($x = 0$) and ending at the direct trap ($x = l$). In Fig. 17(c), we show the quasiparticle density for various values of $\sigma_T$. The lowest $\sigma_T$ corresponds to the case where the quasiparticles diffuse only through the wire and then relax at the direct contact. At higher transparencies, the oxide trap starts to help for the relaxation as well. If we use parameter values $d = 50 \text{ nm}$, $w = 100 \text{ nm}$, $l = 1 \mu \text{m}$, and $T_{\text{qg}} = 130 \text{ mK}$, which are typical for fabricated samples, we see that a typical injection power of $P_{\text{inj}} = 2 \text{ fW}$ yields $n_{\text{qp}} - n_{\text{qgq}} = 10 \mu \text{m}^{-3}$ without the oxide trap and even with the highest transparency $\sigma_T = (100 \Omega \mu \text{m}^2)^{-1}$ that is possible to fabricate without pinholes (Bennings, Kubatkin, and Delsing, 2004), we get only an order of magnitude improvement.

To decrease the quasiparticle density to the acceptable level discussed, one needs to optimize the lead geometry as well. Therefore, we consider a lead that widens as shown in Fig. 17(b). In this case, we can solve a one-dimensional diffusion equation in polar coordinates. The junction is assumed to be located at radius $r = r_0$, and the direct contact trap to begin at radius $r = r_f$. Thickness of the lead and the overlaid trap are as in the previous example. The solution of Eqs. (38) and (39) can be expressed with modified Bessel functions $I_n$ and $K_n$ as

$$ n_{\text{qp}}(r) = n_{\text{qgq}} + \frac{P_{\text{inj}}}{D \sqrt{k} \theta r_0 d} \left\{ \left( \frac{I_n(\sqrt{k} r_0)}{I_0(\sqrt{k} r_0)} \right) - 1 \frac{K_0(\sqrt{k} r_f)}{K_0(\sqrt{k} r_f)} \right\} + \left( \frac{I_1(\sqrt{k} r_0)}{I_0(\sqrt{k} r_0)} \right) \frac{I_n(\sqrt{k} r_f)}{I_0(\sqrt{k} r_f)}.$$

In Fig. 17(d), we show $n_{\text{qp}}(r)$ for various transparencies of the oxide trap. The lowest transparencies, again, correspond to a pure diffusion limit. Note that the quasiparticle density at the junction depends only weakly on the transparency of the trap: Because of the logarithmic dependence, changing the transparency by several orders of magnitude makes less than an order of magnitude difference to $n_{\text{qp}}(r_0)$. In a widening lead, heat sinking is made efficient by spreading the heat to a larger volume, and the area of the trap contact is also increased. By using realistic parameter values $d = 50 \text{ nm}$, $\theta = \pi/2$, $r_0 = 50 \text{ nm}$, $r_f = 5 \mu \text{m}$, $T_{\text{qg}} = 130 \text{ mK}$, $\rho_0 = 10 \text{ n\Omega m}$, and $P_{\text{inj}} = 2 \text{ fW}$, we see that it is possible to reach $n_{\text{qp}} \leq 1 \mu \text{m}^{-3}$ at the junction even without an oxide trap. Increasing the thickness of the electrode by a factor of 10 would then start to be sufficient for the metrological accuracy requirements.

Several experiments (Ullom, Fisher, and Nahum, 1998; Rajauria, Courtois, and Pannetier, 2009; O’Neill et al., 2011; Knowles, Maisi, and Pekola, 2012) show that the above diffusion model is valid for quasiparticle densities of the order of $n_{\text{qp}} \sim 10 \mu \text{m}^{-3}$. A smaller quasiparticle density required for metrological applications implies that the absolute number of quasiparticles in the conductors becomes very small. With a typical volume of a lead $100 \text{ nm} \times 100 \mu \text{m}^2$,
the quasiparticle number is $N < 1$ with $n_{qp} < 0.1 \, \mu m^{-3}$. It is not currently obvious if such a situation can be treated with the diffusion model or whether a more elaborated theory is required. Pumping experiments on metrological accuracy can provide a way to shed light on such a situation.

C. Quantum-dot-based single-electron pumps and turnstiles

In this section, we introduce semiconducting quantum dots and review their applications as single-electron current sources, concentrating on the experimental developments. For an overview of the related theory, we refer to Sec. II.E.

1. Introduction to quantum dots as electron pumps

In contrast to conventional three-dimensional bulk conductors or more exotic two-dimensional conductors such as quantum Hall systems or graphene, semiconducting quantum dots can be regarded as zero-dimensional conductors, for which the electrons are tightly confined in all three spatial dimensions. Thus quantum dots show truly discrete excitation spectra that are reminiscent of those of natural atoms. One of the early key experiments on these artificial atoms (Kastner, 1993) was the observation of discrete quantum levels (Reed et al., 1988; Johnson et al., 1992; Su, Goldman, and Cunningham, 1992) and the shell structure in the filling of the electron states (Tarucha et al., 1996).

As discussed in Sec. II.E, the conceptual difference between small metallic islands studied in the previous sections and quantum dots is that the Fermi level and hence the conduction electron density in the metallic islands is high, making the energy spacing between the spatially excited electron states extremely small. The metallic system can be typically described by a constant density of states as opposed to the strongly peaked density of states in quantum dots. Furthermore, quantum dots can contain a low number of electrons in the conduction band ranging from zero (Ashoori et al., 1993; Elzerman et al., 2003; Lim et al., 2009) to more than hundreds, similar to natural atoms, whereas the corresponding number is orders of magnitude higher for metallic systems. In fact, the sharp potential created by a single donor atom in silicon can also be considered to be an ultrasmall quantum dot. By connecting such natural atoms to electron reservoirs, for example, SETs (Lansbergen et al., 2008; Tan et al., 2010; Fuechsle et al., 2012) and electron pumps (Lansbergen, Ono, and Fujiwara, 2012; Roche et al., 2012) have been fabricated.

Figure 18 shows different types of quantum-dot architectures. The most conventional quantum dots are based on a two-dimensional degenerate electron gas (2DEG) that either forms naturally, for example, at the interface between AlGaAs and GaAs (Chang, Esaki, and Tsu, 1974) or is induced at the interface between silicon and silicon oxide by an external gate (Ando, Fowler, and Stern, 1982). Alternatively, quantum dots can be fabricated from epitaxially grown nanowires (Ohiisson et al., 2002; Fasth et al., 2007; Nadj-Perge et al., 2010) or from lithographically defined graphene islands (Connolly et al., 2012). In the conventional dots, the confinement is very strong in the direction perpendicular to the interface. Etching techniques, local anodic oxidation (Held et al., 1997), pattern-sensitive oxidation (Takahashi et al., 1995), or metallic electrodes (see Figs. 18(c) and 18(d)) can be employed to provide the electrostatic potential defining the well for the electrons in the plane of the interface. In the in-plane diameter of this type of dot can vary from tens of nanometers to several micrometers. Thus there are plenty of atoms and electrons in the region of the dot but most of them lie in the valence band and require an energy of the order of 1 eV to be excited. Since the relevant energy scales for the spatial excitations and the single-electron charging effects are orders of magnitude lower, the occupation of the valence states can be taken fixed.

In the effective mass approximation (Ando, Fowler, and Stern, 1982), the details of the electrostatic potential and the effects of the valence electrons in the solid are coarse grained such that only the electrons in the conduction band are taken into account, and these electrons are treated as particles in the smooth potential defining the dot. This description has proved to reproduce several important experimental findings both qualitatively and quantitatively (Ando, Fowler, and Stern, 1982), and it provides insight into the single-electron phenomena in quantum dots. In particular, the potential barriers arising from the gates defining the dot can be visualized just for the small number of electrons in the conduction band.

2. Pioneering experiments

Although quantum dots hold a much smaller number of electrons than metallic islands, probably their greatest benefit is that the tunnel barriers can be formed by electrostatic potentials and controlled externally by gate voltages. Thus the height of the potential barrier, through which the electrons tunnel to the source and drain reservoirs, can be controlled in situ. This property provides fruitful grounds for electron pumping since the dependence of the tunneling rate on the barrier height and hence on the voltage of the gate electrode is typically exponential.

The first experiments employing quantum dots for frequency-locked single-electron transport were reported by Kouwenhoven et al. (1991a, 1991b) [see also Kouwenhoven (1992)]. Here they used surface-gated GaAs dots as shown in Fig. 19(a). The negative voltages on gates $C$, $F$, 1, and 2 deplete the 2DEG that is located 100 nm below the surface, thus defining the quantum dot in the center with a radius of about 300 nm and charging energy $2E_C = e^2/C_S = 0.67$ meV. (Gates 3 and 4 are grounded and do not deplete the 2DEG.)
The first single-electron current source based on quantum dots by Kouwenhoven et al. (1991b, 1991a). (a) SEM image of the device from the top; (b), (c) operation principle; and (d) measured I-V curves reported. The gate configuration corresponds to the case in Fig. 18(d). The different I-V curves are measured while driving the turnstile with different center-gate [gate C in (a)] voltages, rf amplitudes, and phase differences. The curves are not offset and the dashed lines show the current levels $n\text{ef}$ with $n = -5, \ldots, 5$. Adapted from Kouwenhoven et al., 1991b, 1991a.

In addition to dc voltages defining the dot, 180°-phase-shifted sinusoidal rf drive is superimposed on gates 1 and 2, lowering one barrier at a time. This rf drive induces a turnstile operation as shown in Figs. 19(b) and 19(c) for negative bias voltage on the left side of the dot: when the voltage at gate 1 is high (low tunnel barrier) and low at gate 2 (high tunnel barrier), an excess electron enters the dot through the left barrier [see Fig. 19(b)], and when the voltage at gate 1 is low and high at gate 2, the electron escapes through the right barrier [see Fig. 19(c)]. Thus the average dc current through the device in the ideal case is given by $I_p = ef$, where $f$ is the operation frequency. For bias voltages greater than the charging energy $|eV| \geq EC$, more than a single electron can be transported in a cycle yielding ideally $I_p = n\text{ef}$, where $n$ is an integer. Signatures of this type of current quantization were observed in the experiments (Kouwenhoven et al., 1991a, 1991b; Kouwenhoven, 1992) and are illustrated in Fig. 19(d). The current through the device as a function of the bias voltage tends clearly to form a staircaselike pattern with the step height $ef$. This was the first experimental demonstration of current quantization in quantum-dot structures. Note that in addition to the turnstile operation, Fig. 19(d) also shows the pumping of electrons against the bias voltage for certain phase differences of the driving signals. The error in the pumped current is a few percent, falling somewhat behind the first experiments on metallic structures reported by Geerligs et al. (1990).

The second set of experiments on single-electron turnstiles based on quantum dots was published by Nagamune et al. (1994). Here the quantum dot forms a gallium arsenide 2DEG that is wet etched into the shape of a 460-nm-wide wire as illustrated in Fig. 20(a). Two 230-nm-wide metallic gates are deposited perpendicular to the wire at a distance of 330 nm. This different barrier gate configuration and the higher charging energy of $2EC = 1.7$ meV resulted in a clear improvement of the staircase structure as shown in Fig. 20(b). However, they reported that a parallel channel forms due to the rf operation and the effect of this channel is subtracted from Fig. 20(b). They estimated the accuracy of their device to be about 0.4% if the correction from the parallel channel is taken into account.

In 1997–2001, a series of experiments was carried out on so-called multiple-tunnel junction devices as electron pumps (Tsukagoshi et al., 1997; Tsukagoshi, Alphenaar, and Nakazato, 1998; Altebaeumer and Ahmed, 2001; Altebaeumer, Amakawa, and Ahmed, 2001). Here the most common device was based on either $\delta$-doped GaAs or phosphorus-doped silicon that was etched such that a central region is connected to source and drain reservoirs by narrow strips as shown in Fig. 21(a). The side gates near the strips are set to a constant potential and an rf drive on the central side gate induces a current that depends linearly on frequency as shown in Fig. 21(b). The explanation of this type of operation is that the dopants and disorder in the strips function as Coulomb-blockade devices themselves rather than as single tunnel junctions, which gives rise to the term multiple-tunnel junction. Since these experiments were more motivated by applications in information processing (Ono et al., 2005) with only a few electrons rather than finding a metrological current source, the accuracy of the device was not studied in detail.

3. Experiments on silicon quantum dots

The first step toward single-electron pumping in silicon was taken by Fujiwara and Takahashi (2001) as they presented an ultrasmall charge-coupled device and demonstrated that it could be used to trap and move individual holes controllably at the temperature of 25 K. This device was fabricated with silicon-on-insulator techniques (Takahashi et al., 1995) and had two adjacent polysilicon gates acting as metal-oxide-semiconductor field-effect transistors (MOSFETs). Subsequently, a rather similar device with charging energy $2EC = 30$ meV shown in Fig. 22(a) was utilized for electron pumping by Ono and Takahashi (2003).
at the temperature of 25 K. They obtained an accuracy of the order of $10^{-2}$ up to 1 MHz pumping that was also the limitation set by the calibration of their measurement equipment. Here the electron pumping was based on two sinusoidal driving signals that are offset by less than 180°, which causes the chemical potential of the dot to move during the cycle.

In addition to pumping, Ono et al. (2003) utilized the device shown in Fig. 22(a) as a single-electron turnstile. The operational principle is the same as in the pioneering experiments with GaAs quantum dots described in Figs. 19(b) and 19(c). Ono et al. (2003) observed current steps of $e f$ up to $f = 1$ MHz operation frequencies [see Fig. 23(a)]. The flatness of the plateaus was of the order of $10^{-2}$ measured at 25 K. In these experiments, the tunnel barrier was formed by the combination of the gate voltages and the oxidation process developed by Takahashi et al. (1995) limiting the pumping frequencies. The first fully gate-tunable turnstile in silicon was demonstrated by Fujiwara et al. (2004) at 20 K with $2E_C = 16$ meV and the relative uncertainty in the pumped current of the order $10^{-2}$ at the maximum applied pumping frequency 100 MHz. This was a clear improvement in the speed of quantum-dot electron pumps.

Chan et al. (2011) used metallic aluminum gates to define a silicon quantum dot in the electron accumulation layer of the device as shown in Fig. 23(b). Although the relative variation of the current at the plateau they measured was below $10^{-5}$ for a broad range of source-drain voltages [see Fig. 23(c)], they could not strictly claim lower than 2% relative uncertainty in the current at 60 MHz pumping frequency due to the inaccurate calibration of the gain of the transimpedance amplifier employed. These experiments were carried out with $2E_C = 2.8$ meV at 300 mK phonon temperature but the sequential tunneling model used to fit the data by Chan et al. (2011) suggested that the electron temperature of the dot rose up to 1.5 K. It is to be studied whether 1.5 K was due to power dissipated at the surface mount resistors in the vicinity of the sample or due to the direct heating of the 2DEG from the electrostatic coupling to the driven gate potentials.

Fujiwara, Nishiguchi, and Ono (2008) introduced a single-electron ratchet based on a silicon nanowire quantum dot with two polysilicon gates working as MOSFETs [Fujiwara et al. (2004) also employed this type of device]. In general, ratchets generate directional flow from a non-directional drive due to the asymmetry of the device. Here an oscillating voltage is applied to one of the gates such that an electron is captured through it near the maximum voltage, i.e., minimum barrier height, and ejected through the other barrier near the minimum voltage. In fact, the number of electrons pumped per cycle depends on the applied dc voltages and current plateaus up to $5e f$ were reported. Furthermore, a nanoampere pumped current was observed at the $3e f$ plateau with the pumping frequency $f = 2.3$ GHz. The error in the current was estimated to be of the order of $10^{-2}$ for the experiment carried out at 20 K temperature. Whereas in the conventional multiparameter pumps, the pumping errors arise mostly due to missed or excess tunneling events in a quasistatic Coulomb-blockade regime, the errors in the single-parameter pumps are taken to be dominated by a dynamic process, in which electrons tunnel out of the dot to the source lead.

