On the observability of Coulomb blockade and single-electron tunneling

M.H. Devoret [a], D. Esteve [a], H. Grabert [b], G.-L. Ingold [b], H. Pothier [a] and C. Urbina [a]

[a] *Service de Physique de l'Etat Condensé, CEA-Saclay, F-91191 Gif-sur-Yvette, France*

[b] *Fachbereich Physik, Universität-GH Essen, W-4300 Essen, Germany*

Coulomb blockade and single-electron tunneling are manifestations, at the macroscopic level, of the granularity of charge. These effects can occur in small capacitance tunnel junction systems, under conditions which are discussed in this article. We examine in particular how the zero-point electromagnetic fluctuations can eventually wash out the single-electron effects.

1. Introduction

Although the charge of the electron was measured for the first time more than eighty years ago [1], the direct manifestation, in a solid-state electronic device, of the discreteness of charge has only been observed a few years ago [2]. One obvious reason is that typical charges in electronic circuits are much larger than the electron charge: for instance, a charge packet in a charge-coupled device (CCD) contains more than 10^6 electrons [3]. A more fundamental reason is that, because of their quantum delocalization, electrons in metals and in semiconductors form a continuous charged fluid. Two ingredients are needed to observe, on macroscopic quantities such as currents and voltages, the effect of single-charge quanta. The first ingredient is the tunnel barrier, which, under appropriate conditions, provides a conduction path for charges in and out of a metallic electrode while preserving the quantization of the total charge of the electrode. The second ingredient is the Coulomb interaction between electrons, which, for sufficiently small electrodes placed at sufficiently low temperature, provides an energy barrier against thermal charge fluctuations. We will see that quantum electromagnetic fluctuations play also an important role since they determine the effective energy barrier entering in the expression of the tunneling rate of electrons across the junctions [4]. These quantum fluctuations wash out single-electron effects in a single, current-biased, tunnel junction unless a very elaborate biasing scheme is adopted. However, for devices in which electrons pass through at least one "island" (a metallic electrode connected electrically to the rest of the circuit through tunnel barriers only), the electromagnetic fluctuations do not prevent the occurrence of single-electron effects. In the following we will first describe the phenomenon of Coulomb blockade in a one-island system without addressing the question of quantum electromagnetic fluctuations and by postulating the "global rule" for the calculation of the Coulomb gap. We will then examine the case of the current-biased single junction and show how the quantum electromagnetic fluctuations induced on the junction by the rest of the circuit (referred to as the junction electromagnetic environment) affect Coulomb blockade. Finally, we will return to the one-island system and discuss how these fluctuations make the global rule a valid approximation. It is obviously not the purpose of this article to review the field of single-electron effects. Detailed lists of references to earlier studies can be found in the review articles by Averin and Likharev [5] and by Schön and Zaikin [6]. We refer the reader to the special issue of Zeitschrift für Physik B [7] for a snapshot of the leading edge of current research on the subject.

2. Coulomb blockade in a one-island system

Consider a charge transport experiment in which a voltage difference is applied to two electrodes (a "source" and a "drain", see fig. 1) separated by an insulating gap. In the middle of the gap lies a third metallic electrode, surrounded everywhere by insulating material. To travel from the source to the drain the electrons must go through this third electrode, which, as mentioned above, we call an "island". We assume that the conduction of electrons through the insulating gaps between the source and the island and between the island and the drain occurs by quantum tunneling. This one-island system can be formed by two metallic tunnel junctions in series [2], a STM tip over a metallic droplet deposited on an oxidized conductive substrate [8] or a 2D electron gas (2DEG) confined by Schottky gates [9].

The tunneling process is so short that we can consider that the electrons are traversing the insulating gaps one at a time. The key point is that during its journey from the source to the drain the electron necessarily makes the charge of the island vary by e. We have indicated in the introduction the general requirements for single-electron effects to occur. Let us manifest them for the one-island system. Firstly, the tunnel barriers must have a tunneling resistance R_T that exceeds the resistance quantum $R_K = h/e^2 \simeq 25.8$ kΩ, i.e.,

$$R_T \gg R_K. \tag{1}$$

Fig. 1. The quantum tunneling of electrons between a "source" and a "drain" electrode through an intermediate "island" electrode can be blocked if the electrostatic energy of a single excess electron on the island is large compared with the energy of thermal fluctuations.

