Pekola, Jukka P.; Koski, Jonne V.; Averin, Dmitri V.

Refrigerator based on the Coulomb barrier for single-electron tunneling

Published in:
Physical Review B

DOI:
10.1103/PhysRevB.89.081309

Published: 24/02/2014

Please cite the original version:
Refrigerator based on the Coulomb barrier for single-electron tunneling

Jukka P. Pekola,1 Jonne V. Koski,1 and Dmitri V. Averin2

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

(Received 5 October 2013; revised manuscript received 10 December 2013; published 24 February 2014)

We propose a remarkably simple electronic refrigerator based on the Coulomb barrier for single-electron tunneling. A fully normal single-electron transistor is voltage 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.

DOI: 10.1103/PhysRevB.89.081309

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 the single-electron transistor [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 while the other one is heating. 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 \((+\rangle\) or out \((\langle-\rangle\) of the island through junction \( k = 1, 2 \) in the charge state \( n \) are given by \( \Gamma^+_{k}(n) = (e^2 R_T)^{-1} \Delta E^+_k(n)/(e^2 \Delta E^+_k(n) - 1) \), 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 \( \Delta E^+_k(n) = \pm(\pm 1)eV/2 \pm 2E_C(n - n_g \pm 1/2) \) are the energy costs for various processes. \( E_C = e^2/2C_g \) 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 \( C = \Sigma_{k=1}^{\infty} \) is the capacitance of one junction (again assuming a symmetric structure). The occupation probabilities \( p(n) \) obey the steady-state equation \( \Gamma^+_{k}(n+1) \Gamma^-_{k}(n-1) p(n+1) = \Gamma^-_{k}(n) \Gamma^+_{k}(n) p(n) \), and are normalized by \( \sum_{-\infty}^{\infty} p(n) = 1 \).

The heat currents in tunneling processes are

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

(1)

for the partial cooling power of the island by the tunneling into \( (+\rangle\) or out \((\langle-\rangle\) of the cooling junction. Here, \( f_{L,k}(E) \) 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

\[
\dot{Q}^+_k(n) = \frac{1}{2e^2 R_T} \frac{\left[\Delta E^+_k(n)^2\right]}{e^{\Delta E^+_k(n)} - 1}.
\]

(2)
The one side of the optimally biased junction is then
equal temperature of all the electrodes, are given in Fig. 2
the cooling power diminishes quickly. The dashed horizontal line is the analytic
in the two-state regime realized at $k_B T \ll E_C$ in the gate interval $0 < n_g < 1$, also assuming that the bias
and it compares favorably with the numerically obtained peak
the single-electron transistor as a cooler. In (a) the
capacitances $C_1$ and resistances $R_{1,2}$, and gate at voltage $V_g$ and capacitance $C_g$. In the text we mostly assume the structure to
be symmetric with $R_{1,2} = R_T$ and $C_1 = C$ 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 $2k_B T$, and $-(eV + 2k_B T)$ through the
second one.

Next we focus on the two-state regime realized at $k_B T \ll E_C$ in the gate interval $0 < n_g < 1$, also assuming that the bias
voltage is large enough, $eV \gg k_B T$, for the tunneling to occur
only in the “forward” direction. Then we need to consider only
two processes, $+$ for the $n = 0 \rightarrow 1$ and $-$ for the $n = 1 \rightarrow 0$ transition, respectively, with energy costs $\Delta E^\pm = -\frac{eV}{2} \pm 2E_C\left(\frac{1}{2} - n_g\right)$ and occupations $p(1) = 1 - p(0) = \Gamma^+/\left(\Gamma^+ + \Gamma^-ight)$. 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
$\Delta E^+ \simeq 2k_B T$ as drawn in Fig. 1(b). The cooling power of
one side of the optimally biased junction is then

$$\dot{Q}_1 \simeq 0.31 \frac{(k_B T)^2}{e^2 R_T}, \quad (3)$$

and the gate position for this maximum cooling is

$$n_g^\text{opt} = \frac{1}{2} = \frac{k_B T}{E_C} + \frac{1}{4} \frac{eV}{E_C}. \quad (4)$$

Also, in the two-state approximation, the total dissipation in the transistor, $P \equiv -2\dot{Q} = -2p[\hat{Q}^+ p(1)\hat{Q}^-]$, 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:

$$\dot{Q}_1 = p(0) \hat{Q}^+ - \frac{1}{2} I \frac{\Gamma^-}{\Gamma^+}. \quad (5)$$

The efficiency of the cooler obtains then a natural value

$$\eta = \frac{\dot{Q}_1}{IV} = \frac{1}{2} \frac{\Delta E^+}{eV}, \quad (6)$$

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
$n_g^\text{opt} = k_B T/eV$.

