

Effects of overheating in a single-electron transistor

A. N. Korotkov^{a)}

Institute of Nuclear Physics, Moscow State University, Moscow 119899 GSP, Russia

M. R. Samuelsen

Physics Department, The Technical University of Denmark, DK-2800, Lyngby, Denmark

S. A. Vasenko

Physics Department, Moscow State University, Moscow 119899 GSP, Russia

(Received 18 October 1993; accepted for publication 8 June 1994)

Heating of a single-electron transistor (SET) caused by the current flowing through it is considered. The current and the temperature increase should be calculated self-consistently taking into account various paths of the heat drain. Even if there is no heat drain from the central electrode of the SET due to transfer of phonons, the temperature of this electrode remains finite because electron tunneling decreases the temperature difference between the central and outer electrodes. Overheating effects can cause hysteresis in the I - V curve of the SET in the vicinity of the Coulomb blockade threshold.

I. INTRODUCTION

It is well known now that the charge of even one electron can significantly affect the behavior of submicron tunnel structures.¹⁻³ These single-electron effects appear to be the physical basis of a new branch of nanoelectronic devices.

During the last years much attention was focused on the simplest of such devices, the so-called single-electron transistor (SET).⁴⁻¹⁰ It consists of two small-capacitance tunnel junctions in series (Fig. 1). The subelectron charge q_0 induced on the central electrode of the SET controls the current through this device. There are three main ways that this double-junction structure is experimentally realized.³ It can be made using thin metal films, semiconductor heterostructures, or scanning tunneling microscope (STM).

The basis of the theoretical consideration of the SET is given by the relatively simple "orthodox" theory,⁵ which is in surprisingly good agreement with most of the experimental results. Some other theories improve the orthodox theory, taking into account various effects, for example, cotunneling processes,^{11,12} the influence of the electromagnetic environment,¹³ and the discreteness of the electron spectrum.^{14,15}

One of the assumptions of the orthodox theory is that the energy relaxation of the tunneling electrons is fast enough so that one can use the Fermi distribution for the electrons. The corresponding temperature is assumed to be constant and equal to the temperature of the substrate.

However, these assumptions can be inapplicable to real experiments because the rates of electron-electron and electron-phonon interactions are very small at low temperatures.

The opposite limiting case of negligible energy relaxation in the central electrode of the SET was considered in Ref. 14. If the relaxation time is larger than the average time the tunneling electron spends on the central electrode, then the electron energy distribution function in this electrode is

determined only by the kinetics of the tunneling. It was shown in Ref. 14 that in this case (extreme overheating) the Coulomb staircase of the I - V curve is partially suppressed; however, for typical values of experimental parameters this effect is small.

Between the orthodox theory and the model of Ref. 14 there is the case in which the energy relaxation due to electron-electron and electron-phonon interactions is fast enough to produce a Fermi distribution; however, the corresponding electron temperature differs from that of the substrate and should be calculated self-consistently. It was found that this case is realized in experiments.^{16,17}

The detailed investigation of overheating was carried out recently in Ref. 17 for the thin-film SET. It was found that the electron temperature of the middle electrode (island) differs from that of the outer electrodes and substrate. The bottleneck for heat drain from the island was taken to be the heat transfer from the electron gas to the phonons in the middle electrode.

In the present article we consider overheating effects in various types of the SET (the metal films case, the semiconductor case, and the STM case). Several models of heat drain are discussed. In particular, we consider the case when the heat drain from the central electrode due to phonon flow is negligible; then the energy is transferred from the hot central electrode to the cold outer electrodes only by the tunneling electrons.

We present some analytical and numerical results for the I - V curve and $T(V)$ dependence. The important consequence of overheating is the possibility of hysteresis in the dc I - V curve.

The article is organized as follows. In the following section the basic equations for the modification of the orthodox theory which takes into account overheating of the SET are presented. The various models for heat drain are discussed in the Sec. III. The analytical and numerical results for the I - V curve and the temperature-voltage dependence are presented in Secs. IV and V. We consider the case of negligible heat transfer between the central electrode and the substrate in

^{a)}Present address: Department of Physics, State University of New York, Stony Brook, NY 11794-3800.

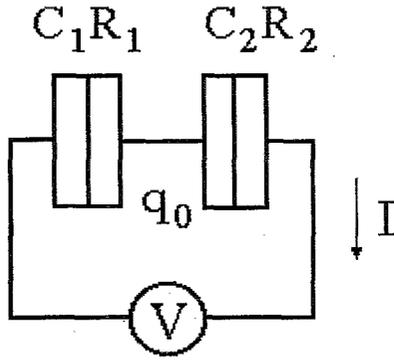


FIG. 1. The single-electron transistor (SET) consisting of two tunnel junctions in series biased by a voltage source V . The current I through the SET can be controlled by subelectron variation of the charge q_0 of the central electrode.

Sec. IV and the model of considerable heat drain in Sec. V. We discuss in detail the hysteresis caused by the overheating effects.

II. BASIC EQUATIONS

Before consideration of the double-junction system, let us discuss the simpler case of a single junction biased by a voltage \tilde{V} (this voltage can be either positive or negative). The rate Γ of electron tunneling in the positive direction (from the first electrode to the second one) can be calculated using the golden rule

$$\Gamma = \frac{2\pi}{\hbar} \int \rho_1(\epsilon_1) \rho_2(\epsilon_2) |T_{\epsilon_1 \epsilon_2}|^2 n_1(\epsilon_1) [1 - n_2(\epsilon_2)] \times \delta(\epsilon_1 - \epsilon_2 + W) d\epsilon_1 d\epsilon_2, \quad (1)$$

where indices (1) and 2 correspond to the electrodes, $T_{\epsilon_1 \epsilon_2}$ is the tunneling matrix element, ρ_i are the densities of electron states in the electrodes, $n_i(\epsilon_i)$ are the electron distribution functions, and $W = e\tilde{V}$ is the energy gain due to the tunneling event.

We consider the case when the energy relaxation in the electrodes is fast enough to produce the Fermi distribution $n_i(\epsilon_i) = f(\epsilon_i, T_i)$, $f(\epsilon, T) = 1/[1 + \exp(\epsilon/T)]$; however, the corresponding electron temperatures T_1 and T_2 may be different. Note that the energies ϵ_1, ϵ_2 are measured from the Fermi levels of the electrodes.

Assuming that the densities of states and the matrix element are independent of the energy (this approximation is usual for metal junctions), one obtains

$$\Gamma(W, T_1, T_2, R) = \frac{1}{e^2 R} \int f(\epsilon, T_1) [1 - f(\epsilon + W, T_2)] d\epsilon, \quad (2)$$

where R is the tunnel resistance of the junction, and e is the electron charge. This expression can be simplified in some special cases, for example, if $T_1 = T_2$ or $T_1 T_2 = 0$. The rate given by Eq. (2) is insensitive to the interchange of electrode temperatures $T_1 \leftrightarrow T_2$;¹⁷ however, this is not true in the general case given by Eq. (1). As an example, consider the important case of a low semiconductor tunnel barrier.

