Limitations in Cooling Electrons using Normal-Metal-Superconductor Tunnel Junctions

J. P. Pekola, T. T. Heikkilä, A. M. Savin, and J. T. Flyktman
Low Temperature Laboratory, Helsinki University of Technology, P.O. Box 3500, 02015 HUT, Finland

F. Giazotto
NEST-INFM & Scuola Normale Superiore, I-56126 Pisa, Italy

F. W. J. Hekking
Laboratoire de Physique et Modélisation des Milieux Condensés, CNRS & Université Joseph Fourier, BP 166, 38042 Grenoble CEDEX 9, France

(Received 5 September 2003; published 6 February 2004)

We demonstrate both theoretically and experimentally two limiting factors in cooling electrons using biased tunnel junctions to extract heat from a normal metal into a superconductor. First, when the injection rate of electrons exceeds the internal relaxation rate in the metal to be cooled, the electrons do not obey the Fermi-Dirac distribution, and the concept of temperature cannot be applied as such. Second, at low bath temperatures, states within the gap induce anomalous heating and yield a theoretical limit of the achievable minimum temperature.

DOI: 10.1103/PhysRevLett.92.056804 PACS numbers: 73.50.Lw, 05.70.Ln, 72.15.Lh, 74.50.+r

Refrigerators are generally characterized by their cooling power, coefficient of performance, and operating temperature under various working conditions. To assign a temperature to a system, one needs to assume that the energy relaxation within the system is faster than any rate associated with the heat flux between the system in concern and its surroundings. If this condition fails, the energy distribution of the particles of which the system is formed is nonthermal, and applying the concept of temperature is, strictly speaking, inappropriate. Such a limit can be achieved in submicron-size coolers at low temperatures. The structure we study is a symmetric configuration of a NIS refrigerator [1], formed by a series connection of two superconductor (S)-insulator (I)-normal metal (N) tunnel junctions sharing the N island to be cooled in between them (SINIS) [2]. We demonstrate two striking phenomena occurring in these electron microcoolers at low temperatures: evidence of nonthermal energy distribution of the cooled electrons and reentrant behavior with anomalous heating at low-bias voltages. The cooler performance is typically limited by coupling of the electrons to the underlying lattice (electron-phonon, e-p coupling). This has, however, strong dependence on temperature T: relaxation rate $\tau_{\mathrm{e\text{-}p}}^{-1}$ slows down on lowering T typically as $\tau_{\mathrm{e\text{-}p}}^{-1} \propto T^3$ [3]. Consequently, at low enough temperature, characteristically around 100 mK, the behavior of the cooler can be described as if the lattice would not exist at all. The interplay of the rates for e-p, electron-electron (e-e), and the injection through the junctions determines the distribution in the normal metal. If e-p or e-e rates are fast, the system assumes Fermi-Dirac energy (E) distribution $f_0(E, T_e)$:

$$f_0(E, T_e) = \frac{1}{1 + \exp(E/k_{\mathrm{B}}T_e)}. \quad (1)$$

In the limit of very strong e-p relaxation, T_e equals the temperature of the lattice. We call this *equilibrium*. On the other hand, if e-p relaxation is slow, and e-e relaxation is fast, T_e is, in general, different from the lattice temperature (*quasiequilibrium*). Finally, the slow e-e relaxation as compared to the injection rates implies that the electrons assume a *nonequilibrium* energy distribution $f(E)$, which cannot be written as Eq. (1) [4,5], and it is not possible to assign a true temperature to them.

