Refrigerator based on the Coulomb barrier for single-electron tunneling

Jukka P. Pekola,[1] Jonne V. Koski,[1] and Dmitri V. Averin[2]

[1]*Low Temperature Laboratory (OVLL), Aalto University School of Science, P.O. Box 13500, 00076 Aalto, Finland*
[2]*Department of Physics and Astronomy, Stony Brook University, SUNY, Stony Brook, New York 11794-3800, USA*

We propose a remarkably simple electronic refrigerator based on the Coulomb barrier for single-electron tunneling. A fully normal single-electron transistor is voltage V biased at a gate position such that tunneling through one of the junctions costs an energy of about $k_B T \ll eV, E_C$, where T is the temperature and E_C is the transistor charging energy. The tunneling in the junction with positive energy cost cools both leads attached to it. Immediate practical realizations of such a refrigerator make use of Andreev mirrors which suppress heat current while maintaining full electric contact.

Thermal transport properties of nanocircuits are receiving increased attention [1,2]. Overheating due to dissipative currents is a concern for applications with either dense architecture or when operating in a regime where thermal relaxation becomes weak, for instance at low temperatures. Active cooling below the bath temperature is one of the available strategies against overheating and can be achieved directly by electric means on a chip. The practical realizations employ energy-selective transport either with the help of a superconducting gap [1,3–5] or via a discrete level in a quantum dot [6–8]. Here we present a basic, till now overlooked alternative method based on the Coulomb gap in a single-electron transistor with metallic electrodes. The overall dissipation of a biased normal single-electron transistor is naturally positive, but one can find regimes where one of the junctions cools the lead and the island while the other one is dissipative. This provides an interesting possibility for realizing a Coulomb blockade enabled refrigerator ("SET cooler"), if the charge and energy can be controlled independently, e.g., if the transistor island can be split thermally in two halves by a superconducting inclusion while maintaining its electric unity. Operation of such a cooler is based on the Coulomb gap for electron transport similarly to the superconducting gap and quantum dot coolers, which also use the energy gaps. However, the nature of this gap makes the SET cooler different from them in one important respect. While those coolers can be viewed in some respects as Peltier-effect refrigerators (see, e.g., [9]) in which only one electrode of the tunnel junction is cooling down while the other one is heating, the removed heat in the SET cooler is split equally between the two sides of the cooling junction. Another attractive feature of the SET cooler is the possibility to adjust the gap by gate voltage to optimize the operation at a given temperature. We discuss the performance of the refrigerator in detail and potential ways to realize it in practice. It turns out that the SET cooler is most suitable for very low temperatures, where the standard superconducting gap based electronic coolers become inefficient [1,2].

Figure 1 shows the basic scheme, where a standard single-electron transistor [10,11] is biased at voltage V, and its gate position is $n_g \equiv -C_g V_g/e$, where C_g and V_g are the gate capacitance and voltage, respectively. We analyze the energetics of the single-electron transistor, in particular in the low-temperature regime $k_B T \ll E_C$, where

only two charge states $n = 0$ and $n = 1$ are possible. For optimal operation in this regime, the gate voltage is adjusted to a value where the in-tunneling electron experiences a barrier $\sim k_B T \ll eV$, T is the temperature of the electrodes, and only the out-tunneling electron gains energy $\sim eV$. Under these conditions, the electrodes of the in-tunneling junction are cooled down and those of the other junction heat up. Due to simple symmetries, the roles of the two junctions are interchanged when operating at the gate-voltage position $1 - n_g$ instead of n_g within the range $0 < n_g < 1$.

We write first the equations governing the charge and energy dynamics of the single-electron transistor, but here limiting to equal temperatures in all electrodes. The rates of single-electron tunneling into $(+)$ or out $(-)$ of the island through junction $k = 1,2$ in the charge state n are given by $\Gamma_k^\pm(n) = (e^2 R_T)^{-1} \Delta E_k^\pm(n)/(e^{\beta \Delta E_k^\pm(n)} - 1)$, where R_T is the tunnel resistance of the junctions that is for the moment assumed to be the same for the two junctions, and $\Delta E_k^\pm(n) = \pm(-1)^k eV/2 + 2E_C(n - n_g \pm 1/2)$ are the energy costs for various processes. $E_C = e^2/2C_\Sigma$ is the charging energy, and the common temperature is $T = (k_B \beta)^{-1}$. Here, $C_\Sigma = 2C + C_g$ is the total capacitance of the island, and C is the capacitance of one junction (again assuming a symmetric structure). The occupation probabilities $p(n)$ obey the steady-state equation $[\Gamma_1^+(n-1) + \Gamma_2^+(n-1)]p(n-1) = [\Gamma_1^-(n) + \Gamma_2^-(n)]p(n)$, and are normalized by $\sum_{-\infty}^\infty p(n) = 1$.