Following Refs. 10 and 18 let us use the WKB approximation for the tunneling matrix element,

$$|T_{\epsilon_1 \epsilon_2}|^2 \propto \exp\left(-\frac{(8m)^{1/2}}{\hbar} \int_0^d \left[U(x) - \frac{e\tilde{V}x}{d} - \epsilon_1\right]^{1/2} dx\right), \quad (3)$$

where m is the effective electron mass, and $U(x)$ and d are the shape and the thickness of the tunnel barrier. Simple calculations to first order lead to the expression

$$\Gamma(W, T_1, T_2, R) = \frac{1}{e^2 R} \left[\exp\left(\frac{2W\tau^+}{\hbar}\right) \frac{T_1 \pi}{\sin(2\tau\pi T_1/\hbar)} - \exp\left(-\frac{2W\tau^-}{\hbar}\right) \int f(\epsilon - W, T_1) f(\epsilon, T_2) \times \exp\left(\frac{2\epsilon\tau}{\hbar}\right) d\epsilon \right], \quad (4)$$

where

$$\tau = \int [2U(x)/m]^{-1/2} dx,$$

$$\tau^+ = \int (x/d) [2U(x)/m]^{-1/2} dx,$$

$$\tau^- = \tau - \tau^+,$$

and R is the low voltage resistance at zero temperature.

The tunneling through the junction causes heat generation (or absorption) in the electrodes. Let us consider a process consisting of two steps: the tunneling conserving total energy and then the relaxation of the electron and the "hole" created in the electrodes. If the tunneling electron brings some energy ϵ_2 to the second electrode, then after energy relaxation ϵ_2 transforms into heat. Similarly, if the electron having energy ϵ_1 leaves the electrode, an amount of heat $-\epsilon_1$ is produced in it. Hence, the powers P_1, P_2 of heat generation in the electrodes are

$$P_j = \frac{2\pi}{\hbar} \int (-1)^j \epsilon_j \rho_1(\epsilon_1) \rho_2(\epsilon_2) |T_{\epsilon_1 \epsilon_2}|^2 n_1(\epsilon_1) [1 - n_2(\epsilon_2)] \times \delta(\epsilon_1 - \epsilon_2 + W) d\epsilon_1 d\epsilon_2, \quad j=1,2. \quad (5)$$

Adding these two expressions and taking the relation $\epsilon_2 - \epsilon_1 = W = e\tilde{V}$ and Eq. (1) into account, one obtains the obvious equation

$$P_1 + P_2 = W\Gamma = I^+ \tilde{V}, \quad (6)$$

where $I^+ = e\Gamma$ is the part of the current flowing in the positive direction.

Assuming the WKB approximation for the matrix element [as for Eq. (4)], one obtains from Eq. (5)

$$P_1(W, T_1, T_2, R) = \frac{1}{e^2 R} \exp\left(\frac{2W\tau^+}{\hbar}\right) \int (-\epsilon) f(\epsilon, T_1) \times [1 - f(\epsilon + W, T_2)] \exp\left(\frac{2\epsilon\tau}{\hbar}\right) d\epsilon. \quad (7)$$

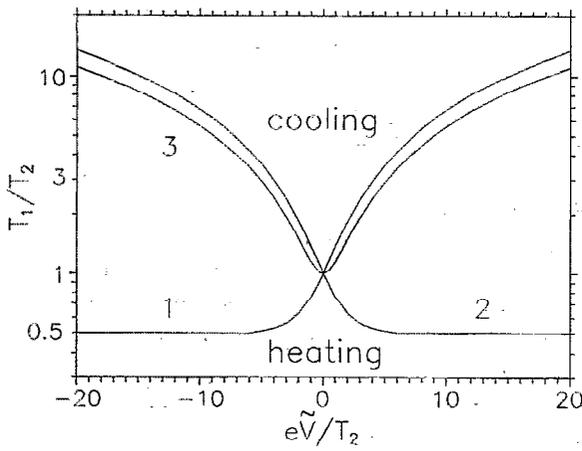


FIG. 2. Illustration of the possibility of cooling due to the tunneling. The curves separate regions of heating and cooling of the "first" electrode of the tunnel junction biased by the voltage \tilde{V} . Curve 1 corresponds to tunneling only from the first electrode to the second one, curve 2 corresponds to tunneling in the opposite direction, and curve 3 takes both processes into account.

A further simplification assuming a constant matrix element ($\tau^+ = \tau^- = 0$), gives

$$P_1(W, T_1, T_2, R) = \left(\frac{1}{e^2 R} \right) \int (-\epsilon) f(\epsilon, T_1) \times [1 - f(\epsilon + W, T_2)] d\epsilon. \quad (8)$$

If the temperatures are also equal, $T_1 = T_2 = T$, then the total heat generated in the junction is evenly divided between two electrodes,

$$P_1 = P_2 = \frac{1}{2} W \Gamma, \quad \Gamma = \frac{1}{e^2 R} \frac{W}{1 - \exp(-W/T)}. \quad (9)$$

When the temperatures of the electrodes are different, the corresponding heating powers are different. Moreover, heat can be generated in one electrode and absorbed in the other one. The reason is that in accordance with general principles of thermodynamics the tunneling of electrons provides a heat flow from the hot side to the cold one. If the temperature difference is large enough, the hot electrode will be cooled by tunneling. This effect is illustrated in Fig. 2 for the simplest case of constant matrix element [see Eq. (8)]. The curves separate regions of heating and cooling of the first electrode. Curve 1 corresponds to the heating or cooling due to tunneling only from the first electrode to the second one, curve 2 corresponds to tunneling in the opposite direction, and curve 3 takes both processes into account.