Our discussion is motivated by a puzzling experimental observation in many coolers at temperatures below or around 100 mK [6]. Figure 1 shows in (a) a typical SINIS cooler. The device has been fabricated by standard electron beam lithography. The central part forms the N island of copper (purity nominally 99.9999%). The

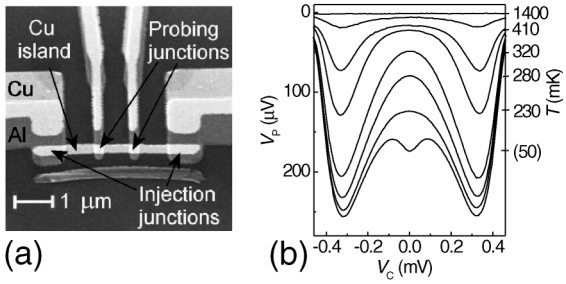

FIG. 1. Scanning electron micrograph of a typical cooler sample in (a), and cooling data in (b), where voltage V_{P} across the probe junctions in a constant current bias (28 pA) is shown against voltage V_{C} across the two injection junctions. Cryostat temperature, corresponding to the electron temperature on the N island at $V_{\mathrm{C}} = 0$ is indicated on the right vertical axis. Below 100 mK this correspondence is uncertain, because of the lack of calibration and several competing effects to be discussed in the text.

injecting tunnel junctions, with normal-state resistances R_T (100 Ω–2 kΩ), contact the island symmetrically into the two superconducting aluminum reservoirs at the two ends of the N island. The overlapping extra copper shadows outside the SINIS structure provide better thermalization of the Al reservoirs. Two additional NIS probe junctions, with normal-state resistances $\gg 1$ kΩ in the center are used to measure the temperature of the N island or to probe the distribution through the differential conductance of the nominally symmetric series connection. $R_T \simeq 200$ Ω and thickness of the copper film $\simeq 30$ nm are the essential parameters of the sample (S1) whose data are presented in this paper. The data in Fig. 1(b) were taken by applying a constant current I_0 through the probe, which, due to the thermal rounding of the current-voltage characteristics, provides, by detecting voltage V_P across it, a measure of the electron temperature on the N island [7]. We calibrated this dependence by varying the bath temperature of the cryostat. Typically one applies I_0 such that the voltage remains within the gap region of the superconductor, $V_P < 2\Delta/e$, whereby excess heating or cooling by the probes is avoided. Here Δ is the energy gap of the superconductor and e is the electron charge ($\Delta/e \simeq 0.2$ mV). Figure 1(b) shows V_P against injection voltage V_C; the several curves represent different bath temperatures. The approximate temperature calibration is given on the right vertical axis. At all but the lowest bath temperature of about 50 mK, the curves show the expected refrigeration behavior symmetrically around $V_C = 0$, with optimum cooling at about $V_C \simeq \pm 2\Delta/e$ [2]. The lowest-temperature curve demonstrates, however, a feature which appears as heating at low values of V_C and reentrant cooling again close to $2\Delta/e$. This behavior is common with many similar samples, and it appears only below 200 mK. Moreover, the measured conductance curves of the probe junctions [Fig. 4 (below)] indicate that at low temperatures the actual temperature we assign in a V_P measurement depends on the choice of I_0.

To understand the observed behavior, we consider the properties of the kinetic equation describing the dominant processes in our system. We assume that the two reservoirs are identical and the quasiparticles have a thermal distribution of Eq. (1), with $T_e \equiv T_S$ on them. In steady state $f(E)$ on the N island is then determined by [8]

$$\frac{\delta}{e^2 R_T}\{n(E_R)[f_0(E_R, T_S) - f(E)] + n(E_L)[f_0(E_L, T_S) - f(E)]\} = I_{coll}[f; E]. \quad (2)$$

Here $I_{coll}[f; E]$ is the collision integral discussed below, δ is the level spacing on the island, $E_{L,R} = E \pm eV_C/2$ are energies on the left (L) and right (R) reservoirs, and

$$n(E) = \left| \mathrm{Re}\left(\frac{E + i\Gamma}{\sqrt{(E + i\Gamma)^2 - \Delta^2}} \right) \right| \quad (3)$$

is the broadened BCS density of states (DOS) of the superconductor. Γ smears the DOS singularity at $E = \pm\Delta$ and allows for states within the gap, e.g., due to inelastic electron scattering in the superconductor [9] or by inverse proximity effect due to nearby normal metals. A more phenomenological choice is to add a nonzero constant Γ/Δ to the ideal singular DOS ($\Gamma \equiv 0$). This choice would not essentially affect our conclusions.

