

Cooling electrons in semiconductor devices: A model of evaporative emission

Thushari Jayasekera, Kieran Mullen, and Michael A. Morrison*

Homer L. Dodge Department of Physics and Astronomy, The University of Oklahoma, 440 West Brooks Street,
Norman, Oklahoma 73019-0225, USA

(Received 4 April 2006; revised manuscript received 2 November 2006; published 11 January 2007)

We discuss the theory of cooling electrons in solid-state devices via “evaporative emission.” Our model is based on filtering electron subbands in a quantum-wire device. When incident electrons in a higher-energy subband scatter out of the initial electron distribution, the system equilibrates to a different chemical potential and temperature than those of the incident electron distribution. We show that this re-equilibration can cause considerable cooling of the system. We discuss how the device geometry affects the final electron temperatures, and consider factors relevant to possible experiments. We demonstrate that one can therefore induce substantial electron cooling due to quantum effects in a *room-temperature device*. The resulting cooled electron population could be used for photodetection of optical frequencies corresponding to thermal energies near room temperature.

DOI: [10.1103/PhysRevB.75.035316](https://doi.org/10.1103/PhysRevB.75.035316)

PACS number(s): 73.50.Lw, 73.23.-b

I. INTRODUCTION

As electronic devices become smaller, they leave the regime of classical physics and enter the realm of quantum physics. Many classical quantities such as resistance must be reinterpreted for systems on a mesoscopic scale. One such classical concept is that of the refrigerator: a device that uses an external source of work to cool a gas. In this paper we consider whether this classical concept can be applied to an electron gas so that one could cool such a gas by applying a voltage to a device.

There are many ways to cool electrons in a condensed-matter system. For example, thermoelectric coolers based on the Peltier effect¹ are available commercially. A different kind of electron-cooling mechanism proposed in semiconductor devices is based on a quasistatic expansion of a two-dimensional electron gas.² Still other possibilities include taking advantage of many-body effects that can lead to liquid and gas phase transitions in the electron population in a semiconductor quantum well.^{3,4}

In this paper we investigate electron cooling in a mesoscopic solid-state device using evaporative emission. This method entails removal of (“filtering”) electrons from a high-energy subband of a many-electron system, followed by relaxation of the remaining electrons to a temperature lower than that of the initial system. Evaporative cooling is widely used in bosonic systems,⁵ but this method is harder to implement for fermionic systems, as we shall discuss below.

In Sec. II we describe the theory of a two-dimensional device that can cool electrons in quantum wires. We use the Landauer formula^{6–8} to analyze the cooling properties of such a device. This formula was originally developed to explain the transport properties of electrons in a quantum device. It relates these properties to the quantum-mechanical scattering amplitudes for electrons that pass through the device. To calculate these amplitudes, we use an extension of *R*-matrix theory^{9–11} summarized in the Appendix. In Sec. III we use this theory to calculate cooling properties of several two-dimensional devices. We begin with a simple T-junction device and show that by optimizing its design we can

achieve electron cooling. We improve upon this result by switching to a “plus-junction” design, which improves the cooling characteristics of the device. In Sec. IV we discuss applications and realistic parameters for a device to cool electrons, and in Sec. V we summarize our key results and describe future research.

II. ELECTRON COOLING IN TWO-DIMENSIONAL QUANTUM DEVICES

Our theoretical approach is analogous to the working principle of the classical Hilsch vortex tube,^{12–14} which uses a T-shaped assembly of pipes to separate high-pressure air into a high-temperature system and a low-temperature system. This separation does not violate the second law of thermodynamics, because the system is driven by an external force.

We use a similar idea to cool electrons in a quantum-mechanical system. The simplest such device uses a T-shaped assembly of quantum wires to remove higher-energy electrons from an electron gas at fixed temperature. Figure 1 is a schematic of such a configuration that defines the regions of the device. We assume that our device is formed from a quantum well whose thickness is sufficiently small that the device can be considered two-dimensional. That is, we assume that the device confines electrons to a layer of thickness z_0 such that the confinement energy associated with motion in the z direction is much larger than any other energy in the problem. (This confinement energy is $\hbar^2 \pi^2 / 2m^* z_0^2$, where m^* is the electron effective mass.)