The tunneling resistance is a phenomenological quantity which is defined in the situation where a fixed voltage difference V is imposed onto the two electrodes on either side of the tunnel barrier. The tunneling rate Γ of an electron through the barrier is then proportional to V: $\Gamma = eV/R_T$. The tunneling resistance can be expressed in terms of the microscopic quantity $\mathscr{T}$, which is the barrier transmission coefficient at the Fermi energy: $R_T^{-1} = 4\pi N \mathscr{T} R_K^{-1}$, where N is the number of independent electron channels through the barrier. Condition (1) is obtained by requiring that the energy uncertainty associated with the lifetime due to tunneling $\tau_r = R_T C$ of an excess charge on the island is much smaller than the Coulomb energy $E_c = e^2/2C$. Condition (1) ensures that the wave function of an excess electron or Cooper pair on an island is essentially localized there. It is generally believed that, in systems with tunneling resistances that are small on the scale provided by R_K, charging effects will be suppressed since delocalized states in which electrons flow through an island without charging it are available for charge transport, although the exact circumstances are not precisely known at the time of this writing [10]. Secondly, the total capacitance C of the island must be small enough and the temperature low enough that the energy $E_c = e^2/2C$ required to add a charge carrier to an island far exceeds the available energy of thermal fluctuations, i.e.,

$$E_c \gg k_B T. \tag{2}$$

Conditions (1) and (2) ensure that the transport of charge through the island is governed by the Coulomb charging energy.

In practice, only islands having capacitances not much below a femtofarad can be reliably designed with metallic tunnel junctions or confined 2DEG, thus imposing experiments done at a few tens of millikelvins, now routinely attainable with a dilution refrigerator. However, in experiments involving an STM, much smaller islands can be used and very sharp single-electron effects have been observed at 4 K [8].

Let us now show that there is a threshold voltage for the passage of current through this

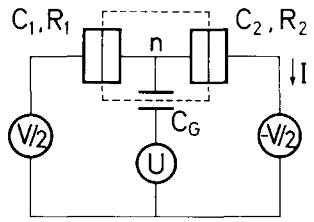

Fig. 2. The one-island system depicted in fig. 1 can be modelled as two tunnel junctions in series (box symbols) whose common electrode forms an island shown inside a dashed line. The electrostatic potential of the island is acted upon by the gate voltage U through the gate capacitance C_G. The bias voltage V induces a net flow of charge through this "single-electron transistor", the value of the current I being controlled by the gate voltage U. The tunnel junction capacitances and tunnel resistances are denoted by C_i and R_i, where i is the junction number.

one-island system. This voltage is called the Coulomb gap voltage. We show in fig. 2 the lumped element representation of the one-island system biased with a voltage V. The island is submitted through the "gate" capacitance C_G to a gate voltage U. In the case of the metallic junction system or the 2DEG system, U can be controlled externally, a possibility which leads to a device called the "single-electron transistor". In the case of the STM/droplet/substrate system, U can be thought of as representing charged impurities in the vicinity of the droplet and is basically uncontrolled.

Between tunneling events, all charges on the capacitances having relaxed to equilibrium, the state of the entire circuit shown on fig. 2 is described by the numbers n_1 and n_2 of electrons having passed through junctions 1 and 2, respectively, in the direction of the current I (see fig. 2). The equilibrium energy of the circuit which includes the electrostatic energy of the capacitances and the work performed by the voltage source is

$$E_{eq} = \frac{(-ne + C_G U + C_1 V/2 - C_2 V/2)^2}{2(C_G + C_1 + C_2)} + peV/2, \tag{3}$$

where $n = n_1 - n_2$ and $p = n_1 + n_2$. The rates Γ^i ($i = 1$ or 2) of electron tunneling through junctions 1 and 2 are given by the following expression [11]

$$\Gamma^i = \frac{1}{e^2 R_i} \frac{\Delta E^i_{eq}}{1 - \exp\left(-\Delta E^i_{eq}/k_B T\right)}, \tag{4}$$

where ΔE^i_{eq} is the difference between the equilibrium energy (3) before and after an electron has tunneled across junction i. The tunneling rate (4) reduces at zero temperature to:

$$\Gamma^i = \frac{\Delta E^i_{eq}}{e^2 R_i} \quad \text{for } \Delta E^i_{eq} > 0, \tag{5a}$$

$$\Gamma^i = 0 \qquad \text{for } \Delta E^i_{eq} < 0. \tag{5b}$$

Expression (4), together with (3), has been nicknamed the "global rule" [12] since the tunneling rate across a particular junction involves the change of energy of the entire circuit. We will see in section 6 how the quantum electromagnetic fluctuations of the circuit explain this seemingly paradoxical property. The global rule can easily be generalized to a circuit containing an arbitrary number of junctions [11].

A straightforward analysis of the rate formula (5) shows that at zero temperature the state with

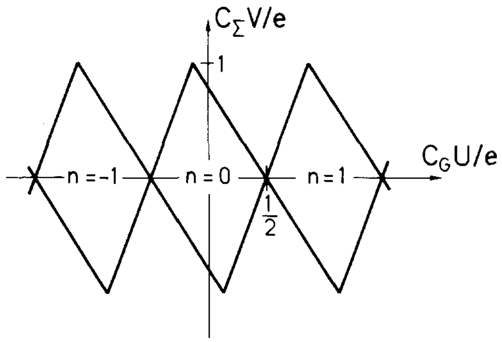

Fig. 3. The blockade diagram of a single-electron transistor with $2C_2 = 10C_G = C_1$. The capacitance C_Σ is the total capacitance of the island: $C_\Sigma = C_1 + C_2 + C_G$. The transistor conducts only outside the rhombic-shaped regions. Inside these regions, there is a constant number n of electrons on the island.