Now let us consider single-electron transport in a system of two tunnel junctions in series (Fig. 1), and extend the previous consideration to this case. The orthodox theory⁵ enables us to calculate all characteristics of the double junction system if one knows the rates $\Gamma_{1,2}^\pm(n)$ of tunneling through the first and second junctions leading to an increase (+) or decrease (-) in the number n of excess electrons on the central electrode. In particular, the stationary probability distribution of the different charge states $\sigma(n)$ obeys the equation¹

$$\sigma(n)[\Gamma_1^+(n) + \Gamma_2^+(n)] = \sigma(n+1)[\Gamma_1^-(n+1) + \Gamma_2^-(n+1)], \quad (10)$$

$$\sum_n \sigma(n) = 1,$$

and the average current through the system can be calculated as

$$I = e \sum_n \sigma(n) [\Gamma_1^+(n) - \Gamma_1^-(n)]. \quad (11)$$

The only modification of the orthodox theory needed is the use of Eqs. (1), (2), (4) for calculation of $\Gamma_{1,2}^\pm(n)$ and self-consistent calculation of temperatures. Let T_1, T_2, T_m be the electron temperatures of the outer electrodes and the middle electrode (we assume that the parts of the middle electrode adjacent to different junctions have equal temperatures). Then [see Eqs. (2) and (4)]

$$\Gamma_j^+(n) = \Gamma[W_j^+(n), T_j, T_m, R_j], \quad j=1,2, \quad (12)$$

$$\Gamma_j^-(n) = \Gamma[W_j^-(n), T_m, T_j, R_j],$$

where the energy gain $W_j^\pm(n)$ for the tunnel event is the decrease in the Gibbs energy of the whole system,¹

$$W_j^\pm(n) = \mp \frac{e}{C_\Sigma} [q_0 + e(n \pm \frac{1}{2})] \mp (-1)^j \frac{C_1 C_2}{C_j C_\Sigma} eV. \quad (13)$$

Here $R_{1,2}$ and $C_{1,2}$ are the tunnel resistances and capacitances of two tunnel junctions, $C_\Sigma = C_1 + C_2$, q_0 is the background charge of the central electrode, and V is the total voltage (Fig. 1).

The temperatures T_1, T_2, T_m should be calculated self-consistently to ensure the balance in each electrode between the heat generated by tunneling and the heat drain P_{out} by thermoconductivity, energy exchange between electrons and phonons, etc.,

$$\sum_n \sigma(n) \{ P_1[W_1^+(n), T_1, T_m, R_1] + P_2[W_1^-(n), T_m, T_1, R_1] \} = P_{\text{out},1}, \quad (14a)$$

$$\sum_n \sigma(n) \{ P_1[W_2^+(n), T_m, T_2, R_2] + P_2[W_2^-(n), T_2, T_m, R_2] \} = P_{\text{out},2}, \quad (14b)$$

$$\sum_n \sigma(n) \sum_j \{ P_1[W_j^-(n), T_m, T_j, R_j] + P_2[W_j^+(n), T_j, T_m, R_j] \} = P_{\text{out},m}. \quad (14c)$$

Equations (10)–(14) and the equations for heat drain P_{out} form a complete set for the calculation of overheating of the SET. In contrast to equations above (which accurately describe heat generation in a SET), we can use only rough estimates of P_{out} because of the lack of detailed experimental research in the interesting range of parameters. We consider several models for heat drain in the following section.

At the end of this section let us consider briefly heat generation in the case when the voltage is smaller than the

blockade threshold, and the current is caused by cotunneling processes¹¹ only (this case is not a subject for the present article; however, it is important for experiments). In “elastic” cotunneling,¹² only the outer electrodes are heated because the electron distribution in the middle one is not affected by the tunneling event. The expressions for heat generation in the outer electrodes are similar to Eqs. (5) and (8) for a single junction if one uses the effective matrix element and the effective resistance $R_{\text{eff}}=V/I$. In the case of “inelastic” cotunneling¹² one electron-hole pair appears in the middle electrode after each tunneling event. Following Ref. 12, one can calculate the corresponding heating power,

$$P_m = \frac{\hbar}{2\pi e^4 R_1 R_2} \int d\epsilon_1 d\epsilon_2 d\epsilon_3 d\epsilon_4 (\epsilon_2 - \epsilon_3) f(\epsilon_1, T_1) \times [1 - f(\epsilon_2, T_m)] f(\epsilon_3, T_m) [1 - f(\epsilon_4, T_2)] \left(\frac{1}{\epsilon_2 - \epsilon_1 + E_1} + \frac{1}{\epsilon_4 - \epsilon_3 + E_2} \right)^2 \delta(eV + \epsilon_1 - \epsilon_2 + \epsilon_3 - \epsilon_4), \quad (15)$$

where the energies of the intermediate (virtual) states are $E_1 = -W_1^+(0)$, $E_2 = -W_2^-(0)$. The heat generation in the first and second outer electrodes is given by similar expressions with the substitution of $(\epsilon_2 - \epsilon_3)$ with $-\epsilon_1$ and ϵ_4 , respectively.

III. MODELS FOR HEAT DRAIN

For a crude estimate of overheating in the different types of SETs, let us first estimate the total heating power $P=IV$. Considering a SET made of thin metal films, assume that the voltage is on the order of the blockade threshold voltage $V_t \sim 1$ mV, and that the current is about V_t/R_Σ , where $R_\Sigma \sim 10^5 \Omega$. Then the heating power is on the order of 10^{-11} W. The use of voltages in the vicinity of the blockade threshold and the use of larger resistances or capacitances decrease this estimate; however, the heating power is usually not less than 10^{-14} – 10^{-13} W.

The semiconductor SET (in which electrons tunnel through a small island of a 2D electron gas) usually has capacitances and resistances on the same order as in the metallic case. Hence, heat up to 10^{-11} W can be generated in these experiments; however, bias voltages much smaller than the maximum blockade threshold are usually used. Let us take $V=3 \mu\text{V}$, then for an “open” SET (peak of the conductance) the heating power is about 10^{-16} W.

The double-junction system in STM experiments can have a threshold voltage as high as 10–100 mV. So, if the resistance of this system is on the order of $10^6 \Omega$, the heating power is 10^{-10} – 10^{-8} W. Although this value is much larger than in the previous cases, one should note that the substrate temperatures are usually much higher in the STM case.

The heat generated in the SET causes an increase in the electron temperature of the electrodes. In order to relate this increase to the heating powers in the electrodes, and to obtain the expressions for the right-hand side of Eqs. (14), one should consider the heat drain from the electrodes.

The estimates above show that the heat generated in one or several SETs cannot increase the temperature of the whole

substrate considerably, even if the experiment is carried out in a dilution refrigerator (this effect would be important only on large integration levels). Hence, the substrate temperature T_0 far from the SET region is constant.

For the thin metal films case, in the vicinity of the SET one should consider separately the temperature of the substrate, the lattice temperature of the electrodes, and the electron temperatures in the electrodes. It is necessary to take into account the following: electron thermal conductivity in the leads; heat exchange between electron gas and lattice in each electrode; the Kapitza resistance^{19,20} for heat flow between electrodes and substrate; the loss of the energy of the tunneling electron because of either electron-electron ($e-e$) or electron-phonon ($e-ph$) interactions; the escape of created phonons through thin tunnel barriers into metallic leads; heat spread in the substrate and leads.

