Cryogenic cooling using tunneling structures with sharp energy features

H. L. Edwards,* Q. Niu, G. A. Georgakis, and A. L. de Lozanne
Department of Physics, The University of Texas, Austin, Texas 78712
(Received 27 October 1994; revised manuscript received 18 April 1995)

Thermoelectric cooling, based upon the extraction of hot electrons and holes from a metallic electron gas, holds unrealized potential for refrigeration at cryogenic temperatures. We discuss the performance of two such electronic refrigerators: the quantum-dot refrigerator (QDR) and the normal-insulator-superconductor (NIS) refrigerator. We obtain the QDR base temperature using a numerical simulation and verify the validity of certain simplifying assumptions which allow refrigerating performance to be summarized on a diagram of ambient temperature versus electronic temperature. In this way, we find that the best refrigeration is obtained with the electronic distribution far from the equilibrium Fermi-Dirac function and the temperature reduction achieved is limited by the rate at which phonons are absorbed. We predict that, with sufficient thermal isolation, electronic devices could be cooled to a small fraction of the ambient temperature using these solid-state refrigerators. The NIS refrigerator should be capable of cooling thin-film devices from above 300 mK to below 100 mK; the QDR will cool macroscopic metallic samples in the μK or nK range. We also discuss topics related to thermoelectric refrigeration including other cryogenic thermoelectric cooling schemes, the validity of the linear-response theory of thermoelectric effects, the refrigerating efficiency of an optimized thermoelectric refrigerator, and the overall cooling power of thermoelectric refrigeration.

I. INTRODUCTION

The electronic heat capacity exceeds the phonon heat capacity in metallic solids at temperatures below about 1 K.[1] This occurs because the phonon heat capacity is proportional to the third power of the temperature, whereas the electronic heat capacity falls as only the first power. Electronic refrigeration thus allows bulk cooling at low enough temperatures.

Thermoelectric effects can be used to cool electrons. The Peltier effect, in which a heat current accompanies an electric current, is used for this purpose in semiconductor thermoelectric refrigerators.[2] The application of Peltier refrigeration to the cryogenic cooling of metallic electrons can be understood by examining a generic thermoelectric refrigerator.

Consider the following hypothetical device, shown schematically in Fig. 1(a). At the center is a gas of metallic electrons (the reservoir) with Fermi energy E_F. The electronic temperature T_0 is manifest only in the thermal smearing of the electron distribution by $k_B T_0$ about E_F—in other words, its departure from a step function. Now imagine that, by some type of tunneling or other process, it is possible to extract electrons above E_F and holes from below E_F. These processes would modify the electronic distribution function, with the result that the electronic distribution sharpens and the electronic temperature is lowered.

How well does this electronic refrigerator perform? The range of energies over which electrons are excited above and below E_F is proportional to T_0. Each electron extracted (and hole filled) removes an amount of heat which is proportional to T_0 as well. Thus the cooling power of this refrigerator drops as the square of the electronic temperature.

The electrons also exchange heat with the lattice. From the acoustic phonons which dominate at low temperatures, metallic electrons absorb an amount of heat proportional to $(T^5 - T_0^5)$, where T is the ambient temperature.[3,4] This phonon heat leak drops faster with temperature than does the T_0^2-dependent cooling power. Thus, at lower temperatures, the refrigerator is increasingly adept at cooling the electrons, as is shown in Fig. 1(b). In addition, because the heat capacity of the electrons exceeds that of the phonons at low temperatures, the phonons will be cooled as well; again, this effect improves with temperature reduction. At first glance, the potential for cooling seems limitless.

One possible limit to the efficacy of such a refrigerator is the rate of electron-electron scattering in the metal. According to the Fermi-liquid theory, the inverse lifetime (decay rate) of an electron near E_F in a clean metal is proportional to T_0^2.[1] Because the fraction of the electrons which are thermally excited is proportional to T_0, the rate at which electron-electron scattering can replenish the electron distribution under the action of this refrigerator is proportional to the third power of T_0. Because each electron (or hole) removed from near E_F carries away a quantity of heat proportional to T_0, the cooling power of our hypothetical refrigerator is proportional to the fourth power of T_0 in this regime. Nevertheless, this cooling power decreases more slowly with temperature than the T^5 phonon heat leak does. In fact, the scattering is enhanced by disorder in the metal and the exponent may be smaller than 4. Therefore, our hypothetical thermoelectric refrigerator performs better overall as the temperature is decreased, even when the cooling power is limited by electron-electron scattering.

0163-1829/95/52(8)/5714(23)/$06.00

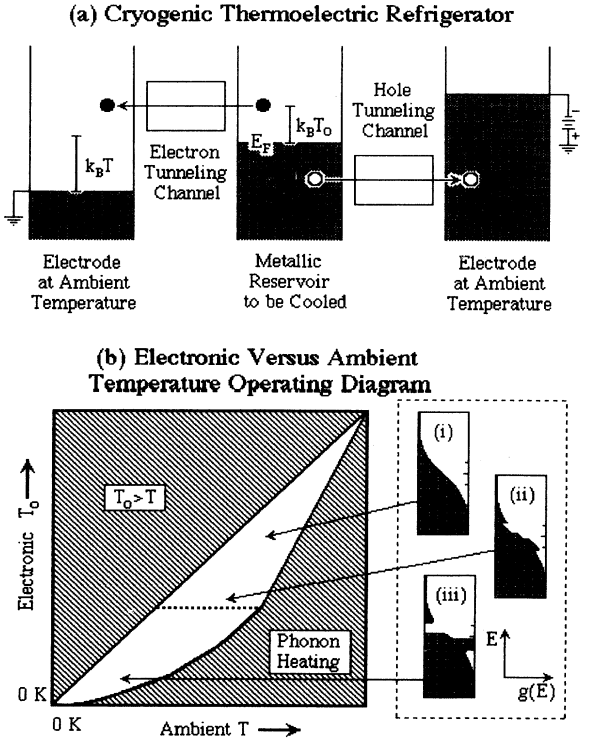

FIG. 1. (a) A hypothetical electronic cryogenic refrigerator which uses the Peltier effect. Electrons and holes are extracted through tunneling channels to cool the reservoir, center. (b) Operating diagram for the electronic refrigerator. At large ambient temperatures T, phonon absorption prevents operation. At electronic temperatures T_0 below the dotted line, electron-electron scattering limits the tunneling rate, and the electrons are driven out of thermal equilibrium (see diagrams at right).

These effects are shown schematically in Fig. 1(b). At too-high ambient temperatures, the refrigerator may not operate due to phonon heating (hatched area to the right). The hatched area to the left corresponds to heating because there the electronic temperature T_0 exceeds the ambient temperature T. Actually, the refrigerator may be useful in this regime—cooling electrons which have been heated above the ambient temperature—but we restrict our discussion to cooling the electrons to below the lattice temperature.

The unhatched region represents the regime of refrigerator operation. Above the dotted line, the electrons are in thermal equilibrium. This is represented schematically in diagram (i), where the electronic distribution $g(E)$ is sketched as a function of electron energy E. In the case of equilibrium, $g(E)$ is equal to $f(E)$, the Fermi-Dirac distribution. At lower T_0, the electrons in the reservoir depart from thermal equilibrium because the electron-electron-scattering rate ($\propto T_0^3$) falls more quickly with temperature than the extraction rate ($\propto T_0$) does [diagram (ii)]. At extremely low T_0, the electrons are far from equilibrium [diagram (iii)]; there is a range of electron (hole) energies above (below) E_F which is entirely depleted. However, the remainder of the reservoir's electronic distribution is distributed thermally, so that the thermal expressions for the electron-electron-scattering rate and the phonon heating still apply approximately. In this limit, the refrigerator's electron and hole extraction rate is fixed by electron-electron scattering rather than the conductance of the extraction mechanism (e.g., tunneling conductance) as it is in the equilibrium limit.

An electronic refrigerator seems to be capable of cooling the electrons in the reservoir. In some cases, the reservoir's crystal lattice will be cooled as well. For instance, if the thermal coupling between the electrons and the reservoir's crystal lattice is stronger than the thermal coupling of the reservoir's crystal lattice to ambient-temperature components such as the electrodes or a substrate, then the crystal lattice will be cooled by the refrigerator.

The above arguments are based upon general properties of metals; we arrived at a favorable conclusion regarding the performance of our hypothetical refrigerator without reference to specific mechanisms for extracting electrons and holes near E_F. In a previous paper, we proposed a specific device, the quantum-dot refrigerator (QDR), which would use resonant tunneling through the discrete electronic states of quantum dots to provide these tunneling channels.[5] In this device, the electrons in a two-dimensional electron gas (2DEG) in a $GaAs/Al_xGa_{1-x}As$ heterostructure would be cooled, whereas the crystal lattice would remain at the ambient temperature. Other cryogenic electronic refrigerators based upon the extraction of hot electrons and holes have been studied by other authors. These use the sharp superconducting gap edge to provide the tunneling channels to remove hot excitations.

This idea was first proposed in 1961: removing quasiparticles from a superconductor can enhance its superconductivity.[6] An increase in the superconducting energy gap by quasiparticle extraction was observed some years later.[7] The use of superconductor-superconductor tunneling to cool one superconducting electrode was proposed in 1981.[8] This line of research has recently come to fruition in an experiment by Blamire *et al.*; they used a $Nb/AlO_x/Al/AlO_x/Nb$ tunneling structure (hereafter called *SIS'IS*) to drive the Al layer superconducting from a temperature several times greater than its equilibrium T_c.[9] The ability of quasiparticle extraction to enhance superconductivity has been clearly demonstrated.

Martinis and co-workers at National Institute of Standards and Technology (NIST) have recently demonstrated a similar device, with two important differences: their device cooled a normal metal film rather than another superconductor, and the temperature was measured independently.[10] This demonstrated unambiguously that the normal-insulator-superconductor (NIS) tunneling was refrigerating the metal. This NIS junction refrigerator cooled the electrons in a Cu thin film from the ambient temperature of $T=100$ mK to $T_0=85$ mK. This result provides experimental support for our idea of extracting hot electrons and holes to cool metallic electrons. In all these devices, the electrons were cooled but the crystal

lattice remained at the ambient temperature. The performance of these refrigerators could be improved substantially if the reservoir were thermally isolated from ambient-temperature components. Although cryogenic thermoelectric refrigeration has been demonstrated, we believe that the potential thermoelectric cooling holds for cryogenic refrigeration has gone widely unrecognized.

In this paper, we argue that the exploitation of thermoelectric cooling for the cryogenic refrigeration of metallic and other samples should allow the exploration of whole new temperature regimes and hence new physics. To this end, we describe the refrigerating performance of the QDR.[5] The QDR is a direct implementation of the above hypothetical refrigerator because it relies upon individual tunneling channels to extract electrons from above E_F and holes from below E_F in a two-dimensional electron gas (2DEG). These tunneling channels are provided by resonant tunneling through the discrete electronic states of quantum dots. Due to its analytical simplicity, the QDR model can be solved in detail to reveal how thermoelectric cooling improves at extremely low temperatures. And, as we discuss below, the QDR offers scientific and technological promise in its own right.

We also argue that a refrigerator similar to that built by the NIST group should be capable of much greater cooling than was achieved by their prototype if certain parameters are optimized. Specifically, our model indicates that by increasing the NIS junction area and fabricating the device on a thin membrane for thermal isolation from ambient-temperature components, it should be possible to cool a metallic thin film (and hence small electronic devices such as radiation detectors) from above 300 mK to below 100 mK. This is an important technological goal, since 100 mK requires a dilution refrigerator whereas 300 mK can be attained by pumping on liquid ^{3}He. Our model also indicates that, with a proper material choice, it may be possible to cool a metal thin-film device from above 1 K to below 100 mK with this type of refrigerator.

The outline of the paper is as follows. In Sec. II we describe the operation of the QDR, first presenting equations for the electric current and cooling power, then determining its base temperature numerically. The latter is done in a numerical simulation based upon the condition that the cooling power match the flow of input heat. With this simulation, we verify the accuracy of certain simplifying assumptions, which we use in Sec. III to estimate how the phonon heat leak and electron-electron scattering limit QDR operation. In Sec. III we also consider the performance and refrigerating efficiency of a QDR with a specific set of operating parameters. In Sec. IV, we perform similar calculations for the NIS refrigerator, finding agreement with recent experiments. We predict that large reductions in electronic temperature should be possible for an optimized NIS refrigerator. In Sec. V we discuss various issues related to cryogenic thermoelectric cooling. We conclude in Sec. VI.

In the Appendix, we summarize the linear-response theory commonly used to treat thermoelectric effects, and apply it to the QDR.[2] This theory expresses thermoelectric properties in terms of the thermoelectric figure of merit Z, which is used to find the base temperature. We compute the thermoelectric figure of merit for the QDR, finding that the base temperature so obtained disagrees with the results of the numerical simulation of Sec. II D. We then discuss the physical origins of this discrepancy, concluding that the linear-response theory should not be applied without careful consideration of its suitability.

In this paper, we use the following units: temperatures in K, energies in eV, conductances in conductance quanta $(e^2/h)=(26\ \mathrm{k}\Omega)^{-1}$, lengths in μm, areas in $\mu\mathrm{m}^2$, and volumes in $\mu\mathrm{m}^3$, unless otherwise specified.

II. THE QUANTUM-DOT REFRIGERATOR

The QDR offers more than just a pedagogical argument in favor of Peltier cooling; it will also provide scientific and technological advances in its own right. As we see in Sec. III C, the QDR can cool a μm-sized 2DEG reservoir in the mK regime and larger 2DEG devices at lower temperatures. In Sec. V A, we find that the 2DEG heat capacity exceeds that of a mm-thick GaAs substrate below 200 μK. For temperatures in μK and nK, the QDR can cool macroscopic electron gases.

There are applications in the field of mesoscopic physics for a refrigerator which cools μm-sized devices in the mK range. Due to the weak electron-phonon coupling in small metallic components, mesoscopic devices often heat to a temperature far above the ambient. For instance, a recent experiment[11] revealed an electronic temperature (60 mK) which was five times the ambient of 12 mK. Other cryogenic-temperature experiments have found the electronic temperature to be much higher than the ambient as well.[3,4] A QDR could be integrated into the design of mesoscopic circuits, providing local refrigeration.

Potential applications for an ultracold (μK or nK) 2DEG include improved accuracy in the quantum Hall effect, nonequilibrium effects in 2DEG, localization studies, and the discovery of additional low-temperature phenomena. On the technological side, one could imagine constructing a 2DEG bolometer (radiation detector) whose thermal noise is many orders of magnitude less than that of existing bolometers due to the low temperature. The QDR may also provide cooling for certain applications like quantum computing in which quantum coherence (and hence a lack of thermal scattering) is needed. Scientific applications of refrigerating macroscopic metallic samples to the μK and nK regimes include exploring the possibility of superconductivity in gold and other metals which have not been tested below 100 μK.[12]

A. QDR model

The operating principle of the QDR was introduced in another paper,[5] so we will summarize its features briefly. The QDR is designed to be constructed by laterally patterning gates above a 2DEG in a $GaAs/Al_xGa_{1-x}As$ heterostructure, a technology which has been applied successfully to the fabrication and testing of quantum

dots.[13] The 2DEG structure of the QDR is shown in Fig. 2(a). There is a central reservoir R which is cooled by the Peltier effect of resonant tunneling through the discrete electronic states of quantum dots D_L and D_R. The latter lie between R and the electrodes V_L and V_R. The electrochemical potential (Fermi level) of R is μ_0; those of V_L and V_R are μ_L and μ_R. We assume for simplicity that the bias voltage drops equally across the junctions between R and each of the electrodes: $eV_b \equiv \mu_R - \mu_0 = \mu_0 - \mu_L$. The total voltage applied to the QDR is thus $\mu_R - \mu_L = 2eV_b$.