It is illustrative to investigate first the case where relaxation (both e-e and e-p) is very weak, i.e., when $I_{coll}[f; E] \equiv 0$ in Eq. (2). Then we obtain an explicit expression for $f(E)$ as

$$f(E) = \frac{n(E_R) f_0(E_R, T_S) + n(E_L) f_0(E_L, T_S)}{n(E_R) + n(E_L)}. \quad (4)$$

The solution of Eq. (4) is plotted in Fig. 2(a) for five different values of $v_C \equiv eV_C/\Delta$, assuming $\eta \equiv \Gamma/\Delta = 1 \times 10^{-4}$ [10] and $T_S = 0.1 T_C$. The BCS relation $\Delta \simeq 1.764 k_B T_C$ has been assumed for the critical temperature T_C. This solution exhibits some nontrivial features. At low values of the bias voltage up to $v_C \sim 1$, the distribution first broadens whereafter it starts to get narrower again and, at $v_C = 2.0$, it becomes very narrow, effectively demonstrating strong cooling. In each case, except at $v_C = 0$ (equilibrium), the distribution is not thermal. At $v_C > 2.0$, the distribution would become even more unusual [11], but we do not consider this regime in detail for reasons to be explained below. Yet, at those voltages the distribution effectively broadens again, suggesting already the nonmonotonic behavior, heating-cooling-heating upon increasing v_C. The distribution is tested by measuring the differential conductance dI/dV_P of the probe junctions [Fig. 1(a)]. The relation between

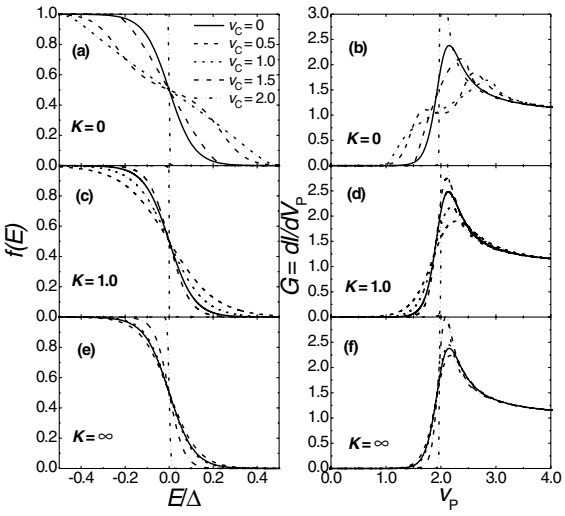

FIG. 2. Energy distribution $f(E)$ against E/Δ in (a), (c), and (e), and differential conductance G against probe voltage v_P in (b), (d), and (f) at three different values of e-e collision parameter K, which present very slow, intermediate, and very fast e-e relaxation from top to bottom, respectively. Parameter values $\eta = 1 \times 10^{-4}$ and $T_S/T_C = 0.1$ have been assumed.

dI/dV_P, V_P, and $f(E)$ is given by

$$\frac{dI}{dV_P} = -\int_{-\infty}^{\infty} n(E - eV_P/2)\frac{df(E)}{dE}dE. \quad (5)$$

Here dI/dV_P has been scaled by the normal-state conductance of the series connection of the probe junctions. $G \equiv (dI)/(dV_P)$ has been plotted in Fig. 2(b) for the collisionless distributions of Fig. 2(a) against $v_P \equiv eV_P/\Delta$. These curves exhibit some similarities to those of biased diffusive metal wires [4], especially at low values of v_C, where quasiconstant DOS at low energies $[n(E) \simeq \Gamma/\Delta]$ mimics resistive normal-metal wire. Another important limit is quasiequilibrium with electrons perfectly decoupled from the phonon bath. The latter condition is well justified at low T_S by a standard estimate of the minimum temperature T_{min} determined by the balance between cooling power and e-p coupling only [12]. In quasiequilibrium fast e-e relaxation forces the distribution into a thermal one [Eq. (1)]. Any temperature satisfies this and we need to determine T by setting the net heat flux from the island, $P(T, T_S)$, equal to zero. In our case, we can write [2]