n electrons on the island is stable with respect to tunneling across the first and second junctions for voltages satisfying

$$e(n - \tfrac{1}{2}) < C_G U + (C_2 + C_G/2)V < e(n + \tfrac{1}{2}), \tag{6}$$

$$e(n - \tfrac{1}{2}) < C_G U - (C_1 + C_G/2)V < e(n + \tfrac{1}{2}), \tag{7}$$

respectively. Hence, in the UV plane there are rhombic-shaped regions within which the island is charged with a fixed number of excess electrons (see fig. 3). Inside these rhombi all transitions are suppressed by a Coulomb blockade, and no current flows through the system.

3. The current-biased single junction

It is apparent from the preceding section that the basic system in which single-electron effects occur is a metallic island connected to electron reservoirs through *two* tunnel barriers. What would happen with only *one* small capacitance tunnel junction? Physicists are attracted to simple systems. Historically, it is this question which started the field a few years ago. Several new effects due to the quantization of charge were predicted to arise in a tunnel junction, both in the superconducting and the normal tunnel state [13–16]. Likharev and co-workers [14,16] gave a major thrust to this new area of low-temperature physics by making detailed predictions of Coulomb blockade phenomena in a single junction and by proposing various applications of the new effects.

The theory of Likharev and co-workers thus considers one tunnel junction which is biased by a current I and whose voltage V is measured by a very-high-impedance voltmeter. The junction is characterized by two parameters: its capacitance C and tunnel resistance R_T. The state of the junction is described by two degrees of freedom whose differing nature is crucial. The first degree of freedom is the charge Q on the junction capacitance. It is a continuous variable since it describes the bodily displacement of the electron density in the electrodes with respect to the positive ionic background. In fact, Q can be an arbitrarily small fraction of the charge quantum. The second degree of freedom is the discrete number n of electrons (or Cooper pairs if the electrodes are in the superconducting state) which have passed through the tunnel barrier. The key hypothesis in the theory is that variables Q and n are assumed to be classical variables with a well defined value at every instant t. Charge conservation is imposed by the relation $\dot{Q}(t) + e^*\dot{n}(t) = i(t)$, where $i(t)$ is the current flowing in the leads of the junction and e^* is the charge e (normal state) or $2e$ (superconducting state) of the carriers. Since the current bias is assumed to be ideal we have $i(t) = I$. During a tunneling event, the charge Q must thus discontinuously jump by the elementary charge e^*. The resulting change in the electrostatic energy of the junction is

$$\Delta E = \frac{Q^2}{2C} - \frac{(Q - e^*)^2}{2C} = \frac{e^*(Q - e^*/2)}{C}. \tag{8}$$

At zero temperature, tunneling can only occur if ΔE is positive. This has two consequences. Firstly, the I–V characteristic should have a $I = 0$ branch:

$$-\frac{e^*}{2C} < V < \frac{e^*}{2C} \quad \text{for } I = 0, \tag{9}$$

where the particular value of V is determined by the history of the current in the junction leads: $CV = \int_{-\infty}^{t} i(t')\,\mathrm{d}t'$ modulo e^*. This is the Coulomb blockade for single junctions. Secondly, when a non-zero current is imposed through the junction assumed to be in the normal state, the junction capacitor charge Q will increase linearly until the threshold charge $e/2$ is reached. Then, a tunneling event occurs, making Q jump to $-e/2$ and a new charging cycle starts over again. This leads to single-electron tunneling (SET) sawtooth oscillations of the junction voltage with the fundamental frequency.

$$f_{\mathrm{SET}} = I/e. \tag{10}$$

By a similar kind of reasoning, one predicts for a superconducting junction (Josephson junction) the so-called Bloch oscillations with the frequency

$$f_{\mathrm{Bloch}} = I/2e. \tag{11}$$

The difference between the SET and Bloch oscillations is that in the normal state the charge tunnels irreversibly as Q goes beyond $e/2$ because it is accompanied by quasi-particle excitations, whereas in the superconducting state the charge tunnels reversibly at $Q = e/2$ because Cooper pairs have no kinetic degrees of freedom.

This analysis rests on Q and n being classical variables. The "classicity" of the variable n is solely determined by the properties of the junction. We can safely assume that condition (1), which is a statement about the tunnel barrier and which translates directly in terms of junction fabrication, is a sufficient condition. However, the classicity of the variable Q depends on the junction electromagnetic environment, and the original predictions concerning Coulomb blockade phenomena did not make very explicit statements about what the characteristics of this environment should be. In this theoretical void, two questions concerning the observability of Coulomb blockade and SET oscillations arose:

Question A: The pads on the junction chip which are needed to make connections to the I–V measuring apparatus have parasitic capacitances in the pF range. How should the junction environment be designed for these parasitic capacitances not to shunt the junction capacitance, which needs to be kept in the fF range to observe the charging effects?