The quantum dots are small enough (hundreds of nanometers across) that the single-electron levels are quantized (the energy-level separation Δ is approximately tenths of a meV). Capacitively coupled gate electrodes can be used to tune the energy levels of D_L and D_R separately, to achieve the energy-level configuration of Fig. 2(b). Energy levels E_{D_L} and E_{D_R} of quantum dots D_L and D_R are tuned so that E_{D_L} is above μ_0, and E_{D_R} is below μ_0: $\varepsilon \equiv E_{D_L} - \mu_0 = \mu_0 - E_{D_R}$. With this arrangement, a current flows from left to right through the QDR as hot electrons resonantly tunnel through E_{D_L} to V_L and hot holes resonantly tunnel through E_{D_R} to V_R. The magnitude of this tunneling current is set by the energy smearing δ of the quantum dots' energy levels. This smearing is due to the finite lifetime of an electron in the quantum dot under tunneling to an electrode or R. Its magnitude is roughly $\delta \approx t\Delta$, where t is the transmission coefficient for tunneling from a quantum dot to R or the electrodes and Δ is the energy-level separation. Thus the quantum-dot energy levels can be sharpened by reducing the quantum dots' coupling to R and the electrodes.

The QDR contains several parameters $\{\Delta, \varepsilon, \delta, k_B T, k_B T_0$, and $eV_b\}$ which must be specified in order to predict its properties. We can make several physically plausible simplifying assumptions. First we assume for rough estimates that $\varepsilon \approx k_B T_0$ gives the best QDR performance. We will show that this is true in the numerical simulation of Sec. II D. This relation can be understood qualitatively by the following argument. If $\varepsilon \ll k_B T_0$, then not enough heat is removed per electron (hole), since the thermal energy removed from the reservoir by an electron (hole) above (below) μ_0 is proportional to ε. On the other hand, if $\varepsilon \gg k_B T_0$, the electron (hole) population at an energy ε above (below) μ_0 will be exponentially small and the tunneling rate will be low.

We also assume $k_B T_0 > \delta$. If the quantum dots' energy-level smearing δ is too great, the electronic thermal conductance of the tunneling junction can produce a heat current into R which is greater than the cooling power of the QDR. The validity of this assumption is also supported by the numerical simulation of Sec. II D.

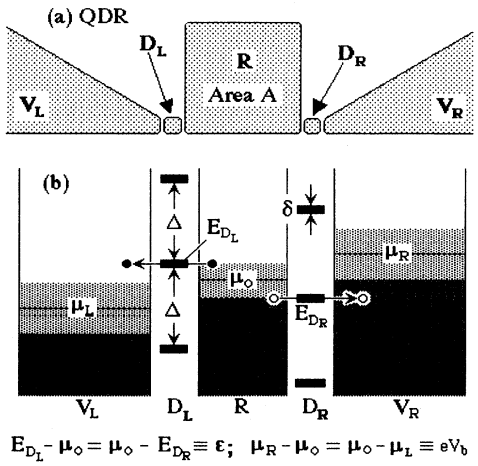

FIG. 2. (a) Schematic and (b) energy-level diagram for a quantum-dot thermoelectric refrigerator (QDR). (a) represents a laterally gated 2DEG in a $GaAs/Al_xGa_{1-x}As$ heterostructure. Quantum dots D_L and D_R are about $(100\ nm)^2$ in area and act as thermoelements (D_L negative, D_R positive) which cool the reservoir R.

B. QDR tunneling current and cooling power

In our previous paper on the QDR,[5] we presented a simplified calculation of the QDR's tunneling current and cooling power. Here we will calculate these currents in more detail, considering the finite smearing of the quantum-dot levels and the effects of tunneling through adjacent quantum-dot levels. The results of this section do not rely upon a linear-response approximation, but do assume thermal equilibrium in the 2DEG of the QDR reservoir. The effects of the departure from equilibrium as the tunneling frequency approaches the finite electron-electron-scattering rate in the 2DEG are discussed in Sec. III.

For simplicity, we take the probability of electrons and holes tunneling between a quantum dot and R, V_L, or V_R to be t, independent of energy. This is an idealization; in reality, the resonant-tunneling conductance will be reduced by a factor of 2 or so due to asymmetric barriers.[13] The tunneling barriers that separate the quantum dots from the other 2DEG regions can be tuned independently with gate electrodes to optimize the symmetry of the barrier.

In order to find the QDR cooling power, we need to specify the heat removed from an electron gas by extracting an electron. The change in internal energy U of an electron gas on removal of an electron is

$$\Delta U = T\Delta S - P\Delta V - \mu\Delta N = \Delta Q - P\Delta V - \mu\ ,$$

since $\Delta N = 1$ and $\Delta Q = T\Delta S$. We assume that the volume V of the electron gas is unaffected by the removal of an electron and that the chemical potential μ is unchanged by the removal of a single electron. Essentially, we assume that the QDR reservoir R is large enough that single-particle charging effects are unimportant. Noting that the internal energy of the electron gas decreases by the energy of the electron, $\Delta U = -E$, we see that the heat removed from the electron gas by removing an electron is $(E - \mu)$. Thus, removing an electron above the Fermi level cools the electron gas. By a similar argument, removing a hole of energy E from an electron gas extracts heat $-(E - \mu)$ from the electron gas. Removing holes

from below the Fermi level also cools the electron gas. Because the heat added or removed is just the difference of the energy with the Fermi energy, we take the zero of energy to be the Fermi energy and equate the heat removed to the electron energy.

With reference to Fig. 2(b) we see that, for positive bias potential V_b, positive electric current should flow to the right so that positive electron flux should flow to the left. The tunneling current is identical in our model for both junctions:[14]

$$I=e\int dE\frac{G(E)}{e^2}[f_{T_0}(E)-f_T(E+eV_b)]\ . \tag{2.1}$$

We define positive heat flow to be out of R—P is the QDR cooling power. In this paper, we refer to a heat current as a cooling power in the context of a thermoelectric refrigerator. When discussing thermoelectric effects in a general manner, we revert to the term heat current. Because each junction cools R equally, we double the expression for the cooling power to obtain

$$P=2\int dE\,E\frac{G(E)}{e^2}[f_{T_0}(E)-f_T(E+eV_b)]\ , \tag{2.2}$$

where the integration extends over the conduction band of R and the Fermi-Dirac function is $f_T(E)=(1+e^{-E/k_BT})^{-1}$. The temperature in R is T_0, and all other circuit elements are assumed to be at the ambient temperature T. This thermal isolation was found to occur in the Coulomb-blockade electrometer, in which the Coulomb island heated itself but the electrodes remained at the ambient temperature.[4]

The conductance of resonant tunneling through a quantum dot is[15]

$$G(E)=\frac{e^2}{h}\left\{2\sum_i\left[1+\left[\frac{E-E_i}{\delta}\right]^2\right]^{-1}\right\}$$
$$=\frac{2\pi\delta e^2}{h}\sum_i\left\{(\pi\delta)^{-1}\left[1+\left[\frac{E-E_i}{\delta}\right]^2\right]^{-1}\right\}, \tag{2.3}$$

where E_i are the energies of the discrete electronic states of the quantum dot, e is the electron charge, and h is Planck's constant. The quantity in curly brackets in the first equality is the transmission coefficient for resonant tunneling through a discrete state (including electron spin), in accord with the Landauer formula $G=(e^2/h)T$ which connects conductance G to the transmission coefficient T.[16] The quantity in curly brackets in the second equality is the Lorentzian energy broadening of a single quantum-dot energy level due to the smearing δ. The expressions for the tunneling current and cooling power become

$$I=2\frac{e}{h}\int dE\sum_i\left[1+\left[\frac{E-E_i}{\delta}\right]^2\right]^{-1}$$
$$\times[f_{T_0}(E)-F_T(E+eV_b)]\ , \tag{2.4}$$

$$P=\frac{4}{h}\int dE\,E\sum_i\left[1+\left[\frac{E-E_i}{\delta}\right]^2\right]^{-1}$$
$$\times[f_{T_0}(E)-f_T(E+eV_b)]\ . \tag{2.5}$$

Now we estimate the tunneling current (2.4) for use in Sec. III below. We replace the Lorentzian broadening with constant conductance $G=2(e^2/h)$ at $E=\varepsilon$ over an energy range δ:

$$I\approx e\left[\frac{G}{e^2}\right]\delta f_{T_0}(\varepsilon)$$
$$=\left[\frac{2e}{h}\right](k_BT_0)f_{T_0}(k_BT_0)\approx 1.8\ \text{nA}\ T_0\ . \tag{2.6}$$

We have used the relations $\varepsilon\approx k_BT_0$ and $\delta\approx k_BT_0$ obtained in the numerical simulation of an optimized QDR of Sec. II D below. We also neglect $f_T(\varepsilon+eV_b)$, assuming that the population in the electrodes at E_{D_L} and E_{D_R} is much less than the corresponding population $f_{T_0}(\varepsilon)$ in R. Converting to a rate by dividing out an electron charge,

$$\tau^{-1}_{\text{tunneling}}=\frac{I}{e}\approx 11\ \text{GHz}\ T_0\ , \tag{2.7}$$

we estimate the cooling power (2.5) by assuming that each tunneling electron and hole removes k_BT_0 of heat from R:

$$P\approx 2\left[\frac{G}{e^2}\right]\delta f_{T_0}(\varepsilon)(k_BT_0)$$
$$\approx\left[\frac{4}{h}\right](k_BT_0)^2f_{T_0}(k_BT_0)\approx 0.31\ \text{pW}\ T_0^2\ . \tag{2.8}$$

We have used the relations $\varepsilon\approx k_BT_0$ and $\delta\approx k_BT_0$ because these represent an optimized configuration for the QDR and so can be used to determine intrinsic limitations. We will use these estimates in Sec. III for comparison to the phonon heat leak and electron-electron-scattering rate in the 2DEG. Note that Eqs. (2.6)–(2.8) are only appropriate for an optimized QDR, i.e., a QDR at its base temperature; they do not apply for arbitrary T_0, in which case the integrals in Eqs. (2.4) and (2.5) must be evaluated.

C. Thermoelectric effects in the quantum and Coulomb-blockade regimes

In this section, we explain why we restrict QDR operation to the quantum regime which requires lower temperatures than the Coulomb-blockade regime. We will also briefly review some of the previous theoretical and experimental works on thermoelectric effects in quantum dots. We will find in the Appendix that our theory is consistent with these other results.

Thermoelectric effects have recently been studied in systems of reduced dimensionality including quantum point contacts and quantum dots in the Coulomb-blockade regime. The thermopower, Peltier coefficient, and thermal conductance of quantum point contacts were found to exhibit quantum-size effects.[17] Coulomb-blockade oscillations were observed in the thermopower of a quantum dot.[18] These latter results were predicted by a theory which used a linear-response approximation

to derive the thermoelectric coefficients for a quantum dot[19] and a theory of thermoelectric effects in Coulomb islands.[20]

Our linear-response results of the Appendix agree with the theory of Ref. 19. Specifically, our expression for the quantum-regime thermopower and electric resistance is equivalent to theirs. This agreement lends confidence in the accuracy of our model. However, our treatment differs from theirs in several ways. First, a thermoelectric refrigerator utilizes the Peltier effect, rather than the Seebeck effect. The Seebeck effect occurs when a temperature difference ΔT is applied across a junction; a thermoelectric voltage $V=\alpha\Delta T$ develops across the junction. In the Peltier effect, the junction is current biased and a heat current $P=\pi I$ flows with the electric current I. In reversible junctions the Peltier and Seebeck coefficients π and α are proportional, $\pi=T\alpha$, due to Onsager's reciprocity relations. However, this relation does not hold for mesoscopic junctions[20] and is not defined out of the linear-response regime. Because the best operation of the QDR occurs in the nonlinear regime, the measured Seebeck effect does not serve to predict the striking temperature-reduction capabilities of the QDR.

The transport behavior of a quantum dot with an energy-level spacing of roughly Δ, a capacitance C, and a Coulomb charging energy $e^2/2C$ has several characteristic temperature regimes. If $k_BT>e^2/2C$ and $k_BT>\Delta$, the quantum dot behaves like a macroscopic, albeit small, metal island. In the Coulomb-blockade regime, $k_BT<e^2/2C$; the Coulomb-blockade effect dominates transport through the quantum dot. Because $e^2/2C$ generally exceeds Δ, the Coulomb-blockade regime is usually $\Delta<k_BT<e^2/2C$. The lowest-temperature regime, for which $k_BT<\Delta$, is the quantum regime. The effects of both the Coulomb blockade and the discrete electronic states of the quantum dot are active at these low temperatures. The QDR operates in the quantum regime; its refrigerating properties are based upon the discrete electronic states of the quantum dot rather than the Coulomb-blockade effect.

Equations (2.4) and (2.5) are based upon a model of the QDR which ignores the Coulomb-blockade effect. Its inclusion would modify Fig. 2(b) by introducing a Coulomb-blockade gap between the empty and filled states of the quantum dot. The result is that the discrete electronic state at $E_{D_L}-\Delta$ would be lowered to $E_{D_L}-\Delta-e^2/2C$ by the Coulomb-blockade gap. Similarly, the discrete electronic state at $E_{D_R}+\Delta$ would be raised to $E_{D_R}+\Delta+e^2/2C$. This would effectively remove tunneling channels which, in the current model, provide a heat leak between the electrodes and R. In this sense, the neglect of the Coulomb-blockade effect renders the results based upon our QDR model conservative. Specifically, including the Coulomb-blockade effect in our QDR model should lower the case temperature obtained by the numerical simulation of Sec. II D. Note that the Coulomb-blockade gap is only beneficial in the quantum regime ($k_BT<\Delta$); at higher temperatures ($k_BT>\Delta$) the self-heating mechanisms discussed below become prevalent and the QDR cooling power is washed out.

In the QDR, electrons will primarily traverse a single channel through each quantum dot: the one which extracts hot electrons or holes from R. Other channels are suppressed because there are few electrons or holes thermally excited further than Δ from the electrodes' Fermi levels. In the Coulomb-blockade regime, however, other channels—including those which heat R—are probable as well. Since in the Coulomb-blockade regime $k_BT>\Delta$, thermally excited electrons and holes will tunnel through states other than E_{D_L} and E_{D_R}, depositing hot electrons and holes in R. Essentially, in the Coulomb-blockade regime, heat streams back into the reservoir from the electrodes.

The effect of this heat leak can be understood within the context of the linear-response theory discussed in the Appendix. The following argument is valid because the linear-response theory does give qualitatively correct results, even if they are quantitatively inaccurate. The performance of a thermoelectric refrigerator is parametrized by the dimensionless figure of merit Z as given by Eq. (A2); large Z corresponds to good refrigeration. The Seebeck coefficient (or thermopower) α is proportional to the cooling power. If the Coulomb-blockade gap were used for refrigeration instead of the gap Δ between the discrete states, α would be increased by the larger energy scale, about a factor of 3–10 for semiconductor quantum dots.[13]

However, the thermal conductance κ would be increased by a much larger factor, roughly $e^{\Delta/kT}$. In the quantum regime, the electrodes' population of electrons at levels other than E_{D_L} and E_{D_R} is roughly $e^{-\Delta/kT}$. In the Coulomb-blockade regime, this population would be increased to finite values (say, 0.2) because of the higher temperature $k_BT>\Delta$. The heat leak and hence the thermal conductance that it causes would increase by roughly $e^{\Delta/kT}$. As we will see in Sec. II D below, the best QDR operation occurs for $\Delta/k_BT>10$. Thus the heat leak is $e^{10}\approx 20\,000$ times worse in the Coulomb-blockade regime than in the quantum regime. Overall, the dimensionless figure of merit would be reduced by $20\,000/10^2=200$. The QDR refrigerating performance would suffer in the Coulomb-blockade regime.