According to the Wiedemann–Franz law, high electrical conductivity of metal wires causes high thermal conductivity. In Ref. 17 it was shown that the electron temperature of the outer electrodes is close to the refrigeration temperature T_0 . This fact should be valid also for the STM realization of the SET because the outer electrodes are bulk and therefore have better thermal conductivity.

The heat flow from the middle electrode can be considered as a process consisting of several steps. If the incoming electrons and holes lose their energy mainly in $e-e$ interactions, the first step is a thermalization of the electron gas. Then the energy is transferred from the electron gas to phonons. The created phonon can leave the middle electrode without any scattering or be involved in $e-ph$ or $ph-ph$ interactions. In the latter case the temperature of the phonon gas in the middle electrode is well defined. As a next step, phonons can either escape through tunnel barriers into metallic leads or cross the boundary between electrode and substrate and cause heat spread in the substrate.

If the incoming electrons and holes lose energy mainly in $e-ph$ interactions, the first step mentioned above is absent. The increase in the electron temperature in this case is caused by secondary $e-ph$ processes.

For a crude estimation of $e-e$ and $e-ph$ interaction rates, one can use expressions²¹ $\tau_{e-e}^{-1} \approx (\epsilon - \epsilon_F)^2 / \hbar \epsilon_F$, $\tau_{e-ph}^{-1} \approx \omega_D [(\epsilon - \epsilon_F) / 2\hbar k_F c]^3$ where ω_D is Debye frequency, c is the sound velocity, ϵ_F is the Fermi energy, and k_F is the Fermi wave vector. Since the $e-e$ interaction rate scales roughly as $(\epsilon - \epsilon_F)^2$ and the $e-ph$ rate scales roughly as $(\epsilon - \epsilon_F)^3$ (there are of course, deviations from these dependencies in experiments), phonon emission should prevail at large electron and hole energies. In thin metal films it occurs for energies larger than approximately $4 \text{ K} \approx 0.4 \text{ meV}$.²² Hence, if the voltage across the SET is above several millivolts (this is sometimes true in the thin-film case and always true in the STM case), the $e-ph$ interaction probably plays the dominant role in the energy relaxation.

Consider first this case and assume that the electron temperature T_m in the middle electrode is equal to the phonon temperature. If the thermoconductivity of the substrate is high, then the heat flow from the middle electrode to the substrate is restricted by the Kapitza resistance. The heat

drain $P_{\text{out},m}$ in this case follows approximately the dependence^{19,20}

$$P_{\text{out},m} = \alpha S (T_m^4 - T_0^4), \quad (16)$$

where S is the thermal contact area between the middle electrode and the substrate, and the parameter αk_B^4 (k_B is Boltzmann's constant) is on the order of $10^2 \text{ W/m}^2 \text{ K}^4$.^{20,23} Note that there are deviations from this expression in experiments; however, it is quite sufficient for estimates.

We should also take into account the heat drain through the thin tunnel barriers into the metallic leads. In the limit when the barriers are transparent to phonons, the heat drain satisfies again Eq. (16) with S being the area of tunnel contact and²⁴

$$\alpha k_B^4 = \pi^2 / (120 \hbar^3) (1/c_1^2 + 2/c_t^2),$$

which is about $500 \text{ W/m}^2 \text{ K}^4$ for Al (c_1 and c_t are velocities of sound).

Adding the two channels of heat drain, we can still use Eq. (16) by adding the corresponding terms αS . If the area of tunnel contacts is much less than the total middle electrode area, the correction due to the second channel is small.

Now let us consider the case when the e - e interaction is stronger than the e -ph interaction. Then the temperature of the electron gas T_m differs from the lattice temperature T_{ph} of the middle electrode and the heat flow can be calculated as^{25,26}

$$P_{\text{out},m} = \Sigma \Omega (T_m^5 - T_{\text{ph}}^5) = \alpha S (T_{\text{ph}}^4 - T_0^4), \quad (17)$$

where Ω is the volume of the middle electrode and the constant $k_B^4 \Sigma$ is about 0.2 – $2 \text{ nW/K}^5 \mu\text{m}^3$.^{17,25,26} The last part of Eq. (17) means that the Kapitza resistance also should be taken into account (we use the effective value for αS as discussed above).

If

$$(T_0^5 + P_{\text{out},m}/\Sigma\Omega)^{1/5} - T_0 \ll (T_0^4 + P_{\text{out},m}/\alpha S)^{1/4} - T_0, \quad (18)$$

the electron and phonon temperatures practically coincide and Eq. (16) is valid. In the opposite case (small $P_{\text{out},m}$ and T_0), the Kapitza resistance can be neglected and

$$P_{\text{out},m} = \Sigma \Omega (T_m^5 - T_0^5). \quad (19)$$

For numerical estimates let us take the parameters $\alpha = 100 \text{ W/m}^2 \text{ K}^4$, $\Sigma = 2 \text{ nW/K}^5 \mu\text{m}^3$, $S = 1 \times 0.1 \mu\text{m}^2$, $\Omega = S \times 0.05 \mu\text{m}$, and $T_0 = 50 \text{ mK}$. Then the heating power which separates cases (16) and (19) is about 10^{-11} W , and corresponds to overheating of about 1 K . Hence, Eq. (16) is appropriate only for very strong overheating. However, there is not much reason to discuss the distinction between models (16) and (19) because both models are crude approximations, and they show formal similarity.

Using the parameters above, one can conclude that overheating in thin-film SETs is considerable ($T_m - T_0 > T_0$) for a heat generation more than 10^{-16} W , i.e., practically always.

Now let us estimate overheating of the semiconductor SET. If the 2D electron gas has a resistance on the order of $10 \Omega/\square$; then from the Wiedemann–Franz law the thermal conductivity at $T_0 = 50 \text{ mK}$ is about 10^{-10} W/K per square. Hence, overheating of the outer electrodes is negligible for a

heat generation less than 10^{-12} W ; however, the electron temperature in the middle electrode (island of the 2D electron gas) can differ considerably from T_0 . The problem of hot electrons in the 2D gas has been extensively investigated for the temperature range 2 – 20 K (see, for example, Refs. 27–29 and references therein). Extrapolating the experimental data of Refs. 27 and 28, one can conclude that a heating power of about 10^{-18} – 10^{-17} W per electron is sufficient to increase the electron temperature from the mK range up to 0.5 K . The same degree of overheating should appear in island containing 10^2 – 10^3 electrons for heating power roughly on the order of 10^{-15} W . If one is interested in the case when the temperature increases from $T_0 = 50 \text{ mK}$ to, say, $T_0 = 100 \text{ mK}$, one should reduce further the heating power probably by two orders of magnitude. Hence, even for bias voltages as small as $3 \mu\text{V}$, overheating of the open SET is considerable. Note that the temperature of the island can exceed the bias voltage, $T_m > eV$; in principle, T_m can exceed eV by many times. In this case the heat transfer due to tunneling discussed in the following section is very important.