Question B: Each mode in the environment is coupled to the charge Q and its zero point energy induces $i(t)$ and Q to fluctuate. How should the environment be designed for these quantum mechanical fluctuations not to affect Coulomb blockade? In other words, how perfect does the current biasing need to be?

We will see below that these two questions are, in fact, closely related and that their answer can be obtained by a full quantum mechanical analysis of the influence of the junction environment on the tunneling probability. Before presenting the results of this analysis we have to discuss the various time scales of the problem, both the time scales pertaining to the junction itself and those pertaining to its environment.

4. Time scales

The junction is characterized by three time scales. The two longer ones can be deduced from quantities we have already mentioned. The longest time scale is set by the tunneling resistance and the capacitance: $\tau_r = R_T C$. It is the reciprocal of the rate of tunnel events for a junction biased at the Coulomb voltage e/C. The intermediate time scale is the uncertainty time associated with the Coulomb energy $\tau_c = h/(e^2/C)^{-1} = R_K C$. The shortest time scale is the tunneling time τ_t of the junction which is given by

$$\tau_t = \hbar \left(\frac{\partial \log \mathscr{T}}{\partial E} \right)_{E=E_F}, \tag{12}$$

where, as previously, $\mathscr{T}(E)$ is the transmission probability through the tunnel barrier of an electron with energy E. This tunneling time, whose importance has been stressed by Büttiker and Landauer [17] (see also ref. [18], and references therein) can be loosely described as the time spent by the tunneling electron under the barrier. In metallic tunnel junctions it is of the order of 10^{-15} s. Here it may be worth pointing out that electron tunneling in a metallic junction is in fact a complex process, at least much more so than what elementary textbooks might lead one to suppose: the electrons in the metallic electrodes travel as quasi-particles, i.e. bare electrons dressed by a positive cloud of charge. When a tunneling quasi-particle impinges on the tunnel barrier it has to undress, leaving the positive charge cloud behind as it travels through the barrier. When this bare electron arrives in the other electrode it attracts a new cloud of positive charge and dresses again to form a quasi-particle. The characteristic time scale for the undressing and dressing processes is the inverse of the plasma frequency. These processes have to be taken into account in the computation of the effective tunneling time, which is the one of interest here. The tunneling rate of a quasi-particle will be quite different from the tunneling rate of a bare electron if the effective tunneling time is notably longer than the inverse of the plasma frequency [19].

Let us now discuss the time scales of the environment. We have first to indicate how one should model the junction electromagnetic environment which includes not only the I–V characteristic measuring apparatus at high temperature but also the leads close to the junction. A priori, we need to consider the response of the environment up to the frequency τ_t^{-1}. Although this natural cut-off provided by the tunneling time is a frequency in the optical domain, the junction, whose dimensions have to be of the order of 100 nm or less to ensure a capacitance in the fF range, is small enough to be treated as a lumped element. The electromagnetic environment, as seen from the location of the junction, can thus be completely described in electrical engineering terms by the relationship between the complex amplitudes $v(\omega) = Q(\omega)/C$ and $i(\omega)$ at frequency ω of the voltage across the junction and the current in the first few hundred nanometers of its leads. Assuming the environment to be linear, we arrive at the general lumped element model of fig. 4a for the electromagnetic environment as seen from the location of the junction. The bias circuitry, which includes the room-temperature electronics, the filters and the leads down to the pads on the junction chip, is modelled by a bias resistor $R_b \gg R_T$ in series with a voltage source V_b. There is also, in parallel with the resistor and the source, a capacitance C_b which models the parasitic capacitances in the bias circuitry. This three-element model of the bias circuitry accounts for the low-frequency response of the environment and is placed in series with a complex impedance $Z(\omega)$ which represents the impedance of the last few mm of leads on the junction chip. The impedance $Z(\omega)$ accounts for the high-frequency response of the environment. This general model of the environment of the junction can be somewhat further simplified, however. One has to note that the parasitic capacitance C_b is larger than the junction capacitance C by orders of magnitude (typically $C_b \simeq 10^4 C$). This means that, although the current through the junction is composed of pulses corresponding to each tunnel event, the voltage on the capacitance C_b is essentially time independent. One can thus replace the model of fig. 4a by the model of fig. 4b in which the impedance $Z(\omega)$ is simply in series with a voltage source V. Of course, V has to be determined self-consistently from the time-averaged current $I(V)$ through the junction by the relation $V = V_b - R_b I(V)$; but this is not a problem. If we know how to calculate the I–V characteristics of the junction for the model of fig. 4b, the junction voltage as a function of the bias current V_b/R_b for the model of fig. 4a can be reconstructed. An important remark is now in order. The value of the impedance $Z(\omega)$ at moderately high frequencies can be made large by making the leads on the junction chip very narrow and by using a resistive material like NiCr. However, at frequencies corresponding to