The heat currents in tunneling processes are

$$\dot{Q}_k^\pm(n) = \mp \frac{1}{e^2 R_T} \int dE \, E \, f_{L,k}(\pm E + \Delta E_k^\pm(n)) \times [1 - f_I(\pm E)] \tag{1}$$

for the partial cooling power of the island by the tunneling into $(+)$ and out of $(-)$ it. Here, $f_{I/L,k}(E)$ are the energy distributions (typically Fermi functions) of the island I and the leads L,k, respectively. The total heat current out of the island $(=$ cooling power$)$ at each junction is then given by $\dot{Q}_k = \sum_{n=-\infty}^\infty p(n)[\dot{Q}_k^-(n) + \dot{Q}_k^+(n)]$. This is also the cooling power for the corresponding lead of the junction k: the heat extracted or released is the same for both sides of the junction, provided $f_I(E) = f_{L,k}(E)$, in particular for Fermi distributions with the same temperature T. Equation (1) gives then

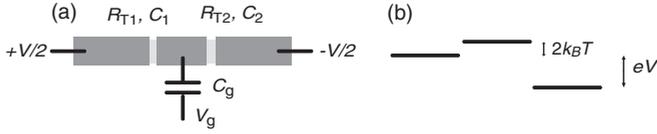$$\dot{Q}_k^\pm(n) = \frac{1}{2e^2 R_T} \frac{[\Delta E_k^\pm(n)]^2}{e^{\beta \Delta E_k^\pm(n)} - 1}. \tag{2}$$

FIG. 1. The single-electron transistor as a cooler. In (a) the basic structure is shown with bias voltage V, tunnel junctions with capacitances C_i and resistances R_{Ti}, and gate at voltage V_g and capacitance C_g. In the text we mostly assume the structure to be symmetric with $R_{Ti} = R_T$ and $C_i = C$ for both junctions. In (b) we demonstrate the biasing for optimum cooler operation in the two-state approximation, where the energy cost to tunnel through the first (cooling) junction is $2k_BT$, and $-(eV + 2k_BT)$ through the second one.

Next we focus on the two-state regime realized at $k_BT \ll E_C$ in the gate interval $0 < n_g < 1$, also assuming that the bias voltage is large enough, $eV \gg k_BT$, for the tunneling to occur only in the "forward" direction. Then we need to consider only two processes, $+$ for the $n = 0 \to 1$ and $-$ for the $n = 1 \to 0$ transition, respectively, with energy costs $\Delta E^{\pm} = -\frac{eV}{2} \pm 2E_C(\frac{1}{2} - n_g)$ and occupations $p(1) = 1 - p(0) = \Gamma^+/(\Gamma^+ + \Gamma^-)$. We do not write the redundant indices when discussing the two-state approximation. Based on Eq. (2), the cooling power of the first junction reaches maximum when the barrier is $\Delta E^+ \simeq 2k_BT$ as drawn in Fig. 1(b). The cooling power of one side of the optimally biased junction is then

$$\dot{Q}_{\text{opt}} \simeq 0.31 \frac{(k_BT)^2}{e^2 R_T}, \tag{3}$$

and the gate position for this maximum cooling is

$$n_g^{\text{opt}} - 1/2 = \mp \left(\frac{k_BT}{E_C} + \frac{1}{4} \frac{eV}{E_C} \right). \tag{4}$$

Also, in the two-state approximation, the total dissipation in the transistor, $P \equiv -2\dot{Q} = -2[p(0)\dot{Q}^+ + p(1)\dot{Q}^-]$, equals IV, independent of the gate position. At an arbitrary position within the given gate interval, the cooling power of each side of junction 1 can be expressed directly in terms of the current $I = e\Gamma^+\Gamma^-/[\Gamma^+ + \Gamma^-]$ through the transistor:

$$\dot{Q}_1 = p(0)\dot{Q}^+ = \frac{1}{2}\frac{I}{e}\Delta E^+. \tag{5}$$

The efficiency of the cooler obtains then a natural value

$$\eta = \frac{\dot{Q}_1}{IV} = \frac{1}{2}\frac{\Delta E^+}{eV} \tag{6}$$

for one side of the cooling junction, and twice this value for the entire junction. At the optimum working point of Eq. (4), we have $\eta_{\text{opt}} = k_BT/eV$.