Recently, Jehl et al. (2012) reported on frequency-locked single-electron pumping with a small quantum dot formed in metallic NiSi nanowire interrupted by two MOSFETs controlled by barrier gates. With rf drives on the barrier
gates, they were able to pump currents beyond 1 nA but the accuracy of the pump was not studied in detail. The MOSFET channels in this device had very sharp turn-on characteristics requiring only about 4.2 mV of gate voltage to change the conductivity of the channel by a decade, which can be important in reducing unwanted effects from the gate-voltage drive such as heating.

The first error-counting experiments in silicon were carried out by Yamahata, Nishiguchi, and Fujiwara (2011) [see also Nishiguchi et al. (2006)]. In contrast to the pioneering error-counting experiments by Keller et al. (1996), only a single silicon nanowire quantum dot was used as the current source and the electrons were steered into and out of a quantum dot coupled to a charge sensor. By opening the MOSFET separating the node from the drain reservoir, it was possible to use the same device as a dc current source. The observed pumping error was of the order of $10^{-2}$ and was reported to be dominated by thermal errors at the 17 K temperature of the experiments. Furthermore, electron-counting experiments were recently carried out by Fricke et al. (2013) in a quantum-dot array. Further details of error-counting schemes are discussed in Sec. III.H.2.

4. Experiments on gallium arsenide quantum dots

After the pioneering experiments discussed in Sec. III.C.2, the focus on single-electron sources based on gallium arsenide moved toward the idea of using SAWs to drive the single electrons in a one-dimensional channel—a topic to be discussed in Sec. III.D. In this section, we focus on gate-controlled GaAs pumps for dc current. A similar device to the ones discussed here has also been applied in the search for an ac-current standard which is the topic of Sec. III.G.1.

The seminal work by Blumenthal et al. (2007) took gate-controlled GaAs quantum dots a leap closer to a metrological current source, namely, they reported 547 MHz (87.64 pA) single-electron pumping with one-standard-deviation ($\sigma$) relative uncertainty of $10^{-4}$ (see Fig. 24). However, they did not report the full dependence of the pumping errors as functions of all control parameters. As the device, they employed a chemically etched AlGaAs-GaAs wire with overlapping metallic gates as shown in Fig. 24. Only the three leftmost gates $L$, $M$, and $R$ were used such that 180°-phase-shifted sinusoidal driving signals were applied to gates $L$ and $R$ in addition to dc voltages applied to all three gates. The amplitudes of the rf signals were chosen asymmetric such that the device can work as a pump rather than a turnstile. The charging energy of the device was estimated to be $2E_C = 1$ meV, and the experiments were carried out at the bath temperature of 300 mK.

With a similar device architecture as shown in Fig. 24 but using only two gates instead of three, Kaestner, Kashcheyevs, Amakawa et al. (2008) demonstrated that frequency-locked single-electron pumping can be carried out with a single sinusoidal driving voltage, thus decreasing the complexity of the scheme. This type of single-parameter pumping with two gates is employed in the remainder of the works discussed in this section. Maire et al. (2008) studied the current noise of a similar single-parameter pump at $f = 400$ MHz and estimated based on the noise level that the relative pumping error was below 4%. Kaestner, Kashcheyevs, Hein et al., 2008 studied the robustness of the current plateaus as functions of all control parameters of the pump except the source-drain bias. They showed that single-parameter pumping is robust in the sense that wide current plateaus appear in the parameter space but their measurement uncertainty was limited to about $10^{-2}$, and hence a detailed study of the behavior of the accuracy as a function of these parameters was not available.

Wright et al. (2008) made an important empirical observation that the accuracy of the single-parameter pump can be improved by an application of perpendicular-to-plane magnetic field [see also Wright et al. (2009) and Fig. 25(b)]. They applied fields up to 2.5 T and demonstrated that the $n=1$ plateau as a function of the dc voltage on the nondriven gate widens noticeably with increasing magnetic field. In further studies by Kaestner et al. (2009) and Leicht et al. (2011) up to magnetic fields of 30 T, a great widening on the plateau was observed, but it essentially stopped at 5 T. On the contrary, high-resolution measurements on the pumped current up to 14 T by Fletcher et al. (2011) showed a continuous improvement on the pumping accuracy with increasing field [see also Fig. 25(b)]. This discrepancy is possibly explained by the different samples used in the different sets of experiments.

Giblin et al. (2010) employed a magnetic field of 5 T and reported 54 pA of pumped current with $1\sigma = 15$ ppm relative uncertainty with a single-parameter sinusoidal drive. They were able to measure at such a low uncertainty with a room-temperature current amplifier since they subtracted a reference current from the pumped current and passed less than 100 fA through the amplifier. Thus the uncertainty in the gain of the amplifier did not play a role. The reference current was created by charging a low-loss capacitor and was traceable to primary standards of capacitance.

To date, the most impressive results on single-electron pumping with quantum dots have been reported by
in this field moved toward the idea of using SAWs to drive single electrons in a one-dimensional channel (Shilton, Mace et al., 1996). Here the sinusoidal potential created for the electrons in the piezoelectric GaAs by a SAW forms a moving well that can trap an integer number of electrons and transport them in a one-dimensional channel.

The first experiments of this kind of SAW electron pumps were carried out by Shilton, Talyanskii et al., (1996). They employed a SAW frequency of 2.7 GHz and observed a corresponding \( n = 1 \) current plateau at 433 pA with the uncertainty of the order of \( 10^{-2} \) at 1 K temperature. Talyanskii et al. (1997) carried out more detailed experiments on similar samples at two different SAW frequencies and the results were in agreement with the \( ef \) scaling law. Furthermore, several current plateaus were observed as a function of the gate voltage corresponding to different integer values of pumped electrons per cycle. However, the experimental uncertainty at the plateau was again of the order of \( 10^{-2} \) and sharp current peaks were observed at various gate-voltage values.

After these first experiments, Janssen and Hartland (2000a, 2000b) studied the accuracy of the SAW pump and reported a 431 pA current at the center of the plateau with 200 ppm relative deviation from the ideal value. Ebbecke et al. (2000) demonstrated SAW pumping up to 4.7 GHz frequencies and with two parallel channels to increase the current, but the measurement accuracy was rather limited here. To improve the quality of the plateau Janssen and Hartland (2001) decreased the width of the one-dimensional channel, which helps in general. However, they observed that the required \( rf \) power to drive the electrons increases with decreasing channel width, causing severe \( rf \) heating of the sample. This heating caused the quality of the plateau to drop and the conclusion was that materials with lower losses due to \( rf \) are needed [see also Utko, Lidelof, and Gloos (2006)]. In fact, Flensberg, Niu, and Pustilnik (1999) and Ebbecke et al. (2003) reported that the accuracy of the SAW current is fundamentally limited in one-dimensional channels because of tunneling of electrons out from a moving dot.

To overcome the limitation pointed out by Ebbecke et al. (2003), the charging energy was increased in the system by defining a quantum dot with surface gates rather than utilizing an open one-dimensional channel (Ebbecke et al., 2004). Thus the applied SAWs modulate both the tunnel barriers between the dot and the reservoirs and the electrochemical potential at the dot. With this technique, current plateaus were observed at a SAW frequency of 3 GHz and the reported relative deviation from the ideal value was of the order of \( 10^{-3} \). Although the results by Janssen and Hartland (2000a, 2000b) remain the most accurate ones reported to date with SAW electron pumps, and hence are not valuable for a metrological current source, single-electron transfer with SAWs can be useful in other applications. For example, McNeil et al. (2011) showed that an electron taken by SAWs from a quantum dot can be captured by another dot at a distance. This kind of electron transport can potentially be used to transport single spins working as quantum bits in a spin-based quantum computer (Hanson et al., 2007; Morello et al., 2010).

FIG. 25 (color online). (a) SEM image of the device with a schematic measurement setup. (b) Current plateaus obtained by using a sine wave drive at different frequencies and magnetic fields. (c) Relative difference of the pumped current from \( ef \) using a sine wave form and a tailored arbitrary wave form at different frequencies. The rightmost data point denoted by an asterisk shows the result with the potential of the entrance gate shifted by 10 meV from the optimal operation point. From Giblin et al., 2012.

Giblin et al. (2012). Compared with the previous results of Giblin et al. (2010), they made several changes to improve the results. They used a higher magnetic field of 14 T and an advanced generation of samples with a lithographically defined place for the quantum dot in both directions in the plane [see Fig. 25(a)]. Instead of using sinusoidal wave forms, they also tailored the drive voltage so that the cycle time was distributed more evenly for the different parts of the cycle. To make traceable measurements, a reference current was created by an accurate temperature-controlled 1 GΩ resistor, a voltage source, and a high-precision voltmeter. The voltmeter and the resistor were calibrated through intermediate steps against the Josephson voltage standard and the quantum Hall resistance (QHR) standard, respectively. In this work, Giblin et al. (2012) reported 150 pA pumped current with relative 1σ uncertainty of 1.2 ppm [see Fig. 25(c)]. Most of the uncertainty, 0.8 ppm, arose from the calibration of the 1 GΩ resistor. Thus it is possible that the electron pumping was actually even more accurate, as suggested by fitting the results to a so-called decay-cascade model (Kashcheyevs and Kaestner, 2010). However, there can be processes that are neglected by the model and since there is no experimental evidence on lower than 1.2 ppm uncertainty, it remains the lowest demonstrated upper bound for relative pumping errors for quantum-dot single-electron pumps. Error counting, as demonstrated in silicon by Yamahata, Nishiguchi, and Fujiwara (2011) and in aluminum by Keller et al. (1996), is a way to measure the pumping errors to a very high precision independent of the other electrical standards and remains to be carried out in the future for the GaAs quantum-dot pumps.

D. Surface-acoustic-wave-based charge pumping

After the pioneering experiments on single-electron sources based on GaAs discussed in Sec. III.C.2, the focus...
The envisioned advantage in pumping Cooper pairs instead of electrons is that the supercurrent produced by the Cooper-pair pumps is inherently dissipationless and the BCS gap protects the system from microscopic excitations. Thus the operation frequency of the pump can possibly be high with the system still remaining at very low temperature. Another advantage of the supercurrent is that it can sustain its coherence, and hence be virtually noiseless, in contrast to the single-electron current that is based on probabilistic tunneling. Furthermore, since the charge of a single Cooper pair is \(2e\), single-Cooper-pair pumps yield twice the current compared with single-electron pumps operated at the same frequency. Despite these advantages, the lowest uncertainties in the achieved Cooper-pair current is at the percent level (Vartiainen et al., 2007; Gasparinetti et al., 2012). One reason for this is the low impedance of the device, rendering it susceptible to current noise.

Two types of Cooper-pair pumps exist in the literature: arrays of superconducting islands (Geerligs et al., 1991) with source and drain leads, all separated by single Josephson junctions with fixed tunnel couplings, and a so-called sluice (Niskanen, Pekola, and Seppä, 2003; Niskanen et al., 2005) that is composed of a single island connected to the leads by two SQUIDs that function as tunable Josephson junctions; see Fig. 26. As in the case of single-electron pumps, the device operation is based on Coulomb-blockade effects allowing the controlled transfer of individual Cooper pairs, which means in the case of array pumps that the fixed Josephson energies of the junctions must be much lower than the Cooper-pair charging energy of the corresponding islands. In the sluice, it is sufficient that the minimum obtainable Josephson energy is much lower than the charging energy. For the arrays, the thermal energy \(k_B T\) must be much lower than the Josephson energy that defines the energy gap between the ground state and the excited state of the quantum system at charge degeneracy. For the sluice, the maximum Josephson energy of the SQUIDs yields the minimum energy gap of the system, thus relaxing the constraint on temperature.

The first experiment demonstrating Cooper-pair pumping was performed by Geerligs et al. (1991). The device is a linear array of three Josephson tunnel junctions. The two superconducting islands separated by the junctions are capacitively coupled to individual gate electrodes. Except in the vicinity of the charge degeneracy points in the gate-voltage space, the number of Cooper pairs on these islands is rather well defined by the gate voltages because the Coulomb-blockade regime is employed. By biasing the device and applying sinusoidal ac voltages with appropriate amplitudes to the gates, one obtains a continuously repeated cycle, during which a Cooper pair is transferred through the device, i.e., Cooper pairs are pumped one by one. Ideally, this yields a dc current \(I = 2ef\) that is proportional to the pumping frequency \(f\). The driving voltage at each gate should have the same frequency and a phase difference of \(\pi/2\). The pumping direction can be reversed if the difference is changed by \(\pi\). Thus the pumping principle is the same as for a normal pump discussed in Sec. III.A. The height of the measured current plateau follows rather well the predicted relation \(I = 2ef\) at low pumping frequencies, but deviates strongly at higher frequencies. This is explained by several mechanisms. The pumping uncertainty of the device was not assessed in detail but it seems to lie at least on the percent level with picoampere currents. One of the error mechanisms is the Landau-Zener tunneling when the system is excited to the higher-energy state without transferring a Cooper pair. This was the dominant mechanism at the high end of the studied pumping frequencies in the experiment by Geerligs et al. (1991) thus imposing an upper limit on the operation frequency of the device. Other error sources in the device are the tunneling of nonequilibrium quasiparticles, photon-excited tunneling, and relaxation of the excited states produced by Landau-Zener tunneling. In addition, cotunneling of Cooper pairs through the two junctions produces a steplike feature in the current plateaus, thus reducing the pumping accuracy. Later, a similar three-junction Cooper-pair pump was studied by Toppari et al. (2004) and essentially the same conclusions on the pumping accuracy were made. In both experiments, no \(2e\) periodicity was observed in the dc measurements, which suggests a substantial presence of nonequilibrium quasiparticles in the system.
The effect of quasiparticles on Cooper-pair pumping was also observed in the seven-junction Cooper-pair pump (Aumentado, Keller, and Martinis, 2003). The device is basically the same as the one used for pumping single electrons in the earlier experiments in the normal state (Keller et al., 1996). The pump consists of 6 μm-scale aluminum islands linked by aluminum-oxide tunnel barriers. Investigation of this circuit in the hold and pumping modes revealed that besides 2e tunneling events, there is a significant number of 1e events associated with the quasiparticle tunneling. All these experiments show that in order to obtain accurate Cooper-pair pumping, one must suppress unwanted quasiparticle tunneling. Leone and Lévy (2008) and Leone, Lévy, and Lafarge (2008) proposed topological protection in pumping Cooper pairs. The charge is expected to be strictly quantized determined by a Chern index. To our knowledge, this idea has not been tested experimentally.