There are three leads in the T junction: the input lead, the output lead, and the sidarm, as shown in Fig. 1. We shall label physical quantities by subscripts i , o , and s , accordingly; for example, we denote the widths of the leads by w_i , w_o , and w_s . Electrons are injected into this device through the input lead, and in the input region are in thermal equilibrium at an initial temperature T_i and chemical potential μ_i . Filtering of higher-energy electrons from the initial electron gas occurs in the *scattering region*. The rate of scattering into the sidarm depends upon the electron energy: if the subband

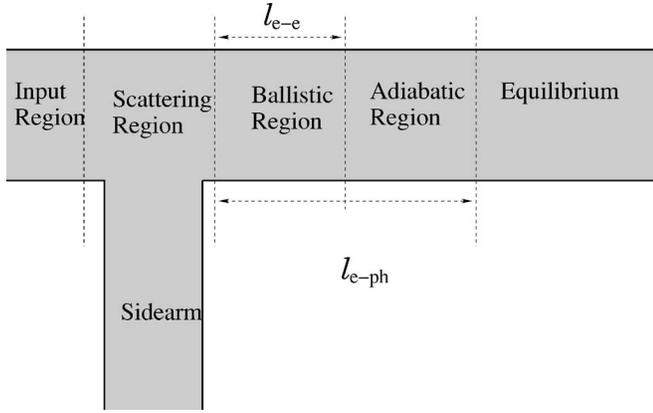


FIG. 1. Regions in a two-dimensional T-junction device. The “ballistic region” is the part of the device within a distance comparable to the electron-electron scattering length (ℓ_{e-e}) of the junction. The “adiabatic region” is within the electron-phonon scattering length (ℓ_{e-ph}) of the junction. It is in the adiabatic region that we achieve cooling. At larger distances, in the “equilibrium region,” the electrons have returned to the temperature of the lattice.

energy in lead of width w is given by $E_n = \hbar^2 \pi^2 n^2 / 2m^* w^2$, then the “force” that electrons exert on the sides of the lead, $F = -dE/dw$, is larger for larger n . Thus when an electron encounters the sidearm, the higher-subband states are more likely to scatter into the sidearm. Alternatively, one can see that the higher-subband electrons scatter preferentially to the sidearm because these electrons are in states with wave functions that are linear combinations of plane waves with larger transverse momenta.

Electrons that scatter forward into the output lead proceed into the *ballistic region*, where the electron population is determined entirely by the product of the initial electron distribution and the scattering probability. Next these electrons enter the *adiabatic region*, where they exchange energy among themselves and so relax to a temperature T_o and a chemical potential μ_o . In principle, the values of T_o and μ_o can be calculated from conservation laws for energy and particle flux. In Sec. III we report such calculations and show that for some device geometries the temperature T_o is less than the initial temperature: *this temperature decrease is the desired cooling effect*.

Finally, at large distances (as determined by the electron-phonon scattering rate) the electrons will return to equilibrium with the lattice at the initial temperature T_i ; this re-equilibration occurs in the *equilibrium region*. To see cooling, one must measure the temperature of electrons before they get to this region.

Below we define our notation and describe how we calculate the electron distributions in the input and output leads. We also define and describe the calculation of a quantitative measure of electron cooling in the device.

A. Input electron densities and populations

Provided the electrons in each lead are in thermodynamic equilibrium, we can treat them as an ideal Fermi gas. For lead ℓ therefore the subband electron density (per unit vol-

ume) in an open (energetically accessible) subband n is

$$\rho_n^{(\ell)}(E, T_\ell, \mu_\ell) = f(E; T_\ell, \mu_\ell) \mathcal{D}_n^{(\ell)}(E) \quad (1)$$

for n such that the subband energy obeys $\epsilon_n^{(\ell)} \leq E$. In this equation, the density of states is