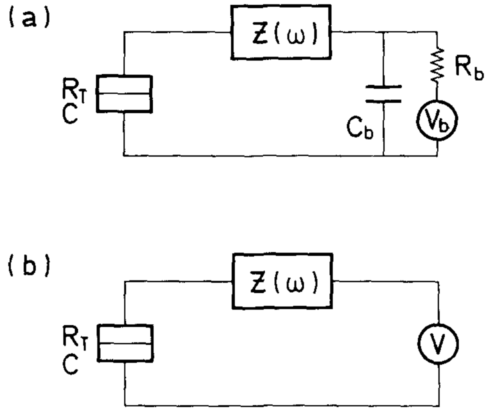

Fig. 4. Lumped element model of the electromagnetic environment of a current-biased tunnel junction, which is represented by a double box symbol. The capacitance and tunnel resistance of the junction are C and R_T, respectively. The impedance $Z(\omega)$ models the high-frequency response of the environment which is dominated by the effect of the leads attached to the junction. The environment low-frequency response, which is dominated by the bias circuitry, is modelled in (a) by a resistance R_b, a capacitance C_b and a voltage source V_b. In all practical cases $C_b \gg C$, and one can use the simplified model (b) in which the junction is biased by an effective voltage source V which is a function of the time-averaged current through the junction.

micron wavelengths, no matter how careful one is in the fabrication of the leads, the impedance $Z(\omega)$ will be dominated by radiation phenomena and will be of the order of the impedance of free space $Z_V \simeq 377\ \Omega$. The modulus $|Z(\omega)|$ of the impedance is thus a decreasing function of frequency. This behavior can be crudely understood by considering the parasitic capacitance between the leads, whose shorting effect on the junction becomes more and more pronounced as the frequency gets higher. A more precise understanding of the frequency dependence of the environment is provided by a resistive transmission line model of the function $Z(\omega)$. This model is analogous to the model described by Martinis and Kautz for experiments on the phase diffusion of a small Josephson junction [20]. The resistive transmission line can be thought of as a ladder of discrete components R_{ln}, C_{ln} and L_{ln}, as shown on fig. 5. The total capacitance and resistance of the transmission line are respectively $C_l = \sum_{n=1}^{N} C_{ln}$ and $R_l = \sum_{n=1}^{N} R_{ln}$ while the characteristic impedance of the line is $Z_l = \sqrt{L_{ln}/C_{ln}}$. In practice, Z_l is always a fraction of the vacuum impedance Z_V while C_l, which one tries to get as small as possible, is not much below 0.1 pF. As we mentioned above, by using very narrow leads made from NiCr, values of the order of 100 kΩ can be obtained for R_l (we refer to this case as "extreme"). If no special effort is put into making high-resistance lead resistors, typical values for R_l are in the 100 Ω – 1 kΩ range (hereafter referred to as the "standard" case). A log–log plot of the function $|Z(\omega)|$ is shown schematically in fig. 6. If one is in the standard case where the lead total resistance is comparable to the line impedance, the environment behaves then as a resistor Z_l. If one is in the extreme case where

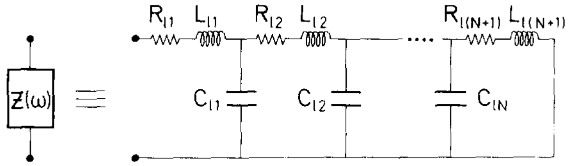

Fig. 5. Resistive transmission line model for the impedance $Z(\omega)$.

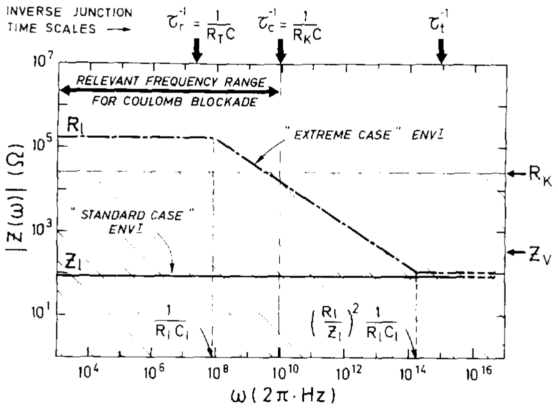

Fig. 6. Schematic behavior of the modulus $|Z(\omega)|$ of the environment impedance, as a function of the frequency ω. We have shown for comparison (i) the inverse junction time scales on the frequency axis and (ii) the resistance quantum $R_K = h/e^2$ and the impedance of the vacuum $Z_V = \sqrt{\mu_0/\epsilon_0}$ on the resistance axis.

the lead total resistance is much higher than the line impedance, the environment behaves as a resistor R_l until a roll-off frequency given by $(R_l C_l)^{-1}$ is reached. One then enters an *RC* line regime where the leads behave as an impedance with equal reactive and dissipative parts falling off as $\omega^{-1/2}$. Finally, at the final saturation frequency $\omega_s = (R_l/Z_l)^2 (R_l C_l)^{-1}$, one recovers the frequency-independent behavior of the standard case. Note that the final saturation frequency ω_s is independent of the length of the leads provided that $C_{ln} = C_l/N$ and $R_{ln} = R_l/N$, which is a realistic assumption. As an example, for leads with distributed resistance, capacitance and inductance of 100 Ω/μm, 0.5×10^{-16} F/μm and 0.5×10^{-12} H/μm, respectively, one finds that $\omega_s/2\pi = 1.6 \times 10^{13}$ Hz. These values correspond to NiCr resistors that are 60 nm thick and 1 μm wide.