In order to increase the output dc current and accuracy of a single pump, the sluice pump was introduced by Niskanen, Pekola, and Seppä (2003) and Niskanen et al. (2005). In the pumping cycle, the two SQUIDs separating the single island work in analogy with valves of a classical pump and the gate voltage controlling the island potential is analogous to a piston. At each moment of time, at least one SQUID is closed, i.e., set to minimum critical current. The gate voltage is used to move Cooper pairs through open SQUIDs with maximum critical current. If the pairs are taken into the island through the left SQUID and out of the island through the right SQUID, the resulting dc current is ideally \( I = N2ef \), where the number of pairs transported per cycle \( N \) is determined by the span of the gate-voltage ramp. In practice, the critical current of the SQUIDs is controlled by flux pulses generated by superconducting on-chip coils. Since each operation cycle can transfer up to several hundreds of Cooper pairs, Vartiainen et al. (2007) managed to pump roughly 1 nA current with uncertainty less than 2% and pumping frequency of 10 MHz; see Fig. 26. The investigated high-current Cooper-pair pump demonstrated steplike behavior of the pumped current on the gate voltage; however, its accuracy was affected by the residual leakage in the tunnel junctions and the fact that the SQUIDs did not close completely due to unequal Josephson junctions in the structure.

The leakage current in the sluice can be suppressed by working with a phase bias instead of a voltage bias, as was applied by Niskanen et al. (2005) and Vartiainen et al. (2007). The only experiment reported for a phase-biased pump was carried out by Möttönen, Vartiainen, and Pekola (2008). They connected a sluice in a superconducting loop with another Josephson junction. By measuring the switching behavior of this junction from the superconducting state to the normal branch with forward and backward pumping, they were able to extract the pumped current of the sluice. However, this type of current detection did not turn out to be as sensitive as the direct measurement with a transimpedance amplifier used in the case of voltage bias. A potential way to improve the sensitivity is, instead of the switching junction, to use a cryogenic current comparator (CCC) coupled inductively to the superconducting loop. This type of an experiment has not been carried out to date. Instead, Gasparinetti et al. (2012) measured a sluice in the vicinity of vanishing voltage bias, where they demonstrated single-Cooper-pair pumping plateaus in both the bias voltage and dc level of the gate voltage; see Fig. 27. The quasiparticle poisoning was reported to be suppressed compared with the previous experiments, and hence they observed clear \( 2ef \) spacing of the current plateaus.

In addition to the above-mentioned pumping schemes, Nguyen et al. (2007) studied how a superconducting quantum bit referred to as a quantronium can be used to detect the gate charge ramp arising from a current bias on the gate electrode of the island of the device. The accuracy of this technique in converting the bias current into frequency remains to be studied in detail. Hoehne et al. (2012) studied another type of quantum bit, a charge qubit, for pumping Cooper pairs nonadiabatically. The aim here was to increase the pumping speed compared to adiabatic schemes but due to the accumulation of the errors from one pumping cycle to another, a waiting period between the cycles needed to be added. Furthermore, Giazotto et al. (2011) showed experimentally how phase oscillations can drive Cooper pairs in a system with no tunnel junctions. However, this kind of pumping was found to be very inaccurate in this proof-of-the-concept experiment.

F. Quantum phase slip pump

There is a proposal to build a source of quantized current based on the effect of QPSs in nanowires made of disordered superconductors (Mooij and Nazarov, 2006). Phase slip events occur in thin superconducting wires where thermodynamic fluctuations of the order parameter become significant (Arutyunov, Golubev, and Zaikin, 2008). During the phase slip, the superconducting order parameter vanishes at a certain instance and position in the wire, and the phase difference between the wire ends changes by \( 2\pi \). This gives rise to a voltage pulse in accordance with the Josephson relation. If the phase slips happen frequently, they produce a finite dc voltage or a finite resistance.

Phase slips caused by thermal activation broaden the temperature range of superconducting phase transitions and
produce a resistive tail below the critical temperature of a superconductor (Tinkham, 1996). At sufficiently low temperatures quantum fluctuations take over, and the residual temperature-independent resistivity of a nanowire can be attributed to the quantum phase slips [see Arutyunov, Golubev, and Zaikin (2008) and references therein]. Thermally activated phase slips are inherently incoherent. Quantum phase slips may be coherent provided dissipation associated with every switching event is suppressed. This can be achieved in superconductors with strong disorder in which Cooper pairs localize before the superconducting transition takes place (Feigel’man et al., 2007). Such a localization behavior has been observed by scanning tunneling microscopy in amorphous TiN and InOx films (Sacépé et al., 2010, 2011), which are believed to be the most promising materials for the observation of QPSs.

The key parameter describing a nanowire in the quantum phase slip regime is the QPS energy \( E_{\text{QPS}} = \hbar \Gamma_{\text{QPS}} \), where \( \Gamma_{\text{QPS}} \) is the QPS rate. Consider a superconducting nanowire of length \( L \) and sheet resistance \( R_{\square} \), and made of a superconductor with the superconducting transition temperature \( T_c \), and coherence length \( \xi = (\xi_0 \ell)^{1/2} \), where \( \xi_0 \) is the BCS coherence length and \( \ell \) is the electron mean free path (\( \ell \ll \xi_0 \)). Although there is no commonly accepted expression for \( \Gamma_{\text{QPS}} \), it is agreed (Mooij and Harmans, 2005; Arutyunov, Golubev, and Zaikin, 2008) that

\[
\Gamma_{\text{QPS}} \approx \exp\left(-0.15A \frac{R_K}{2R_{\square} \xi}\right)
\]

where \( A \) is a constant of order unity and \( R_K = \hbar/e^2 \). Clearly the exponential dependence of \( \Gamma_{\text{QPS}} \) on the wire resistance on the scale of \( \xi, R_\ell = R_{\square} \xi \), requires extremely good control of the film resistivity as well as the wire cross-sectional dimensions. For the nanowires to be in the quantum phase slip regime rather than in the thermally activated regime, \( E_{\text{QPS}} \) should exceed the energy of thermal fluctuations \( k_B T \). For a typical measurement temperature of 50 mK, \( \Gamma_{\text{QPS}}/2\pi \) should be higher than 1 GHz. Although the exact estimation of \( R_\ell \) is rather difficult, especially in the case of strongly disordered films, the experimental data presented by Astafiev et al. (2012) for InOx films agrees with the following values: \( \Gamma_{\text{QPS}}/2\pi = 5 \) GHz, \( R_\ell = 1 \) k\( \Omega \), and \( \xi = 10 \) nm.

The first experiment reporting the indirect observation of coherent QPS in nanowires was performed by Hongisto and Zorin (2012). They studied a transistorlike circuit consisting of two superconducting nanowires connected in series and separated by a wider gated segment. The circuit was made of amorphous NbSi and embedded in a network of on-chip 30-nm-thick Cr microresistors ensuring a high external electromagnetic impedance. The NbSi film had a superconducting transition temperature of \( T_c \approx 1 \) K and normal-state sheet resistance of about 550 \( \Omega \) per square. Provided the nanowires are in the regime of QPSs, the circuit is dual to the dc SQUID. The samples demonstrated appreciable Coulomb-blockade voltage (the analog of the critical current of the dc SQUID) and periodic modulation of the blockade by the gate voltage. Such behavior was attributed to the quantum interference of voltages in two nanowires that were in the QPS regime. This is completely analogous to the quantum interference of currents in a dc SQUID.

An unambiguous experimental evidence of a coherent QPS was provided in the work by Astafiev et al. (2012). Coherent properties of quantum phase slips were proven by a spectroscopy measurement of a QPS qubit, which was proposed earlier by Mooij and Harmans (2005). The qubit was a loop that had a 40-nm-wide and about 1-\( \mu \)m-long constriction. The loop was made of a 35-nm-thick superconducting disordered InOx film with \( T_c = 2.7 \) K and a sheet resistance of 1.7 k\( \Omega \) per square slightly above \( T_c \). The qubit was coupled inductively to a step-impedance coplanar waveguide resonator, which was formed due to the impedance mismatch between an indium oxide strip and Au leads to which it was galvanically connected. The ground planes on both sides of the strip were made of Au. At the qubit degeneracy point at a flux bias \( (m + 1/2)\Phi_0 \), where \( m \) is an integer, there is an anticrossing in the qubit energy spectrum with a gap \( E_{\text{QPS}} = \hbar \Gamma_{\text{QPS}} \). At this flux bias, the two quantum states \( |m \rangle \) and \( |m + 1 \rangle \) corresponding to the loop persistent currents circulating in the opposite directions are coupled coherently, which gives a gap \( E_{\text{QPS}} \) between the lowest-energy bands of the qubit. With the flux offset \( \delta \Phi \) from degeneracy, the gap evolves as \( \Delta E = [(2I_\rho \delta \Phi)^2 + E_{\text{QPS}}^2]^{1/2} \), where \( I_\rho \) is the persistent current in the loop. This gap was revealed in the spectroscopy measurements by monitoring the resonator transmission as a function of the external magnetic field and microwave frequency. When the microwave frequency matched the qubit energy gap, a dip in the transmission was observed. The width of the dip \( \sim 260 \) MHz close to the degeneracy point indicated rather strong decoherence whose origin is still to be understood.

Based on the exact duality of the QPS and the Josephson effects, it is argued that it should be possible to build a QPS electric current standard, which is dual to the existing Josephson voltage standard (Mooij and Nazarov, 2006). When biased resistively and irradiated by a high-frequency signal, QPS junctions exhibit current plateaus, which could provide the basis for the fundamental standard of the electric current. When an ac signal of frequency \( f \) is applied to a Josephson junction, Shapiro voltage steps \( V_n = n(h/2e) \), where \( n \) is an integer, are observed. Similarly, when an ac signal is applied to a QPS junction, an equivalent of Shapiro steps will occur in the form of plateaus at constant current levels \( I_n = n2ef \). One should note, however, that error mechanisms have not yet been analyzed for this type of quantized current source: thus it is not clear at the moment how accurate this source will be.

From a practical point of view, the realization of a QPS current source looks rather challenging because it requires fabrication of nanowires with an effective diameter \( \sim 10 \) nm as well as precise control of the sheet resistance \( R_{\square} \) of the nanowire, which is in the exponent of the expression for \( E_{\text{QPS}} \). Various approaches to the nanowire fabrication including the step decoration technique, sputtering of a superconductor on a suspended carbon nanotube, trimming of a nanowire by argon milling, etc. are described by Arutyunov, Golubev, and Zaikin (2008). Another issue is the overheating of the nanowire electron system. Assuming that the phase slip region becomes normal (which is true, for example, for Ti nanowires), for the estimation of the electron temperature, one can use Eq. (34) for the power transfer from electrons.
to phonons. A nanowire with the cross-sectional dimensions $20 \times 20$ nm$^2$, sheet resistance 1 kΩ per square, and carrying a dc current of 100 pA will have the effective electron temperature of the order of 250 mK, which is high enough to smear the current plateaus.

One of the first attempts to observe current plateaus on the current-voltage characteristics of superconducting nanowires under rf radiation was reported by Lehtinen, Zakharov, and Arutyunov (2012). The nanowires were made of Ti and had length up to 20 μm and effective diameter from 40 nm down to about 15 nm. The nanowire sheet resistance varied from about 20 Ω up to 1.9 kΩ per square. They were biased through high-Ohmic Ti or Bi leads having total resistance of 15 kΩ and 20 MΩ, respectively. The low-Ohmic samples biased through 15 kΩ exhibited a weak Coulomb blockade. The estimated $E_{\text{QPS}}$ was ≈ 0.1 μeV only. More resistive nanowires ($R_{\text{d}} = 180$ Ω, effective diameter = 24 nm) biased through 20 MΩ leads had a pronounced Coulomb blockade with a critical voltage of up to 0.4 mV. The thinnest nanowires ($R_{\text{d}} = 1.9$ kΩ, effective diameter ≈ 18 nm) exhibited a Coulomb gap of a few hundred millivolts with the largest gap exceeding 600 mV. These gaps did not vanish above $T_c$ of Ti, from which the authors concluded that some weak links were unintentionally formed in the thinnest nanowires. Despite the fact that the nanowires had large variations of parameters, all their current-voltage or $dV/dI$ characteristics exhibited some quasiregular features under the external rf radiation. Those features were interpreted as being current steps formed due to the phase locking of intrinsic oscillations by the external signal.

It is interesting to note that the physics of QPSs in superconducting nanowires resembles the physics of QPSs in Josephson junction arrays (Fisher, 1986). A nanowire can be modeled as a 1D array of small superconducting islands connected by Josephson junctions. The formation of isolated superconducting regions within a nominally uniform disordered film was confirmed experimentally (Sacépé et al., 2010, 2011). Such a weakly connected array of superconducting islands is characterized by the junction Josephson energy $E_J$ and the island charging energy $E_c$. The phase and charge dynamics of the 1D array depends on the ratio $E_J/E_c$. In the experiment by Pop et al. (2010) $E_J/E_c$ in a SQUID array was tuned in situ by applying a uniform magnetic flux through all SQUIDs. The state of the array was detected by an extra shunt Josephson junction. They deduced the effect of the quantum phase slips on the ground state of the array by measuring the switching current distribution of the entire Josephson circuit as a function of the external magnetic flux for different values of $E_J/E_c$.

G. Other realizations and proposals

In this section we cover various ideas that have been brought up for experimental demonstration. Although their metrological relevance is still to be proven, we present them for their complementarity, potential, and for completeness.

1. ac-current sources

The current pumps described in Secs. III.A–III.C can be considered as single-electron injectors generating dc current. Coulomb blockade ensures a good control of the electron number on an island during the charge transfer.

A time-controlled single-electron source generating ac current was reported by Fève et al. (2007). The source was made of a GaAlAs/GaAs quantum dot tunnel coupled to a large conductor through a quantum point contact (see Fig. 28). A magnetic field $B \approx 1.3$ T was applied to the sample so as to work in the quantum Hall regime with no spin degeneracy. The discrete energy levels of the quantum dot were controlled by the pulse voltage $V_{\text{exc}}$ applied to the top gate and by the QPC dc gate voltage $V_g$, which also controlled the transmission $D$ of a single edge state. The dot addition energy $\Delta + e^2/C \approx 2.5$ K was determined by the energy-level spacing $\Delta$ as the Coulomb energy $e^2/2C$ was negligibly small. As proposed by Gabelli et al. (2006), this circuit constitutes an effective quantum-coherent RC circuit with the effective quantum resistance $R$ and capacitance $C$ defined as $R = \hbar/2e^2$ and $C = e^2/(dN/dx)$, where $dN/dx$ is the local density of states of the mode propagating in the dot, taken at the Fermi energy (Prêtre, Thomas, and Büttiker, 1996).

The single-charge injection was achieved by the application of a high-amplitude excitation voltage $V_{\text{exc}} \sim \Delta/e$ to the top gate, which leads to the electron escape from the dot at a typical tunnel rate $\tau^{-1} = D\Delta/\hbar$, where $\Delta/\hbar$ is the attempt frequency. Typically, the tunnel rates are in the nanosecond time scales, and this makes single-shot charge detection a challenging task. To increase the signal-to-noise ratio, a statistical average over many individual events was used by repeating cycles of single-electron emission with period $T$ followed by single-electron absorption (or hole emission) as shown in Fig. 28. This was done by applying a periodic square-wave voltage of amplitude $\approx \Delta/e$ to the top gate.