Pure numerical evaluation of the equations above in the general situation, not limited to two charge states only, is straightforward. The resulting cooling powers, still assuming equal temperature of all the electrodes, are given in Fig. 2 for a realistic set of parameters. The optimum cooling power in the two-state model, Eq. (3), is shown by the dashed line, and it compares favorably with the numerically obtained peak cooling power.

The higher order tunneling processes in principle can have a detrimental effect on the cooling discussed above. Inelastic cotunneling through the transistor [12] creates excitations with
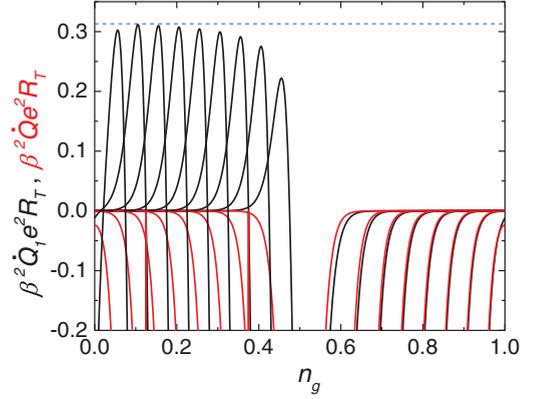

FIG. 2. (Color online) The normalized cooling power of each side of junction 1, $\dot{Q}_1/[(k_BT)^2/(e^2 R_T)]$ (black), and the normalized total cooling power on the island by the two junctions (red), $\dot{Q}/[(k_BT)^2/(e^2 R_T)]$. The latter quantity is naturally always negative with the value $\dot{Q} = -IV/2$. The parameters of the system are $k_BT/E_C = 0.025$, $eV/E_C = 0.1, 0.3, 0.5, 0.7, 0.9, 1.1, 1.3, 1.5, 1.7$ for curves with cooling maxima in the black curves shifting from right to left. Above this value of voltage, the achievable cooling power diminishes quickly. The dashed horizontal line is the analytic prediction of the optimum cooling power, Eq. (3).

energies $\sim eV$ in all junction electrodes, heating them up. We analyze the most relevant regime close to the optimum cooling bias for junction 1, when $\beta E_1 \sim 2$, while $E_2 \sim E_C \gg \beta^{-1}, E_1$, and $eV \gg \beta^{-1}, E_1$, where $E_1 \equiv \Delta E_1^+(0)$ and $E_2 \equiv \Delta E_2^-(0)$. Quantitative description of this regime is complicated by the fact that for such a small energy barrier E_1, sequential "first-order" classical tunneling *over* the barrier cannot be clearly separated from the cotunneling, which is the "second-order" tunneling *through* the barrier (cf. Fig. 1). In general, coexistence of the tunneling events of different order requires taking into account the nonperturbative effect of broadening of the charge states by tunneling [13]. In the situation of the optimum cooling bias, $eV \gg \beta^{-1}, E_1$, the broadening of the relevant charge state E_1 is dominated by tunneling in the second junction. On the other hand, for small E_1, the cotunneling goes predominantly through this intermediate charge state, making it possible to neglect the processes through the other charge state with energy E_2.

Quantitatively, employing the usual tunnel Hamiltonian H_T, we express the average of the cooling power $\dot{Q}_1$ as

$$\langle \dot{Q}_1 \rangle = \langle U^{\dagger}(t)\dot{Q}_1(t)U(t) \rangle,$$
$$U(t) = \mathcal{T} \exp\left\{ \frac{-i}{\hbar} \int^t dt' H_T(t') \right\}. \tag{7}$$

Here the time dependence of all operators is due to the charging energy of the transistor and internal energy of the electrodes, the average $\langle ... \rangle$ is taken over the assumed equilibrium state of the electrodes, $\mathcal{T}$ denotes time ordering, and, in the standard notations,

$$\dot{Q}_1 = \frac{i}{2\hbar} \sum_{k,p} (\epsilon_k - \epsilon_p)[t_{kp}^{(1)} c_k^{\dagger} c_p - \text{H.c.}], \quad H_T = H_1 + H_2,$$

where $H_2 = \sum_{q,l} [t_{q,l}^{(2)} c_q^{\dagger} c_l + \text{H.c.}]$, and a similar expression for the tunneling Hamiltonian H_1 of the first junction.