The imperfection of the current bias scheme, on one hand, and the parasitic and the parasitic capacitances with which the environment shunts the junction, on the other hand, are thus just two aspects of the properties of the function $Z(\omega)$. Questions A and B can now be unified into a single one: What values should $|Z(\omega)|$ have at the junction characteristic frequencies in order to observe Coulomb blockade? It is clear that a *sufficient* condition written in the spirit of (1),

$$|Z(\omega)| \gg R_K \quad \text{for } \omega < \tau_T^{-1}, \tag{13}$$

to ensure that Q is a classical variable is impossible to satisfy. This is a fundamental limitation since the ratio between the impedance of free space Z_V, which controls the asymptotic behavior of $Z(\omega)$ at high frequencies, and the resistance quantum R_K is equal to twice the fine structure constant 1/137.0! One thus cannot avoid the problem of finding the tunneling rate as a function of V for a junction coupled to an arbitrary $Z(\omega)$. The environment needs to be treated quantum-mechanically since over most of the relevant frequency range, thermal fluctuations are smaller than quantum fluctuations, i.e $\hbar\omega \gg k_B T$. In the following we will just emphasize the main features of the theory [21,4,22].

5. Theory of the effect of the quantum nature of the junction electromagnetic environment

The theory first assumes a clear separation of time scales:

$$\tau_t \ll \tau_c \ll \tau_r. \tag{14}$$

The first inequality states that the tunneling time is negligible while the second one restates the classicity of n. The theory then considers the modes of the linear circuit formed by the environment impedance $Z(\omega)$ in series with the junction capacitance C. Of course, for a dissipative environment, the mode frequencies form a continuous spectrum. It is assumed that, before a tunnel event, the environment modes are in their ground state. A tunnel event excites them. This process is described by the function $P(E)$, which gives the probability that the tunneling electron transfers the energy E to the distribution of modes of the circuit. One finds, from a quantum calculation, that $P(E)$ is a distribution function which is given, through a mathematical transformation, by the density of modes given by the real part of the total circuit impedance $Z_t(\omega) = [\mathrm{i}C\omega + Z(\omega)]^{-1}$. The probability $P(E)$ is given by

$$P(E) = \frac{1}{2\pi\hbar}\int_{-\infty}^{+\infty} \mathrm{d}t \, \exp[J(t) + \mathrm{i}Et/\hbar], \tag{15}$$

with the function $J(t)$ given by

$$J(t) = 2\int_0^{\infty} \frac{\mathrm{d}\omega}{\omega} \frac{\mathrm{Re}[Z_t(\omega)]}{R_K} (\coth(\beta\hbar\omega/2) \times [\cos(\omega t) - 1] - \mathrm{i}\sin(\omega t)), \tag{16}$$

where $\beta = 1/k_B T$ is the inverse temperature. Finally, the tunnel rate in the direction imposed by V is computed from $P(E)$ using

$$\Gamma = \frac{\tau_r^{-1}}{E_c}\int_{-\infty}^{+\infty} \mathrm{d}E \int_{-\infty}^{+\infty} \mathrm{d}E' \, f(E)[1 - f(E')] \times P(E + eV - E'), \tag{17}$$

which reflects the fact that only a part of the energy eV of the voltage source is used to excite the environment, the rest being used to excite one hole and one electron on either side of the barrier. At temperatures much lower than the Coulomb gap, i.e. $\beta E_c \gg 1$, one finds:

(i) For impedances $Z(\omega)$ such that, whatever the frequency, $|Z(\omega)| \ll R_K$, the tunneling electrons are not well coupled to lower-frequency modes of the environment. The function $P(E)$ is sharply peaked at $E = 0$, i.e. $P(E) \simeq \delta(E)$, and we find from eq. (17) a straight I–V characteristic with no Coulomb blockade. This result means that most tunneling transitions leave the environment modes undisturbed except for the $\omega = 0$ mode. The charge transferred through the junction is thus removed instantaneously from the voltage source constituted by the pads even though it is physically located a few mm away. In a way, for most tunnel events, *the environment acts as a perfect voltage source*. This is a purely quantum-mechanical effect. It seems to defy locality since one would expect the charge to propagate at the speed of light from the reservoir of charge constituted by the pads to the junction. This expectation, which seemed based on good relativistic common sense, was actually the basis for an argument in favor of the existence of