When the charge on the dot is well defined, repeatable single-electron injection leads to quantization of the ac...
current. $|I_{\omega}|$ as a function of $V_{\text{exc}}$ for two values of the dc dot potential at $D = 0.2$ and $D = 0.9$ is shown in Fig. 29. The transmission $D = 0.2$ is low enough and the electronic states in the dot are well resolved, as shown in the inset of Fig. 29 (left). On the other hand, the transmission is large enough for the escape time to be shorter than $T/2$. When the Fermi energy lies exactly in the middle of the density-of-states valley (rightmost vertical line in the left inset), a well-pronounced $|I_{\omega}| = 2ef$ current plateau is observed centered at $2eV_{\text{exc}}/\Delta = 1$. It is claimed that the current uncertainty at the plateau is 5% due to the systematic calibration error. In contrast, if, with the same transmission, the Fermi energy lies on the peak (middle vertical line in the left inset), there is still a current plateau, but it is not as flat and it is sensitive to parameter variations. When transmission is increased, the charge fluctuations become stronger and the plateau gets narrow and finally nearly vanishes at $D = 0.9$ even for the optimal working point, as seen from Fig. 29 (right). The experimental results (dots) are compared with the theoretical model, the solid lines [1D modeling of the circuit described by Gabelli et al. (2006) was used], showing excellent agreement between the two.

The device above has been described as an electron analog of the single-photon gun. It is not a source of quantized dc current as the dot emitting the electron can be recharged only though the reverse process of electron absorption. Using a similar technique of electron emission with fast pulses, but adding one more lead, one can produce a highly accurate dc current (Giblin et al., 2012) as described in Sec. III.C.4. Recent correlation experiments on electron guns have been reported in Bocquillon et al. (2012).

2. Self-assembled quantum dots in charge pumping

The idea of using self-assembled quantum dots for charge pumping is based on conversion of optical excitation into deterministic electric current; see Nevou et al. (2011). In the experiment of Nevou et al. (2011) a plane of self-assembled InAs quantum dots is coupled to an InGaAs quantum-well reservoir through an AlGaAs barrier [see Figs. 30(a)–30(c)]. The structure is sandwiched between two n-doped GaAs regions. The device basically works as a strongly asymmetric quantum-dot infrared photodetector (Nevou et al., 2010). In the absence of any optical excitation, electrical conduction is inhibited by the AlGaAs barrier. When a laser pulse ionizes the quantum dots, a fixed number of electrons are excited out of the dot and then swept away by the applied bias voltage, giving rise to a photocurrent. After that, the dots will be refilled from the electron reservoir by tunneling through the AlGaAs barrier. If the process is repeated at a frequency $f$, the current will be given by $I = nef$, where $n$ is determined by the number of dots and the number of electrons per dot. To obtain the desired pumping accuracy, the laser pulse duration time must be much shorter than the refilling time, and $f$ should not exceed the kilohertz range. Even with such a low repetition frequency, currents in the nanoampere range can be generated as the number of dots running in parallel in a typical device is tens of millions. There are error sources arising from the uncertainty in the number of quantum dots contained in one device, as well as...
the variability in the quantum-dot transition energy. These errors were of the order of 10% of the pumped current in the experiment of Nevou et al. (2011).

3. Mechanical single-electron shuttles

Besides charge pumps with entirely electronic control, there is a group of devices in which a mechanical degree of freedom is involved. They are called mechanical charge shuttles, because they transfer either single charges (electrons or Cooper pairs) or portions of charges between the two electrodes due to the mechanical back-and-forth motion of a small island between the electrodes. This results in current flow, either incoherent or coherent. The concept of the mechanical electron shuttle was introduced by Gorelik et al. (1998) and Isacsson et al. (1998).

The proposed device has a small conducting island, which is mechanically attached to electrical leads with the help of an elastic insulator. The dc voltage applied between the leads and elastic properties of the insulator together with charging and discharging of the island creates instability and makes the island oscillate; see Fig. 31(a). For the proper operation of the shuttle, two assumptions were made: the amplitude of the mechanical oscillations is much larger than the electron tunneling distance, and the number of electrons on the island is limited. With these assumptions the island motion and charge fluctuations become strongly coupled. Depending on the shuttle details, two regimes can be distinguished: classical (Gorelik et al., 1998; Isacsson et al., 1998; Weiss and Zwerger, 1999) and quantum mechanical (Armour and MacKinnon, 2002; Fedorets et al., 2004; Johansson et al., 2008; Cohen, Fleurov, and Kikoin, 2009). The shuttle, when made superconducting, can transfer not only electrons but also Cooper pairs (Gorelik et al., 2001; Shekhter et al., 2003).

The first experimental realization of a mechanical charge shuttle that operated due to a shuttle instability was reported by Tuominen, Krotkov, and Breuer (1999). This was a rather bulky device even though it was scaled down considerably in size and operating voltage in comparison to the earlier electrostatic bell versions. The observed jumps of the current as a function of the bias voltage as well as hysteresis in the transport characteristics were the main indications of the shuttling regime of the device. A nanoscale version of the instability-based electron shuttle was implemented by Kim, Qin, and Blick (2010); see Fig. 31(b). The device was a Si pillar covered on top with a thick gold layer and placed in the gap between two electrodes, the source and the drain, of the central line of a coplanar waveguide. For reference, they also fabricated and measured a similar device without a pillar in the gap. The samples were measured at room temperature in vacuum. The pillar was actuated by applying a small rf signal together with a dc bias voltage across the source and drain electrodes. A clear frequency dependence was observed for the sample with a pillar in the gap, with the resonance frequency of 10.5 MHz and quality factor of about 2.5. It was estimated that the device shuttles on average 100 electrons per cycle.

Another realization of the nanoelectronic shuttling device was reported by Moskalenko et al. (2009a, 2009b). It had the configuration of a single-electron transistor, whose island was a gold nanoparticle placed in between the Au source and drain electrodes by means of an atomic force microscope. Current-voltage characteristics of the devices were measured at room temperature, and characteristic current jumps in the current-voltage curves were observed, which were attributed to the shuttling effect. They also compared characteristics of the working shuttle device from which the nanoparticle was removed. After this procedure, the current through the device dropped below the noise level.

The effect of the mechanical vibrational modes on charge transport in a nanoelectronic device was observed in a C60 single-electron transistor (Park et al., 2000). In this device, a single C60 molecule was placed in the narrow gap between the two gold electrodes. It was found that the current flowing...
through the device increases sharply whenever the applied voltage was sufficient to excite vibrations of the molecule. Although mechanical vibrations were observed in this structure, they are not related to the shuttling of electrons.

Externally driven resonant shuttles may be easier to implement in comparison to the instability-based shuttles, because a much larger displacement amplitude can be achieved. However, the drawback of using a mechanical resonator is the discrete set of eigenfrequencies, which are determined by geometry and materials. Therefore, only a limited number of frequencies are available for electron transfer. In the experiment performed by Erbe et al. (1998) a Si nanomechanical resonator was placed in between two contacts. The whole device, being a scaled-down version of the classical bell, was functioning as a mechanically flexible tunneling contact operating at radio frequencies. The contact was driven by $\pi$-shifted oscillating voltages applied on two gates. The current-frequency dependence of the device contained strong peaks, which were interpreted as being due to the mechanical resonances of the beam, indicating that shuttling was occurring. The peaks had low quality factors, ranging from 100 to 15 only. The number of electrons $N$ shuttled per cycle was estimated from the current peak height $I$ using the $N = I/\text{ef}$. Below 20 MHz, $10^3$–$10^4$ electrons were shuttled in each cycle. On the 73 MHz peak the number was decreased to about 130 electrons per cycle.

The same group (Erbe et al., 2001) fabricated a singly clamped beam with a metal island on its end (see Fig. 31(c)). It was found at 300 K that there was no detectable current through the device unless a driving ac voltage ($\pm 3 \text{ V}$) was applied to the driving gates. Under the external drive, the current exhibited several peaks, similar to those in the earlier device (Erbe et al., 1998), which was attributed to the beam motion. The background current was explained by the thermal motion of the beam. At 4.2 K all the current peaks were suppressed except one at about 120 MHz with much smaller height (only 2.3 pA). This corresponded to shuttling on average of 0.11 electron per cycle.

Koenig, Weig, and Kotthaus (2008) implemented electromechanical single-electron transistors with a metallic island placed on a doubly clamped SiN beam (see Fig. 31(d)), which was measured at a temperature of 20 K. The observed resonance features in the SET dc current were attributed to the mechanical resonances. It was argued that the mechanical motion of the resonator was strongly nonlinear. This was imposed by the side electrodes constituting the impacting boundary conditions. The nonlinear nature of the system resulted in a shape of the resonance curves different from Lorentzian. Although the expected steplike dependence of the SET current on the source-drain voltage was not observed (because of the high measurement temperature as compared to the charging energy), they made an optimistic conclusion that the device may be useful for quantum metrology.

### 4. Electron pumping with graphene mechanical resonators

An electron pump based on a graphene mechanical resonator in the fundamental flexural mode was introduced by Low et al. (2012). The resonator is actuated electrostatically by a gate electrode. Time-varying deformation of graphene modifies its electronic energy spectrum and in-plane strain. Cyclic variation of these two properties constitutes the scheme for quantum pumping. To have a nonzero pumping current, spatial asymmetry must be introduced. It is assumed that the contacts between the graphene layer and the left and right electrodes are not equivalent, which is modeled by different densities of states. This can be achieved in the experiment by using different materials for the two electrodes. It is emphasized that Coulomb-blockade effects will favor the transfer of an integer number of electrons per cycle, so that the relation between current and frequency will be quantized. This is just a proposal and the applicability of this approach for quantum metrology is still to be verified.

### 5. Magnetic- field-driven single-electron pump

Another proposal, not implemented though, is based on using a ferromagnetic three-tunnel-junction device for electron pumping (Shimada and Ootuka, 2001). Its islands and leads are made of ferromagnetic metals with different coercive forces. Such a device can be operated as a single-electron pump if controlled by ac magnetic fields, and not by the gate voltages. In addition to the charging effects, it makes use of the magnetic-field-induced shift of the chemical potential and magnetization reversal in the ferromagnetic electrode.

The proposed device has intrinsic limitations of the pumping speed, which are determined by the physical time constraints of the ferromagnet. The pump operation frequency must be much lower than the characteristic relaxation times. The prospects of this type of an electron pump for quantum metrology are still to be understood.

### 6. Device parallelization

As discussed in Sec. III.A, it is possible to reach precise electron pumping with a normal-metal single-electron pump consisting of a sufficiently long array of islands. With six islands and seven junctions, the accuracy of the pumped current is at the $10^{-7}$ level. However, the maximum current is limited to a few picoamperes. To get the current scaled up to the 100 pA level, a requirement for practical metrological applications (see Sec. IV.D.1), approximately 100 pumps should be operated in parallel. The main reason why parallelization is impractical for normal-metal devices is the tuning of the offset charges (Keller et al., 1996; Camarota et al., 2012). Each island has an individual offset charge that has to be compensated separately. Therefore, a metrological current source implemented as parallelized normal-metal pumps would require of the order of 1000 dc lines.

Compared to normal-metal pumps, quantum-dot-based pumps allow for higher pumping speeds using fewer control lines thanks to their tunability; see Sec. III.C. Accuracy of 1.2 ppm at an output current of 150 pA has already been demonstrated with a single quantum dot (Giblin et al., 2012). Therefore, parallelization of such pumps may not even be required if the accuracy can be improved without a loss in speed. Nevertheless, parallelization of semiconducting pumps has been considered in the literature. With two pumps, invariance with respect to gate variations has been shown to be below the 20 ppm level (Wright et al., 2009) with output current exceeding 100 pA. In this case, all signals were
individually tuned for each device requiring two dc and one rf signals per device. However, it is possible to use common signals for rf drive and for the barrier voltages (Mirovsky et al., 2010). In this case, only one dc voltage per device is required for tuning the other barrier and possible offset charges. The obtainable accuracy, depending on device uniformity, is still an open question for this approach.

For the hybrid NIS turnstiles, the maximum current per device is limited to a few tens of picoamperes, as discussed in Sec. III.B.2. Hence, at least ten devices are to be run in parallel, which has been shown to be experimentally feasible (Maisi et al., 2009). In Fig. 32 we show a scanning electron micrograph of a sample used in that work and the main experimental findings. The turnstiles in these experiments suffered from photon-assisted tunneling due to insufficient electromagnetic protection (see Sec. II.F), and hence the quantization accuracy was only on the $10^{-3}$ level. Improved accuracy is expected for a new generation of turnstile devices (Pekola et al., 2010). For parallel turnstiles, a common bias voltage can be used if the devices have roughly equal output impedance. The curves are taken at different rf drive and for the barrier voltages (Mirovsky et al., 2010). In this case, only one dc voltage per device is required for tuning the other barrier and possible offset charges. The obtainable accuracy, depending on device uniformity, is still an open question for this approach.

II. Single-electron readout and error correction schemes

1. Techniques for electrometry

The electrometer used to detect the presence or absence of individual charge quanta is a central component in schemes for assessing pumping errors and error correction. Figure 33(a) introduces the essential components of an electron-counting setup. In order to observe proper charge quantization, the counting island is connected to other conductors only via low-transparency tunnel contacts. The electrometer is capacitively coupled to the counting island and biased such a manner that the small voltage drop of the counting island due to change of its charge state by one

FIG. 33 (color online). (a) Circuit diagram of a charge-counting device. Electric charge $Q$ on the island on the left is monitored. The island is coupled to an electrometer island via capacitor $C_j$ and also tunnel coupled to an external conductor. The single-electron box configuration illustrated here requires only one tunnel junction with capacitance $C_f$. In addition, there is capacitance $C_0$ to ground, which accounts also for gate electrodes and any parasitic capacitances. The probing current $I_{det}$ through the detector is sensitive to the charge on the coupling capacitor, which is a fraction $C_x/C_j$ of the total charge $Q$, where $C_x = C_j + C_0$. The detector is a single-electron tunneling transistor based on Coulomb blockade, and hence the total capacitance of the detector island $C_{det}$ is of the order of 1 fF or less. (b) Circuit diagram of a general noisy electrical amplifier that can also be adapted to describe the electrometers of single-electron experiments. From Devoret and Schoelkopf, 2000. For the configuration shown in (a), one has for input impedance $Z_{in}(\omega) = 1/j\omega C_{in}$, where $C_{in}^{-1} = C_x^{-1} + C_{det}^{-1}$. The input voltage is related to the island charge $Q$ through $V_{in} = Q/C_x$. The noise source $I_N$ represents backaction and $V_N$ is the noise added by the electrometer at the output referred to the input. The gain of the amplifier is given by $G$. The output impedance $Z_{out}$ equals the differential resistance at the amplifier operation point.

FIG. 32 (color online). (a) Scanning electron micrograph of parallel turnstiles. The turnstiles are biased with a common bias $V_b$ and driven with a common rf gate voltage $V_{rf}$. Gate offset charges are compensated by individual gate voltages $V_{g,i}$. (b) Output current $I$ for ten parallel devices tuned to the same operating point producing current plateaus at $I = 10Nef$. The curves are taken at different $V_b$ shown in the top left part of the panel. From Maisi et al., 2009.
such a detector is modeled as a linear voltage amplifier, fluctuations at signal frequency and the quantum theory limit for the spectral density of noise limit, where $\kappa$ is the intrinsic shot noise of the electrometer was comparable to the noise of the readout electronics (Brenning et al., 2006; Kafanov and Delsing, 2009). For the normal-state SET, sensitivities of the order of $10^{-7} e/\sqrt{Hz}$ are attainable in theory with present-day fabrication technology, where the intrinsic noise is due to stochastic character of the tunneling processes and includes both shot and thermal noise (Korotkov, 1994; Korotkov and Paulanen, 1999). The best charge sensitivities reported to date for a single-electron transistor by Brenning et al. (2006) were almost identical in normal and superconducting states, namely, 1.0 and $0.9 \times 10^{-6} e/\sqrt{Hz}$, respectively, at a signal frequency of 1.5 MHz. Xue et al. (2009) also measured the backaction of a superconducting SET and the product of noise and backaction was found to be 3.6 times the quantum limit. For QD charge detection with QPCs, charge sensitivity of $2 \times 10^{-4} e/\sqrt{Hz}$ referred to the QD charge has been demonstrated (Cassidy et al., 2007). It appears to be easier to realize large charge coupling fraction $\kappa$ with metallic SETs than with QPCs (Yuan et al., 2011).