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
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 $E_2 \gg E_C \gg \beta^{-1} E_1$, 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 $\text{over the barrier cannot be clearly separated from the cotunneling, which is the “second-order” tunneling through the barrier (cf. Fig. 1).}$

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 $\langle \dot{Q}_1 \rangle$ as

$$\langle \dot{Q}_1 \rangle = \left \langle U(t) \dot{Q}_1(t) U(t)^\dagger \right \rangle \quad (7)$$

$$U(t) = T \exp \left \{-i \hbar \int dt' H_T(t') \right \}.$$ 

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

$$\dot{Q}_1 = \frac{i}{\hbar} \sum_{\epsilon, l} \langle \hat{c}^\dagger_l \epsilon \hat{c}_l \rangle \left [ \hat{c}^\dagger_l \epsilon \hat{c}_l - H.c. \right ], \quad H_T = H_1 + H_2,$$

where $H_2 = \sum_{q, l} \epsilon_q \hat{c}^\dagger_q \hat{c}_l + H.c.$, 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 $G_{1,2} = 1/R_{T_{1,2}}$ to be in general different.) This gives, dropping (...) out for simplicity in notation,

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

where

$$\xi = \frac{hG_2}{2\pi e^2} \int d\epsilon' \frac{1}{1 - e^{-\beta (\epsilon + E_1 + i0^\pm)}}.$$

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

$$\text{Im} \xi = \frac{hG_2}{2\pi e^2} \frac{eV - \epsilon}{1 - e^{-\beta(eV - \epsilon)}} \equiv \gamma(\epsilon);$$

i.e., the level is broadened to the width $\gamma$ 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:

$$\dot{Q}_1 = \frac{hG_1G_2}{2\pi e^4} \int_{-\infty}^{eV} d\epsilon \, \frac{\epsilon^2}{1 - e^{-\beta(\epsilon + E_1 + i0)}} = \frac{\epsilon - eV}{\beta(\epsilon + E_1 + i0)}.$$

In the limit of interest, $eV \gg \gamma, \beta^{-1}, E,$ the energy dependence of $\gamma$ can be neglected, $\gamma \approx \gamma(\epsilon = 0) = hG_2/2e,$ and the integral in Eq. (9) can be evaluated in terms of the digamma function $\psi(z)$ as

$$\dot{Q}_1 = \frac{hG_1G_2}{2\pi e^4} \left[ eVE \left( \ln \frac{\beta eV}{2\pi} - 1 \right) - \frac{(eV)^2}{4} \right] + \frac{G_1}{4e^2} \left[ \frac{1}{2} \ln \left( E - i\gamma \right)^2 \psi \left( \frac{\beta (E + i\gamma)}{2\pi} \right) \right] - E^2 - 2\beta^{-1}E + \gamma^2.$$

This result is plotted in Fig. 3. For $\gamma \rightarrow 0,$ Eq. (10) reproduces the classical result of Eq. (2), $\dot{Q}_1 = (G_1E^2/2e^2)/(e^{\beta E} - 1),$ 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 $\gamma$ (the second line). Direct cotunneling heating is small as long as $\gamma \ll (e\beta)^{-1},$ 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...
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 $n_q$ 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 $(T_{1}/T_2)_{\text{min}}$ 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 $E_C/k_B \sim 1 - 3 \text{ K}$ can be achieved in a single-electron transistor whose island is several $\mu$m 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 $\sim 1 \mu$m, 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 $\Sigma V(T_p^5 - T^5)$ from the phonons, where $T_p$ is the temperature of the phonon bath, $\Sigma = 2 \times 10^9 \text{ W K}^{-5} \text{ m}^{-3}$ for copper as the normal metal [1], and $V = 10^{-21} \text{ m}^3$ is the volume of the cooled electrode. With these parameters, it should be possible to reach $T_1$ as low as 10 mK with $R_T = 1 \Omega$ at the bath temperature of $T_p = 50$ mK.

On the other hand, the island near junction 2 would warm up to a temperature $T_2 \simeq [P/(\Sigma V)]^{1/5}$, where $P \simeq IV/2$ is the Joule power due to dissipative tunneling in junction 2 and $V$ 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 $eV = 0.1E_C$). 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].

The work has been supported partially by the Academy of Finland through its LTQ CoE grant (Project No. 250280) and the European Union FP7 project INFERNOS (Grant Agreement No. 308850).