In the regime described qualitatively above, Eq. (7) can be evaluated expanding the evolution operator $U(t)$ to the lowest power in H_1, but summing the main terms that correspond to broadening of E_1 to all powers in H_2. (In this calculation, we allow the two junction conductances $G_{1,2} = 1/R_{T1,2}$ to be in general different.) This gives, dropping $\langle ... \rangle$ out for simplicity in notation,

$$\dot{Q}_1 = \frac{G_1}{2\pi e^2} \int d\epsilon \, \frac{\epsilon^2}{1 - e^{-\beta\epsilon}} \, \text{Im}\left[\Sigma_{n=0}^{\infty} \frac{\xi^n}{(\epsilon + E_1 + i0)^{n+1}} \right],$$

where

$$\xi = \frac{\hbar G_2}{2\pi e^2} \int d\epsilon' \, \frac{\epsilon'}{1 - e^{-\beta\epsilon'}} \frac{1}{\epsilon + \epsilon' - eV + i0}.$$

The real part of ξ contributes to the tunneling-induced shift of the energy of the intermediate charge state. Incorporating it into the actual energy of this state, $E_1 \to E$, one is left with the imaginary part of ξ,

$$\text{Im}\,\xi = \frac{\hbar G_2}{2e^2} \frac{eV - \epsilon}{1 - e^{-\beta(eV-\epsilon)}} \equiv \gamma(\epsilon); \quad (8)$$

i.e., the level is broadened to the width γ which coincides to only half of the tunneling rate in the second junction at bias $eV - \epsilon$. Taking into account that $\beta eV \gg 1$, one obtains then the following final expression for the cooling power:

$$\dot{Q}_1 = \frac{\hbar G_1 G_2}{4\pi e^4} \int_{-\infty}^{eV} d\epsilon \, \frac{\epsilon^2}{1 - e^{-\beta\epsilon}} \frac{\epsilon - eV}{(\epsilon + E)^2 + \gamma^2(\epsilon)}. \quad (9)$$

In the limit of interest, $eV \gg \gamma, \beta^{-1}, E$, the energy dependence of γ can be neglected, $\gamma = \gamma(\epsilon = 0) = \hbar G_2 V/2e$, and the integral in Eq. (9) can be evaluated in terms of the digamma function $\psi(z)$ as

$$\dot{Q}_1 = \frac{\hbar G_1 G_2}{2\pi e^4} \left[eVE\left(\ln \frac{\beta eV}{2\pi} - 1 \right) - \frac{(eV)^2}{4} \right]$$
$$+ \frac{G_1}{4e^2}\left\{ \frac{2}{\pi}\text{Im}\left[(E - i\gamma)^2 \psi\left(\frac{\beta(E + i\gamma)}{2\pi} \right) \right] \right.$$
$$\left. - E^2 - 2\beta^{-1}E + \gamma^2 \right\}. \quad (10)$$

This result is plotted in Fig. 3. For $\gamma \to 0$, Eq. (10) reproduces the classical result of Eq. (2), $\dot{Q}_1 = (G_1 E^2/2e^2)/(e^{\beta E} - 1)$, which also closely approximates the top numerical curve in Fig. 3. We see, both from Eq. (10) and Fig. 3, that the effect of the higher order tunneling processes on cooling includes direct cotunneling-induced heating [the first line in Eq. (10)] and broadening and suppression of the classical cooling peak by the level width γ (the second line). Direct cotunneling heating is small as long as $\gamma \ll (eV\beta^2)^{-1}$, while the broadening is almost negligible for $\beta\gamma < 0.1$. Elsewhere in this Rapid Communication we assume that these conditions are satisfied and we can use the classical description of cooling.

Next we turn to the practical realization of the cooler. In general, the cooling effect is unnoticeable in a standard single-electron transistor, because the lead electrodes are reservoirs thermalized by large volume and by effective heat conduction near the junction, and, on the other hand, the total power on the island is positive. However, it is quite straightforward to realize a configuration, where the charge and heat currents
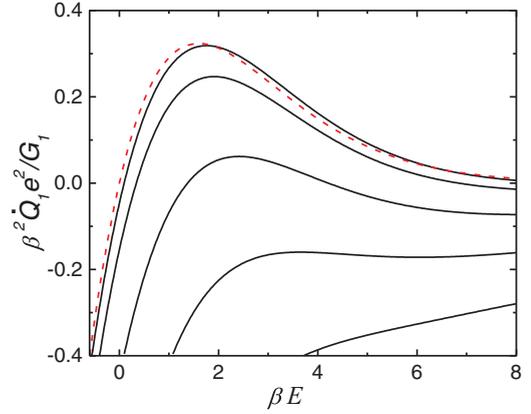