D. QDR base temperature

In this section we calculate the base temperature of the QDR in a numerical simulation. Our primary motivation is to justify the approximations that we used in Sec. II B to estimate the QDR tunneling current and cooling power. Another motivation is to show that the QDR can achieve a large enough temperature reduction that the linear-response theory of thermoelectrics (see the Appendix) is not quantitatively accurate.

The base temperature of the QDR is the temperature T_0 for which the cooling power Eq. (2.5) just balances any external heat leaks. To find this temperature, we simulated the QDR by numerically integrating Eq. (2.5). Once we can find the QDR base temperature for a given set of operating parameters $\{\Delta,\varepsilon,\delta,k_BT,eV_b\}$, we vary these parameters to obtain the maximum reservoir cool-

ing ratio (T/T_0).

We can gain insight into the physics of the QDR by considering the case in which there are no external heat leaks. Then the QDR base temperature T_0 is the root of the equation $P=0$, with P defined in Eq. (2.5). It might seem that, in the absence of external heat leaks, the QDR should cool R to zero temperature. The reason that this procedure produces a finite value T_0 is that the QDR has intrinsic heat leaks. Heat can leak back into R by resonant tunneling through states other than E_{D_L} and E_{D_R} or by off-resonance tunneling through the wide Lorentzian tails of E_{D_L} and E_{D_R}.

In order to compute the base temperature we evaluated Eq. (2.5) numerically. By choosing Δ as the energy scale, it is eliminated as a parameter. For given values of ε, δ, k_BT, and eV_b, we compute the reservoir temperature k_BT_0 by finding the root of $P=0$, assuming no external heat leaks as we discussed above. Once our computer program could find $k_BT_0(\varepsilon,\delta,k_BT,eV_b)$, we wrote another program which would maximize the temperature reduction (T/T_0) with respect to k_BT, ε, and eV_b. That a maximum with respect to these three variables exists can be seen physically by the following considerations based upon the diagram of Fig. 2(b).

If ε is too large, there will not be enough hot excitations in R to produce substantial tunneling. If ε is too small, then each electron and hole tunneling through E_{D_L} and E_{D_R} will not carry much heat. In fact, as we will see below, the optimum value for ε is always near k_BT_0.

If k_BT is too large, electrons and holes will tunnel from the electrodes to R via quantum-dot energy levels other than E_{D_L} and E_{D_R}; R will be heated. However, if T is too small, the ratio (T/T_0) will suffer; thus the optimum ambient temperature k_BT is the maximum possible without heating R.

If eV_b is too small, then tunneling cannot occur because the electron population at E_{D_L} in V_L will be higher than the corresponding electron population in R and electrons will tunnel backwards (from V_L to R). On the other hand, if eV_b is too large, then electrons and holes will tunnel from the electrodes to R via quantum-dot states other than E_{D_L} and E_{D_R}, heating R.

There is not a value of δ which optimizes the QDR operation. In fact, for a QDR operating in the absence of external heat leaks, the temperature-reduction ratio (T/T_0) improves as δ decreases. Thus we will give the results of the calculation as a function of δ, listing the optimum values for k_BT, ε, and eV_b.

We performed the numerical integration for a varying number of energy levels other than E_{D_L} and E_{D_R}. Aside from the energy levels at $E_{D_L}-\Delta$ and $E_{D_R}+\Delta$, the inclusion of more levels made only a small difference. These two levels are important to include since they represent a major heat leak that determines the optimum eV_b. The numerical results we quote here (and in the Appendix) were computed using six energy levels of each quantum dot (e.g., $E_{D_L}-3\Delta$, $E_{D_L}-2\Delta,\ldots,E_{D_L}+2\Delta$), to be certain of accurate results.

Table I lists the maximum value of (T/T_0) as a function of δ. Also listed in Table I are the corresponding values of k_BT_0, k_BT, ε, eV_b, (ε/k_BT_0), and (k_BT_0/δ). Note that (k_BT_0/δ), which is near unity for large δ, is a slowly increasing function of $-\log_{10}\delta$. We can summarize this behavior approximately by the expression $k_BT_0>\delta$, which we took as an assumption in our previous paper.[5] Also, the rule $\varepsilon\approx k_BT_0$ is obeyed very accurately in these results. These accurate numerical results justify the approximations used below in Sec. III C to delimit QDR operation. Specifically, we assume that the lowest value that k_BT_0 can achieve is δ, and that the QDR operation is optimized for $\varepsilon\approx k_BT_0$.

The results of this section account accurately for the nonlinear dependence of the QDR cooling power on temperature difference $\Delta T=T-T_0$, but assume that the electron distribution in R and the electrodes is at thermal equilibrium, i.e., is well described by the Fermi-Dirac function. In Sec. III, we will specify the regime of QDR operation for which this assumption is justified, and also estimate the performance of the QDR in the nonequilibrium regime.

In the Appendix, we will compute the base temperature using the linear-response theory of thermoelectric effects. The base temperature of any refrigerator is determined by the condition that the cooling power match the heat load (or heat leak). This basic approach is shared by the accurate numerical simulation of the present section and the linear-response theory of the Appendix. Thus

TABLE I. Results of a numerical simulation of the QDR. These numbers represent an optimized QDR. The base temperature T_0 is obtained by finding the roots of $P=0$ [using Eq. (2.5) for P], then varying the parameters ε, eV_b, and T until a maximum of (T/T_0) is found. Note the difference between this estimate of (T/T_0) and that obtained in Sec. III (see Table I); the linear-response approximation does not describe the QDR well.

δ/Δ	(T/T_0)	k_BT_0/Δ	k_BT/Δ	ε/Δ	eV_b/Δ	ε/k_BT_0	k_BT_0/δ
10^{-1}	1.22	0.072 1	0.0881	0.158	0.103	2.19	0.721
10^{-2}	3.60	0.024 9	0.0896	0.039 5	0.237	1.59	2.49
10^{-3}	16.6	0.004 65	0.0772	0.006 66	0.257	1.43	4.65
10^{-4}	94.9	0.000 695	0.0660	0.000 953	0.243	1.37	6.95
10^{-5}	615	0.000 093 1	0.0572	0.000 125	0.224	1.34	9.31
10^{-6}	4310	0.000 011 7	0.0505	0.000 015 5	0.206	1.32	11.7

calculations based upon an accurate numerical simulation and the linear-response theory should, in principle, be equivalent. However, for most cases of interest, the temperature reduction $\Delta T = T - T_0$ is comparable to T, invalidating the linear-response approximation. Furthermore, there are also more subtle reasons why linear-response theory breaks down; these are discussed in Sec. 7 of the Appendix. Thus it may seem fruitless to apply the linear-response theory of thermoelectric effects to the QDR. Our motivation for doing so in the Appendix is partly to compare our theory to that of Ref. 19, partly to show by counter-example that the linear-response theory of thermoelectric effects does not apply well to all thermoelectric refrigerators, and should only be applied with attention to the underlying physical assumptions.

III. QDR PERFORMANCE

The QDR can achieve a large temperature reduction in the absence of external heat leaks. Heat leaks will constrain this cooling and set the temperature regimes in which the QDR can operate. In Sec. III A, we incorporate the effects of phonon absorption on QDR performance. In Sec. III B we explore the limit in which the electrons in R are driven far from thermal equilibrium by resonant tunneling. These effects are summarized in Sec. III C as regions in diagrams of electronic versus ambient temperature. In Sec. III D we discuss ways of measuring the electronic temperature in the QDR. In Sec. III E, we present an example of the performance of a QDR of specified properties. Finally, in Sec. III F, we compute the refrigerating efficiency of this example QDR, finding that an optimized QDR approaches a value near the Carnot efficiency limit.

A. Phonon heat leak for a 2DEG

Electrons at temperature T_0 residing in a solid at temperature T will absorb and emit phonons. If $T > T_0$, the electron gas will absorb heat from the phonons. For the low temperatures considered here, optical phonons are frozen out and the dominant acoustic-phonon wavelength is comparable to the dimensions of the reservoir R. Thus we expect the heat leak in the QDR due to phonon absorption to be small. Nonetheless, phonon heating constrains the regime of the QDR operation.

The heat absorbed by electrons from acoustic phonons at low temperatures follows the universal form $P_{\rm ph} = \Sigma\Omega(T^5 - T_0^5)$.[4] This form holds for bulk metals such as aluminum and copper as well as other electron gases such as the 2DEG of which the QDR is composed.[4,21] The constant Σ contains all of the intensive material parameters. For a bulk metal, Ω is the volume that the electrons occupy; for a 2DEG, Ω is the area that they cover.

Liu and Niu have computed Σ for bulk metals and 2DEG layers in a $GaAs/Al_xGa_{1-x}As$ heterostructure, finding good agreement with experiment.[21] Their calculation is based upon a transition rate obtained from Fermi's golden rule; it includes the electron-phonon coupling due to both the deformation potential and the screened piezoelectric effect. For a 2DEG in a $GaAs/Al_xGa_{1-x}As$ heterostructure, they find $\Sigma = 30$ fW K^{-5} μm^{-2}, so[21]

$$P_{\rm ph} = 30 \text{ fW } A[T^5 - T_0^5] . \tag{3.1}$$

The QDR base temperature is determined by balancing the QDR cooling power P against the phonon heat leak $P_{\rm ph}$. Using Eq. (2.8) for P and setting it equal to $P_{\rm ph}$, we obtain a relation which determines the QDR base temperature T_0 for a given ambient temperature T and reservoir area A:

$$10.3T_0^2 > A[T^5 - T_0^5] . \tag{3.2}$$

B. Departure from thermal equilibrium

The calculations in Sec. II assumed that the electrons in the reservoir of the QDR are in thermal equilibrium. For some regimes of QDR operation, this is a good approximation. However, when the tunneling frequency approaches the rate at which electron-electron scattering can replenish the 2DEG states, the electronic distribution in R departs from equilibrium. In this case, the electron-electron-scattering rate, rather than the resonant tunneling rate Eq. (2.7), determines the QDR tunneling rate and hence its cooling power. Here we estimate the electron-electron-scattering rate, considering also the enhancement due to disorder.

Electron-phonon scattering is another potential thermalizing mechanism for a 2DEG. However, for the regime of QDR operation, electron-phonon scattering is much slower than electron-electron scattering. This is true because the average phonon has a large amount of energy—$k_B T$—and could heat the 2DEG. Thus the QDR only operates when the electron-phonon-scattering rate is slower than the tunneling rate which is always less than or comparable to the electron-electron-scattering rate. Because electron-phonon scattering is always slower than electron-electron scattering during QDR operation, we neglect the former as a mechanism for 2DEG thermalization in our analysis.

The departure of the QDR 2DEG from thermal equilibrium should be similar to that in diagrams (i)–(iii) of Fig. 1(b). That is, whereas the electron states near the tunneling channels are saturated (empty above the Fermi level and filled below), the rest of the electron distribution should be thermal as in diagram (iii). Thus the rate at which electrons and holes scatter into the tunneling channels is approximated by the expression for thermal electron-electron scattering.

To obtain the actual tunneling rate when limited by electron-electron scattering, we compute the rate at which electrons scatter into states which lie in a range of δ about $\mu_0 + \varepsilon$. By the principle of detailed balance, this rate is identical to the rate at which they are scattered out of this range when the 2DEG is at equilibrium. Since the latter rate is easier to compute, we will do so.

Electron-electron-scattering rates are quoted in terms of the inverse lifetime (decay rate or scattering rate) of a given electron. To obtain the scattering-limited QDR tunneling rate we multiply the scattering rate by the

number of initial states N_δ in each tunneling channel:

$$N_\delta = N_F \delta A f_{T_0}(\varepsilon) \approx 6.5 T_0 A \tag{3.3}$$

where the energy width of each tunneling channel is $\delta \approx k_B T_0$, $N_F = 2.8 \times 10^5\ \mu\text{m}^{-2}\,\text{eV}^{-1}$ is the 2DEG density of states, and we have used the relation $\varepsilon \approx k_B T_0$.[22]

The electron-electron-scattering rate in a clean metal is given by Fermi-liquid theory:[1]

$$\tau_{\text{FL}}^{-1} = \frac{(k_B T_0)^2}{h E_F} \approx 1.3\ \text{GHz}\ T_0^2 \ . \tag{3.4}$$

For a 2DEG, there is a geometrical correction to the electron-electron-scattering rate.[23] Liu and Niu[21] calculated the matrix element necessary to obtain an absolute scattering rate from the equations of Hodges, Smith, and Wilkins.[23] The result for an electron $\varepsilon \approx k_B T_0$ above the Fermi level is[21]

$$\tau_{\text{clean}}^{-1} = 0.199\ \text{GHz}\ T_0^2 \left[1 + 0.53 \ln \frac{99}{T_0} \right] . \tag{3.5}$$

In the presence of disorder such as impurity ions, the electron-electron-scattering rate is enhanced. This is due to the modification of electronic eigenfunctions by the elastic scattering. This correction depends upon the dimensions of the electron gas relative to a disorder length scale[21,24] $L_{\delta E} = \sqrt{\hbar v_F l / \delta E}$, where l is the elastic mean free path and $\delta E = k_B T_0$. Mobilities as high as 2×10^6 $\text{cm}^2/\text{V s}$ have been observed for 2DEG electrons in a $\text{GaAs/Al}_x\text{Ga}_{1-x}\text{As}$ heterostructure.[25] This mobility is equivalent to a mean free path of $l = 16\ \mu\text{m}$. For this value, $l_{\delta E} = 5.1\ \mu\text{m} T_0^{-1/2}$, where we have used $v_F = 2.11 \times 10^7$ cm/s.[22] For this value of l, $L_{\delta E}$ exceeds the dimensions of R for the temperatures at which electron-electron scattering dominates the tunneling rate. Thus we use the zero-dimensional expression for the disorder-enhanced electron-electron scattering:[21]

$$\tau_{\text{disorder}}^{-1} = (4 h N_F A)^{-1} = 0.216\ \text{GHz}\ A^{-1} \ . \tag{3.6}$$

The two-dimensional form (appropriate when $L_{\delta E}^2 < A$) is obtained by replacing A in Eq. (3.6) by $L_{\delta E}^2$. The disorder-enhanced electron-electron-scattering rate in Eq. (3.6) is independent of temperature because the physical boundaries of the 2DEG have replaced the temperature-dependent disorder length $L_{\delta E}$.

Making the 2DEG dirtier (i.e., decreasing l) would increase the disorder enhancement to the electron-electron-scattering rate by decreasing $L_{\delta E}$. However, for length scales above $L_{\delta E}$, localization effects may occur; thus it is best if impurity level is low enough that $L_{\delta E}^2 > A$.

The electron-electron-scattering-limited tunneling frequency is

$$\begin{aligned} \tau_{\text{ee}}^{-1} &= N_\delta (\tau_{\text{clean}}^{-1} + \tau_{\text{disorder}}^{-1}) \\ &= 1.4\ \text{GHz}\ T_0 \left\{ 1 + 0.9234 A T_0^2 \right. \\ &\qquad \left. \times \left[1 + 0.53 \ln \frac{99}{T_0} \right] \right\} . \end{aligned} \tag{3.7}$$

When the tunneling frequency of the QDR is limited by electron-electron scattering, the cooling power Eq. (2.8) must be replaced by

$$\begin{aligned} P_{\text{scat}} &= 2 (k_B T_0) \tau_{\text{ee}}^{-1} \\ &= 39\ \text{fW} T_0^2 \left\{ 1 + 0.9234 A T_0^2 \right. \\ &\qquad \left. \times \left[1 + 0.53 \ln \frac{99}{T_0} \right] \right\} . \end{aligned} \tag{3.8}$$

We have multiplied by 2 because there are two tunneling channels: E_{D_L} and E_{D_R} [cf. Fig. 2(b)]. We can find the temperature below which the QDR tunneling current is limited by electron-electron scattering by setting Eq. (3.7) equal to the QDR tunneling rate Eq. (2.7):

$$0.773 = A T_0^2 \left[1 + 0.53 \ln \frac{99}{T_0} \right] . \tag{3.9}$$

For a larger reservoir, the greater number of initial states allows the QDR to maintain thermal equilibrium at lower reservoir temperatures. In a QDR with a $(10\text{-}\mu\text{m})^2$ reservoir, the 2DEG departs from equilibrium when T_0 dips below 130 mK. For a 1-cm^2 reservoir, however, T_0 can be as low as 95 μK before thermal equilibrium is lost. The disadvantage to a large reservoir is that it absorbs more heat from phonons; it must be operated at a lower ambient temperature. It is the competition between these effects which defines the regime of QDR operation.