$$P(T, T_S) = \frac{2}{e^2 R_T}\int_{-\infty}^{\infty} n(E)[f_0(E_R, T) - f_0(E, T_S)]E_R\, dE = 0, \quad (6)$$

where factor 2 on the right-hand side takes into account the flux through the two identical junctions L and R. The solutions of Eq. (6) allow us to plot the distributions and differential conductance at different values of v_C as in Figs. 2(e) and 2(f). The reentrant heating-cooling-heating behavior survives but less pronounced than in the collisionless case. It transforms into more conventional cooling-heating behavior above $T_S = T^*$, given by

$$(\Delta/k_B T^*)^{3/2}\exp(-\Delta/k_B T^*) \simeq \eta/\sqrt{2\pi}. \quad (7)$$

Thus, T^* depends approximately logarithmically on η and assumes a value $T^*/T_C \simeq 0.125$ when $\eta = 1 \times 10^{-4}$. Equation (7) is obtained by equating at low bias the quadratic in v_C heating and cooling terms arising from the states inside and outside the gap, respectively. For an intermediate strength of e-e interaction, the distribution functions were obtained by Eq. (2) with the e-e collision integral given in Ref. [4]. In the case of a diffusive normal-metal island whose transverse dimensions are smaller than the coherence length $\xi_0 = \sqrt{\hbar D/\Delta}$, one finds $I_{coll} = \kappa\sqrt{\Delta}\tilde{I}_{coll}$, where

$$\tilde{I}_{coll} = \int \frac{d\omega d\epsilon}{\omega^{3/2}}[\, f(E)f(\epsilon\Delta)(1 - f(E - \omega\Delta)) \times [1 - f((\epsilon + \omega)\Delta)] - f(E - \omega\Delta) \times f((\epsilon + \omega)\Delta)(1 - f(E))(1 - f(\epsilon\Delta))] \quad (8)$$

is the dimensionless collision integral, $\kappa = (\sqrt{2}L\delta)/(\pi\sqrt{D}\hbar^{3/2})$, and D and L are the diffusion coefficient and the length of the island. Dividing Eq. (2) by $\delta/(e^2 R_T)$, we obtain a dimensionless equation where the strength of e-e scattering is governed by $K \equiv 2\sqrt{2}[(R_T)/(R_K)]\sqrt{\Delta/E_T}$ [13]. Here $R_K \equiv h/e^2$ and $E_T \equiv \hbar D/L^2$ are the resistance quantum and the Thouless energy of the island, respectively.

Although temperature is not a valid concept in nonequilibrium, we can define an effective temperature T_{eff}. A natural choice is to require that T_{eff} satisfies the standard relation between the temperature and the thermal energy density of electrons in Sommerfeld expansion, which yields

$$k_B T_{eff} = \frac{\sqrt{6}}{\pi}\sqrt{\int_{-\infty}^{\infty}[f(E) - 1 + \theta(E)]E\, dE}. \quad (9)$$

Here $\theta(E)$ is the Heaviside step function. This T_{eff} coincides with the true temperature in (quasi)equilibrium, and it is not affected by the strength of e-e relaxation as such. Yet in a biased SINIS, T_{eff} depends on the strength of e-e relaxation, because of the heat exchange with reservoirs with nonconstant DOS. Figure 3 shows T_{eff} as a function of v_C at $T_S/T_C = 0.1$ and $\eta = 1 \times 10^{-4}$ for different rates K. In the collisionless limit, the rise of T_{eff} is largest (almost threefold) and the maximum is reached at $v_C \simeq 0.8$. On increasing the collision rate, the maximum T_{eff} gets lower and it is reached at a lower value of v_C and ultimately, in quasiequilibrium ($K = \infty$), the maximum is reached at $v_C \simeq 0.4$ and its value is about $0.12T_C$. The minimum temperature in quasiequilibrium