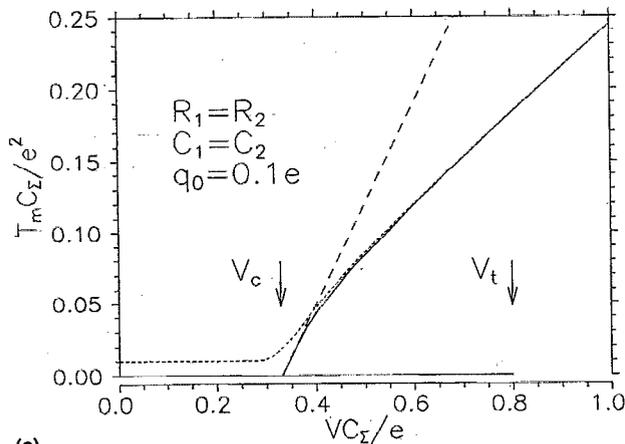
IV. COOLING DUE TO TUNNELING

Let us consider the situation when the temperatures of the outer electrodes are equal to the temperature of the substrate, $T_1 = T_2 = T_0$, and only the central electrode is heated. In Sec. II we noted that the tunneling provides heat flow from the hot electrode to the cold one. The tunneling leads not only to heat generation in the central electrode, but also to heat flow from it. Of course, the distinction between these processes is rather arbitrary (although it has a clear physical meaning), and Eqs. (14) and (5)–(8) take both effects into account simultaneously.

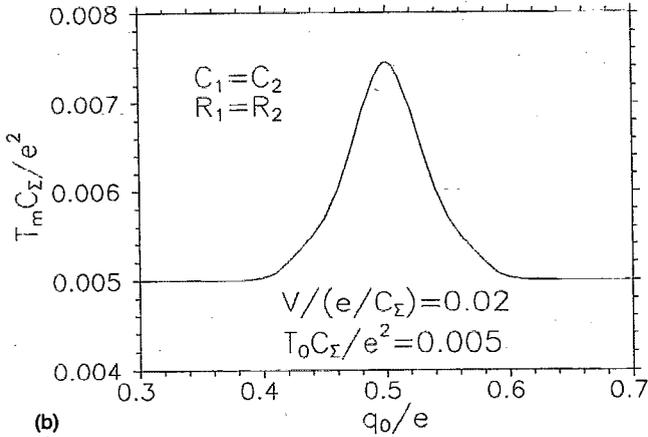
In order to make clear the role of heat flow due to tunneling and to estimate the effectiveness of it, let us assume that there is no other process of heat drain from the middle electrode, i.e., $P_{\text{out},m} = 0$ in Eq. (14c) (a similar assumption was used in Ref. 30 for a single junction). The temperature T_m of the middle electrode and the current through the double junction system can be calculated using Eqs. (2), (6), (8), (10)–(13), and (14c) (for simplicity we consider only the case of constant tunneling matrix element).

The numerical results for the dependence $T_m(V)$ are presented in Fig. 3(a) (solid line) for the symmetrical system $R_1 = R_2$, $C_1 = C_2$ at $T_0 = 0$. When the voltage is less than some critical value V_c , only the trivial solution $T_m = 0$ exists and no current flows through the system. In the voltage region between V_c and V_t , formally two solutions exist: the trivial one and another solution with $T_m \neq 0$ corresponding to nonzero current through the system. One can show, however, that only the nontrivial solution is stable.

It is possible to carry out analytical calculations in the vicinity of V_c because T_m is small and only two charge states, say, $n = 0$ and $n = -1$ should be taken into account [this implies $(C_2 - C_1)V_c < 2q_0 < e$]. Using the low-temperature approximation of Eqs. (2) and (10) one obtains the probabilities



(a)



(b)

FIG. 3. Electron temperature T_m of the thermally isolated middle electrode of the SET as a function of (a) the voltage V , and (b) the background charge q_0 . In (a) the solid line corresponds to $T_0=0$, the dotted line to $T_0=0.01e^2/C_\Sigma$, and the dashed line shows the analytical result [Eq. (22)].

$$\sigma(-1) = \frac{R_1 T_m}{R_2 e V_t} \frac{\exp[C_1 e(V - V_t)/C_\Sigma T_m]}{1 + (V/V_t - 1)(C_2/C_\Sigma - R_1 C_1/R_2 C_\Sigma)},$$

$$\sigma(0) = 1 - \sigma(-1) \gg \sigma(-1). \quad (20)$$

In the case under consideration, tunneling events which increase charge states have $W > 0$ and heat the central electrode; events which decrease n have $W < 0$ and cool it. The energy balance [Eq. (14c)] has the form

$$\begin{aligned} & [\sigma(-1)/2R_1] [V_t^2 + 2C_2 V_t (V - V_t)/C_\Sigma \\ & + (V - V_t)^2 (C_1^2 R_1 / C_\Sigma^2 R_2 + C_2^2 / C_\Sigma^2)] \\ & = (T_m^2 / e^2 R_2) [1 - C_1 e(V - V_t)/C_\Sigma T_m] \\ & \times \exp[C_1 e(V - V_t)/C_\Sigma T_m]. \end{aligned} \quad (21)$$

The nontrivial solution ($T_m \neq 0$) of this equation in the vicinity of V_c is

$$T_m(V) = [1 + (C_1/C_\Sigma)(R_\Sigma/R_2)^{1/2}] \times [1 + (R_\Sigma/R_2)^{1/2}]^{-1} e(V - V_c), \quad (22)$$

where

$$V_c = V_t / [1 + (C_\Sigma/C_1)(R_2/R_\Sigma)^{1/2}]. \quad (23)$$

This dependence is shown in Fig. 3(a) by the dashed line.

If the temperature of the outer electrodes T_0 is not strictly zero, the $T_m(V)$ dependence [the dotted line in Fig. 3(a)] has no lower part between V_c and V_t because any small current heats the central electrode until the cooling due to tunneling becomes considerable. It is worth noting that the real bistability, leading to hysteresis, appears if the additional process of heat drain is taken into consideration (see following section).

Thus, the main consequence of the model of negligible heat drain from the central electrode is a strong shift of the threshold voltage $V_t \rightarrow V_c$ and a relatively high temperature of the central electrode. For example, in the case shown in Fig. 3(a), we have $T_m C_\Sigma / e^2 = 0.18$ at $V = V_t$. Such overheating was never observed experimentally. This means that in this case the transfer of heat to the lattice plays the main role in the cooling of the central electrode. However, if the background charge q_0 is close to $e/2$ and $V \approx V_t \ll e/C_\Sigma$ (this case is typical for semiconductor SET), then the heat balance provided by cooling due to tunneling only requires not very high overheating of the middle electrode [see Eq. (22) and Fig. 3(b)]. In this case the cooling due to tunneling can be dominant in comparison to the weak heat transfer to the lattice at low temperature.