C. Regime of QDR operation

In this section we show how the physical phenomena described in Secs. III A and III B constrain the operational regime of the QDR. The phonon heat leak limits the ambient temperature T for a given reservoir size; the electron-electron-scattering rate determines at what reservoir temperature T_0 the 2DEG electrons are no longer able to maintain thermal equilibrium.

At high enough electronic temperatures T_0 [above the cutoff given by Eq. (3.9)] the 2DEG is in thermal equilibrium, and Eq. (2.8) describes the QDR cooling power. In this case, Eq. (3.2) determines the QDR base electronic temperature T_0 for a given ambient temperature T.

For low T_0, the 2DEG departs from thermal equilibrium and the QDR cooling power is given by the electron-electron-scattering-limited form Eq. (3.8). Setting this equal to the phonon heat leak Eq. (3.1), we obtain a relation between T_0 and T in the nonequilibrium regime:

$$T=T_0\left\{1+1.29T_0^{-3}\left[A^{-1}+0.923T_0^2\left[1+0.53\ \ln\frac{99}{T_0}\right]\right]\right\}^{1/5}. \tag{3.10}$$

In using Eqs. (2.7) and (2.8) to represent QDR tunneling properties, we have assumed that the QDR has been optimized as was done in the numerical simulation of Sec. II D. Thus the QDR base temperatures T_0 determined by Eqs. (3.2) and (3.10) represent the limits of an optimized QDR. For T_0 above the base temperature, the QDR may still operate, albeit nonoptimally; in this case, its cooling power and tunneling current are not described by Eqs. (2.6) and (2.8).

The performance of the QDR is best summarized by marking regions where the QDR may operate on a diagram of electronic temperature T_0 versus ambient temperature T as we did for a generic cryogenic thermoelectric refrigerator in Sec. I. In Fig. 3, we show diagrams of this sort. In region (1), the QDR operates with the 2DEG in thermal equilibrium. The dashed line shows the cutoff between thermal equilibrium and nonequilibrium given by Eq. (3.9). In region (2), the QDR acts as a heater ($T_0 > T$). In region (3), the ambient temperature T is so high that the QDR cooling is overcome by phonon heating. Finally, the QDR operates in the nonequilibrium limit in region (4). Note that the QDR operating diagrams in Fig. 3 differ from the generic operating diagram of Fig. 1(b) by their scale. Figure 1(b) is a linear-linear plot, with temperatures ranging from 0 K to a finite value, whereas the diagrams in Fig. 3 are log-log plots. Thus the region of QDR operation appears in Fig. 3 to increase without limit as the temperature is reduced, whereas in Fig. 1(b) the operating region shrinks as the temperatures approach 0 K. This is just a difference in the plotting, not a difference in the physics. Plotted on a linear-linear scale, the QDR operating regime shrinks as the temperature is reduced; unlimited cooling is not possible. We used a log-log format in Fig. 3 because the QDR can operate over several orders of magnitude in temperature.

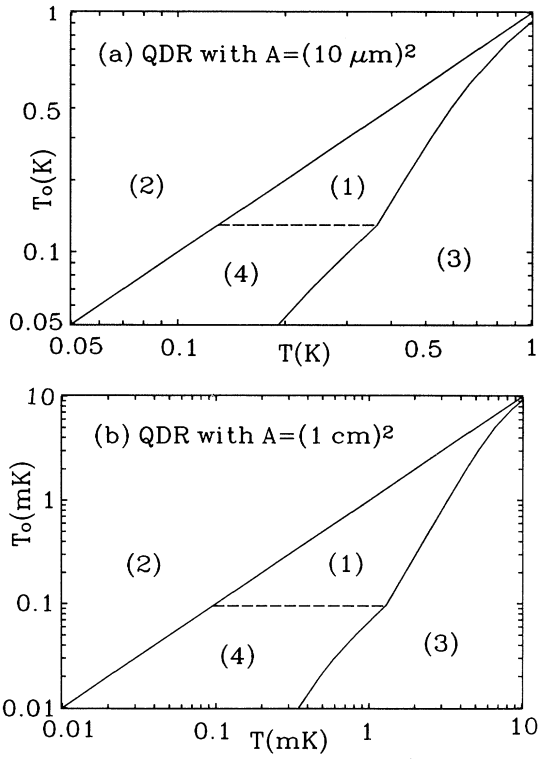

FIG. 3. Ambient temperature T—QDR reservoir temperature T_0 diagram indicating regimes of QDR behavior. (a) A small reservoir area $A=1\ \mu m^2$ operates at $T=600$ mK, cooling R to 200 mK. (b) A large reservoir area $A=1\ mm^2$ allows lower-temperature operation with a base temperature of $T_0=200\ \mu K$ from an ambient of 3 mK. Regimes of QDR behavior are as follows. (1) Refrigeration. (2) Cooling hot electrons ($T_0 > T$). (3) QDR cannot overcome phonon heat leak. (4) Current limited by electron-electron scattering. (5) Point with highest fractional temperature reduction (T/T_0); max from Eq. (3.11).

The small QDR [with a (10-μm)^2 reservoir] of Fig. 3(a) begins refrigerating at temperatures of hundreds of mK; this device could serve as a good proof of concept since dilution refrigerators can easily access this temperature range. The larger QDR [$A=(1\ cm)^2$] of Fig. 3(b) demonstrates the potential of the QDR for ultracryogenic refrigeration. This device is best suited to μK-range operation and should be capable of cooling the 2DEG to nK temperatures.

In general, a QDR refrigerates better at lower temperatures. There may be a limit to this behavior as thermodynamic fluctuations destroy our steady-state assumption of heat balance. In other words, if phonon absorption occurs only once an hour but tunneling occurs once a second, it does not make sense to talk about a steady-state electronic temperature. Each time a phonon is absorbed, the 2DEG is heated to a high temperature, afterwards cooling until essentially all of the excited electrons and holes are removed. This behavior should not dominate until (T/T_0) becomes very large.

The above limitations are mostly based upon the physics of the 2DEG in R. What do the quantum-dot-imposed physical restrictions imply about the QDR operational regimes? The main constraint is that the QDR must operate in the quantum regime, with $k_B T_0$, $k_B T$, and eV_b less than Δ. The size of currently available quantum dots limits the ambient temperature to below a few hundred mK; this limit is not restrictive since even the small QDR in Fig. 3(a) can only achieve significant cooling for a few hundred mK and lower. Another quantum-dot property which is critical for QDR performance is the energy-level smearing δ, which we determined in Sec. II D to be $\delta \approx k_B T_0$ for an optimized QDR. δ can be adjusted by changing the coupling between the

quantum dots and the electrodes and R; this is accomplished by changing the gate voltages.

D. Measuring the temperature in the QDR

The QDR is designed to cool an electron-gas reservoir which is small in size and is embedded in a GaAs chip. One of the most difficult aspects of building and testing a QDR will be measuring its temperature. We propose several techniques.

The line shape (conductance as a function of gate voltage) of resonant tunneling through the quantum states of a quantum dot is in some cases given by the temperature. If the quantum dot couples two 2DEG regions which differ in temperature, the line shape is generally thought to be determined by the greater temperature. Conceptually, this notion requires that the bias eV_b be negligible compared to the thermal smearing. For higher bias values, one should see a step up to maximum conductance when the quantum-state energy lies between the 2DEG Fermi levels and a step back down to zero conductance when the quantum-state energy leaves this range. The step up would be smeared by one temperature, and the step down would be smeared by the other temperature. Thus it should be possible to infer the temperatures of both 2DEG regions coupled by a quantum dot is the bias eV_b exceeds the thermal smearing on both sides.

This method could be applied to a QDR in which a third quantum dot is coupled to the 2DEG reservoir. This extra quantum dot would be coupled more weakly to the reservoir than the other two, so that it would not modify the electronic distribution in the reservoir. Its asymmetric resonant-tunneling line shape should allow a measurement of the electronic temperature in R.

This approach relies upon the assumption that the conductance line shape of tunneling through a quantum dot reveals the true thermodynamic temperature of the electrons in the 2DEG. A more fundamental measurement of the QDR temperature would directly take it from some physical property of R. One possibility is offered by finding the minima of the quantum Hall resistivity with respect to magnetic field,[26] i.e., constructing a QDR with additional voltage-sensing electrodes on the sides of the reservoir to measure the Hall voltage.

Alternatively, a superconducting quantum interference device (SQUID) magnetometer could be used to measure thermal current noise in a resistor coupled to a QDR. This would provide another temperature measurement which is independent of quantum-dot physics. Both of these techniques have the drawback that any electrical coupling to an external circuit could provide a strong heat leak back into the QDR reservoir. In all, thermometry is probably the biggest experimental challenge in testing a QDR, as it is in any low-temperature experiment.

E. QDR example

In this section, we present numerical results for a QDR of specific parameters. This is intended to illuminate the capabilities and realm of usefulness of the QDR. In addition, we show just how far the QDR lies from the linear-response regime.

The results of the Appendix rely on the linear-response approximation. Specifically, the linear-response theory of thermoelectrics assumes that the cooling power P and the tunneling current I depend linearly upon the temperature difference $\Delta T = T - T_0$. To check the validity of this assumption, we have evaluated the integrals in Eqs. (2.4) and (2.5) numerically. Figure 4 shows the variation of I and P with reservoir temperature T_0 for $\Delta = 0.1$ meV and $\delta = 10^{-3}\Delta = 10^{-4}$ meV. The values of $k_B T$, eV_b, and ε (listed in Table I) were chosen in the numerical simulation of Sec. II D to maximize (T/T_0), yielding $(T/T_0)_{\max} \approx 16$. The optimum ambient temperature so obtained is $T = (0.0772\Delta)/k_B \approx 90$ mK. Note that I and P do not vary linearly with ΔT at the QDR base temperature $T_0 \approx 5$ mK. The linear-response approximation is not adequate to describe the operation of the QDR.

The calculations leading to Fig. 4 do not include the phonon heat leak or the electron-electron-scattering effects. In addition, we have not yet specified the area A of the reservoir. We do so now to illustrate these constraints on QDR operation. Suppose that we wish to use a $(10\text{-}\mu\text{m})^2$ reservoir, as in the QDR of Fig. 3(a). For a QDR this size, the electron-electron-scattering limit sets in for electronic temperatures below 40 mK. According to Eq. (3.10), the base temperature at $T = 90$ mK is $T_0 = 10$ mK. For temperatures below 40 mK, Fig. 4 does not give the accurate tunneling current and cooling power because the QDR is operating in the nonequilibrium regime, where the tunneling rate is determined by the electron-electron-scattering rate.

Suppose the QDR had a larger reservoir, $(100\ \mu\text{m})^2$ in area. Then [again from Eq. (3.9)] the electron-electron-scattering limit sets in at a lower temperature, $T_0 = 3.5$ mK; this is because the larger reservoir has more total states aligned with the QDR tunneling channel. This QDR operates in the equilibrium regime, for which the

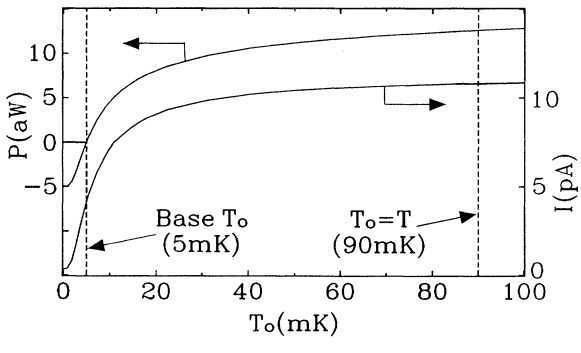

FIG. 4. Numerical evaluation of Eqs. (2.4) and (2.5) for I and P vs reservoir temperature $k_B T_0$, for $\Delta = 0.1$ meV and $\delta = 10^{-3}\Delta = 10^{-4}$ meV. Here $k_B T = 0.0772\Delta = 90$ mK (shown by dashed line), $eV_b = 0.255\Delta = 0.0255$ meV, and $\varepsilon = 0.006\,66\Delta = 0.666\ \mu$eV were chosen (as listed in Table I) to maximize (T/T_0), giving $T_0 = 4.65$ mK (shown by dotted line). The base temperature is set by the condition $P = 0$.

cooling power is given by Eqs. (2.5) or (2.8). Then the base temperature is determined by Eq. (3.2) to be $T_0=70$ mK. The larger reservoir allows greater electron-electron scattering but also absorbs more phonons, increasing the base temperature substantially.

We can also use this QDR example to check our simplifying assumptions from Sec. II B. Equation (2.8) gives $P\approx 8$ aW for a QDR with a base temperature of $T_0=5$ mK. This agrees well with the magnitude of the cooling powers shown in Fig. 4. It is important to note that Eq. (2.8) assumes optimized parameters, as we have chosen for this QDR example. As we found in the QDR simulation of Sec. II D, the optimum QDR parameters can be summarized by the relations $\varepsilon\approx k_B T_0$ and $\delta\approx k_B T_0$. In the case of the QDR example with $\delta=10^{-3}\Delta$, we see from Table I that the QDR behavior is optimized for $\varepsilon=1.43k_B T_0$ and $\delta=4.65k_B T_0$. This indicates that the simplifying assumptions $\varepsilon\approx k_B T_0$ and $\delta\approx k_B T_0$ are obeyed within a factor of 5.

Equation (2.8) is not appropriate for a QDR with nonoptimal parameters. For instance, consider the (thermal-equilibrium limit) QDR with a $(100\text{-}\mu\text{m})^2$ reservoir. With its base temperature of 70 mK, Eq. (2.8) predicts a cooling power of 1.5 fW, about 200 times greater than the values shown in Fig. 4; since this QDR operates with the reservoir in thermal equilibrium, however, the values in Fig. 4 should be valid. Equation (2.8) predicts the wrong value because the $(100\text{-}\mu\text{m})^2$-reservoir QDR is not optimized for the temperature regime in which it is being placed. Specifically, for this QDR, $\delta \ll k_B T_0$.

F. QDR coefficient of performance

The refrigerating coefficient of performance ϕ is defined in thermodynamics to be the ratio of the cooling power P to the rate P_W at which work is performed in order to operate the refrigerator. For a Carnot refrigerator cooling a reservoir to T_0 from an ambient temperature of T, the coefficient of performance is[27]

$$\phi_{\text{Carnot}}=\frac{T_0}{T-T_0}\ ,$$

where P is the cooling power and P_W is the rate at which work is performed to operate the refrigerator. According to the second law of thermodynamics, the coefficient of performance for any real refrigerator cannot exceed the Carnot value. Note, however, that the refrigerating coefficient of performance can exceed 1; this is because its definition differs from the efficiency of an engine, for which the Carnot value is $\Delta T/T$.