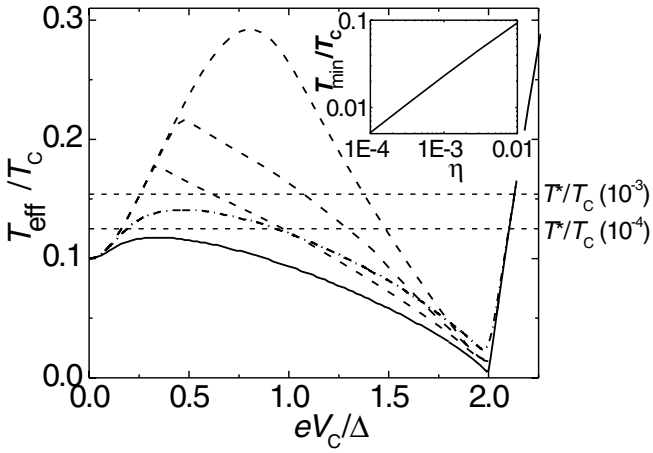

FIG. 3. Effective electron temperature in the cooler against the injection voltage. The dashed lines correspond to $K = 0$ (extreme nonequilibrium), $K = 0.1$, and $K = 1.0$, from top to bottom, and the solid line to $K = \infty$ (quasiequilibrium), all with $T_S/T_C = 0.1$ and $\eta = 1 \times 10^{-4}$. The dash-dotted line is the result for quasiequilibrium but for $\eta = 1 \times 10^{-3}$. The horizontal dashed lines indicate the crossover temperature T^* of reentrant behavior in quasiequilibrium for $\eta = 1 \times 10^{-4}$ and 1×10^{-3}. The inset shows the dependence of the ultimate minimum temperature of the cooler against η.

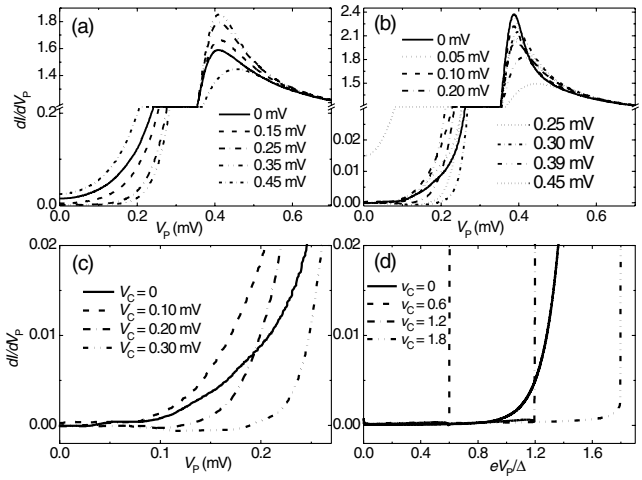

FIG. 4. Measured differential conductance of S1 at $T_S = 340$ mK in (a), and at $T_S \simeq 100$ mK in (b). Zoom-up of a few data sets of (b) are shown in (c), and the corresponding theoretical lines with $K = 0.05$ in (d).

at $\upsilon_C \simeq 2$ is given approximately by $T_{min}/T_C \simeq 2.5\eta^{2/3}$ (inset of Fig. 3). Although the experimental working temperatures well below 100 mK imply that $\eta < 0.01$, we cannot make a direct comparison to this theoretical result in the present device. For this, we would need a thermometer calibration down to the lowest temperatures, the electron system should be forced to quasiequilibrium (magnetic impurities would help), and the S reservoirs should be well thermalized.