$$\mathcal{D}_n^{(\ell)}(E) = [E - \epsilon_n^{(\ell)}]^{-1/2}. \quad (2)$$

The quantity E is the dimensionless total electron energy measured in units of the lowest subband energy of the input channel, $\mathcal{E}_1^i \equiv \hbar^2 \pi^2 / 2m^* w_i^2$. We choose the zero of energy at the energy of the ground transverse state in the input lead. We measure all other subband thresholds $\epsilon_n^{(\ell)}$ relative to this energy and in units of \mathcal{E}_1^i , so that

$$\epsilon_n^{(\ell)} = n^2 \frac{w_i^2}{w_\ell^2} - 1. \quad (3)$$

The occupation probability for lead ℓ is given by the Fermi-Dirac distribution function for total electron energy $E = E_n^{(\ell)} + \epsilon_n^{(\ell)}$,

$$f(E; T, \mu_\ell) = \frac{1}{e^{(E - \mu_\ell)/k_B T} + 1}, \quad (4)$$

where μ_ℓ is the electrochemical potential in lead ℓ , $E_n^{(\ell)}$ is the *longitudinal kinetic energy* of the electron with longitudinal wave number k , and k_B is Boltzmann’s constant.

The corresponding subband number flux in lead ℓ is the integral of the subband density [Eq. (1)] times the electron velocity $v_n^{(\ell)}(E)$ over all allowed total electron energies E :

$$J_n^{(\ell)}(T_\ell, \mu_\ell) = \int_{\epsilon_n^{(\ell)}}^{\infty} \rho_n^{(\ell)}(E; T_\ell, \mu_\ell) v_n^{(\ell)}(E) dE, \quad (5a)$$

where $v_n^{(\ell)}$ is the velocity of electrons in subband n of lead ℓ with energy E , this velocity being measured in units of $\sqrt{2m^* \mathcal{E}_1^i}$,

$$v_n^{(\ell)}(E) = \sqrt{E - \epsilon_n^{(\ell)}}. \quad (5b)$$

Note that the dimensionless velocity is the inverse of the one-dimensional density of states. The total *input number flux* in lead ℓ is the sum of the subband number fluxes for that lead,

$$J^{(\ell)}(T_\ell, \mu_\ell) = \sum_{n=1}^{\infty} J_n^{(\ell)}(T_\ell, \mu_\ell). \quad (6)$$

The total *energy flux* past a given point, $K^{(\ell)}(T_\ell, \mu_\ell)$, can be calculated in a similar fashion:

$$K^{(\ell)}(T_\ell, \mu_\ell) = \sum_{n=1}^{\infty} K_n^{(\ell)}(T_\ell, \mu_\ell) \quad (7a)$$

$$= \sum_{n=1}^{\infty} \int_{\epsilon_n^{(\ell)}}^{\infty} E \rho_n^{(\ell)}(E; T_\ell, \mu_\ell) v_n^{(\ell)}(E) dE, \quad (7b)$$

where $K_n^{(\ell)}(T_\ell, \mu_\ell)$ is the energy flux for the n th subband of lead ℓ .

B. Transmitted-electron densities and populations

Immediately upon leaving the scattering region, transmitted electrons are in a highly nonequilibrium distribution and cannot be characterized by a temperature or a chemical potential. By the time these electrons have traveled a distance along the lead comparable to several times their relaxation length, they have come to equilibrium at T_o and μ_o , and it is meaningful to describe them by a Fermi-Dirac distribution function $f(E; T_o, \mu_o)$. The output-lead properties T_o and μ_o refer to the electron population in the adiabatic region, where the output electrons are in thermodynamic equilibrium, although they are not in equilibrium with the lattice. Given T_i and μ_i for the *input* lead, our goal is to determine the values of T_o and μ_o for the *output* lead that yields the lowest $T_o < T_i$; i.e., that maximizes cooling of transmitted electrons.