For the QDR, $P_W=I(2V_b)$. Using I and P from Eqs. (2.6) and (2.8), which represent an optimized QDR,

$$\phi_{\text{QDR}}\approx\frac{k_B T_0}{eV_b}\ .$$

To compare this to the Carnot value, we find limits for eV_b in terms of T and T_0. To extract hot electrons, the electron population in R at E_{D_L} must be greater than that in V_L:

$$f_T(E_{D_1}-\mu_L)<f_{T_0}(E_{D_1}-\mu_0)$$
$$\Rightarrow eV_b>\epsilon\left[1-\frac{T_0}{T}\right]\approx k_B(T-T_0)\ .$$

This gives an upper limit to the QDR coefficient of performance:

$$\phi_{\text{QDR}}\equiv\frac{P}{P_w}<\frac{T_0}{T-T_0}=\phi_{\text{Carnot}}\ . \tag{3.11}$$

This rough calculation indicates that the QDR coefficient of performance is bounded above by the Carnot value. This must be true for any real refrigerator. The result Eq. (3.11) was derived for a model in which only the tunneling channel is considered. In reality, the QDR departs from this model. The energies of the quantum dots' discrete electronic states are smeared by δ, and some tunneling occurs through states other than E_{D_L} and E_{D_R}. These effects can be accounted for by using a computer simulation similar to those employed in Sec. II D to compute the QDR base temperature.

We computed the QDR coefficient of performance by numerically integrating Eqs. (2.4) and (2.5) to find I and P; we substituted these values into the definition of the coefficient of performance $\phi_{\text{QDR}}=P/P_W=P/(2IV_b)$. We then varied ε, eV_b, and $k_B T$ to obtain the maximum ϕ_{QDR}. Figure 5 shows ϕ_{QDR} for a QDR with $\delta/\Delta=10^{-3}$, plotted as a function of reservoir temperature reduction from the ambient, T_0/T. Note that ϕ_{QDR} is very close to the Carnot limit—the ratio exceeds 70% for large T_0/T. For small (T_0/T), ϕ_{QDR} goes to zero. The temperature at which this occurs is the base temperature of the QDR. This occurs at $T/T_0\approx 16$, in agreement with the QDR example of Sec. III E.

Under a heat load such as the phonon heat leak, the coefficient of performance of the QDR is reduced. This is because P is replaced by the net cooling power $P\text{-}P_{\text{ph}}$. Nonetheless, the intrinsic efficiency of the QDR is high

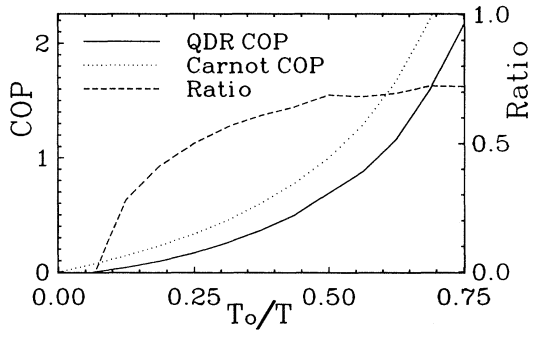

FIG. 5. Comparison of the refrigerating coefficient of performance ϕ_{QDR} of a QDR with $\delta/\Delta=10^{-3}$ and that of an ideal Carnot refrigerator $[\phi_{\text{Carnot}}=T_0/(T-T_0)]$. The coefficient of performance is plotted vs the ratio of reservoir temperature T_0 to ambient temperature T. Note that ϕ_{QDR} exceeds 70% of ϕ_{Carnot} for high temperatures. For smaller δ/Δ, the QDR is expected to approach the Carnot limit even more closely.

and may present an advantage over other cryogenic thermoelectric refrigerators, such as the NIS refrigerator, discussed in Sec. IV below. Each electron removed from the Cu reservoir carries away $k_B T_0$ in heat; as a quasiparticle in the Al electrode, it can recombine with another quasiparticle to release a phonon with $2\Delta_{Al}$ in heat, where Δ_{Al} is the superconducting energy gap of Al. Δ_{Al} is ten times or greater than $k_B T_0$ so the refrigerating efficiency of the NIS refrigerator is very low. However, as we will see in Sec. IV B, these quasiparticles can travel a long way before recombining. The recombination heat can be deposited in a warmer or dirtier region of the Al—a heat sink—which is far away and thermally decoupled from the NIS refrigerator. Thus, the relatively large amount of work performed to remove a given amount of heat does not cause a disproportionate heat leak to the refrigerator.

Refrigerating efficiency would only be important in two situations. If the heat deposited in the electrodes or thermoelements could leak back into the reservoir, low efficiency would hurt refrigerating performance. This is probably not the case for either the QDR or the NIS refrigerator. If it is desired to cascade several refrigerators so that the reservoirs of the refrigerators in each stage cool the electrodes of the next stage, low efficiency requires an inordinately large cooling power of the outer stages. For this type of application, it would be appropriate to use high-cooling-power NIS refrigerators for the outer stages and high-efficiency QDR for inner, colder stages.

IV. NORMAL-SUPERCONDUCTOR TUNNELING REFRIGERATOR

In this section, we use the methods of Sec. III to analyze the refrigerating performance of the NIS refrigerator built by Martinis and his group NIST (Boulder). This device operates by using a normal metal-insulator-superconductor (NIS) tunneling junction to remove electrons above the Fermi level from the normal metal N.[10] This prototype NIS refrigerator cooled a Cu strip to 85 mK from an ambient temperature of 100 mK. Although the temperature reduction achieved so far is not large, we show in this section that a much greater temperature reduction should be possible with an optimized, thermally isolated NIS refrigerator.

The NIST group's refrigerator was very small and was fabricated on a solid substrate; the electrons in the Cu reservoir were strongly coupled to the ambient temperature. With a larger NIS refrigerator fabricated on a silicon nitride membrane, it should be possible to achieve temperatures below 100 mK from ambient temperatures in excess of 300 mK.[10] This refrigeration would be technologically useful because the most sensitive bolometers (radiation detectors) operate best below 100 mK.[28]

In Sec. IV A, we describe our NIS-refrigerator model, including an estimate of the NIS cooling power. Then in Sec. IV B, we discuss the heat leaks provided by phonon absorption, quasiparticle recombination, and thermal conduction through the silicon nitride membrane. We also estimate the temperature at which electron-electron scattering limits the NIS tunneling rate and the NIS cooling power in this nonequilibrium limit. In Sec. IV C, we summarize the NIS-refrigerator performance, comparing our model to the NIST group's experiment. In Sec. IV D, we calculate the performance of a similar device, the *SIS'IS* refrigerator built by Blamire *et al.*[9] The latter device provides a test of thermoelectric refrigeration in the nonequilibrium regime, for which electron-electron scattering limits the tunneling rate. We find that our theory agrees with the results from both the NIS and *SIS'IS* devices.

A. NIS refrigerator model

In a normal-insulating-superconducting (NIS) tunneling junction, heat flows from the normal electrode N to the superconducting electrode S if the bias voltage eV_b is less than but near the energy gap Δ_{Al} ($\Delta_{Al} - k_B T_0 < |eV_b| < \Delta_{Al}$) and the ambient temperature T is much less than the critical temperature T_c. For $T \ll T_c$, there are few thermally excited quasiparticles in S so the tunneling of quasiparticles in S to states in N may be neglected. However, with the above bias condition, there will be enough thermal excitations in N to provide an appreciable current of electrons and holes tunneling from near E_F in N to quasiparticle states in S. If the bias is too high ($|eV_b| > \Delta_{Al}$), N will be heated.

If N is biased negatively with respect to $S[-(\Delta_{Al} - k_B T_0) > eV_b > -\Delta_{Al}]$, the tunneling current will be composed of electrons tunneling from above E_F in N to electron quasiparticle states in S [cf. Fig. 6(b)]; if N is biased positively with respect to $S(\Delta_{Al} - k_B T_0 < eV_b < \Delta_{Al})$, holes below E_F in N will tunnel to hole quasiparticle states in S. As we discussed in Sec. II B, both of these processes remove heat from $N(\approx k_B T_0$ per electron or hole).

To model the NIS refrigerator, we use the configuration of Fig. 6(a). A circular silicon nitride membrane of radius r_2 is attached to a substrate which is at the ambient temperature T. The NIS refrigerator is fabricated in the center of the membrane and is confined to a circle of radius r_1. The NIS refrigerator contains a Cu thin film of thickness t, area A, and volume $V = tA$. Superconducting films are deposited on the Cu film and extend out to connect to an external circuit. One superconducting film (Al) connects to the Cu reservoir via a tunneling barrier so that an NIS tunneling junction is formed.

The other superconducting film (Pb) makes a metallic junction with the Cu film so that it injects electrons at the Fermi level of the Cu reservoir by Andreev reflection. In Andreev reflection, a hole at E_F in the normal metal combines with a Cooper pair in the superconductor, leaving an electron at E_F in the normal metal. We show this schematically as the injection of an electron at E_F in Fig. 6(b); there is a hole (not shown) traveling in the opposite direction as well. Andreev reflection is elastic so it does not affect heat flow.

The silicon-nitride membrane is used as a substrate for the NIS refrigerator because, due to its low thermal con-

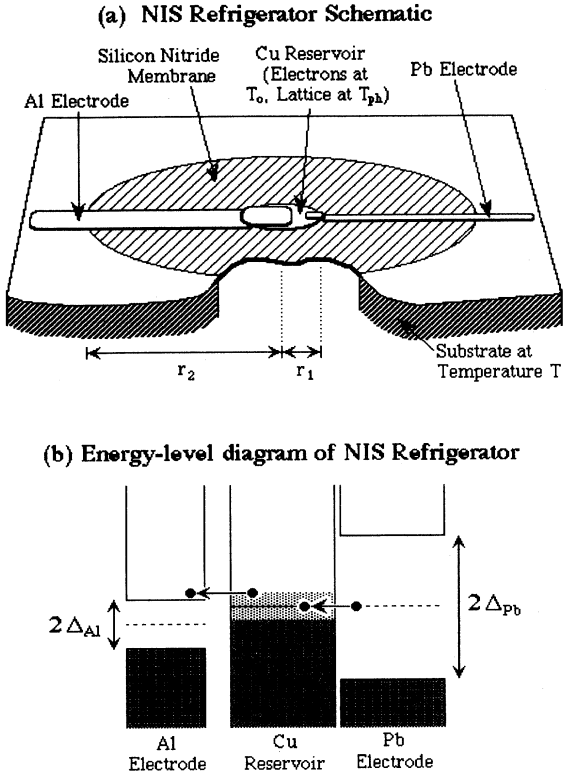

FIG. 6. (a) Schematic and (b) electronic energy diagram of the NIS refrigerator. A Cu thin film is cooled as hot electrons are removed by NIS tunneling to the Al electrode. The circuit is completed by a NS contact to the Pb electrode which injects electrons at E_F. The silicon nitride membrane provides thermal isolation for the NIS refrigerator.

ductance, it allows thermal isolation from the ambient temperature. We assume that the Cu film and the portion of the membrane upon which it is deposited are at T_{ph}, a temperature intermediate between T_0 and T. The goal of our NIS-refrigerator model is to find the highest ambient temperature T at which a given T_0 may be achieved.

The NIS junction is biased [as was the case for the NIST group's device; see Fig. 6(b)] so that E_F in the Cu reservoir is aligned in energy with the quasielectron states above the gap in the superconductor S. In this configuration, the cooling occurs as follows. Electrons are injected at E_F in the Cu reservoir by Andreev reflection from the Pb electrode. These electrons scatter with other electrons in the Cu reservoir; some absorb energy from this scattering and are promoted to states $k_B T_0$ above E_F, from which they can tunnel to the Al electrode. In this process, each electron traversing the entire structure removes about $k_B T_0$ of heat from the Cu reservoir.

The cooling power is approximately[10]

$$P=\left[\frac{G}{e^2}\right](k_B T_0)^2=0.29\ \text{pW}T_0^2 G\ , \tag{4.1}$$

where G is the conductance of the NIS junction in units of the conductance quantum $(e^2/h)=(26\ \text{k}\Omega)^{-1}$, and T_0 is in K. For the NIST group's device, $G=(10\ \text{k}\Omega)^{-1}=2.6(e^2/h)$.[10] Equation (4.1) may be obtained by noting that the cooling power is due to extracting electrons from above E_F in the Cu reservoir; on average, each electron removes $k_B T_0$ from the Cu reservoir. The tunneling rate is the product of the conductance per energy (G/e^2) and the energy range $k_B T_0$ of initial states. We arrive at Eq. (4.1) by noting that the cooling power is the product of the tunneling rate and the heat carried per electron.

An alternate design would use two NIS junctions, one to extract electrons from above E_F in the Cu, and one to extract holes from below E_F in the Cu, analogous to our generic cryogenic thermoelectric refrigerator of Sec. I. To extend our model to such an SINIS (SIN+NIS) refrigerator, we would double the cooling power in Eq. (4.1) to account for the combined effect of the two tunneling junctions. We do this in Sec. IV D to compare our theory to the *SIS'IS* devices of Blamire *et al.*[9]

B. NIS-refrigerator heat leaks and electron-electron-scattering limit

The electrons in the Cu reservoir are heated by absorbing phonons. This heat leak is[10]

$$P_{ph}=2\ \text{nW}V(T_{ph}^5-T_0^5)\ . \tag{4.2}$$

This phonon heating is of the same form as the 2DEG expression Eq. (3.1). However, it is much stronger (nW instead of fW) due to the thin (100 Å thick) 2DEG used in Eq. (3.1) and the stronger electron-phonon coupling in metals.

Another heat leak comes from lattice thermal conduction through the silicon nitride membrane. The thermal conductance of a silicon nitride membrane is $\kappa=162$ $(\mu\text{W/K})\ \text{cm}^{-1}\ T_m^{5/2}$, where T_m is the local membrane temperature in K.[10] Because the temperature varies radially on the membrane (from $T_m=T_{ph}$ at r_1 up to $T_m=T$ at r_2), we integrate radially to obtain the overall thermal conductance from the outer edge to the NIS refrigerator. The heat transfer equation to be integrated is[27]

$$\nabla\cdot(\kappa\nabla T_m)=0\ ,$$

whose solution gives a heat leak of

$$P_{cond}=(2\pi r_1 t)\kappa(T)\left.\frac{dT_m}{dr}\right|_{r=r_1}$$

$$=\frac{4\pi(162\ \mu\text{W})t(T^{7/2}-T_{ph}^{7/2})}{7\ \ln(r_2/r_1)}$$

$$=6.3\ \text{nW}(T^{7/2}-T_{ph}^{7/2})\ , \tag{4.3}$$

where we have assumed in the final expression a membrane thickness of $t=0.5\ \mu\text{m}$, an outer radius of $r_2=1$ cm, and an inner radius of $r_1=1$ mm.

There are two other possible heat leaks: thermal conduction through the Al and Pb films, and quasiparticle

recombination in the Al. We neglect these for the following reasons. First, the electronic thermal conductivity of a superconductor is very low.[1] The lattice thermal conductance of the Al and Pb thin films should be less than that of the silicon nitride membrane because these films have a much smaller cross section (<1 μm thick and 1 mm wide) than the membrane (0.5 μm thick and several cm wide). Thus we can neglect thermal conduction through the Pb and Al films.

Another potential heat leak is provided by quasiparticle recombination in the Al. When an electron is extracted from above E_F in the Cu reservoir, it enters an electron quasiparticle state in the superconducting Al electrode. This quasiparticle can combine with another quasiparticle to form a Cooper pair, emitting a phonon of energy $2\Delta_{\rm Al}$. If this occurs near the NIS junction, these hot (because $2\Delta_{\rm Al} \gg k_B T_0$) phonons can travel back to the Cu reservoir where they could be absorbed by electrons, causing a heat leak.

Quasiparticle scattering rates in the BCS superconductors were published in a comprehensive study by Kaplan *et al.*[29] Of the different scattering processes in their calculations, there are two which release a phonon of energy $2\Delta_{\rm Al}$. Equilibrium recombination of quasiparticles, according to their Eq. (14), occurs at a rate of $\tau_r^{-1}=80$ MHz $\sqrt{T}e^{-2/T}$ in Al. At 1 K, for instance, $\tau_r^{-1}=10$ MHz, giving a mean free path $v_F\tau$ of several cm; at lower temperatures, it is much greater. At 300 mK, $\tau_r^{-1}=60$ kHz, and the quasiparticles collide with the boundaries of the thin film far more often than they recombine. At 100 mK, this rate is negligible due to the exponential temperature dependence.