To assess the degree of nonequilibrium, we measured dI/dV_P at various values of V_C, which allows for a semiquantitative comparison between the experiment and theory [see Eq. (5)]. In the data, especially below about 200 mK, we concentrate on the values of both V_C and V_P below the gap ($<2\Delta/e$) because of the excessive heating of the reservoirs when current increases abruptly at the gap edge. The data at $T_S = 340$ mK in Fig. 4(a) exhibit cooling behavior in quasiequilibrium: The conductance curves measured get first sharper monotonically on increasing V_C from 0 to 0.35 mV, whereas data at $V_C = 0.45$ mV are more smeared, i.e., hotter than any other curve. All data sets conform in shape without crossing, demonstrating near-to-quasiequilibrium behavior. The data in Fig. 4(b), taken at the base temperature of the cryostat of about 50 mK (best fit to dI/dV_P would yield $T_S \simeq 100$ mK), show, on the contrary, that the energy distribution in this case deviates from the thermal one when applying bias V_C. Data at $dI/dV_P \leq 0.03$ first indicate that the low-bias conductance becomes larger when increasing V_C from 0 up to 0.1 mV (unlike at 340 mK), whereafter the curves start to get sharper, but they heavily cross each other in this regime. At 0.45 mV, the data present significant heating again. The mere fact that the data sets corresponding to different values of V_C crisscross in the regime below 0.4 mV is a demonstration of nonequilibrium. A few data sets from (b) are magnified in (c), and the corresponding theoretical results, assuming $K = 0.05$, have been shown in (d). The resemblance is obvious, although the theoretical lines show abrupt rise from $dI/dV_P = 0$ at $V_P = V_C$ due to the influence of the gap edge. This feature is smeared in experiment most likely because of noise and finite excitation level (25 μV p-p) in the measurement.

In summary, we have shown that slow electron-electron relaxation restricts the use of the concept of temperature in electron coolers at low temperatures, and that the nonzero DOS within the gap of the superconducting reservoirs gives rise to anomalous heating and determines the ultimate minimum temperature that can be achieved.

We thank Academy of Finland (J. P. P.) and Institut Universitaire de France (F. W. J. H.) for financial support, and H. Courtois, L. I. Glazman, J. M. Kivioja, L. S. Kuzmin, J. E. Mooij, M. A. Paalanen, and J. N. Ullom for discussions.

[1] M. Nahum, T. M. Eiles, and J. M. Martinis, Appl. Phys. Lett. **65**, 3123 (1994).
[2] M. M. Leivo, J. P. Pekola, and D. V. Averin, Appl. Phys. Lett. **68**, 1996 (1996).
[3] M. L. Roukes, M. R. Freeman, R. S. Germain, R. C. Richardson, and M. B. Ketchen, Phys. Rev. Lett. **55**, 422 (1985).
[4] H. Pothier, S. Guéron, N. O. Birge, D. Esteve, and M. H. Devoret, Z. Phys. B **104**, 178 (1997).
[5] J. J. A. Baselmans, A F. Morpurgo, B. J. van Wees, and T M. Klapwijk, Nature (London) **397**, 43 (1999).
[6] See, e.g., P. A. Fisher, J. N. Ullom, and M. Nahum, Appl. Phys. Lett. **74**, 2705 (1999).
[7] J. M. Rowell and D. C. Tsui, Phys. Rev. B **14**, 2456 (1976).
[8] We assume resistive enough tunnel barriers, such that Andreev reflection can be neglected. We expect significant modifications of the results presented here in the case of highly transparent interfaces.
[9] R. C. Dynes, J. P. Garno, G. B. Hertel, and T. P. Orlando, Phys. Rev. Lett. **53**, 2437 (1984).
[10] The precise value of η is not important, as soon as it provides the dominant coupling of the island to the environment as is the case at low temperature. We use $\eta = 1 \times 10^{-4}$ based on the strength of subgap heating [Fig. 1(b)], on the extrapolation of data in Ref. [9], and on the ratio ($\simeq \eta$) of the conductances at low and high bias, respectively, of the measured junctions at low temperature.
[11] W. M. van Huffelen, T. M. Klapwijk, D. R. Heslinga, M. J. de Boer, and N. van der Post, Phys. Rev. B **47**, 5170 (1993).
[12] As an example, we obtain $T_{min} \simeq 2$ mK with $T_S/T_C = 0.1$ in S1, a value well below any temperature measured here.
[13] Δ appears in K because we chose it as a suitable energy scale for writing Eq. (8) and to describe the whole process including injection: The true collision integral I_{coll} depends on N island properties only (not on Δ).