We now discuss the backaction mechanisms in more detail; see a schematic illustration in Fig. 34. Despite the above quantum theory result connecting backaction and noise, the electronic backaction of the electron counter can be addressed in principle independently of its charge noise, as the readout bandwidth (at most 100 MHz, see below) is much below the microwave frequencies that can promote charge transfer errors: Overcoming even a modest 100 $\mu$V energy barrier requires photon frequencies above 24 GHz if multiphoton processes are neglected. Nevertheless, voltage fluctuations induced by the shot noise of the detector usually have a non-negligible spectral density at microwave frequencies. A fraction $\kappa$ of the voltage fluctuations of the SET island are coupled back to the counting node. This mechanism can dominate the equilibrium thermal...

![FIG. 34. Detector backaction mechanisms. The backaction can originate by direct electromagnetic (EM) coupling either by variations in pump biasing or by high-frequency photon-assisted tunneling (PAT). Another source of backaction is via heat conduction. The detector located in proximity of the device typically heats up. The heat can then be conducted to the device by either phononic or photonic coupling.](image-url)
noise from resistive components at the sample stage; cf. Martinis and Nahum (1993). A full quantum calculation of the backaction of an SET electrometer on a Cooper-pair box has been presented by Johansson, Kääck, and Wendin (2002). Lotkhov and Zorin (2012) measured the effect of photon irradiation by a nearby SET on the hold time of an electron trap. They found that the rate of electron escape events was proportional to the theoretically calculated emission rate of photons having an energy larger than the energy barrier of the trap. Saiara, Kemppinen et al. (2012) suggest attenuating the high-frequency backaction by replacing the capacitive coupling by a lossy wire that acts as a low-pass filter for microwaves but does not affect the charge signal. For QPCs, Coulombic backaction can be divided into shot noise, which can be in principle eliminated by circuit design and fundamental charge noise (Aguado and Kouwenhoven, 2000; Young and Clerk, 2010).

The low-frequency part of detector backaction manifests itself as variation of the dc bias of the pump or turnstile device. The case of an SET electrometer coupled to a single-electron box was studied by Turek et al. (2005). In the limit of small coupling capacitance \( C_x \), the voltage swing on the counting island due to loading and unloading the detector island is given by \( \Delta V_h = \kappa e/C_{det} \). We note that this is just a fraction \( C_x/C_{det} < 1 \) of the voltage swing from loading or unloading the actual counting island with an electron. Hence, dc backaction of the detector does not necessarily place an additional constraint on the design of the electron-counting circuit.

In addition to the electronic backaction described above, one needs to consider the phononic heat conduction from the detector to the charge pump. For reaching the ultimate accuracy, the charge pumps typically require temperatures of the order of 100 mK or lower, where small on-chip dissipation can raise the local temperature significantly due to vanishing heat conductivity in the low-temperature limit (Kautz, Zimmerli, and Martinis, 1993; Giazotto et al., 2006); see also Sec. II.G. The average power dissipated by the detector is given by \( P = \langle I_{det} V_{det} \rangle \), and it needs to be transported away by the substrate phonons or electronically via the leads. Requirement for a sufficiently large charge coupling coefficient \( \kappa \) limits the distance by which the detector and charge pump can be separated. The temperature increase by dissipated power has been studied on a silicon substrate by Savin et al. (2006) and they give

\[
T = \left( T_0^4 + \frac{2fP}{\pi^4\nu_0} \right)^{1/4},
\]

where \( T \) is the substrate temperature at distance \( r \) from a point source of heating power \( P \), \( T_0 \) is the bath temperature, \( \nu_0 = 3600 \text{ W m}^{-2} \text{K}^{-4} \) is the material parameter, and \( f = 0.72 \) is a fitting parameter for their experimental observations. For an illustrative example, we estimate that the dissipated power at the electrometer in the original rf-SET paper (Schoelkopf et al., 1998) was 120 fW based on the published numbers. According to Eq. (40), this will heat the substrate underneath nearby junctions \( (r = 200 \text{ nm}) \) to 140 mK, which is high enough to deteriorate the performance of many single-electron devices below the metrological requirements. Sillanpää, Roschier, and Hakonen (2004) coupled the readout to the Josephson inductance of a superconducting SET instead of conductance, reducing the dissipation by 2 orders of magnitude. Usually it is possible to assess the severity of detector backaction effects in the experiment by measuring the tunneling rates using different values of \( I_{det} \) [see, e.g., Kemppinen et al. (2011), Lotkhov et al. (2011), and Saiara, Kemppinen et al. (2012)], so that any variation of the observed rates can be attributed to backaction. The picture is somewhat different in 2DEG systems due to significantly weakened electron-phonon coupling. Experimental study of phononic backaction in 2DEGs is presented by Schinner et al. (2009) and Harbusch et al. (2010).

The bandwidth of the readout, \( B \), is commonly defined as the corner frequency of the gain from gate charge to output voltage (Visscher et al., 1996). The performance requirements for the charge readout depend on the particular charge-counting scheme, but in general the bandwidth \( B \) places a limit on the fastest processes that can be detected and hence constrains the magnitude of the electric current that can be reliably monitored. In practice, \( B \) is limited by the inverse \( RC \) constant of the electrometer’s differential resistance and the capacitive loading on its outputs. Both the QPC and SET electrometers have an impedance of the order of \( R_K = h/e^2 \approx 25.8 \text{ k} \Omega \). For the QPC, \( R \approx R_K \) is required to realize strong Coulomb blockade according to the orthodox theory of single-electron tunneling (Averin and Likharev, 1991; Ingold and Nazarov, 1992). For a QPC, the most charge-sensitive operation point is around a bias point where \( \partial V/\partial I \approx R_K \), midway between the first conductance plateau and pinch-off (Cassidy et al., 2007). As the barrier capacitance is of the order of 1 fF or less for the devices, the intrinsic bandwidth is in the gigahertz range. In practice, the capacitance of the biasing leads and the input capacitance of the preamplifier dominate. When the preamplifier is located at room temperature as in the pioneering experiments (Fulton and Dolan, 1987; Kuzmin et al., 1989), the wiring necessarily contributes a capacitance of the order of 0.1–1 nF and henceforth limits the readout bandwidth to the kilohertz range (Pettersson et al., 1996; Visscher et al., 1996). Readout by a current amplifier from a voltage-biased SET avoids the RC cutoff on the gain, but the usable bandwidth is not substantially altered as current noise increases at high frequencies where the cabling capacitance shorts the current amplifier input (Starmark et al., 1999).

In order to increase the effective readout bandwidth, the SET impedance has to be transformed down toward the cable impedance, which is of the order of 50 Ω. Bandwidths up to 700 kHz have been achieved by utilizing a high-electron-mobility transistor (HEMT) amplifier with a low impedance output at the sample stage (Pettersson et al., 1996; Visscher et al., 1996). The dissipated power at the HEMT in these studies was 1–10 μW depending on the biasing, which can easily result in overheating of the electrometer and/or the coupled single-electron device. The best readout configuration to date is the rf reflectometry technique, applicable to both SETs (Schoelkopf et al., 1998) and QPCs (Qin and Williams, 2006), where the electrometer is embedded in a radio frequency resonant circuit and the readout is achieved by measuring the damping of the resonator. A readout bandwidth of 100 MHz was achieved in the original demonstration.
Requiring \( /C_{14M} < pN \) to be satisfied. They also note that their charge sensitivity of \( 1 \times 10^{-5} e/\text{Hz} \) yields \( \Delta q = 0.1e \) for the full detection bandwidth, i.e., electron counting at 100 MHz would have been possible in a scenario where the charge coupling fraction \( \kappa \) was close to unity.

2. Electron-counting schemes

Realization of a current standard based on electron counting has been one of the key motivators for development of ultrasensitive electrometry (Schoelkopf et al., 1998; Gustavsson et al., 2008; Keller, 2009). First we see why direct current measurement of uncorrelated tunneling events, like those produced by a voltage-biased tunnel junction, cannot be used for a high-precision current standard: Assume a noise-free charge detector that yields the charge state of the counting island with time resolution \( \tau = 1/B \), and that Markovian (uncorrelated) tunneling events occur at the rate \( \Gamma \ll B \). With probability \( \Gamma \tau \), a single tunneling event occurs during the time \( \tau \) and is correctly counted by the detector. With probability \( (\Gamma \tau)^2/2 \), two tunneling events occur within \( \tau \) and constitute a counting error. Hence, to achieve a relative error rate \( p \), one needs \( \Gamma < 2pB \). Even with a noiseless 100 MHz rf SET, one could not measure a direct current greater than \( 2e/\text{s} \) at metrological accuracy \( p = 10^{-8} \) in this manner. Would it be practical to account statistically for the missed events in a manner similar to Naaman and Aumentado (2006) assuming truly Poissonian tunneling statistics and a well-characterized detector? The answer is unfortunately negative: If \( N \) tunneling events are observed, the number of missed events \( M \) is a Poissonian variable with a mean of \( N \Gamma \tau /2 \) and standard deviation \( \delta M = \sqrt{N \Gamma \tau /2} \). Requiring \( \delta M < pN \) gives \( N > \Gamma \tau / (2p^2) \). For \( \Gamma = 1 \text{ MHz} \) and \( \tau \) and \( p \) as above, one has to average over \( N > 5 \times 10^{13} \) events, which is impractical. A more detailed calculation based on Bayesian inference presented by Gustavsson et al. (2009) results in the same \( N \) dependence.

Charge transport through a 1D array of tunnel junctions can take place in the form of solitons depending on device parameters (Likharev, 1988; Likharev et al., 1989). Propagation of the solitons promotes time correlation in the electron tunneling events, allowing the accuracy limitations of counting uncorrelated electrons presented above to be lifted. A proof-of-concept experimental realization has been presented by Bylander, Duty, and Delsing (2005). The array is terminated at the middle island of an SET, allowing for unity charge coupling, and a signal centered around frequency \( f_c = I/e \) is expected. They claim a possible accuracy of \( 10^{-6} \) based on the charge sensitivity of their electrometer only. However, the spectral peaks in the experimental data appear too wide for an accurate determination of the center frequency. Factors not included in the accuracy estimate are the instability of the bias current and SET background charge fluctuations.

Single-electron electrometry can be used to count the much rarer pumping errors instead of the total pumping current. Such an approach has been used to study the accuracy of metallic multijunction pumps that are used in the electron-counting capacitance standard (ECCS) (Keller et al., 1999; Keller, Zimmerman, and Eichenberger, 2007). A circuit diagram of an ECCS experiment is shown in Fig. 35. Two cryogenic needle switches are required to operate the device in different modes: determining coupling capacitances and tuning the pump drive signal (NS1 and NS2 closed), operating the pump to charge \( C_{\text{cryo}} \) (NS1 closed, NS2 open), and comparing \( C_{\text{cryo}} \) with an external traceable capacitor (NS1 open, NS2 closed). The SET electrometer is used as part of a feedback loop that maintains the voltage of the node at the end of the pump constant. With NS1 open, the pump can be operated in a shuttle mode: a charge of \( e \) is repeatedly pumped back and forth across the pump at the optimal operation frequency (which is above the detector bandwidth), and pumping errors appear as discrete jumps in the electrometer output.

An error rate of \( 1.5 \times 10^{-8} \) relative to the used shuttling frequency of 5.05 MHz was demonstrated in the NIST ECCS setup (Keller et al., 1996). In this experiment, the seven-junction pump illustrated in Fig. 13 was used. At Physikalisch-Technische Bundesanstalt (PTB), a relative error rate of the order of \( 10^{-7} \) was reported (Camarota et al., 2012) for a five-junction \( R \) pump operated at a shuttling frequency of 0.5 MHz. Recently, the PTB group improved their relative error rate to \( 4.4 \times 10^{-8} \) (Scherer et al., 2012). They also argued that in order to account for the possibility of pumping errors in opposite directions to cancel out each other, the proper measure of the pump accuracy in an ECCS experiment is \( \sqrt{\Gamma_{\text{err}}/\Delta f/f} \), where \( \Gamma_{\text{err}} \) is the absolute rate at which pumping errors occur, \( \Delta f \) is the time spent at charging the capacitor, and \( f \) is the pumping frequency. Using this methodology, they inferred a relative accuracy of \( 1.5 \times 10^{-8} \) for their pump. A complete ECCS experiment has not been performed with a semiconductor pump to date. Yamahata, Nishiguchi, and Fujiwara (2011) described a single-electron shuttling experiment performed on a Si nanowire, but the reported error rates are rather high, at the \( 10^{-2} \) level. In semiconductor realizations, the memory node can be isolated from the rest of the lead by a FET switch that is defined with the same lithography process as the pump, eliminating the need for the needle switch.

Wulf (2012) proposed an error-accounting architecture, where a few modestly accurate charge pumps are connected...
in series, and electrometers are used to observe the charge on memory nodes situated in between the pumps. Assuming that the error rate of the pumps is small in comparison to the readout bandwidth, the sign of individual pumping errors can be reliably inferred from the electrometer response. Hence, the accuracy of the pumped current can then be improved beyond the accuracy of the constituent pumps.

To date, an error-counting or -accounting algorithm integrated with continuously operating series-connected pumps has not been demonstrated experimentally. Measurements of two series-connected semiconductor QD pumps with a QPC electrometer coupled to a node in the middle were presented by Fricke et al. (2011), although quantized pumping errors were not observed. Recently, an initial report of on-demand single-electron transfer in a device consisting of three QD pumps and two metallic SET electrometers was presented (Fricke et al., 2012, 2013). Although no detectable current was produced in this experiment due to a low repetition rate of 5 Hz, they were able to distinguish between different types of pumping errors from the electrometer signature.

I. Device fabrication

Fabrication of charge pumps, regardless of their operational principle, requires advanced nanofabrication methods. These include, for example, electron-beam lithography, various dry etching techniques, and molecular-beam epitaxy growth of semiconductor heterostructures. In general, pumping devices can have small feature sizes in multiple layers that must be accurately aligned with each other. We begin with the description of the fabrication procedure for the metallic single-electron and Cooper-pair pumps and turnstiles described in Secs. III.A, III.B, and III.E. Subsequently, we present the fabrication methods for quantum-dot pumps and turnstiles, the operation of which is discussed in Sec. III.C.

I. Metallic devices

Metallic single-electron and Cooper-pair pumps and turnstiles are typically made by the angle deposition technique, which was first introduced by Dolan (1977) for the photolithography process and then later adapted by Dolan and Dunsmuir (1988) for the electron-beam lithography process. We note that there is a myriad of different ways of fabricating these devices. Below we describe only a certain fabrication process for these devices in great detail instead of giving a thorough study of all possible variations.