The second relevant quasiparticle scattering process described by Kaplan *et al.* is the recombination of injected quasiparticles, in this case those injected into quasiparticle states roughly $k_B T_0$ above the gap edge by the NIS tunneling. This process occurs in Al at a rate of 20 kHz at 300 mK and 500 Hz at 100 mK.

We now estimate the quasiparticle recombination heat leak at 300 mK. If electrons are injected into quasiparticle states in the Al at $k_B T_0$ above the gap edge, the quasiparticle decay rate (including both processes) is 80 kHz. Thus the average quasiparticle will travel 4000 cm before decaying. Or, in other words, only a fraction (1 cm/4000 cm) $= 3\times10^{-4}=0.03\%$ of the quasiparticles will decay near the NIS junction. Since each electron removed from the Cu reservoir carries away $k_B T_0$ of heat, and each quasiparticle recombination generates $\Delta_{\rm Al}\approx 10$–$20 k_B T_0$ of heat, this heat leak is less than 1% as strong as the NIS cooling power. If the action of the NIS refrigerator cools the membrane and the metal thin films, the quasiparticle heat leak will decrease dramatically.

Even though the quasiparticles are slow to decay, it is necessary to get rid of their energy; this must be done far from the Cu reservoir. If a normal metal contact is made to the Al at a point which is not in thermal contact with either the NIS refrigerator or the membrane, this electrode will serve as a heat sink; quasiparticles in the Al will be deposited as hot electrons in the normal metal. Overall, we expect that quasiparticle recombination can be avoided as a heat leak in the NIS refrigerator. We neglect quasiparticle recombination in our analysis of NIS refrigerator performance in Sec. IV C. One possible weakness of this approximation is enhanced quasiparticle recombination at the boundaries of the superconducting films.

If the tunneling rate is too high, the electrons in the Cu reservoir will be driven out of equilibrium; eventually the tunneling rate will be determined by the rate at which electron-electron scattering replenishes the tunneling channels. As we discussed in Sec. III B above, the rate at which a given electron near E_F in a clean metal scatters from other electrons is given by

$$\tau_{\rm FL}^{-1}=\frac{(k_B T_0)^2}{hE_F}\approx 1.3\ \text{GHz}T_0^2\ . \tag{3.4}$$

Disorder enhances this rate by a correction[21,30]

$$\tau_{\rm disorder}^{-1}=\left[hN_F\left[\frac{\hbar v_F l}{k_B T_0}\right]t\right]^{-1}=1100\ \text{Hz}\frac{T_0}{lt}\ , \tag{4.4}$$

where l is the elastic mean free path, t is the film thickness, and we have used the two-dimensional form for the disorder correction.[21] For the sheet resistance of the NIST group's Cu thin film,[10] we obtain an elastic mean free path of $l=30$ Å. As in the case of the QDR, we must multiply by the number of initial states to find the tunneling rate when limited by electron-electron scattering.

$$N_{\rm init}=(k_B T_0)VN_F=1.55\times10^6 T_0 V\ .$$

The tunneling rate as limited by both clean-metal and disorder-enhanced electron-electron scattering is then

$$\begin{aligned}\tau_{\rm ee}^{-1}&=N_{\rm init}(\tau_{\rm disorder}^{-1}+\tau_{\rm FL}^{-1})\\&=0.57\ \text{THz}T_0^2A+0.39\ \text{THz}T_0^3V\ .\end{aligned} \tag{4.5}$$

When the NIS tunneling rate is limited by electron-electron scattering, the NIS-refrigerator cooling power is

$$P_{\rm scat}=(k_B T_0)\tau_{\rm ee}^{-1}=7.9\ \text{pW}T_0^3A[1+0.684tT_0]\ . \tag{4.6}$$

Setting Eq. (4.6) equal to the NIS cooling power Eq. (4.1) gives a temperature

$$\begin{aligned}T_{\rm ee}=0.73K_t^{-1}[\sqrt{1+0.101t^2(G/V)}-1]&\approx 37\ \text{mK}t\left[\frac{G}{V}\right],\quad 0.101t^2\left[\frac{G}{V}\right]\ll 1\\&\approx 230\ \text{mK}\left[\frac{G}{V}\right]^{1/2},\quad 0.101t^2\left[\frac{G}{V}\right]\gg 1\ ,\end{aligned} \tag{4.7}$$

below which the NIS tunneling rate is limited by electron-electron scattering.

C. NIS-refrigerator performance

We can use the above results to find T_{ph} and T for a given T_0. At steady state, heat balance determines the relationships between these temperatures. That is, an equal amount of heat must flow through the NIS junction [via Eq. (4.1)]; from the phonons to the electrons in the Cu reservoir [via Eq. (4.2)]; and from the periphery of the silicon nitride membrane to its center [via Eq. (4.3)]. As we discussed above, we neglect the heat leak due to quasiparticle recombination in the Al electrode. Then heat balance in this thermal circuit gives

$$P=P_{ph}=P_{cond} .$$

This leads to two equations for the two unknown temperatures, T and T_{ph}:

$$\begin{aligned} T&=T_{ph}\{1+4.6\times10^{-5}GT_0^2T_{ph}^{-7/2}\}^{2/7} , \\ T_{ph}&=T_0\left\{1+0.000\,145T_0^{-3}\left[\frac{G}{V}\right]\right\}^{1/5} . \end{aligned} \tag{4.8}$$

Equations (4.8) apply to the case in which the electronic temperature T_0 exceeds T_{ee}. For lower temperatures, the electron-electron-scattering-limited rate P_{scat} from Eq. (4.6) must replace Eq. (4.1), modifying the heat-balance condition

$$P_{scat}=P_{ph}=P_{cond}$$

to give two equations once again:

$$\begin{aligned} T&=T_{ph}\{1+0.317VT_{ph}^{-7/2}[T_{ph}^5-T_0^5]\}^{2/7} , \\ T_{ph}&=T_0\{1+0.003\,95T_0^{-2}[t^{-1}+0.68T_0]\}^{1/5} . \end{aligned} \tag{4.9}$$

Note that these equations do not depend on the NIS tunneling conductance G, verifying that NIS tunneling is not the limiting factor.

For the NIST group's NIS refrigerator, $G=2.6(e^2/h)$, $V=0.4\ \mu m^3$, and $t=0.08\ \mu m$. The area of the NIS junction is about $(0.5\ \mu m)^2=0.25\ \mu m^2$, so that the conductance per junction area is $10(e^2/h)\ \mu m^{-2}$. For these parameters, Eq. (4.7) gives $T_{ee}=20$ mK. Because they operated above this temperature, the NIST group's NIS refrigerator was not limited by electron-electron scattering; the electrons in the Cu reservoir were near equilibrium. If $T_0=85$ mK, then from Eqs. (4.8) $T_{ph}=100$ mK. For these parameters, $T\approx T_{ph}$. This indicates that the NIST group's refrigerator was not capable of cooling the phonons as well as the electrons, so that the thermal isolation of the silicon nitride membrane is not adequate to help the refrigeration in this case. Our model thus agrees well with their results.

For different NIS refrigerator designs, however, the silicon nitride membrane will improve NIS-refrigerator performance. If the conductance of the NIS junction is increased (by increasing its area or decreasing the thickness of the tunneling barrier), the NIS cooling power Eq. (4.1) will increase. The heat leak due to thermal conduction through the silicon nitride membrane is essentially fixed; the NIS cooling power can be increased by orders of magnitude. Consider the following example.

Suppose that the Cu reservoir is $t=1\ \mu m$ thick, $A=1\ mm^2$ in area, and the tunneling conductance per area of the NIS junction is the same as that of the NIST group's refrigerator. If the junction covers most of the surface area of the Cu reservoir then the conductance is $G=10^7(e^2/h)$ and the Cu reservoir volume is $V=10^6\ \mu m^3$. With these parameters, $T_{ee}=300$ mK. At $T_0=50$ mK, then, the NIS tunneling current is electron-electron scattering limited and Eqs. (4.9) give $T_{ph}=61$ mK and $T=600$ mK. This NIS refrigerator can cool the electrons and the lattice of the Cu reservoir to well below 100 mK from an ambient temperature above 500 mK. This refrigeration is technologically interesting because temperatures down to 300 mK can be attained by pumping on ^{3}He. Lower temperatures currently require a dilution refrigerator.

The NIS cooling power for this NIS refrigerator is given by Eq. (4.6) to be 0.8 nW at $T_0=100$ mK. A 1-cm^2 NIS refrigerator (on a several-cm silicon nitride membrane) would provide about a μW of cooling power at 100 mK. This is 1% of the cooling power that the Oxford Instruments model 75 dilution refrigerator provides at the same temperature. This is an impressive amount of cooling power for a thin-film device.

If a material other than Al were chosen for S, the NIS refrigerator could operate at temperatures up to a few K. Consider the above NIS refrigerator with a 1-μm-thick reservoir which is 1 cm^2 in area. Then $V=10^8\ \mu m^3$ and $G=10^9(e^2/h)$. With these parameters, Eq. (4.7) gives $T_{ee}=300$ mK. For $T_0=300$ mK, this device is on the border between equilibrium and nonequilibrium; Eqs. (4.8) and (4.9) both give $T_{ph}=303$ mK and $T=11$ K. For these high temperatures, some assumptions of our model such as the low quasiparticle recombination rate may no longer be valid. However, it is plausible that the NIS refrigerator can operate in an ambient temperature of a few K and cool a thin-film device to 300 mK or below.

D. *SIS'IS* refrigerator: A test of the nonequilibrium limit

Blamire *et al.* fabricated $Nb/AlO_x/Al/AlO_x/Nb$ double-barrier devices, referred to as the *SIS'IS* structure, where $I=AlO_x$, $S=Nb$ ($T_c=9.3$ K), and $S'=Al$ (thin film $T_c=1-2$ K). The devices were operated at 4 K and, because of the extraction of hot quasiparticles, the Al was driven superconducting.[9] Although a realistic interpretation of these experiment requires that the superconducting properties of Al be incorporated, it is illuminating to apply our theory to this structure, treating the Al as a normal metal.

Repeating the above analysis with the parameters for Al,[1,4] we obtain

$$\begin{aligned} P&=0.29\ \text{pW}T_0^2G , \\ P_{ph}&=0.2\ \mu\text{W}V(T^5-T_0^5) , \\ P_{scat}&=12\ \text{pW}VT_0^2(t^{-1}+0.69T_0) . \end{aligned} \tag{4.10}$$

We have assumed an elastic mean free path of $l=30$ Å (we simply chose the smallest film thickness) and used the fact that there were two thermoelements in the *SIS'IS* structure (giving a factor of 2 in P). Setting $P=P_{\rm scat}$, we find an expression for the electronic temperature T_0 below which electron-electron scattering dominates the cooling power:

$$T_0=73\ \text{mK}\,t^{-1}[\sqrt{1+0.13tGA^{-1}}-1]\ . \tag{4.11}$$

For the *SIS'IS* devices, the tunneling conductance was in range $400-1000(e^2/h)\ \mu\text{m}^{-2}$ and the thickness of the Al layers was 3–20 nm. The junctions were conductive (40–100 times more so than the NIST group's) and thin (4–25 times thinner than the NIST group's). According to Eq. (4.11), the tunneling is electron-electron scattering limited for temperatures below $T_0=16-44$ K. These devices, tested at a few K, operated in extreme nonequilibrium, in agreement with Heslinga and Klapwijk's interpretation.[9]

Setting $P_{\rm scat}=P_{\rm ph}$, we obtain an expression for the electronic temperature T_0 in terms of the ambient temperature T:

$$T=T_0\{1+0.060T_0^{-3}(t^{-1}+0.69T_0)\}^{1/5}\ . \tag{4.12}$$

For the devices with the thinnest Al layers (3 nm), Eq. (4.12) gives $T=3$ K for $T_0=2.4$ K, the measured T_c of the Al film in the *SIS'IS* devices. Blamire *et al.* saw a greater temperature reduction than this; even the devices with thicker Al layers (up to 10 nm) were driven superconducting from 4 K.

There are several possible causes for this discrepancy. The actual mean free path may be shorter than 30 Å. It is also possible that the Al lattice was cooled by the electrons, reducing the phonon heat leak. Or perhaps thermoelectric cooling does not account for all of the T_c enhancement; nonequilibrium superconductivity may explain part of the effect. However, it does seem likely that thermoelectric cooling played a role; Heslinga and Klapwijk's simulations of the electronic distribution in the Al film also indicate that the electrons were cooled substantially.[9]

The agreement of our calculations with both the equilibrium (the NIST group's NIS refrigerator) and nonequilibrium (the *SIS'IS* structures of Blamire *et al.*) limits lends confidence in the validity of our arguments in Sec. I on the general properties of a cryogenic thermoelectric refrigerator.

V. DISCUSSION

In this section, we discuss several aspects of cryogenic thermoelectric refrigeration. In Sec. V A, we discuss the cooling power of a thermoelectric refrigerator in relation to that needed for useful cryogenic applications. Then in Sec. V B we review several other schemes for thermoelectric cooling at low temperatures. In Sec. V C we discuss a thermodynamic limit to the temperature dependence of the cooling power of an electronic refrigerator; we find that an optimized thermoelectric refrigerator operates near this limit, consistent with the finding of Sec. III C that a QDR may operate with an efficiency close to that of a Carnot refrigerator.

A. Cooling power of cryogenic thermoelectric refrigerators

Electronic refrigeration schemes such as the QDR and the NIS refrigerator may be accused of providing meager cooling power without much temperature reduction. This criticism has indeed been leveled at cryogenic thermoelectric refrigeration recently, in an article by Washburn.[31] He makes several points: that the temperature reduction so far achieved is small; that the ambient temperature at which it is accomplished is too low to be of any use; and that the cooling power is so small that it will be impossible to cool bulk samples.

The temperature reduction which the NIST group achieved with their prototype refrigerator was small—only 15% of the absolute temperature. However, their experiment was aimed at a proof of the concept rather than an exhaustive test of the method. As we have shown, a NIS refrigerator with an optimized design can accomplish the technologically important goal of bulk cooling from above 300 mK to below 100 mK. Furthermore, the *SIS'IS* devices of Blamire *et al.* achieved as large a temperature reduction from ambient temperatures as high as 4 K. Thus it is fair to say that cryogenic refrigeration by thermoelectric effects is capable of large temperature reduction at useful temperatures.

The total cooling power of these electronic refrigerators is smaller than that of a dilution refrigerator. Thus NIS refrigeration will not replace other refrigeration techniques in applications such as liquid-helium studies, where a large sample volume is required. However, in fields such as high-precision radiation detectors, quantum computing, or any based upon thin-film devices, a compact solid-state refrigerator may be preferable to bulk techniques. In addition, direct electronic cooling may be preferable to bulk cooling for some applications.

The cooling power of the QDR example discussed in Sec. III E was a few aW for refrigeration in the tens of mK. This is orders of magnitude smaller than the cooling power of a dilution refrigerator at the same temperature; the latter is measured in μW. The difference is that, at mK temperatures, the QDR cools only a microscopic portion of a 2DEG. At lower temperatures—below the operating regime of a dilution refrigerator—a QDR should be capable of cooling the electrons of macroscopic samples. In fact, as can be seen from a simple calculation,[1] the 2DEG heat capacity exceeds that of a mm-thick GaAs substrate for temperatures below 200 μK. Thus for μK and nK temperatures, a QDR cools the largest heat capacity of the solid. With strong enough thermal coupling between the electrons and the GaAs crystal lattice (and weak enough thermal coupling to ambient-temperature components), bulk refrigeration would become possible.