V. CONSIDERABLE HEAT DRAIN

Now let us calculate the I - V curves and the temperature-voltage dependencies for the two models of heat drain discussed in Sec. III.

First, let us assume that the electron gas in the outer electrodes has fixed temperature $T_1 = T_2 = T_0$, and the main factor determining the electron temperature in the central electrode is the Kapitza resistance for heat flow from the central electrode to the substrate. Then $P_{out,m}$ in Eq. (14c) is determined by Eq. (16).

It is convenient to characterize the scale of overheating by a dimensionless parameter $\gamma = T_h C_\Sigma / e^2$, where T_h is the electron temperature of the middle electrode if the power $(e/C_\Sigma)^2 / R_\Sigma$ is dissipated in it and the substrate temperature is assumed to be zero. From Eq. (16) it follows that

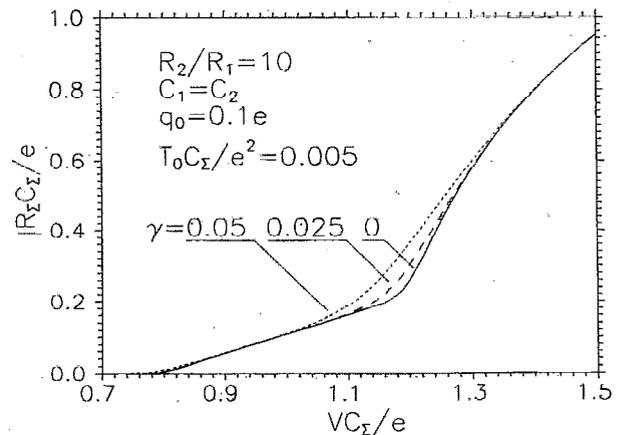


FIG. 4. Large scale I - V curves of the SET for three values of overheating parameter $\gamma=0, 0.025$, and 0.05 . The heat drain is described by Eq. (16).

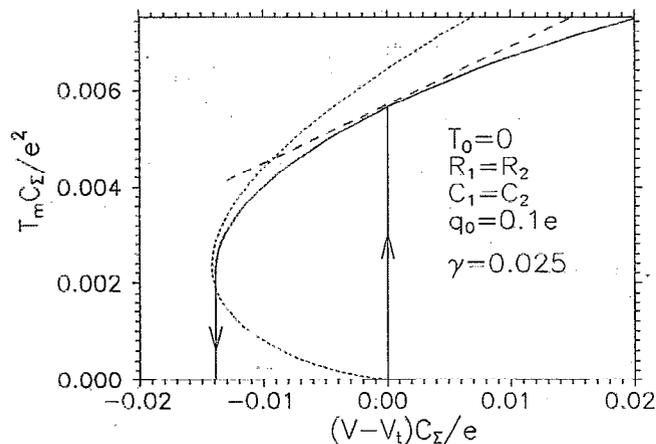


FIG. 5. The hysteretic $T_m(V)$ dependence in the vicinity of the Coulomb blockade threshold V_t for $\gamma=0.025$ and $T_0=0$. The dotted line corresponds to Eq. (25), the dashed line to Eq. (29).

$\gamma = [C_S^2 / (e^6 \alpha S R_S)]^{1/4}$. Taking, for example, typical parameters of the STM SET $C_S = 10^{-17}$ F, $S = (100 \text{ \AA})^2$, $R_S = 10^7 \Omega$, and $k_B^4 \alpha = 100 \text{ W/m}^2 \text{ K}^4$, one obtains $\gamma = 0.04$. It is obvious that the overheating is considerable if $T_0 \ll \gamma e^2 / C_S$, i.e., $T_0 \ll 7$ K in this case.

Results of the numerical calculations of the $I-V$ curve are presented in Fig. 4 for the case $C_1 = C_2$, $R_2 = 10R_1$, $q_0 = 0.1e$, $T_0 C_S / e^2 = 0.005$, and three values of γ : $\gamma = 0, 0.025, 0.05$. Overheating causes additional rounding of the features of the Coulomb staircase. This rounding becomes stronger when the voltage increases. The influence of overheating in the vicinity of the Coulomb blockade threshold leads also to significant deviations from the $I-V$ curve given by the orthodox theory (Fig. 4). Moreover, at small temperatures T_0 hysteresis in the $I-V$ curve is possible in this region.

The origin of the hysteresis is quite clear. Suppose that the temperatures of all electrodes of the SET are small and the voltage V is slightly below the Coulomb blockade threshold V_t . Then no current flows and no heat is produced. Now suppose that the temperature increases, then current appears and produces heat. The amount of heat production can be enough to maintain the increased temperature.

The hysteretic dependencies $T_m(V)$ and $I(V)$ for $T_0 = 0$ near the blockade threshold are shown in Figs. 5 and 6. It is possible to carry out some analytical calculations for $T_0 = 0$ and small γ .

Consider first the case

$$T_m / e \ll (V_t - V) \ll V_t \quad (24)$$

and assume that the blockade threshold for the first junction is larger than that for the second junction. Then, using Eqs. (2), (8), and (6) one can write Eq. (14c) for heat balance in the form

$$(T_m V_t / 2eR_2) \exp[C_1 e(V - V_t) / C_S T_m] = \alpha S T_m^4 \quad (25)$$

The nontrivial solution of this equation is a two-valued function (see the dotted curve in Fig. 5) within the voltage range of width

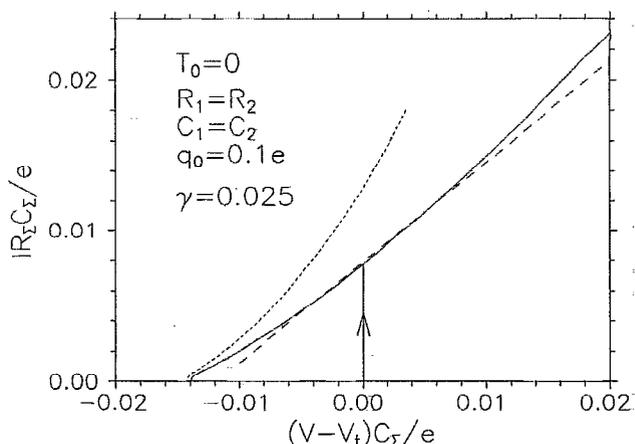


FIG. 6. The hysteretic $I-V$ curve for the parameters of Fig. 5. Dotted line: Eq. (27); dashed line: Eq. (30).

$$\Delta V = (3C_S / eC_1) \exp(-1) \left(\frac{V_t}{2\alpha S e R_2} \right)^{1/3} \quad (26)$$

It is easy to prove that the lower branch is unstable. Hence, only the upper branch and the trivial solution $T_m = 0$ are stable within the hysteresis range. The current in this approximation is given by the expression

$$I = (T_m / eR_2) \exp[C_1 e(V - V_t) / C_S T_m], \quad (27)$$

which coincides well with numerical calculations only near the left-hand edge of the hysteresis region (Fig. 6).