The process starts with the deposition of an Au layer on an Si wafer covered by a native silicon oxide. The Au pattern is formed by a standard photolithography and lift-off process using photore sist S1813 and contains contact pads and on-chip wiring as well as alignment markers for the deposition of the subsequent layers. Next, a trilayer resist structure is built (from bottom to top): copolymer/Ge/poly-methylmethacrylate (PMMA) with the thicknesses 200, 20, and 50 nm, respectively [see Fig. 36(a)]. The polymer layers are spin coated on the wafer and baked in a nitrogen oven, and the Ge layer is deposited in an electron gun evaporator. The wafer is then cleaved into smaller pieces which are exposed and processed separately. After the exposure of the top PMMA layer on one of the pieces in the electron-beam writer, the chips with masks are placed in an electron gun evaporator equipped with a tilting stage. Two consecutive depositions of metal through the same mask are carried out at different angles to create a partial overlap between the metal layers [see Fig. 36(e)]. If the surface of the bottom layer (typically Al) is oxidized by introducing oxygen into the evaporation chamber, after the deposition of the top electrode, the sandwich structure composed of the overlapping metal layers with a thin oxide in between forms small tunnel junctions [see Fig. 36(f)].

The normal-metal or superconducting charge pumps are made entirely of Al, which can be turned normal at low temperatures by an external magnetic field [see Geerligs et al. (1990, 1991), Pothier et al. (1992), Keller et al. (1996), and Vartiainen et al. (2007)]. In the case of the hybrid
structures described in Pekola et al. (2008), Kemppinen, Kafanov et al. (2009), Kemppinen, Meschke et al. (2009), and Maisi et al. (2009), the bottom electrode was Al and the top one was either Cu or AuPd.

2. Quantum dots

The gate structure of the charge pumps based on quantum dots is also fabricated using electron-beam lithography. The main differences in the fabrication compared with metallic devices are the following: Ohmic contacts have to be made between metallic bonding pads on the surface of the chip and the 2DEG located typically \( \sim 100 \) nm below the surface. Furthermore, the 2DEG has to be either depleted with negative gate voltage from the unwanted positions in the case of GaAs devices [see Fig. 18(d)] or accumulated with positive gate voltage in the case of MOS silicon devices [see Fig. 18(c)]. For GaAs, also etching techniques have been employed to dispose of some parts of the 2DEG leading to the formation of MOS silicon quantum dots. We begin with a high-resistivity GaAs devices, typically a single deposition of metal through a monolayer PMMA resist and the use of electron-beam lithography with a 200 nm PMMA resist and an electron gun evaporator is employed to deposit the first layer of aluminum gates inside the window and their bonding pads outside the window. After the lift-off, the gates are passivated by an 8-nm-thick high-resistivity oxide on top. All the following process steps involving metal evaporation with an electron gun evaporator is employed to deposit the first layer of aluminum gates inside the window and their bonding pads outside the window. After the lift-off, the gates are passivated by an Al(100 nm) layer formed by thermal evaporation of 100 nm of aluminum on the Al regions, forming a connection to the bonding pads made by depositing metal on these etched regions. The field oxide is etched selectively above the other ends of the metallic bonding pads on the surface of the chip and the 2DEG located typically \( \sim 100 \) nm away from the metal. The Ohmic contacts and the bonding pads are made by depositing metal on these etched regions, forming a connection to the metallic devices employing angle evaporation.

We now describe in detail a fabrication process for MOS silicon quantum dots. We begin with a high-resistivity \( (\rho > 0 \) kΩ cm at 300 K) near-intrinsic silicon wafer. Phosphorus atoms are deposited on the silicon surface using standard photolithography and they diffuse to a depth of roughly 1.5 μm during the growth of a 200-nm-field silicon oxide on top. All the following process steps involving wet oxidation or deposition have to be aligned with the previous ones with the help of alignment markers, a routine we do not discuss separately. Then a window and their bonding pads outside the window. After the lift-off, the gates are passivated by an Al layer formed by oxidizing the aluminum gates by either oxygen plasma or thermally on a hot plate (150 °C, 5 min). The oxide layer electrically completely insulates the following overlapping layers of aluminum gates that are deposited in the same way with alignment accuracy of \( \sim 20 \) nm.

At least one gate has to overlap with areas where \( n^+ \) regions extend to the thin oxide window. By applying positive voltage on these reservoir gates, the electrons from the \( n^+ \) are attracted to the Si/SiO\(_2\) interface below the reservoir gates, forming the source and drain reservoirs of the device. For example, the device shown in Fig. 23(b) is composed of one or two layers of gates: one top gate that induces the source and drain reservoirs, two barrier gates below the top gate defining the quantum dot, and a plunger gate in the same layer with the barrier gates. Finally, a forming gas (95% N\(_2\), 5%H\(_2\)) anneal is carried out for the sample at 400 °C for 15 min to reduce the Si/SiO\(_2\) interface trap density to a level of \( \sim 5 \times 10^{10} \) cm\(^{-2}\) eV\(^{-1}\) near the conduction band edge. Silicon quantum dots can also be fabricated with an all-silicon process, in which the aluminum gates are replaced by conducting polysilicon gates shown in Fig. 23(a).

IV. QUANTUM STANDARDS OF ELECTRIC QUANTITIES AND THE QUANTUM METROLOGY TRIANGLE

The ampere is one of the seven base units of the International System of Units (SI) (Bureau International des Poids et Mesures, 2006) and is defined as follows: “The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross section, and placed 1 m apart in vacuum, would produce between these conductors a force equal to 2 \( \times 10^{-7} \) newton per meter of length.” The present definition is problematic for several reasons: (i) The experiments required for its realization are beyond the resources of most of the National Metrology Institutes. (ii) The lowest demonstrated uncertainties are not better than about \( 3 \times 10^{-7} \) (Clothey et al., 1989; Funck and Sienknecht, 1991). (iii) The definition involves the unit of newton, kg \( \times \) m/s\(^2\), and thus the prototype of the kilogram, which is shown to drift in time (Quinn, 1991). In practice, electric metrologists are working outside the SI and employing quantum standards of voltage and resistance, based on the Josephson and quantum Hall effects, respectively.

A. The conventional system of electric units

According to the ac Josephson effect, \( V = (h/2e)\hat{\phi}/\hat{t} \), the voltage \( V \) applied over the Josephson junction induces oscillations of the phase difference \( \phi \) over the junction (Josephson, 1962). Phase locking \( \phi \) by a high-frequency \( (f_j) \) signal results in quantized voltage plateaus

\[
V_j = \frac{n_j f_j}{K_j} \approx n_j \frac{h}{2e} f_j, \tag{41}
\]

which are often called Shapiro steps (Shapiro, 1963). Here \( n_j \) is the integer number of cycles of \( 2\pi \) in which \( \phi \) evolves during one period of the high-frequency signal. The proportionality between \( V_j \) and \( f_j \) is denoted by the Josephson constant \( K_j \). According to theory, \( K_j = 2e/h \), but as discussed, this assumption is sometimes relaxed in metrology. The Josephson voltage standards (JVS) have been used in electric metrology since the 1970s; see, e.g., Kohlmann, Behr, and Funck (2003) and Jeanneret and Benz (2009) for reviews. The first standards consisted of a single junction
and generated voltages only up to about 10 mV. Arrays of more than 10,000 junctions with the maximum output of 10 V were developed in the 1980s. They were based on hysteretic junctions where Shapiro steps with different $n_j$ can exist at the same bias current. Since the 1990s, the research has focused on arrays of nonhysteretic junctions where $n_j$ can be chosen by the applied current bias. Arrays divided in sections of $2^m$ junctions ($m = 0, 1, 2, \ldots$) are called program-mable since one can digitally select any multiple of $f_j/K_j$ up to the number of junctions as the output voltage (Hamilton, Burroughs, and Kautz, 1995; Kohlmann et al., 2007). They are practical for dc voltage metrology, but are especially developed for generating digitized ac voltage wave forms up to about 1 kHz, which is an active research topic (Behr et al., 2005). Voltage wave forms at higher frequencies can be generated by pulse-driven Josephson junction arrays where the desired ac wave form is synthesized by the delta-sigma modulation of fast voltage pulses, each having the time integral of one flux quantum $h/(2e)$ (Benz and Hamilton, 1996).

The QHR standard consists of a two-dimensional electron gas, which, when placed in a high perpendicular magnetic field, exhibits plateaus in the Hall voltage $V_H = R_H I$ over the sample in the direction perpendicular to both the field and the bias current $I$. Here

$$R_H = \frac{R_K}{i_K} \approx \frac{1}{i_K} \frac{h}{e^2},$$

is the quantized resistance, which is proportional to the von Klitzing constant $R_K$ and inversely proportional to the integer $i_K$ (von Klitzing, Dorda, and Pepper, 1980). As in the case of the JVS, the theoretical equality $R_K = h/e^2$ is sometimes relaxed in metrology. The plateau index $i_K$ can be chosen by tuning the magnetic field. Usually the best results are obtained at $i_K = 2$.

Quantum Hall standards based on Si MOSFETs or GaAs/AlGaAs heterostructures were harnessed in routine metrology quickly during the 1980s; see, e.g., Jeckelmann and Jeanneret (2001), Poirier and Schooper (2009), and Weis and von Klitzing (2011) and issue 4 of C. R. Physique, Vol. 369 (2011) for reviews. Different resistances can be calibrated against the QHR by using the CVC. It is essentially a transformer with an exact transform ratio due to the Meissner effect of the superconducting loop around the windings (Harvey, 1972; Gallop and Piquemal, 2006). Another way to divide or multiply $R_H$ are parallel or series quantum Hall arrays, respectively, which are permitted by the technique of multiple connections that suppresses the contact resistances (Delahaye, 1993). One rapidly developing research topic is ac quantum Hall techniques, which can be used in impedance standards to expand the traceability to capacitance and inductance (Schurr et al., 2011). An important recent discovery is that graphene can be used to realize an accurate and very robust QHR standard (Zhang et al., 2005; Novoselov et al., 2007; Tzalenchuk et al., 2010; Janssen et al., 2011).

The most precise measurement of $K_J$ within the SI was performed by a device called a liquid-mercury electrometer with the uncertainty $2.7 \times 10^{-7}$ (Clothier et al., 1989). The SI value of $R_K$ can be obtained by comparing the impedance of the QHR and that of the Thompson-Lampard calculable capacitor (Thompson and Lampard, 1956; Bachmair, 2009). The lowest reported uncertainty of such comparison is $2.4 \times 10^{-8}$ (Jeffery et al., 1997). However, both the JVS and QHR are much more reproducible than their uncertainties in the SI; see Sec. IV.B. Therefore, the consistency of electric measurements could be improved by defining conventional values for $R_K$ and $K_J$. Based on the best available data by June 1988, the member states of the Metre Convention made an agreement of the values that came into effect in 1990:

$$K_J = 483 597.9 \text{ GHz/V},$$

$$R_K = 25 812.807 \Omega.$$  \hspace{1cm} (43)

Since then, electric measurements have in practice been performed using this conventional system which is sometimes emphasized by denoting the units by $V_{90}$, $\Omega_{90}$, $A_{90}$, etc., and where the JVS and QHR are called representations of the units.

### B. Universality and exactness of electric quantum standards

A theory can never be proven by theory, but, as argued by Gallop (2005), theories based on very general principles such as thermodynamics and gauge invariance are more convincing than microscopic theories such as the original derivation of the Josephson effect (Josephson, 1962). There are rather strong theoretical arguments for the exactness of the JVS: Bloch has shown that if a Josephson junction is placed in a superconducting ring, the exactness of $K_J$ can be derived from gauge invariance (Bloch, 1968, 1970). Furthermore, Fulton showed that a dependence of $K_T$ on materials would violate Faraday’s law (Fulton, 1973). For quantum Hall devices, early theoretical works argued that the exactness of $R_K$ is a consequence of gauge invariance (Laughlin, 1981; Thouless et al., 1982). However, it is very complicated to model real quantum Hall bars, including dissipation, interactions, etc., and thus the universality and exactness of the QHR has sometimes been described as a “continuous surprise” (Mohr and Taylor, 2005; Keller, 2008). Extensive theoretical work, e.g., on topological Chern numbers, has strengthened the confidence in the exactness of $R_K$; see Avron, Osadczy, and Seiler (2003), Bieri and Fröhlich (2011), and Doucot (2011) for introductory reviews. Recent theoretical work based on quantum electrodynamics (QED) predicts that the vacuum polarization can lead to a magnetic field dependence of both $R_K$ (Penin, 2009, 2010a) and $K_J$ (Penin, 2010b), but only at the level of $10^{-20}$. The case of single-electron transport has been studied much less and there are no such strong theoretical arguments for the lack of any corrections for the transported charge (Gallop, 2005; Stock and Witt, 2006; Keller, 2008).

On the experimental side, comparisons between Si and GaAs quantum Hall bars show no deviations at the experimental uncertainty of $\sim 3 \times 10^{-10}$ (Hartland et al., 1991). Recently, an agreement at the uncertainty of $8.6 \times 10^{-11}$ was found between graphene and GaAs devices (Janssen et al., 2011).

---

1. A condensed-matter correction of $\sim 10^{-10}$ for the charge of the electron was suggested by theory based on QED (Nordtvedt, 1970), but it was refuted by Hartle, Scalapino, and Sugar (1971) and Langenberg and Schrieffer (1971).
FIG. 37. Simplified sketch of the most accurate routes to information on $\epsilon_{J,K}$. Direct measurement of $R_K$ together with an independent measurement of the fine structure constant ($\alpha$) yields a value for $\epsilon_K$. Values for the sum of $\epsilon_J$ and $\epsilon_K$ can be obtained from the combination of the so-called watt balance experiment ($K_0^2 R_K$) and a measurement of the Avogadro constant ($N_A$), or from the combination of $\alpha$ and measurements of low-field gyromagnetic ratios ($y_0$). Less accurate information is provided by measurements of high-field gyromagnetic ratios $K_j$ and the Faraday constant $F = eN_V$, and by the QMT.

This is an extremely important demonstration of the universality of $R_K$ because the physics of the charge carriers is notably different in graphene and semiconductors (Goerbig, 2011). Comparisons between the JVs have been summarized recently by Wood and Solve (2009). The lowest uncertainties obtained in comparisons between two JVs are in the range of $10^{-11}$. Even much smaller uncertainties have been obtained in universality tests of the frequency-to-voltage conversion by applying the same frequency to two different junctions or junction arrays and detecting the voltage difference by a SQUID-based null detector. Several accurate experiments have indicated that the conversion is independent of, e.g., the superconducting material and the junction geometry. The lowest demonstrated uncertainty is astonishing: $3 \times 10^{-19}$ (Clarke, 1968; Tsai, Jain, and Lukens, 1983; Jain, Lukens, and Tsai, 1987; Kautz and Lloyd, 1987).

The reproducibility and universality of the quantum standards are an indication that Eqs. (41) and (42) are exact, but a proof can be obtained only by comparison to other standards. Any one of the electric quantities $V$, $I$, or $R$ can be compared to the other two in a quantum metrology triangle (QMT) experiment (Likharev and Zorin, 1985); see Sec. IV.D. It is a major goal in metrology, but the insufficient performance of single-electron devices has to date prevented the reaching of low uncertainties. However, the exactness of Eqs. (41) and (42) can also be studied in the framework of the adjustment of fundamental constants. The most thorough treatment has been performed by the Committee on Data for Science and Technology (CODATA). Updated papers are nowadays published every four years; see Mohr and Taylor (2000, 2005), and Mohr, Taylor, and Newell (2008, 2012). Karshenboim (2009) provided a useful overview. We review here the most accurate ($< 10^{-7}$) routes to information on the electric quantum standards. They are also illustrated in Fig. 37. Most of the equations in this section assume that Eqs. (41) and (42) are exact, but when referring to possible deviations, we describe them by symbols $\epsilon_{J,K,S}$:

$$K_J = (1 + \epsilon_J) \frac{2e}{\hbar}, \quad R_K = (1 + \epsilon_K) \frac{\hbar}{e^2}, \quad Q_S = (1 + \epsilon_S)e. \quad (44)$$

In this context, the current generated by the single-electron current source is $I_S \equiv (k_3) Q_S f$, where $(k_3)$ is the average number of electrons transported per cycle.