B. Other schemes for cryogenic thermoelectric cooling

In the 1950s and 1960s, there was an extensive research effort devoted to large-scale cooling at noncryogenic temperatures using semiconductor thermoelectric refrigerators (also known as Peltier refrigerators).[2] Aside from

small-scale (e.g., diode-laser cooling) and specialty (e.g., portable cooling of biological specimens) applications, this effort was largely dropped when these refrigerators were found to be inefficient. This inefficiency was due to the large lattice heat leak at noncryogenic temperatures.

These semiconductor thermoelectric refrigerators operated on the following principle. A semiconductor with a small band gap is heavily doped so that hot electrons (holes) above (below) the Fermi level of a metallic reservoir would travel into the conduction (valence) band of the n-type (p-type) semiconductor thermoelement. In theory, this scheme should be possible at cryogenic temperatures. However, there seem to be practical limitations such as finding a stable material with a small enough band gap.[32]

When we first proposed the QDR,[5] we considered (but did not publish) a bulk cooling scheme which would use band-engineered III-V heterostructures. There are many ways of using modulation doping to construct a thermoelectric refrigerator; any device structure which allows the extraction of electrons above E_F and holes below E_F can function in this way. Due to the fast pace of this technology, particularly type-II band-offset heterostructures, this type of device may be feasible in the near future. It should be possible to conceive a structure using this technology which allows a large cooling power by tunneling through a large-area junction or even by ballistic transport across a modulation-doped band edge. The small effective electron mass and low lattice heat capacity make these III-V structures excellent candidates for cryogenic refrigeration of bulk samples.

This idea of using a system of reduced dimensionality to increase the thermoelectric figure of merit ZT has been explored from a different direction by Dresselhaus and co-workers at MIT. They have considered confining carriers in Bi_2Te_3 with a two-dimensional superlattice;[33] confining carriers in Bi_2Te_3 with a one-dimensional superlattice;[34] and the use of two-dimensional superlattices of unconventional thermoelectric materials (they examined the case of Bi) for thermoelements.[35] These studies resulted in predicting that quantum confinement could increase the maximum attainable dimensionless figure of merit ZT from the current value of about 1 to a value of 10 or more. As discussed in the Appendix below, this would allow large temperature reduction—up to a factor of $(T/T_0)=3$. Because they relied on the case of one or at most two electron subbands, this refrigerating scheme seems most likely at low temperatures, although confinement in small enough structures may allow room-temperature operation.

Among the more exotic proposals we found was Kapitulnik's idea of using a metal near a metal-insulator transition as a thermoelement.[36] Kapitulnik computed the thermoelectric figure of merit (see the Appendix) for this material, finding that it can become larger near the transition. This idea is reminiscent of the proposal to use a semimetal as a thermoelement.[2]

C. Thermodynamic limit of electronic refrigeration

In this section we argue that the cooling power of an optimized thermoelectric refrigerator is the best possible for an electronic refrigerator. Specifically, an electronic refrigerator whose cooling power falls more slowly than T^2 with decreasing temperature can be used to achieve $T=0$, a thermodynamic impossibility.

The internal energy U of an electron gas is proportional to the square of the electron temperature T:[1]

$$U=aT^2 .$$

As discussed in Sec. I an optimized cryogenic thermoelectric refrigerator has a cooling power proportional to T^2, as well:

$$P=bT^2 .$$

The identification $dU/dt=-P$ gives the differential equation

$$\frac{dU}{dt}=-\left[\frac{b}{a}\right]U ,$$

with the solution

$$U=U_0e^{-(b/a)t} .$$

U (and hence T) decreases exponentially with time for an optimized thermoelectric refrigerator. Suppose now that $P=aT^n$. If $n>2$, the solution is

$$U\propto t^{-2/(n-2)} .$$

U decreases as a power law of time. This is much slower than the exponential decrease for T^2 cooling. If $n<2$, the solution is

$$U\propto (t_f-t)^{2/(2-n)}$$

so that, after a time t_f, the energy and hence the temperature reach zero. Since it is impossible to use any thermodynamic process to reach absolute zero,[27] this type of cooling is impossible.

For a nonoptimized thermoelectric refrigerator, it is possible to have a cooling power which falls off more slowly than T^2 with decreasing temperature over some temperature ranges. For instance, if we relax the requirement $\delta\approx k_BT_0$ which we used in Sec. III to determine the performance of the QDR, then the cooling power Eq. (2.8) indicates that the cooling power is proportional to T. However, this is only because we have fixed δ at some smaller value than k_BT_0, and the cooling power is less than optimal; for optimal cooling, $\delta\approx k_BT_0$ and the cooling power is again proportional to T_0^2.

Another way of viewing this thermodynamic argument is to consider the coefficient of performance of an electronic refrigerator. As we showed in Sec. III F for the QDR, an optimized thermoelectric refrigerator approaches a value near the Carnot limit for refrigerating efficiency.

VI. CONCLUSIONS

Thermoelectric effects present an unexplored avenue of cryogenic refrigeration. Due to the small lattice heat capacity at low temperatures, an electronic refrigerator can cool bulk samples to low temperatures. We have ex-

amined the effects of phonon absorption and electron-electron scattering on two cryogenic refrigerators: the QDR and the NIS refrigerator. We found that phonon absorption limits the base temperature. Electron-electron scattering limits the cooling rate at the lowest temperatures, but the best refrigeration is obtained in this nonequilibrium regime.

Each of these refrigerators has its own characteristic temperature regime. The QDR can operate over a wide temperature range. In the mK, μm-sized mesoscopic components can be cooled by a QDR; in the μK and below, a cm-sized QDR can provide bulk refrigeration in that the 2DEG electronic heat capacity is greater than that of the semiconductor substrate. An experimental device could be thermally sunk to the reservoir and thus cooled.

NIS-based refrigerators operate from a few K down to tens of mK. The top limit of this range is set by the T_c of the superconducting thermoelement and the thermal isolation of the reservoir. The lower limit is set by the energy smearing of the superconducting gap edge.[10] This lower limit could be decreased by using lower-T_c superconductor since the smearing should scale with T_c. We have shown that the devices of the NIST group and Blamire *et al.* provide experimental tests of our theory in the equilibrium and nonequilibrium limits, respectively; the comparison is favorable. By scaling these microscopic electronic refrigerators up to cm size scales, a cooling power large enough to cool macroscopic thin-film devices could be obtained.

It is conceivable that these refrigerators could be cascaded. For instance, a NIS refrigerator could operate at an ambient temperature of about 10 mK in a dilution refrigerator, cooling the leads of QDR to a few hundred μK. The QDR would cool its reservoir and an attached heat load to a few μK or even nK. Or, for a detector application, a first NIS refrigerator could cool a membrane from 1.2 K (a pumped ^{4}He bath) to 300 mK, and a second NIS refrigerator would cool a bolometer to below 100 mK.

By allowing access to temperatures which have not been explored for metallic samples and mesoscopic devices, additional physics will surely be uncovered. In addition, compact solid-state refrigerators may allow cryogenic components to be used in various technological applications.

ACKNOWLEDGMENTS

We would like to acknowledge helpful discussions with Jonathan Cobb, Ertugrul Demircan, Leo Kouwenhoven, T. M. Klapwijk, John Markert, John Martinis, Michael Nahum, Scott Rubel, and Ken Shih, and to our other colleagues at The University of Texas. We would also like to express our gratitude to Jonathan Cobb, Scott Rubel, and Ertugrul Demircan for carefully reading our manuscript. This work was supported in part by grants from NIST, the Welch Foundation, and the Texas Advanced Research Program.

APPENDIX

In this appendix, we will briefly summarize the linear-response theory of thermoelectric effects and apply this formalism to the QDR. Our motivation for doing so is twofold. First, we wish to compare our theory to the linear-response results of Ref. 19. Second, we want to explore the effects of the departure of the tunneling current and cooling power from linear dependence on the temperature difference $\Delta T = T - T_0$. In other words, we will demonstrate that the linear-response theory is not capable of accurately predicting the QDR base temperature.

As we mentioned in Sec. II D, the general approach to determining the base temperature of a thermoelectric refrigerator is as follows. The base temperature is found by setting the net cooling power to zero, so that the thermoelectric cooling power just balances the heat leaks. In the case of the numerical simulation of Sec. II D, this procedure was carried out by an iterative procedure to find the root of the equation $P(k_B T_0) = 0$. In the linear-response theory, the base temperature is found by setting the coefficient of performance, which is proportional to the net thermoelectric cooling power, to zero. Thus, apart from the linear-response approximation (and the physical processes discussed in Sec. 7 of this Appendix), the two procedures are equivalent.

1. Thermoelectric effects

Thermoelectric effects arise at a junction between two dissimilar materials. The Seebeck effect occurs when a temperature difference is placed across the junction. If no current flows, a thermoelectric voltage $V = \alpha(-\Delta T)$ develops; α is called the Seebeck coefficient or the thermopower. In the Peltier effect, an electric current I is forced through a junction at (initially) uniform temperature. A heat current $P = \pi I$ flows, heating one metal and cooling the other; π is the Peltier coefficient. In some instances, the Peltier coefficient may be related to the Seebeck coefficient via Onsager's reciprocal relations,[37] to obtain $\pi = T\alpha$.

As electrons traverse a junction of dissimilar materials, they will either lose or gain energy on average because the set of electronic states they occupy will generally have different energies in the two materials. This is the simple physical principle which underlies all thermoelectric effects. To cool the reservoir of a thermoelectric refrigerator, the average energy of electrons entering the reservoir must be lower than that of the electrons leaving the reservoir. Because each electron carries away energy on average, heat is removed from the reservoir. This is how the QDR works.

None of the currently available thermoelectric materials are effective at cryogenic refrigeration. The Peltier effect in normal metals is too weak to provide effective cooling at any temperature.[2] Semiconductor thermoelements, whose Fermi level lies within the band gap, suffer carrier freeze-out at low temperatures.

2. Coefficient of performance, base temperature, and figure of merit

Conventional thermoelectric refrigerators cool their reservoir by up to about 20%; ΔT is small relative to T. Since most of the voltage drop occurs resistively across the semiconductor thermoelements, the overall device has a fairly linear $I(V)$ curve. For conventional thermoelectric refrigerators, it is thus appropriate to assume that the electric current and cooling power depend linearly upon the bias V and temperature difference ΔT. For this reason, the theory which is conventionally used to study thermoelectric effects assumes this linear dependence.

The starting point of the linear-response analysis is the refrigerating coefficient of performance ϕ, defined to be the ratio of the net cooling power of the refrigerator to the rate at which work is performed. After maximizing with respect to electric current (in the process fixing the bias voltage), the coefficient of performance for a thermoelectric refrigerator is[2]

$$\phi=\frac{T_0}{T-T_0}\frac{M-T/T_0}{M+1}\ , \tag{A1}$$

where $M=[1+Z(T+T_0)/2]^{1/2}$ and the thermoelectric figure of merit Z is:

$$Z=\frac{\alpha^2}{R\kappa}\ , \tag{A2}$$

which has units of K^{-1}. R is the junction resistance and κ is its thermal conductance. The dimensionless figure of merit ZT, where T is the ambient temperature, is often quoted for thermoelectric devices. Values of ZT for semiconductor thermoelectric materials are generally near unity; a larger value would produce better performance, such as a lower base temperature. Note that the first factor in (A1) is the coefficient of performance of a Carnot refrigerator,[27] and the second factor is between zero and unity, being closer to unity for large ZT and $T_0\approx T$.

The maximum temperature difference $T-T_0=\Delta T_{\max}$ that the refrigerator can achieve (under zero heat load) is obtained from the condition $\phi=0$ that the net cooling power vanish. The result is $\Delta T_{\max}=(1/2)ZT_0^2$. Rearranging,

$$\left[\frac{T}{T_0}\right]_{\max}=\tfrac{1}{2}[1+\sqrt{1+2ZT}\,]\ , \tag{A3}$$

which approaches $1+ZT/2$ for $ZT\ll 1$ and $(ZT/2)^{1/2}$ for $ZT\gg 1$.

Aside from its appearance in the above equations, there is an intuitive motivation for calculating ZT. A large thermopower α implies a large Peltier coefficient $\pi=T\alpha$, hence a large cooling power. Low electrical resistance leads to low resistive heating near the junction and low thermal conductance allows a large temperature gradient to be maintained across the thermoelectric junction. Thus a large ZT should indicate good thermoelectric performance. This is clear for semiconductor thermoelectrics, but what part do resistive losses and thermal conductance play in the QDR? Resistive losses should occur in the QDR structure: the 2DEG has a resistivity, and electrons and holes injected far from the Fermi energies μ_L and μ_R in the electrodes will scatter to thermalize with the 2DEG there, heating the electrodes. However, it is unlikely that this heat current will stream back into the reservoir of the QDR, since the electrodes should be much larger than the reservoir and can conduct the heat current away—we assume that the electrodes stay at the ambient temperature. This thermal isolation was found to be the case for the Coulomb-blockade electrometer, in which the electrodes remained at the ambient temperature even though the Coulomb island was heated to a significantly higher temperature.[4] Below, we will use the expression $(dI/dV)^{-1}$ as the resistance of the QDR to set an upper bound on the resistive heating of the QDR. This renders the analysis conservative; we will see that it is, in fact, very conservative.

A thermal conductance is related to a heat current via Fourier's law $P=\kappa(-\Delta T)$.[27] In semiconductor thermoelectric refrigerators, the appropriate thermal conductance to use in Eq. (A2) is the semiconductor lattice thermal conductance, since this dominates over the electronic conduction of heat at room temperature. For the QDR, however, we will define an electronic thermal conductance κ_e with reference to the heat current which flows through the quantum dots (when no electric current does) under a temperature difference $\Delta T=T-T_0$. The phonon heat leak could be incorporated by adding $\kappa_{\rm ph}\equiv P_{\rm ph}/(-\Delta T)$. Because the goal of this section is to compare the linear-response base temperature to the results obtained in the numerical simulation of Sec. II D, which did not include the phonon heat leak, we neglect $\kappa_{\rm ph}$.

3. Linear-response theory

Thermoelectric properties of a system are intrinsically linked to the system's coupling to the external world; the thermoelectric coefficients are defined differently when different pairs of $\{I,P,V,\Delta T\}$ are chosen as independent variables. Our treatment in this section is similar to that in Chap. 13 of Ashcroft and Mermin.[1] It is easiest to calculate the heat current P and the electrical current I in terms of an applied voltage V and temperature difference ΔT. First, we assume that these quantities have been calculated from a model of the system under study:

$$I=I(V,-\Delta T)\ , \tag{A4}$$

$$P=P(V,-\Delta T)\ . \tag{A5}$$

Next we take the linear-response approximation to these equations:

$$I=\left[\frac{\partial I}{\partial V}\right]V+\left[\frac{\partial I}{\partial(-\Delta T)}\right](-\Delta T)\ , \tag{A6}$$

$$P=\left[\frac{\partial P}{\partial V}\right]V+\left[\frac{\partial P}{\partial(-\Delta T)}\right](-\Delta T)\ , \tag{A7}$$

where $(\partial I/\partial V)$ implies $\lim_{V\to 0}(\partial I/\partial V)_{\Delta T=0}$. We recast

Eqs. (A6) and (A7) in terms of the independent variables I and ΔT:

$$V=\left[\frac{\partial I}{\partial V}\right]^{-1} I-\left[\frac{\partial I}{\partial(-\Delta T)}\right]\left[\frac{\partial I}{\partial V}\right]^{-1}(-\Delta T)\ , \tag{A8}$$

$$P=\left[\frac{\partial P}{\partial V}\right]\left[\frac{\partial I}{\partial V}\right]^{-1} I+\kappa_e(-\Delta T)\ , \tag{A9}$$

where the electronic thermal conductance is

$$\kappa_e=\left[\frac{\partial P}{\partial(-\Delta T)}\right]-\left[\frac{\partial P}{\partial V}\right]\left[\frac{\partial I}{\partial(-\Delta T)}\right]\left[\frac{\partial I}{\partial V}\right]^{-1}.$$

These relations can be used to compute thermoelectric parameters for an electronic device.