Coulomb blockade in a low-impedance environment [23]. There is, in fact, no contradiction between locality and the perfect "quantum rigidity" of the charge along the leads that this quasi-elastic tunneling suggests. In this low-impedance environment case, one calculates that the zero-point motion of the environment modes induces quantum fluctuations of the operator Q which are much larger than e. Remember that the junction and the environment behaves as a whole quantum-mechanically coherent unit. One can thus consider that the transferred charge is entirely provided by the zero-point charge fluctuations of the environment. Even though the charge is removed instantaneously from the voltage source, the source cannot tell when a tunnel event occurs because the charge pulse associated with a tunnel event is buried in the quantum fluctuations. This "charge-less" transfer of charge through the junction is analogous to the Mössbauer effect. Gamma rays can be emitted from a nucleus in a solid without exciting the phonon modes. The conservation of momentum is not violated because the recoil momentum of the nucleus is transferred to the whole crystal ("recoil-less" emission). One can think of our function $P(E)$ as equivalent to the gamma ray energy spectrum.

(ii) For impedances such that $|Z(\omega)| \gg R_K$ for all frequencies $\omega \lesssim \tau_c^{-1} \ll 1$, the tunneling electrons are well coupled to all the environment modes. One finds that the function $P(E)$ is sharply peaked at E_c, i.e. $P(E) \simeq \delta(E - E_c)$. Hence, like in the classical case, an electron can only tunnel when it gains at least E_c from the applied voltage, which leads to a Coulomb blockade of tunneling. The problem is that this limit is very difficult to achieve experimentally. We have represented in fig. 6 the domain where $|Z(\omega)| \ll R_K$ by a shaded area. The $\omega^{-1/2}$ roll-off of the impedance must cross the impedance quantum R_K at low enough frequencies on the scale of the Coulomb frequency τ_c^{-1}. In practice this means that the on-chip lead resistors must have a saturation frequency ω_s as high as possible, which in turn means a resistance and a capacitance per unit length as high as possible. This requirement is unfortunately in conflict with the requirement of no heating in the resistor, and a compromise has to be found. This has been achieved by Cleland et al. [24] for normal junctions and by Kuzmin et al. [25] for superconducting junctions. The theory we have outlined can easily be adapted to the tunneling of Cooper pairs [26]. The function $P(E)$, modified slightly to take into account the charge $2e$ of Cooper pairs, yields directly the I–V characteristic if no quasi-particles are present. As in the normal state, theory predicts that no "Cooper pair gap" exists for a single junction if the environment impedance is less than R_K. This result explains that no Cooper pair gap was found in the Harvard group experiments [27].

6. Validity of the global rule

Let us compute the total equilibrium electrostatic energy of the circuit of fig. 4a as a function of the number n of electrons that went through the junction. We make this calculation as in section 2 for the one-island circuit, taking into account the work performed by the voltage source. One finds:

$$E_{eq} = -ne + \text{terms independent of } n. \tag{18}$$

It is thus always energetically favorable for an electron, as far as equilibrium states are concerned, to tunnel. The charge fluctuations of the junction capacitance, which are much greater than e in standard cases, enable, so to speak, the system to overcome the Coulomb barrier and to find its equilibrium state. Coulomb blockade in a current-biased single jucntion is thus a dynamical effect: one can only slow down the tunneling rates as much as possible by making the environment impedance as high as possible. The global rule (4) together with (18) is equivalent to (17) in the low-environment-impedance limit and predicts that there is no Coulomb blockade.

Consider now the one-island system in the standard case of low-impedance leads between the system and the voltage sources. In addition to the number of electrons n_1 and n_2 having traversed the junctions we should also, in principle, consider the electromagnetic degrees of freedom which are the charges Q_1, Q_2 and Q_G on the capacitances C_1, C_2 and C_G, respectively (the sign convention for these charges is that they should increase when the total charge of the island increases). As in the single-junction case these charges undergo large zero-point fluctuations. However, these fluctuations are not independent. The linear combination $Q_1 + Q_2 + Q_G = -ne$, which is decoupled from the leads, is quantized in units of the elementary charge and can only change during a tunneling event. On the other hand, the two other independent linear combinations which are directly coupled to the leads undergo large quantum fluctuations and thus cannot be excited by the tunnel events, just like the capacitor charge in the single-junction case. The effective electrostatic energy change induced by a tunneling event can thus be computed using the expression of the equilibrium energy (3). In other words, in the standard case of low-impedance leads, the voltage sources can be considered as ideal, and the global rule is a very good approximation for the computation of the tunneling rates. A detailed discussion of the validity of the global rule for the one-island system is given in ref. [28]. As we have seen in section 2, the global rule predicts a clear-cut Coulomb gap, except for particular values of the gate voltage U, because the system energy, even at equilibrium, increases when an electron passes through the island.