In the close vicinity of V_t ,

$$e|V - V_t| \ll T_m, \quad (28)$$

Eqs. (2), (6), (8), and (14c) give the following expressions for the temperature and the current:

$$T_m = T_t + (C_1 / C_S) e(V - V_t) / (6 \ln 2), \quad (29)$$

$$I = \ln 2 (T_t / eR_2) + (2C_1 / 3C_S) (V - V_t) / R_2, \quad (30)$$

$$T_t = \left(\frac{V_t \ln 2}{2\alpha S e R_2} \right)^{1/3}. \quad (31)$$

They are represented by dashed lines in Figs. 5 and 6.

The range of hysteresis is wider when the heat drain is weaker and consequently the parameter γ is larger. The limit $\gamma \rightarrow \infty$ corresponds to the case considered in the previous section. The hysteresis range is small for small γ , $\Delta V \propto \gamma^{4/3}$; however, it always exists for $T_0 = 0$. The increase of the substrate temperature T_0 decreases the hysteresis range (Fig. 7), which eventually disappears.

Now let us consider briefly the situation in which the resistance for heat flow from the electron gas to the phonon gas in the middle electrode of the SET is larger than the Kapitza resistance, and Eq. (19) is close to reality. The only difference from the previous case is the change from the fourth to fifth power of the temperature. It is obvious that qualitatively all results above remain valid. Moreover, in practice it is difficult to distinguish between these two situations. In Fig. 8 we show $I(V)$ curves for T^4 and T^5 models using the equal value of the dimensionless overheating pa-

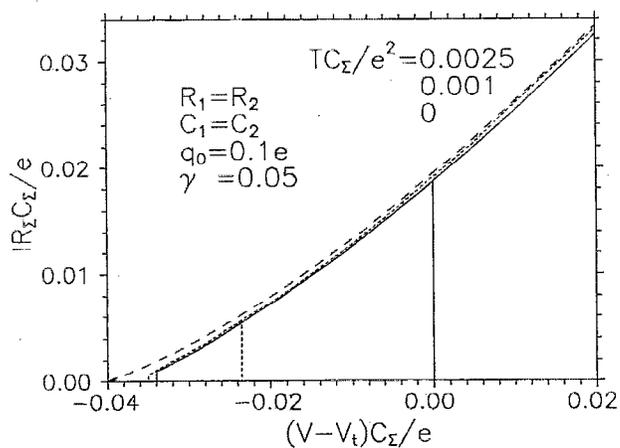


FIG. 7. Hysteretic I - V curves for different temperatures: $T C_{\Sigma} / e^2 = 0$ (solid line); 0.001 (dotted line); 0.0025 (dashed line). The hysteresis disappears at $T C_{\Sigma} / e^2 = 0.0025$.

parameter $\gamma = 0.05$ [for the latter model it should be $\gamma = (C_{\Sigma}^3 / e^8 R_{\Sigma} \Sigma \Omega)^{1/5}$]. Estimating the parameter γ in possible experiments with thin-film SETs, let us choose $C_{\Sigma} = 10^{-16}$ F, $R_{\Sigma} = 10^5 \Omega$, $k_B^5 \Sigma = 2$ nW/K $^5 \mu\text{m}^3$, $\Omega = 0.1 \mu\text{m}^3$. Then $\gamma = 0.04$ and overheating should be considerable for $T_0 \ll \gamma e^2 / C_{\Sigma} \approx 0.7$ K. Analytical expressions in the vicinity of Coulomb blockade can be derived in a way similar to the derivation Eqs. (25)–(31) and show similar form.

At the end of this section let us discuss the possibility of experimental observation of the hysteresis of the I - V curve. The observation is difficult because this effect is very weak. Using the T^4 model for the STM case with parameters $C_{\Sigma} = 10^{-17}$ F, $S = (100 \text{ \AA})^2$, $R_{\Sigma} = 10^7 \Omega$, $k_B^4 \alpha = 100$ W/m 2 K 4 , one obtains from Figs. 6 and 7 that at $T_0 = 0$ the hysteresis has a voltage width of about 0.5 mV and an amplitude of about 20 pA. The hysteresis completely disappears at temperatures higher than about 300 mK (see Fig. 7). However, even when there is no hysteresis, the I - V curve is strongly modified. In particular, one can see that the dashed

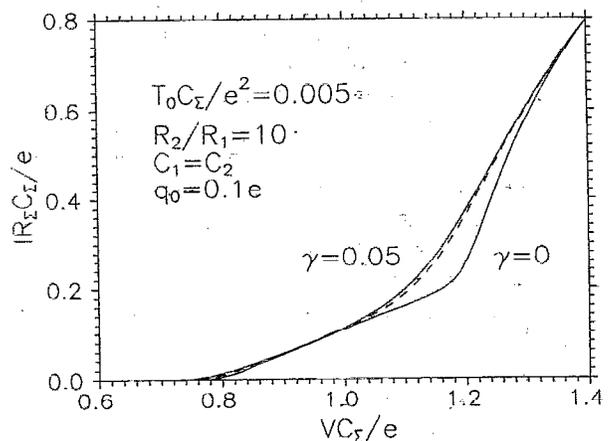


FIG. 8. The I - V curves without overheating ($\gamma = 0$) and for considerable overheating ($\gamma = 0.05$), when the heat drain is described either by Eq. (19) (solid line) or by Eq. (16) (dashed line).

line in Fig. 7 cannot be fitted by the exponential dependence which is the case in the theory without overheating.

Another effect which can destroy the hysteresis is cotunneling processes, not taken into account above. A rough estimate of its contribution to the current in the vicinity of the threshold voltage V_t is 11 $V_t R_Q / R_1 R_2$, where $R_Q = \pi \hbar / 2 e^2$. Since the cotunneling current also heats the middle electrode (see Sec. II), hysteresis can be observable only if the cotunneling current is much smaller than the amplitude of the jump in current due to the hysteresis. This would require resistances at least on the order of $10^7 \Omega$ or more for the case shown in Fig. 7. The use of strongly different resistances $R_1 \gg R_2$ can help to solve this problem. If V_t corresponds to the blockade threshold in the junction having the smaller resistance, then in the vicinity of V_t the current calculated in the orthodox theory is insensitive to the value of the larger resistance. The cotunneling current, however, depends on both resistances, hence, its relative importance can be reduced by the use of considerably different resistances.