There are a number of fundamental constants that are known with much smaller uncertainties than those related to electric metrology. Some constants, e.g., permeability, permittivity, and the speed of light in vacuum, and the molar mass constant, $\mu_0$, $\mu_0$, $c$, and $M_u = 1$ g/mol, respectively, are fixed by the present SI. Examples of constants known with an uncertainty $\leq 10^{-10}$ are the Rydberg constant $R_{\infty}$ and several relative atomic masses, e.g., that of the electron $A_e(e)$. In the past few years, there has been tremendous progress in the determination of the fine structure constant $\alpha$. First the electron magnetic moment anomaly $\alpha_e$ was measured with high accuracy. A separate calculation based on QED gives the function $\alpha(a_e)$. Together these results yield a value for $\alpha$ with an uncertainty $0.37 \times 10^{-9}$ (Hanneke, Fogwell, and Gabrielse, 2008). Soon after, a measurement of the recoil velocity of the rubidium atom, when it absorbs a photon, yielded a value for $\alpha$ with an uncertainty of $0.66 \times 10^{-9}$ (Bouchendira et al., 2011). These two results are in good agreement. Together they give a validity check for QED since the first result is completely dependent and the latter practically independent of that theory.

The fine structure constant is related to $R_K$ by the exact constants $\mu_0$ and $c$:

$$\alpha = \frac{\mu_0 c e^2}{2\hbar} = \frac{\mu_0 c}{2R_K}. \quad (45)$$

This relationship means that when $R_K$ is measured with a calculable capacitor, it also yields an estimate for $\alpha$. Thus, a measurement of $R_K$ could also test QED, but in practice, the atomic recoil measurement is more accurate by about a factor of 30. A metrologically more important interpretation of this relation is that a comparison between $\alpha$ and the weighted mean of the measurements of $R_K$ yields an estimate of $\epsilon_K = (29 \pm 18) \times 10^{-9}$ (Mohr, Taylor, and Newell, 2012). There is thus no proof of a nonzero $\epsilon_K$, but several groups are developing calculable capacitors in order to determine $R_K$ with uncertainty below $10^{-8}$ (Poirier and Schopfer, 2009; Poirier et al., 2011).

The existing data that yield information on $\epsilon_J$ are more discrepant. As described, $\epsilon_J$ is related to measurements of gyromagnetic ratios [see Mohr and Taylor (2000) for a detailed description] and to the efforts for the redefinition of the kilogram (Mohr, Taylor, and Newell, 2008, 2012). The gyromagnetic ratio $\gamma$ determines the spin-flip frequency $f$ of a free particle when it is placed in a magnetic field $B$: $\gamma = 2\pi f/B$. The gyromagnetic ratios of a helium nucleus and a proton are accessible in nuclear and atomic magnetic resonance experiments. These ratios can be related to the gyromagnetic ratio of an electron that is linked to $\alpha$ and $\hbar$. There are two methods to produce the magnetic field: In the low-field method, it is generated by an electric current in a coil and determined from the current and the geometry. In the high-field method, the field is generated by a permanent

---

2The adjustments are named after the deadline for the included data, e.g., CODATA-10 is based on experimental and theoretical results that were available by 31 December 2010. The values of the constants and much more information are available at the Web site physics.nist.gov/const/.
magnet and measured from a current induced in a coil. When the electric current is determined in terms of the JVS and QHR, the product $K_1 R_K = \sqrt{\frac{2\mu_0 c}{\hbar \gamma}}$ appears in either the numerator or the denominator of $\gamma$, depending on which method is used. In the low-field method, $\hbar$ cancels out from the equations of the gyromagnetic ratios, and the experiment yields a value for $\alpha$. The high-field results also depend on $\alpha$, but since it is known much more precisely than $\hbar$, they essentially yield a value for $\hbar$. The high-field results on $\hbar$ are in good agreement with other experiments, albeit their uncertainty is not better than about $10^{-6}$. However, the low-field results are discrepant from the CODATA value of $\alpha$. By substituting Eqs. (44) into the observational equations of the low-field data, one obtains the estimate $\epsilon_j + \epsilon_k = (-254 \pm 93) \times 10^{-9}$ (Cadoret et al., 2011; Mohr, Taylor, and Newell, 2012). Since the measurements of $R_K$ yield a much smaller value for $\epsilon_k$, the gyromagnetic data seem to imply a significant negative $\epsilon_j$. However, as explained below, a positive $\epsilon_j$ can be found from measurements aiming at the redefinition of the kilogram.

There are essentially two candidate methods for the future realization of the kilogram: the watt balance and silicon sphere methods. The first, suggested by Kibble (1975), relates electric power to the time derivative of the gravitational potential energy:

$$mgv = \frac{V^2}{R} = \frac{1}{K_1^2 R_K} = h/4.$$  \tag{46}

When the mass $m$, its velocity $v$, and the gravitational acceleration $g$ are traceable to the SI, the watt balance yields a value for $h$. Watt balance results have already been published by four national metrology institutes, and several devices are under development; see, e.g., Li et al. (2012), Steiner (2013), and Stock (2013) for reviews. The silicon sphere approach is so demanding that it is employed only by the International Avogadro Coordination (IAC). The results were published in 2011; see Andreas et al. (2011a, 2011b) and the entire issue No. 2 of Metrologia, Vol. 48 (2011). This project determines the Avogadro constant $N_A$ by fabricating spheres of enriched $^{28}\text{Si}$ whose mass is compared to the prototype of the kilogram and whose volume is measured by laser interferometry. The lattice parameter and the relative atomic mass of $^{28}\text{Si}$ are measured in different experiments, and the ratio of the relative and absolute mass densities yields $N_A$.

Results for $h$ and $N_A$ can be compared precisely with the help of the molar Planck constant

$$N_A h = \alpha^2 \frac{A_\gamma(e) M_\alpha c}{2R_{\alpha}}.$$  \tag{47}

Its uncertainty is only $0.7 \times 10^{-9}$ (Mohr, Taylor, and Newell, 2012) and depends mainly on those of $A_\gamma(e)$ and $\alpha$. Equation (47) can be derived from the definition of the Rydberg constant by writing the inaccurate absolute mass of the electron in terms of its relative mass and $N_A$ which links microscopic and macroscopic masses. The IAC 2011 result resolved the discrepancy of $1.2 \times 10^{-6}$ between watt balances and the Avogadro constant determined from a sphere of natural Si that had puzzled metrologists since 1998 (Mohr and Taylor, 2000). Especially after the newest results by Steele et al. (2012) there is no longer any clear discrepancy between the two methods, but the two most accurate watt balances deviate by a factor of $260 \times 10^{-9}$ which is 3.5 times the uncertainty of their difference (Steiner et al., 2007; Steele et al., 2012). Also the measurements of the isotope ratio of the silicon sphere spread more than expected (Yang et al., 2012). Nevertheless, by combining the Planck constants obtained from the watt balance ($h_\nu$) and silicon sphere experiments ($h_{\text{Avog}}$), we obtain an estimate of $\epsilon_j + \epsilon_k/2 \approx (h_{\text{Avog}}/h_\nu - 1)/2 = (72 \pm 15) \times 10^{-9}$. Here we neglected correlations between experiments. A more detailed analysis on the existence of $\epsilon_{j,k}$ can be found from the CODATA papers of Mohr, Taylor, and Newell (2008, 2012); see also Keller (2008). They executed the least-squares analysis of fundamental constants several times, allowing either nonzero $\epsilon_k$ or $\epsilon_j$, and including only part of the data. When they excluded the lowest-uncertainty but discrepant data, the remaining higher-uncertainty but consistent data yielded the conservative estimates $\epsilon_k = (28 \pm 18) \times 10^{-9}$ and $\epsilon_j = (150 \pm 490) \times 10^{-9}$. Thus the exactness of the quantum Hall effect is confirmed much better than that of the Josephson effect.

C. The future SI

Modernizing the SI toward a system based on fundamental constants or other true invariants of nature has long been a major goal, tracing back to a proposal by Maxwell in the 19th century; see, e.g., Flowers (2004) and references therein. Atomic clocks and laser interferometry permitted such a revision of the second and the meter. The development of quantum electric standards, watt balance experiments, the Avogadro project, and measurements of the Boltzmann constant have made the reform of the ampere, kilogram, mole, and kelvin realistic in the near future. In particular, suggestions by Mills et al. (2005) launched an active debate among metrologists (Mills et al., 2006; Becker et al., 2007; Milton, Williams, and Bennett, 2007). Soon it was agreed that the SI should not be altered before there are at least three independent experiments (from both watt balance and Avogadro constant calculations) with uncertainties $\leq 50 \times 10^{-9}$ that are consistent within the 95% confidence intervals, and at least one of them has the uncertainty $\leq 20 \times 10^{-9}$ (Gläser et al., 2010). There have also been requests to await better results from single-electron and QMT experiments (Bordé, 2005; Milton, Williams, and Bennett, 2007), and to solve the discrepancy of low-field gyromagnetic experiments (Cadoret et al., 2011).

There is already a draft chapter for the SI brochure that would adopt the new definitions: BIPM (2010); see also the whole issue 1953 in Phil. Trans. Royal Soc., A Vol. 369 (2011), especially Mills et al. (2011). In this draft, the whole system of units is scaled by a single sentence that fixes seven constants. The most substantial changes are that the base units amperc, kilogram, mole, and kelvin are defined by fixed values of $e$, $h$, $N_A$, and $K_B$, respectively. The new definition for the ampere reads “The ampere, A, is the unit of electric current; its magnitude is set by fixing the numerical value of the elementary charge to be equal to exactly $1.60217720 \times 10^{-19}$ when it is expressed in the units of second and ampere,
which is equal to coulomb.” The new definitions do not imply any particular methods for the realizations of the units. They are guided by *mises en pratique*, e.g., the ampere could be realized with the help of the JVS and QHR (CCEM Collaboration, 2012).

The new SI would significantly lower the uncertainties of many fundamental constants; see, e.g., Mills et al. (2011) for evaluations. One should note, however, that choosing the optimal set of fixed constants is always a trade-off. For example, since $\alpha$ is a dimensionless number and thus independent of the choice of units, one can see from Eq. (45) that fixing $e$ and $h$ would make $\mu_0$ and $e_0$ a quantity that is determined by a measurement of $\alpha$. Presently, $\mu_0$ and $e_0$ are fixed by the definition of an ampere. However, their uncertainty would be very low, the same as that of $\alpha$, which is 0.32 $\times$ 10$^{-3}$ (Mohr, Taylor, and Newell, 2012). One alternative suggestion is to fix $h$ and the Planck charge $q_p = \sqrt{2\hbar/e}$, which would keep $\mu_0$ and $e_0$ exact (Stock and Witt, 2006). It is also worth noting that out of $h$, $N_A$, and the molar mass of carbon-12, $M^{(12)} = A^{(12)} M_u$, only two can be fixed. The suggested SI would release the equality $M^{(12)} = 0.012$ kg/mol, which has raised criticism. In particular, there have been claims that the definition of the kilogram based on $h$ would not be understandable for the wider audience, and a definition based on the mass of a number of elementary particles would be better in this respect (Becker et al., 2007; Milton, Williams, and Bennett, 2007; Leonard, 2010; Hill, Miller, and Censullo, 2011). Milton, Williams, and Forbes (2010) studied two alternatives, fixing either $N_A$ and $h$ or $N_A$ and the atomic mass constant $m_u = M_u/N_A$, and showed that this choice has little effect on the uncertainties of fundamental constants, mainly because the ratio $h/m_u$ is well known from atomic recoil experiments.

### D. Quantum metrology triangle

Phase-locked Bloch (Averin, Zorin, and Likharev, 1985) and SET (Averin and Likharev, 1986) oscillations in superconducting and normal-state tunnel junctions, respectively, were proposed as a source of quantized electric current in the mid-1980s, soon after the discovery of the QHR. Already Likharev and Zorin (1985) suggested that the quantum current standard could provide a consistency check for the existing two electric quantum standards in an experiment they named the “quantum metrology triangle.” However, the quantized current turned out to be a much greater challenge than the JVS and QHR. Still, after a quarter of a century, quantum current standards are yet to take their place in metrology. On the other hand, the progress in knowledge of $K_J$ and $R_k$ has also been rather slow: in CODATA-86 the uncertainties were $300 \times 10^{-9}$ and $45 \times 10^{-9}$, respectively (Cohen and Taylor, 1987). These uncertainties are essentially on the same level as in CODATA-10 if the discrepancy of the data is taken into account.

The QMT experiment and its impact has been discussed, e.g., by Piquemal and Geneves (2000), Zimmermann and Keller (2003), Piquemal (2004), Gallop (2005), Keller (2008), Keller et al. (2008), Feltin and Piquemal (2009), and Scherer and Camarota (2012). In this review, we use the terms QMT setup, experiment, or measurement for any experimental setups that pursue a metrological comparison between JVS, QHR, and a quantum current source.

Development of such setups is a challenge in itself and not necessarily related to any specific single-electron source. However, as a closure of the QMT, we consider only experiments which include error counting of the single-electron device. The reason is that they are far more sensitive to errors than the JVS and QHR. Error counting allows one to separate the contribution of the average number of transported charge quanta $\langle k_{S} \rangle$ from the current of the single-electron source $I_S = (k_{S})Q_S$. Only then can the QMT measurement yield information on fundamental constants.

The QMT setups can be divided into two categories: those that apply Ohm’s law $V = RI$ or so-called ECCSs which utilize the definition of capacitance $C = Q/V$. They are sometimes called direct and indirect QMTs, respectively.

#### 1. Triangle by Ohm’s law

Applying Ohm’s law is the most obvious way to compare the three quantum electric standards. It can be realized either as a voltage balance $V_J = R_{HI}S$ or as a current balance $V_J/R_{H} = I_{S}$. In both cases, substituting Eqs. (44) into $V_J = R_{HI}S$ yields

$$\frac{n_j k_q}{2} \frac{f_J}{f_S} \approx 1 + \epsilon_j + \epsilon_k + \epsilon_s.$$  \hspace{1cm} (48)

The major difficulty in QMT experiments is outlined as follows. Consider the ideal case where the noise of the experiment is dominated by the Johnson noise of the resistor. The relative standard deviation of the measurement result is

$$\frac{\delta I_S}{I_S} = \sqrt{\frac{k_q T}{\pi R I_S^2}}.$$

By substituting realistic estimates $t = 24$ h and $T = 100$ mK for the averaging time and the temperature of the resistor, respectively, and by assuming that $R = R_k/2$ and $I_s = 100$ pA, one obtains the uncertainty $\delta I_S/I_S \approx 7 \times 10^{-7}$. In practical experiments, the $1/f$ noise and the noise of the null detection circuit make the measurement even more demanding, but this simple model demonstrates that the magnitude of the current should be at least 100 pA.