We now obtain expressions for the thermoelectric coefficients and the junction resistance in the linear-response approximation. The Seebeck coefficient α is defined by $V=\alpha\Delta T$ when no current flows, resulting in

$$\alpha=-\left[\frac{\partial I}{\partial(-\Delta T)}\right]\left[\frac{\partial I}{\partial V}\right]^{-1}. \tag{A10}$$

The Peltier coefficient is defined by $P=\pi I$ when ΔT is zero:

$$\pi=\left[\frac{\partial P}{\partial V}\right]\left[\frac{\partial I}{\partial V}\right]^{-1}. \tag{A11}$$

The electric resistance R is $\partial V/\partial I$:

$$R=\left[\frac{\partial I}{\partial V}\right]^{-1}. \tag{A12}$$

Using these results, the dimensionless figure of merit [defined in Eq. (A2)] becomes

$$ZT=\frac{T\left[\frac{\partial I}{\partial(-\Delta T)}\right]^2}{\left[\frac{\partial P}{\partial(-\Delta T)}\right]\left[\frac{\partial I}{\partial V}\right]-\left[\frac{\partial P}{\partial V}\right]\left[\frac{\partial I}{\partial(-\Delta T)}\right]}. \tag{A13}$$

ZT may be used in Eq. (A3) to estimate the base temperature of a thermoelectric refrigerator.

The validity of these expressions is limited to the regime in which Eqs. (A4) and (A5) are linear in their dependence on bias voltage and temperature difference. For standard thermoelectric refrigerators, this assumption is good, as discussed above. However, the QDR should be capable to reducing the temperature T_0 to a small fraction of T. In this case, the linear-response theory breaks down. Bearing this in mind, we compute the linear-response approximations to the QDR tunneling the heat currents.

4. Linear approximations to QDR tunneling current and cooling power

To compute the thermopower and Peltier coefficients properly, we must combine the contributions of the R-to-V_R and R-to-V_L junctions. Because Eq. (2.5) includes the contributions of both junctions, this has implicitly been done. To separate the effects of each junction, simply repeat the following analysis using $P/2$ where P is given by Eq. (2.5). The linear-response approximations to Eqs. (2.4) and (2.5) are

$$I=\left[2\frac{e^2}{h}I_0\right]V_b+\left[2\frac{k_Be}{h}I_1\right](-\Delta T)\ , \tag{A14}$$

$$P=\left[4T\frac{k_Be}{h}I_1\right]V_b+\left[4T\frac{k_B^2}{h}I_2\right](-\Delta T)\ , \tag{A15}$$

where we have defined the dimensionless integral

$$I_n=\frac{1}{4}\int\frac{dE}{k_BT}\sum_i\left[1+\frac{(E-E_1)^2}{\delta^2}\right]^{-1}\left[\frac{E}{k_BT}\right]^n \times\operatorname{sech}^2\left[\frac{1}{2}\frac{E}{k_BT}\right]. \tag{A16}$$

The limits of integration are $\pm\infty$ for mathematical convenience; the integrand is small except near the Fermi level in R.

5. Dimensionless figure of merit for QDR

We can now write down the linear-response approximations for the QDR thermoelectric parameters using the definitions of Sec. 3 above:

$$\kappa_e=4k_B\frac{k_BT}{h}\frac{I_0I_2-I_1^2}{I_0}\ , \tag{A17}$$

$$R=\frac{1}{2}\frac{h}{e^2}\frac{1}{I_0}\ , \tag{A18}$$

$$\alpha=2\frac{k_B}{e}\frac{I_1}{I_0}\ , \tag{A19}$$

$$\pi=2\frac{k_B}{e}T\frac{I_1}{I_0}\ , \tag{A20}$$

$$\kappa_eR=2T\left[\frac{k_B}{e}\right]^2\frac{I_0I_2-I_1^2}{I_0}\ , \tag{A21}$$

$$ZT=\frac{1}{2}\frac{I_1^2}{I_0I_2-I_1^2}\ . \tag{A22}$$

Note that Eqs. (A19) and (A20) obey Onsager's reciprocity relation $\pi=T\alpha$.[2] This is because we have not included the Coulomb-blockade effect in our model of the QDR. This neglect is justified because the QDR operates in the quantum regime, with k_BT_0 and eV_b much less than Δ and $e^2/2C$. Because Onsager's reciprocity relation is defined only in the linear-response limit, we cannot test it with the numerical simulation of Sec. II D.

Equations (A18) and (A19) can be compared to the results of Ref. 19, where it is assumed that $\delta\ll k_BT$, in which case Eq. (A16) simplifies to

$$I_n=\frac{\pi}{4}\frac{\delta}{k_BT}\left[\frac{\varepsilon}{k_BT}\right]^n\operatorname{sech}^2\left[\frac{1}{2}\frac{\varepsilon}{k_BT}\right]. \tag{A16'}$$

TABLE II. Results of a numerical determination of the QDR base temperature using the linear-response theory. These numbers represent an optimized QDR. ZT is found using Eq. (A22) and then ε and T are varied until a maximum ZT is obtained. Then we find the base temperature (T/T_0) using Eq. (A3). Comparing this base temperature to the one in Table I demonstrates the inaccuracy of the linear-response theory.

δ/Δ	ZT	(T/T_0)	k_BT_0/Δ	k_BT/Δ	ε/Δ	ε/k_BT_0	k_BT_0/δ
10^{-1}	0.236	1.11	0.074 0	0.0819	0.210	2.84	0.740
10^{-2}	3.89	1.98	0.037 4	0.0741	0.165	4.41	3.74
10^{-3}	36.7	4.81	0.013 1	0.0630	0.139	10.6	13.1
10^{-4}	329	13.3	0.004 07	0.0542	0.122	30.0	40.7
10^{-5}	2960	39.0	0.001 22	0.0476	0.108	88.5	122
10^{-6}	26 800	116	0.000 365	0.0425	0.0969	265	365

We have used a definition of the Dirac δ function and have included only one quantum-dot energy level at ε above the Fermi level μ_0 of reservoir R. Because Ref. 19 quotes results for a single quantum dot, Eq. (A18) must be halved and Eq. (A19) doubled to obtain

$$R=\frac{2}{\pi}\frac{h}{e^2}\frac{k_BT}{\delta}\cosh^2\left[\frac{1}{2}\frac{\varepsilon}{k_BT}\right], \tag{A18'}$$

$$\alpha=\frac{\varepsilon}{eT}. \tag{A19'}$$

With the correspondence between our notation and that of Ref. 19, $\varepsilon \Longleftrightarrow \Delta_{\min}$, $\alpha \Longleftrightarrow S$, $R \Longleftrightarrow G^{-1}$, and $\delta \Longleftrightarrow h\Gamma$; we see that our Eq. (A18') for the tunneling resistance agrees with their Eq. (5.12) for the tunneling conductance, and our Eq. (A19') for the thermopower agrees with their Eq. (5.10).

6. Linear-response base temperature for QDR

The base temperature can be obtained from the dimensionless thermoelectric figure of merit ZT by using Eq. (A3). In this section, we compute the optimum QDR base temperature and tabulate it as a function of δ. As given by Eq. (A22), ZT is a function of δ, ε, and k_BT. eV_b does not enter as it did in the numerical simulation of Sec. II D because the optimum value is already obtained in solving for Eq. (A3). ZT exhibits a maximum $(ZT)_{\max}$ with respect to ε and k_BT.

We computed the numerical integrals in Eq. (A22) using six energy levels for each quantum dot (as in Sec. II D), varying ε and k_BT to obtain $(ZT)_{\max}$. Then we found the base temperature using Eq. (A3). We present these results in Table II. $(ZT)_{\max}$ increases quickly as δ decreases; the values of ZT achieved by the QDR are much greater than 1. This is in accord with the results of Sec. II D, which indicate that the QDR is capable of reducing the temperature by a large factor. However, despite this qualitative agreement, the linear-response calculation and numerical simulation of the QDR base temperature disagree quantitatively.

By comparing the values for (T/T_0) in Tables I and II, we see this sharp disagreement. The linear-response theory clearly underestimates the temperature reduction, and the comparison becomes worse as the temperature reduction increases. This inaccuracy is a result of the fact that the QDR operates best when its response to an applied temperature difference is far from linear. The QDR thus serves as an example of a system for which the standard theory of thermoelectrics does not produce accurate results. Care must be taken in calculating the properties of a thermoelectric refrigerator when the temperature is reduced to a small fraction of the ambient temperature.

7. Physical origins of inaccuracy of linear-response theory

Aside from the nonlinear dependence of I and P on $\Delta T=T-T_0$, there is another, more subtle reason that the linear-response theory of thermoelectrics does not predict the correct base temperature for the QDR. In the linear-response theory of Ref. 2, Ohmic losses in the semiconductor thermoelements give rise to heating of the reservoir. This makes the net cooling power depend nonlinearly on current, as I^2R. This nonlinearity is what causes the base temperature to be finite even in the absence of external heat leaks; without Ohmic losses, the linear-response theory predicts that the base temperature of a thermoelectric refrigerator is always zero. However, the QDR does not suffer from Ohmic losses in this way. While Ohmic losses should occur in the 2DEG, the primary heat leak (aside from phonon absorption) in the QDR is the electronic thermal conductance of the quantum dots. Thus the derivation of the base temperature Eq. (A3) in terms of the dimensionless figure of merit ZT does not apply to the QDR.

We have included the linear-response analysis in this paper in order to demonstrate that, when analyzing the properties of a thermoelectric refrigerator, care must be taken to identify the true heating mechanisms. It does not suffice to quote a high thermopower or even a large thermoelectric figure of merit to prove that a device will be effective as a thermoelectric refrigerator. It is necessary to perform a numerical simulation such as we have done in Sec. II D, or, better, to build a real device and measure its base temperature.

Note added. It has just come to our attention that a device similar to the QDR, in which a quantum dot is cooled rather than a bulk 2DEG, has just been demonstrated [Leo Kouwenhoven, Science **268**, 1440 (1995).

*Present address: MS 147, P. O. Box 655936, Texas Instruments, Inc., Dallas, TX 75265.

[1]N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976).

[2]H. J. Goldsmid, *Thermoelectric Refrigeration* (Plenum, New York, 1964).

[3]M. L. Roukes, M. R. Freeman, R. S. Germain, R. C. Richardson, and M. B. Ketchen, Phys. Rev. Lett. **55**, 422 (1985); F. C. Wellstood, C. Urbina, and J. Clarke, Appl. Phys. Lett. **54**, 2599 (1989).

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

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

[6]R. H. Parmenter, Phys. Rev. Lett. **7**, 274 (1961).

[7]K. E. Gray, Solid State Commun. **26**, 633 (1978); C. C. Chi and J. Clarke, Phys. Rev. B **20**, 4465 (1979).

[8]R. G. Melton, J. L. Paterson, and S. B. Kaplan, Phys. Rev. B **21**, 1858 (1981).

[9]M. G. Blamire, E. C. G. Kirk, J. E. Evetts, and T. M. Klapwijk, Phys. Rev. Lett. **66**, 220 (1991); D. R. Heslinga and T. M. Klapwijk, Phys. Rev. B **47**, 5157 (1993).

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

[11]E. B. Foxman, P. L. McEuen, U. Meirav, N. S. Wingreen, Y. Meir, P. A. Belk, M. A. Kastner, and S. J. Wind, Phys. Rev. B **47**, 10 020 (1993).

[12]F. Pobell, Phys. Today **46**(1), 34 (1993).

[13]See, for instance, P. L. McEuen, E. B. Foxman, U. Meirav, M. A. Kastner, Y. Meir, N. S. Wingreen, and S. J. Wind, Phys. Rev. Lett. **66**, 1926 (1991); A. T. Johnson, L. P. Kouwenhoven, W. de Jong, N. C. van der Vaart, C. J. P. M. Harmans, and C. T. Foxon, *ibid.* **69**, 1592 (1992); Y. Wang and S. Y. Chou, Appl. Phys. Lett. **63**, 2257 (1993).

[14]E. L. Wolf, *Principles of Electron Tunneling Spectroscopy* (Oxford University Press, New York, 1989).

[15]C. W. J. Beenakker and H. van Houten, in *Solid State Physics*, edited by H. Ehrenreich and D. Turnbull (Academic, San Diego, 1991), Vol. 44; W. Xue and P. A. Lee, Phys. Rev. B **38**, 3913 (1988).

[16]R. Landauer, Philos. Mag. **21**, 863 (1970).

[17]L. W. Molenkamp, H. van Houten, C. W. J. Beenakker, R. Eppenga, and C. T. Foxon, Phys. Rev. Lett. **65**, 1052 (1990); L. W. Molenkamp, Th. Gravier, H. van Houten, O. J. A. Buijk, M. A. A. Mabesoone, and C. T. Foxon, *ibid.* **68**, 3765 (1992).

[18]A. A. Staring, L. W. Molenkamp, B. W. Alphenaar, H. van Houten, O. J. A. Buyk, M. A. A. Mabesoone, C. W. J. Beenakker, and C. T. Foxon, Europhys. Lett. **22**, 57 (1993); A. S. Dzurak, C. G. Smith, M. Pepper, D. A. Ritchie, J. E. F. Frost, G. A. C. Jones, and D. G. Hasko, Solid State Commun. **87**, 1145 (1993).

[19]C. W. J. Beenakker and A. A. M. Staring, Phys. Rev. B **46**, 9667 (1992).

[20]M. Amman, E. Ben-Jacob, and J. Cohn, Phys. Lett. A **171**, 389 (1992).

[21]C. Liu and Q. Niu, Phys. Rev. B **48**, 18 322 (1993); C. Liu, Ph.D. thesis, University of Texas, 1993; Chiu Liu and Qian Niu, Phys. Rep. (unpublished).

[22]C. W. J. Beenakker and H. Van Houten, in *Solid State Physics* (Ref. 15).

[23]Christopher Hodges, Henrik Smith, and J. W. Wilkins, Phys. Rev. B **4**, 302 (1971).

[24]B. L. Altshuler and A. G. Aronov, Solid State Commun. **38**, 11 (1981); Pis'ma Zh. Eksp. Teor. Fiz. **32**, 482 (1979) [JETP Lett. **30**, 482 (1979)]; A. Schmid, Z. Phys. **271**, 251 (1974).

[25]Ben G. Streetman, *Solid State Electronic Devices* (Prentice Hall, Englewood Cliffs, NJ, 1990).

[26]Leo Kouwenhoven (private communication).

[27]C. Kittel and H. Kroemer, *Thermal Physics* (Freeman, New York, 1980).

[28]M. Nahum and J. M. Martinis, Appl. Phys. Lett. **63**, 3075 (1993).

[29]S. B. Kaplan, C. C. Chi, D. N. Langenberg, J. J. Chang, S. Jafarey, and D. J. Scalapino, Phys. Rev. B **14**, 4854 (1976).

[30]P. Santhanam and D. E. Prober, Phys. Rev. B **29**, 3733 (1984).

[31]S. Washburn, Nature **373**, 106 (1995).

[32]H. J. Goldsmid and A. S. Gray, Cryogenics **19**, 289 (1979).

[33]L. D. Hicks and M. S. Dresselhaus, Phys. Rev. B **47**, 12 727 (1993).

[34]L. D. Hicks and M. S. Dresselhaus, Phys. Rev. B **47**, 16 631 (1993).

[35]L. D. Hicks, T. C. Harman, and M. S. Dresselhaus, Appl. Phys. Lett. **63**, 3230 (1993).

[36]A. Kapitulnik, Appl. Phys. Lett. **60**, 180 (1992).

[37]H. B. Callen, Phys. Rev. **73**, 1349 (1948).