FIG. 3. (Color online) The total power $\dot{Q}_1$ as a function of the charging energy barrier E for tunneling in this junction, for several values of γ, the energy width of the intermediate charge state created by tunneling in the second junction. From top to bottom, $\beta\gamma = 0.03, 0.1, 0.3, 0.6, 1.0$. For all curves, the bias voltage is $\beta eV = 20$. The dashed red line is the ideal classical cooling power. The plot illustrates increasing cotunneling contribution to heating in junction 1 and the simultaneous broadening of the classical cooling peak with increased width γ.

separate effectively. The most obvious way to do this is to replace parts of the normal electrodes by superconductors (forming Andreev mirrors with direct metal-to-metal contacts) that efficiently isolate the cooled areas without influencing the relevant charge transport [14–16]. This can be done by splitting the island into two halves, and by interrupting one or both the leads this way; see the lower inset in Fig. 4. In this configuration it is more practical to cool and monitor the normal section of the lead outside the transistor island. This makes the thermometry, e.g., by tunnel spectroscopy, and other measurements straightforward, because the potential of
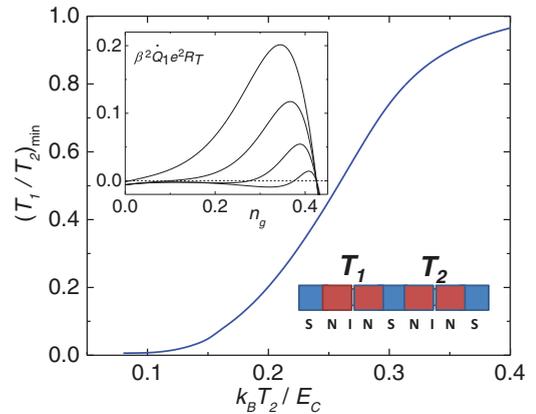

FIG. 4. (Color online) Cooling under the conditions of unequal temperatures. The lower inset shows the sketch of a SET cooler with superconducting (S) mirrors separating normal (N) conductors. The temperatures around junctions 1 and 2 are given by T_1 and T_2, respectively, and they are assumed to be the same on both sides of each junction (I). The upper inset shows the gate voltage dependence of the cooling power for $eV = 0.3E_C$, $k_B T_2/E_C = 0.1$, and $k_B T_1/E_C = 0.1, 0.075, 0.05, 0.025$ from top to bottom. The main frame shows the temperature drop $(T_1/T_2)_{\min}$ for zero cooling power for $eV = 0.3E_C$ as a function of T_2.

the cooled area does not vary then in response to individual tunneling events.

The fundamental limitation of the performance of the SET cooler in terms of the minimum temperature is given by the temperature T_2 of the "hot" junction. The cooling of junction 1 (at temperature T_1) diminishes, as more charge states become available due to tunneling in the higher temperature junction, and eventually there will be power $IV/4$ deposited to all the four electrodes when the Coulomb effects become negligible. Naturally this is not the only limitation on cooling; other mechanisms include heat load from the phonon bath and through the superconducting lead to the cooled area, but the latter contributions can be made small by operating at low temperatures and by proper choice of the geometries of the device. The second inset in Fig. 4 shows as an example a set of cooling powers of junction 1 at various values of $T_1 \ll T_2$, plotted again as a function of n_g at a fixed value of V. Naturally the power gets smaller on reducing T_1 because of the backflow of heat from the hot bath, and since the (cooling) current of the device decreases on decreasing T_1. The main frame of Fig. 4 shows the ultimate achievable temperature reduction $(T_1/T_2)_{\min}$ as a function of T_2, given by the minimum value of T_1 where the cooling power gets positive values at the optimum point. We see that temperature reductions by an order of magnitude seem feasible from this point of view.