Another problem is that the product $R_{HI}I_S$ yields a very small voltage, e.g., $12.9$ kΩ $\times 100$ pA = 1.29 μV.\(^3\) Even the voltage of a JVS with only one junction is typically of the order of 70 GHz/$K_J \approx 140$ μV. Such low voltages are also vulnerable to thermoelectric effects. One way to overcome this problem is to multiply the current of the SET by a CCC with a very high winding ratio $\approx 10000$ as suggested by Hartland et al. (1991), Sese et al. (1999), and Piquemal and Geneves (2000); see Fig. 38(a). It allows room-temperature detection, and that JVS, SET, and QHR can be operated in different refrigerators. This type of effort has been described by Piquemal (2004), Feltin and Piquemal (2009), Feltin et al. (2011), and Devoille et al. (2012). Another approach is to use a high-value cryogenic resistor that is calibrated against the QHR with the help of a CCC (Elmquist, Zimmerman, and Huber, 2003; Manninen et al., 2008). All parts of Ohm’s law

---

\(^3\)A quantum voltage standard based on integrating a semiconducting pump with the QHR was pioneered by Hohls et al. (2011).
are in the same cryostat which can reduce thermoelectric effects; see Fig. 38(b). Only the difference current \(V_J/R_{\text{cryo}} - I_S\) needs to be amplified. Despite persistent efforts, the experimental realizations of these approaches have so far produced only preliminary results. Recently, a promising outcome was obtained by Giblin et al. (2012) who used a CCC with high winding ratio to calibrate a precision ECCS experiment. In the first phase (A), the electron pump charges the cryocapacitor \(C \approx 2 \text{ pF}\). An SET electrometer \((E)\) is used to generate a feedback voltage \((V)\) that maintains the potential of the island at zero. Hence all the charge is accumulated to the cryocapacitor and not to the stray capacitance. The feedback voltage constitutes the third part of the \(Q = CV\) type triangle. In the second phase (B), the cryocapacitor is calibrated against the reference \(C_{\text{ref}}\) which is traceable to a calculable capacitor. From Keller et al., 1999.

2. Electron-counting capacitance standard

The ECCS experiment was first suggested by Williams, Ghosh, and Martinis (1992). A single-electron current source is used to charge a cryogenic capacitor \(C_{\text{cryo}}\) by a known number \(N_S\) of electrons. The generated voltage is compared to the JVS. The result

\[ C_{\text{cryo}} = \frac{N_S Q_S}{V} \]  

thus yields a quantum capacitance standard. The ECCS experiment was pioneered by Keller et al. (1999) [see Fig. 38(d)], where an uncertainty of \(0.3 \times 10^{-6}\) was obtained for the ECCS capacitance. In this approach, the ECCS was compared to a calculable capacitor. Then the observational equation corresponding to Eq. (48),

\[ \frac{\mu_0 e N_f f_j C}{4\pi N_S} = 1 + \epsilon_f + \epsilon_S, \]

does not include \(\epsilon_K\). However, calculable capacitors have been compared to QHR with very low uncertainty, and ac QHR techniques (Schurr et al., 2011; Camarota et al., 2012) allow \(C_{\text{cryo}}\) to be compared directly against \(R_J\). One should thus obtain an uncertainty of \(\sim 10^{-6}\) before there is any significant difference between the implications of the two QMT versions. An important strength of this QMT version is that the feedback electrometer also allows error counting through shuttle pumping; see Sec. III.H.2.

A major weakness of the ECCS is that it calibrates \(C_{\text{cryo}}\) at \(\sim 0.01\) Hz, but commercial capacitance bridges that are used to compare \(C_{\text{cryo}}\) to the calculable capacitor (and also ac QHR) operate at \(\sim 1000\) Hz. Zimmerman, Simonds, and Wang (2006) presented a model for the dielectric dispersion of insulating films at the surface of the electrodes of the capacitor. They fit this model to measurements of the frequency dependence and its temperature dependence in the ranges 100–3000 Hz and 4–300 K. The frequency dependence decreases at low temperatures. They evaluate that it yields an uncertainty component of \(0.2 \times 10^{-6}\) for the QMT. Keller, Zimmerman, and Eichenberger (2007) used this estimate to finish the uncertainty budget of the NIST ECCS experiment that closes the QMT at the uncertainty of \(0.9 \times 10^{-6}\).

Recently, PTB reached the uncertainty of \(1.7 \times 10^{-6}\) in an ECCS experiment (Camarota et al., 2012); see Sec. III.H.2. PTB presented their result as “preliminary” and planned both a more detailed uncertainty budget and several improvements to the experiment. Besides NIST and PTB, the ECCS has been pursued at METAS (Rüfenacht, Jeanneret, and Lotkhov, 2010).

3. Metrological implications of single-electron transport and QMT

So far, QMT has been closed with a reasonable uncertainty \((\sim 10^{-6})\) only in the ECCS experiments of NIST (Keller, Zimmerman, and Eichenberger, 2007) and PTB (Camarota et al., 2012). As shown in Sec. IV.B, an uncertainty of \(\lesssim 0.02 \times 10^{-6}\) is required to yield information on \(\epsilon_K\), and an uncertainty of \(\sim 0.1 \times 10^{-6}\) would strengthen the knowledge of \(\epsilon_f\). Thus, the NIST and PTB results can be expressed in terms of \(\epsilon_S\) only: \(\epsilon_S = (-0.10 \pm 0.92) \times 10^{-6}\) and \(\epsilon_S = (-0.3 \pm 1.7) \times 10^{-6}\), respectively (Keller, 2008).

Milton, Williams, and Forbes (2010) analyzed a scenario where \(\epsilon_f\) is an adjusted parameter and \(\epsilon_S = \epsilon_K = 0\). They studied the effect of QMT on the uncertainties of fundamental constants and showed that when the QMT is inaccurate, the uncertainties of \(h, e,\) and \(m_n\) are mainly determined by the Avogadro experiment. When the QMT is improved, their uncertainties will be dominated by those of the watt balance and the direct measurement of \(R_K\).

One problem of the QMT is that it gives only a value for the sum of the errors of the quantum standards, and, in principle, they could cancel each other. It is thus useful to have independent tests for each standard, and those for the JVS and QHR are discussed in Sec. IV.B. A test for the current standard only, i.e., an SI value for \(Q_S\), can be obtained by combining results from three experiments: QMT, a
measurement of $R_K$ by a calculable capacitor, and watt balance (Keller et al., 2008). Applying Eqs. (46) and (50), and substituting $R$ by $1/\omega C$, one obtains

$$Q_S = \frac{1}{N_S} \sqrt{\frac{m g v C}{\omega}}$$

(52)

Also the Ohm’s law triangle can be used to yield a similar result, but in a less direct way. One should note that JVS and QHR are used here only as transfer standards. Keller et al. (2008) derived a result based on the NIST ECCS: $Q_S = 1.602 176 3 \times 10^{-19} \pm 1.5 \times 10^{-25}$ C. This could be compared to the CODATA value for $e$, which, however, depends strongly on $h$ and the exactness of $K_f$ and $R_K$. Instead, it is better to compare $Q_S$ to another value of $e$ that is independent of JVS and QHR: $e = \sqrt{\alpha^2 A(e) M_v / \mu_0 R_e N_A}$ (Feltin and Piquemal, 2009). Its uncertainty $\sim 0.015 \times 10^{-6}$ is dominated by that of $N_A$. Using the NIST ECCS result and the NRC or IAC values for $N_A$, one obtains $\epsilon_S = (-0.2 \pm 0.9) \times 10^{-6}$.

We note that according to Eq. (51), the QMT also yields a value for $\alpha$ independently of the QHR, which was one of the early motivations for the ECCS (Williams, Ghosh, and Martinis, 1992). This fact, however, has little importance until the uncertainty is competitive with the atomic recoil experiments ($< 10^{-9}$). Then the QMT would strengthen the verification of QED.

Although single-electron transport would be conceptually the most straightforward realization of the ampere in the future SI, it is not likely that it would replace the JVS and QHR as the typical realization in the near future. The exception is naturally the growing field of metrology for small electric currents, where single electronics is expected to yield major improvements of uncertainty. On the other hand, when the accuracy of single-electron transport improves, it can yield vital information on other standards and fundamental constants.

V. PERSPECTIVES AND OTHER APPLICATIONS

The quantum-dot pump (Kaestner, Kashcheyevs, Hein et al., 2008; Giblin et al., 2012) discussed in Sec. III.C has definitely proven its potential to be the basis of the future quantum standard of the ampere. The verified uncertainty of the 150 pA output current on the level of 1 ppm and the theoretically predicted 0.01 ppm uncertainty of the present device are truly remarkable figures of merit. On the other hand, a few important questions remain to be answered before one can realize the ampere with the quantum-dot pump: Superior device performance depends critically on applying a strong $\approx 10$ T magnetic field on it. This dependence is not fully understood, and the exact magnetic field characteristics seem sample dependent. The reproducibility of the highly accurate pumping results with samples from different fabrication runs remains to be shown. Importantly, error-counting experiments on the dot samples have not been carried out, which also prevents one from studying possible errors of other quantum standards in the QMT [see, however, recent results in Fricke et al. (2013)]. Future experiments will likely show whether all the relevant error processes have been accounted for in predicting the obtainable accuracy to be on the level of $10^{-8}$. However, even if not in the case of a bare device, the quantum-dot pump may perhaps be applicable to the realization of the ampere, if the error correction techniques that were described in Sec. III.H become feasible experimentally.

Another important development and potential future realization of the ampere is the SINIS turnstile introduced in Sec. III.B. Although presently inferior to the quantum-dot pump in the level of current output, and consequently with less definite assessment of proven accuracy (present verified uncertainty below $10^{-4}$), this device does not suffer from known obstacles in the way of achieving the required accuracy. Currently, the main error mechanisms have been assessed theoretically and experimentally, including photon-assisted tunneling, Andreev current, cotunneling, residual and generated quasiparticles, and possible residual density of states in a superconductor. Positive conclusions can be drawn from individual experiments with respect to suppressing them in an optimized device. Sample fabrication and reproducibility is currently on a high level, and it has been demonstrated that the requested magnitude of current can be achieved by running many turnstiles in parallel. For the SINIS turnstile, as for the quantum-dot pump, the ultimate test would be an error-counting experiment and the quantum metrological triangle. Currently, such experiments have not been performed. As a summary of the high-accuracy pumps, we present Table I where the obtainable output current, the accuracy, and the possibility for parallelization are compared. Presently several other new proposals are being pushed toward critical tests to study their applicability in current metrology: these include superconducting phase-slip wires, Josephson junction arrays, and mechanical shuttles, just to mention a few less conventional ideas. Although it is not on the horizon at present, it is possible that eventually one of these devices will beat the present Coulomb-blockade-based realizations both in current yield and in their robustness against transfer errors.

Developing ever more accurate current sources has constantly been a driving force for understanding the underlying physical phenomena. On the other hand, the studies for the precise control of single electrons and Cooper pairs have created special expertise that is also applicable in a variety of other research topics.

In addition to the charge degrees of freedom, the electrons hold information in their spin states which have been

<table>
<thead>
<tr>
<th>Name</th>
<th>$I_{\text{exp}}$ (pA)</th>
<th>$\delta I_{\text{exp}}/I$</th>
<th>$\delta I_{\text{theory}}/I$</th>
<th>Parallelization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain of normal metallic islands</td>
<td>1.5</td>
<td>$1.5 \times 10^{-8}$</td>
<td>$\ll 10^{-8}$</td>
<td>Not feasible</td>
</tr>
<tr>
<td>Quantum-dot pump</td>
<td>150</td>
<td>$&lt; 2 \times 10^{-8}$</td>
<td>$\cdots$</td>
<td>Not needed</td>
</tr>
<tr>
<td>SINIS turnstile</td>
<td>3</td>
<td>$&lt; 1 \times 10^{-4}$</td>
<td>$10^{-8}$</td>
<td>Possible</td>
</tr>
</tbody>
</table>

Rev. Mod. Phys., Vol. 85, No. 4, October–December 2013
envisioned (Kane, 1998; Hollenberg et al., 2006) to be utilized (Morello et al., 2010) for quantum information processing. Although the electron transport is typically incoherent in the electron pumps, the spin-encoded information can potentially remain coherent, and hence this information can possibly be transported from the memory cell of the computer to the qubit-qubit interaction cell and back. The transport cycle has to be carried out with high accuracy for fault-tolerant computing to be possible, which creates a close connection to the metrological electron pumps.

Geometric phases (Shapere and Wilczek, 1989) in quantum mechanics have been studied extensively due to both fundamental scientific curiosity and their applications in geometric quantum computing (Zanardi and Rasetti, 1999). The simplest geometric quantum phase, the Berry phase, has already been measured in the superconducting sluice pump (Möttönen et al., 2006; Mottonen, Vartiainen, and Pekola, 2008) thanks to the development of the sluice for metrology. Some theoretical work on the more complex phases referred to as holonomies has been put forward in the framework of Cooper-pair pumps (Pirkkalainen et al., 2010; Solinas, Pirkkalainen, and Mottonen, 2010) but it remains to be seen if these ideas will be implemented experimentally. The main obstacle in practice is perhaps the high level of precision required for the control signals of the pumps, a problem that can possibly be solved with the help of the work on the metrological current source.

Detecting single electrons and Cooper pairs by single-electron transistors and quantum point contacts has been largely motivated by the need for tests of the charge-transport errors in metrology. During the past decade, these techniques have also been successfully implemented, e.g., in experiments on full counting statistics and noise of charge transport. The experiments on the full counting statistics of current fluctuations in a semiconductor quantum dot by real-time detection of single-electron tunneling with a quantum point contact have been successfully performed for instance by Gustavsson et al. (2006, 2007). In these experiments, moments of current up to the fifth and beyond could be reliably measured. Recently, single-charge-counting experiments have been applied to study energy fluctuation relations (Evans, Cohen, and Morriss, 1993; Jarzynski, 1997; Crooks, 1999; Averin and Pekola, 2011) in statistical mechanics. Experiments in steady-state nonequilibrium were performed by Küng et al. (2012), and the Jarzynski and Crooks relations were recently tested by Saia, Yoon et al. (2012). Single-charge-counting experiments allow one to test fundamental statistical mechanics and thermodynamics of classical and quantum systems.

The variety of spin-offs from the development of single-charge current sources for metrology is certainly expanding. In this way the benefits of this research will be obvious not only for the community interested in the system of units and in traceable measurements, but also for other researchers working in basic and applied sciences looking for new tools for measurements that need precise control.

ACKNOWLEDGMENTS

We thank Simone Gasparinetti, Stephen Giblin, Juha Hassel, Panu Helistö, Antti Manninen, Matthias Meschke, Mika Prunnila, Alexander Savin, Heikki Seppä, and Jaw Shen Tsai for useful discussions. We also thank the Academy of Finland through its Centre of Excellence Programs (Projects No. 250280 and No. 251748), MEXT Grant-in-Aid “Quantum Cybernetics,” the FIRST Project from JSPS, and the Finnish National Graduate School in Nanoscience for financial support during the preparation of this manuscript. Y.P. acknowledges partial support from EPSRC, the Royal Society, and Wolfson Foundation. A.K. and M.M. thank the Academy of Finland for financial support (Grants No. 259030, No. 135794, and No. 272806).

REFERENCES

Arias, B., et al., 2011a, Metrologia 48, S1.
Makhlin, Y., G. Schön, and A. Shnirman, 2001, Rev. Mod. Phys. 73, 357.
Millikan, R., 1911, Phys. Rev. (Ser. I) 904.

Rev. Mod. Phys., Vol. 85, No. 4, October–December 2013


Zaikin, A. D., 1994, Physica (Amsterdam) 203B, 255.