Using the global rule, one can show that SET oscillations do not exist for the one-island system. However, long one-dimensional arrays of tunnel junction can be traversed by charge solitons and exhibit a behavior resembling that of SET oscillations [11,29]. A relation between frequency and current such as (10) can be best observed, however, using devices with a small number of islands such as the single-electron turnstile [30] or the single-electron pump [31]. The dynamics of these devices in which an external radio-frequency signal applied to gate electrodes clock electrons one at a time is found to be very accurately described by the global-rule equations.

7. Concluding summary

Coulomb blockade is the suppression of the zero voltage conductance of a tunnel junction caused by the Coulomb interaction of the tunneling electrons with the electromagnetic environment of the junction, i.e. the circuit to which it is electrically connected. Coulomb blockade only occurs if the electrostatic energy $e^2/2C$ of a *single* electron on the junction capacitance C is much greater than the characteristic energy k_BT of thermal fluctuations. Another condition for the occurrence of Coulomb blockade is that the impedance with which the environment shunts the junction capacitance be much larger than the resistance quantum $R_K = h/e^2 \simeq 26\ \mathrm{k}\Omega$ at frequencies smaller than the charging energy frequency $e^2/C\hbar$. In general, unless a very special experimental scheme is adopted, a single junction connected to an I–V measuring apparatus by resistive leads cannot meet this condition. There is a profound reason for this: the ratio between the impedance of the vacuum $Z \simeq 377\ \Omega$ and the resistance quantum R_K, which is fixed by the fine structure constant, is much less than unity. Coulomb blockade can readily manifest itself, however, if the environment of the junction consists of other small capacitance junctions.

References

[1] R.A. Millikan, Phys. Rev. 32 (1911) 359.
[2] T.A. Fulton and G.J. Dolan, Phys. Rev. Lett. 59 (1987) 109.
[3] D.A. Fraser, The Physics of Semiconductor Devices (Clarendon, Oxford, 1986).
[4] M.H. Devoret et al., Phys. Rev. Lett. 64 (1990) 1824.
[5] D.V. Averin and K.K. Likharev, in: Quantum Effects in Small Disordered Systems, Eds. B.I. Altshuler, P.A. Lee and R.A. Webb (Elsevier, Amsterdam, 1991).
[6] G. Schön and A.D. Zaikin, Phys. Rep. 198 (1990) 237.
[7] Special Issue on Single Charge Tunneling, Z. Phys. B 85 (1991) 317–468.
[8] R. Wilkins, E. Ben-Jacob and R.C. Jaklevic, Phys. Rev. Lett. 63 (1989) 801.
[9] U. Meirav et al., Phys. Rev. B 40 (1989) 5871.
[10] K.A. Matveev, Sov. Phys. JETP 72 (1991) 892.
[11] K.K. Likharev et al., IEEE Trans. Mag. 25 (1989) 1436.
[12] U. Geigenmüller and G. Schön, Europhys. Lett. 10 (1989) 765.
[13] A. Widom et al., J. Phys. A 15 (1982) 3877.
[14] K.K. Likharev and A.B. Zorin, J. Low-Temp. Phys. 59 (1985) 347.
[15] E. Ben-Jacob and Y. Gefen, Phys. Lett. A 108 (1985) 289.

[16] D.V. Averin and K.K. Likharev, J. Low-Temp. Phys. 62 (1986) 345.
[17] M. Büttiker and R. Landauer, Phys. Rev. Lett. 49 (1982) 1739.
[18] E.H. Hauge and J.A. Støvneng, Rev. Mod. Phys. 61 (1989) 917.
[19] B.N.J. Personn and A. Baratoff, Phys. Rev. B 38 (1988) 9616.
[20] J.M. Martinis and R.L. Kautz, Phys. Rev. Lett. 63 (1989) 1507.
[21] Yu.V. Nazarov, Sov. Phys. JETP Lett. 49 (1989) 126.
[22] S.M. Girvin et al., Phys. Rev. Lett. 64 (1990) 3183.
[23] M. Büttiker and R. Landauer, IBM J. Res. Dev. 30 (1986) 451.
[24] A.N. Cleland, J.M. Schmidt and J. Clarke, Phys. Rev. Lett. 64 (1990) 1565.
[25] L.S. Kuzmin et al., Phys. Rev. Lett. 67 (1991) 1161.
[26] D.V. Averin, Yu.V. Nazarov and A.A. Odintsov, Physica B 165/166 (1990) 945.
[27] A.T. Johnson, C.J. Lobb and M. Tinkham, Phys. Rev. Lett. 65 (1990) 1263.
[28] G.-L. Ingold, P. Wyrowski and H. Grabert, Z. Phys. B 85 (1991) 443.
[29] P. Delsing et al., Phys. Rev. Lett. 63 (1989) 1861.
[30] L.J. Geerligs et al., Phys. Rev. Lett. 64 (1990) 2691.
[31] H. Pothier et al., Europhys. Lett. 17 (1992) 249.