Finally we give a few practical remarks. It is favorable to increase E_C as high as is practical in order to keep the device in the SET regime with just two charge states. With the conventional metallic realization of the circuit, values of $E_C/k_B \sim 1-3$ K can be achieved in a single-electron transistor whose island is several μm long. This, in turn, allows for the insertion of the superconducting mirror and sufficient volume near junction 2 on the island to avoid excessive overheating. To make these arguments more concrete, we consider the various heat currents briefly. When a superconducting Al wire is longer than $\sim$1 μm, the adjacent island is better coupled to the phonon bath than through the wire electronically at operating temperatures $\sim$100 mK, as was demonstrated in Ref. [16]. Thus the cooling properties are not much affected by the heat leak through the Al wire. We equate the ideal cooling power (3) and the standard heat load $\Sigma \mathcal{V}(T_p^5 - T^5)$ from the phonons, where T_P is the temperature of the phonon bath, $\Sigma = 2 \times 10^9$ W K^{-5} m^{-3} for copper as the normal metal [1], and $\mathcal{V} = 10^{-21}$ m^3 is the volume of the cooled electrode. With these parameters, it should be possible to reach T_1 as low as 10 mK with $R_T = 1$ MΩ at the bath temperature of $T_p = 50$ mK. On the other hand, the island near junction 2 would warm up to a temperature $T_2 \simeq [P/(\Sigma \mathcal{V}_2)]^{1/5}$, where $P \simeq IV/2$ is the Joule power due to dissipative tunneling in junction 2 and $\mathcal{V}_2$ is the volume of the normal island near this junction. We obtain $T_2 \sim 100$ mK, still compatible with $T_1 = 10$ mK based on Fig. 4. The cotunneling heating is low when eV is chosen properly (at such low temperatures the V dependence of cooling is weak even below $eV = 0.1 E_C$). Finally, based on the data of Ref. [17] we estimate that the Joule heating in the cooler does not cause substantial overheating of the phonons at the site of the cooled junction. The phonon temperature remains at its equilibrium value within better than 0.1 mK at $T_p = 50$ mK.

In summary, we have proposed and analyzed an electronic cooler based on the Coulomb gap in a single-electron transistor. The adjustable (by gate voltage) gap makes the presented cooler attractive for very low temperature operation and possibly as a low-temperature stage in a cascade cooler in combination with a fixed-gap superconducting refrigerator [18].

[1] F. Giazotto, T. T. Heikkilä, A. Luukanen, A. M. Savin, and J. P. Pekola, Rev. Mod. Phys. **78**, 217 (2006).

[2] J. T. Muhonen, M. Meschke, and J. P. Pekola, Rep. Prog. Phys. **75**, 046501 (2012).

[3] M. Nahum, T. M. Eiles, and John M. Martinis, Appl. Phys. Lett. **65**, 3123 (1994).

[4] M. M. Leivo, J. P. Pekola, and D. V. Averin, Appl. Phys. Lett. **68**, 1996 (1996).

[5] P. J. Lowell, G. C. O'Neil, J. M. Underwood, and J. N. Ullom, Appl. Phys. Lett. **102**, 082601 (2013).

[6] H. L. Edwards, Q. Niu, and A. L. de Lozanne, Appl. Phys. Lett. **63**, 1815 (1993).

[7] J. R. Prance, C. G. Smith, J. P. Griffiths, S. J. Chorley, D. Anderson, G. A. C. Jones, I. Farrer, and D. A. Ritchie, Phys. Rev. Lett. **102**, 146602 (2009).

[8] S. Gasparinetti, F. Deon, G. Biasiol, L. Sorba, F. Beltram, and F. Giazotto, Phys. Rev. B **83**, 201306 (2011).

[9] D. M. Rowe (ed.), *Thermoelectrics Handbook: Macro to Nano* (CRC Press, Taylor & Francis, Boca Raton, 2006).

[10] D. V. Averin and K. K. Likharev, J. Low Temp. Phys. **62**, 345 (1986).

[11] *Single Charge Tunneling*, NATO Advanced Study Institute, Ser. B, Vol. 294, edited by H. Grabert and M. H. Devoret (Plenum Press, New York, 1992).

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

[13] D. V. Averin, Physica B **194–196**, 979 (1994); H. Schoeller and G. Schön, Phys. Rev. B **50**, 18436 (1994).

[14] M. Nahum, P. L. Richards, and C. A. Mears, IEEE Trans. Appl. Supercond. **3**, 2124 (1993).

[15] D. Golubev and L. Kuzmin, J. Appl. Phys. **89**, 6464 (2001).

[16] J. T. Peltonen, P. Virtanen, M. Meschke, J. V. Koski, T. T. Heikkilä, and J. P. Pekola, Phys. Rev. Lett. **105**, 097004 (2010).

[17] A. M. Savin, J. P. Pekola, D. V. Averin, and V. K. Semenov, J. Appl. Phys. **99**, 084501 (2006).

[18] O. Quaranta, P. Spathis, F. Beltram, and F. Giazotto, Appl. Phys. Lett. **98**, 032501 (2011).