VI. CONCLUSION

The effect of overheating is important for all types of single-electron transistors. The difference in electron temperatures among electrodes requires some modification of the orthodox theory. The heat generated in the middle electrode depends on its temperature, and energy balance can be achieved even if the heat drain due to phonon processes is negligible (this effect is important for a SET operating at low bias voltage and at low temperature). The correct consideration of heat drain is difficult, in particular, because of the lack of detailed experimental data in the interesting range of dimensions and temperatures. We have considered two simple models which both show strong dependence (T^4 and T^5) on the temperature. It leads to a considerable change in the temperature in the vicinity of the Coulomb blockade threshold and weak dependence at higher voltages. The interesting feature is the possibility of hysteresis.

Overheating seems to be an important problem in the applications of SETs, especially at temperatures below 0.1 K. A simple way to reduce the temperature of the central electrode is by increasing its area and thickness (this increases the heat flow from electron gas to phonons and reduces the Kapitza resistance). Obviously this will lead to performance improvement at least until the stray capacitance C_s of the central electrode becomes comparable to the capacitances $C_{1,2}$ of the tunnel junctions.

For a rough estimate of the optimal C_s let us optimize the parameter $T_m (C_1 + C_2 + C_s) / e^2$ using Eq. (16) or Eq. (19), and assuming that the heat drain is proportional to C_s (for example, we alter the length of the central electrode). This gives C_s about one-half of $C_{1,2}$ in the case of strong overheating. We believe that the search for the optimal size of the central electrode is an important experimental problem. The shape of the middle electrode should also be subject to optimization. It is clear that thin narrow electrode geometry (which is used often in experiments) is not the best choice. Note that if overheating is considerable ($T_m / T_0 > 2$) then T_m depends weakly on the heat drain. Hence, a consid-

erable change of parameters is needed to achieve a noticeable gain.

Another important way to reduce overheating is the use of bias voltages much less than $e/(C_1 + C_2)$ (tuning the background charge q_0 close to $e/2$). This reduces also the cotunneling processes and can lead to considerable improvement of the charge sensitivity⁹ of the SET.

ACKNOWLEDGMENTS

Fruitful discussions with D. V. Averin, R. Chen, K. K. Likharev, and Yu. V. Nazarov are gratefully acknowledged. We acknowledge the help of M. H. Joergensen and I. V. Vernik connected with our numerical calculations. One of us (S. A. V.) wants to express his gratitude to the Danish Technical University for the hospitality during his stay. The work was partly supported by the Russian Fund for Fundamental Research, Grant No. 93-02-14136 and U.S. AFOSR, Grant No. 91-0445.

¹K. K. Likharev, IBM J. Res. Dev. **32**, 144 (1988).

²D. V. Averin and K. K. Likharev, in *Mesoscopic Phenomena in Solids*, edited by B. L. Altshuler, P. A. Lee, and R. A. Webb (Elsevier, Amsterdam, 1991), p. 173.

³Z. Phys. B **85**, No. 3 (1991) (special issue on single charge tunneling).

⁴I. O. Kulik and R. I. Shekhter, Zh. Exp. Teor. Fiz. **68**, 623 (1975) [Sov. Phys. JETP **41**, 308 (1975)].

⁵K. K. Likharev, IEEE Trans. Magn. **MAG-23**, 1142 (1987).

⁶H. Tamura, S. Hasuo, and Y. Okabe, Jpn. J. Appl. Phys. **26**, Suppl. 3, 1381 (1987).

⁷K. Mullen, E. Ben-Jacob, and S. Ruggiero, Phys. Rev. B **38**, 5150 (1988).

⁸M. Amman, R. Wilkins, E. Ben-Jacob, P. D. Maker, and R. C. Jaklevic, Phys. Rev. B **43**, 1146 (1991).

⁹A. N. Korotkov, D. V. Averin, K. K. Likharev, and S. A. Vasenko, in *Single-Electron Tunneling and Mesoscopic Devices*, edited by H. Koch and H. Lübbig (Springer, Berlin, 1992), p. 45.

¹⁰A. N. Korotkov and Yu. V. Nazarov, Physica B **173**, 217 (1991).

¹¹D. V. Averin and A. A. Odintsov, Phys. Lett. A **140**, 251 (1989).

¹²D. V. Averin and Yu. V. Nazarov, Phys. Rev. Lett. **65**, 2446 (1990).

¹³A. A. Odintsov, G. Falci, and G. Schön, Phys. Rev. B **44**, 13 089 (1991).

¹⁴D. V. Averin and A. N. Korotkov, Zh. Eksp. Teor. Fiz. **97**, 1661 (1990) [Sov. Phys. JETP **70**, 937 (1990)].

¹⁵D. V. Averin, A. N. Korotkov, and K. K. Likharev, Phys. Rev. B **44**, 6199 (1991).

¹⁶P. Lafarge, H. Pothier, E. R. Williams, D. Esteve, C. Urbina, and M. H. Devoret, Z. Phys. B **85**, 327 (1991).

¹⁷R. L. Kautz, G. Zimmerli, and J. M. Martinis, J. Appl. Phys. **73**, 2386 (1993).

¹⁸R. Stratton, J. Phys. Chem. Solids **23**, 1177 (1962).

¹⁹W. A. Little, Can. J. Phys. **37**, 334 (1959).

²⁰R. E. Peterson and A. C. Anderson, J. Low Temp. Phys. **11**, 639 (1973).

²¹V. F. Gantmakher and Y. B. Levinson, *Carrier Scattering in Metals and Semiconductors* (North-Holland, New York, 1987).

²²G. Bergman, Phys. Rep. **107**, 1 (1984).

²³C. Schmidt and E. Umlauf, J. Low Temp. Phys. **22**, 597 (1976).

²⁴H. Haug, K. Weiss, and M. Van Hove, J. Low Temp. Phys. **4**, 263 (1971).

²⁵M. L. Roukes, M. R. Freeman, R. S. Germain, R. C. Richardson, and M. B. Ketchen, Phys. Rev. Lett. **55**, 422 (1985).

²⁶F. C. Wellstood, C. Urbina, and J. Clarke, Appl. Phys. Lett. **54**, 2599 (1989).

²⁷U. Bockelmann, Phys. Rev. B **42**, 8947 (1990).

²⁸F. Kuchar, J. Lutz, and M. Aichinger, Physica Scripta **T39**, 295 (1991).

²⁹Y. Okuyama and N. Tokuda, Phys. Rev. B **40**, 9744 (1989).

³⁰N. Garcia and F. Guinea, Phys. Rev. B **46**, 571 (1992).