For a given T_i and μ_i , we can determine the temperature in the output lead T_o by requiring that the flux of electrons in the output lead at equilibrium equals the flux of electrons transmitted into this lead (conservation of flux). To set up equations to implement this strategy, we must define subband and lead populations in terms of electrons transmitted from the input lead into the output lead. The state-to-state density of electrons transmitted from an open subband n of the input lead i into an open subband n' of a lead ℓ' is

$$\rho_{n',n}^{\ell',i}(E; T_i, \mu_i) = \rho_n^{(i)}(E, T_i, \mu_i) \mathcal{T}_{n',n}^{\ell',i}(E) \quad (8)$$

for $E \geq \epsilon_{n',n}^{\max}$, where

$$\epsilon_{n',n}^{\max} \equiv \max\{\epsilon_n^{(i)}, \epsilon_{n'}^{(\ell')}\}. \quad (9)$$

In Eq. (8), $\mathcal{T}_{n',n}^{\ell',i}(E)$ is the state-to-state transmission coefficient from subband n in lead i to subband n' in lead ℓ' . The restriction that the total electron energy E be greater than or equal to $\epsilon_{n',n}^{\max}$ ensures that both subbands n and n' are open; were this restriction violated, the transmission coefficient $\mathcal{T}_{n',n}^{\ell',i}$ would be undefined.

The flux of electrons transmitted into subband n' of the output lead, the transmitted subband flux, is

$$J_{n'}^{o,i}(T_i, \mu_i) = \sum_{n=1}^{\infty} \int_{\epsilon_{n',n}^{\max}}^{\infty} \rho_{n',n}^{o,i}(E; T_i, \mu_i) v_n^{(\ell)}(E) dE. \quad (10)$$

Hence the total flux of electrons transmitted into the output lead is

$$J^{o,i}(T_i, \mu_i) = \sum_{n'=1}^{\infty} J_{n'}^{o,i}(T_i, \mu_i). \quad (11)$$

Similarly, for the energy flux we have

$$K^{o,i}(T_i, \mu_i) = \sum_{n'=1}^{\infty} K_{n'}^{o,i}(T_i, \mu_i) \quad (12a)$$

$$= \sum_{n=1}^{\infty} \sum_{n'=1}^{\infty} \int_{\epsilon_{n,n'}^{\max}}^{\infty} E \rho_{n,n'}^{o,i}(E; T_i, \mu_i) v_n^{(\ell)}(E) dE. \quad (12b)$$

C. Cooling parameter

For any given T_i and μ_i we can determine the equilibrium temperature in the adiabatic region of the output lead as follows. Consider a surface that crosses the wire in the ballistic region. The number flux and energy flux through that surface are given by Eqs. (11) and (12). If we draw a second such surface in the adiabatic region, then the number flux and energy flux through this surface will be given by Eqs. (5) and (7) evaluated at the local (output-lead) values of the electronic chemical potential μ_o and temperature T_o . The values of μ_o and T_o will reach steady state only when the fluxes through these surfaces balance. If these fluxes do not balance, then a net charge or energy will flow into the region between the surfaces until these fluxes do balance. Therefore in steady state we have

$$J^{(o)}(T_o, \mu_o) = J^{o,i}(T_i, \mu_i), \quad (13a)$$

$$K^{(o)}(T_o, \mu_o) = K^{o,i}(T_i, \mu_i). \quad (13b)$$

Using Eqs. (12) we calculate the energy flux into the ballistic region of the output lead. We then calculate the chemical potential and temperature that would give the same fluxes in the output region. This calculation produces values of T_o and μ_o for the output lead.

As a measure of the effectiveness of a given device for cooling electrons, we define the *cooling parameter*

$$\eta(T_i, \mu_i) \equiv \frac{T_o(T_i, \mu_i)}{T_i}. \quad (14)$$

If $\eta > 1$, the device *heats* electrons. Our goal therefore is to determine the device geometry and initial electron properties T_i and μ_i that produce the smallest possible $\eta < 1$ —a situation we characterize as “maximum cooling.”

III. RESULTS

To set the stage for presentation of our primary results we briefly consider a toy model in which we replace the Fermi-Dirac distribution of Eq. (4) by a *classical* Boltzmann distribution. In this idealized device the widths of the input and output leads are equal. The input and output electron densities then have the form

$$\rho_{\text{toy}}^{(i)}(E, \tau_i, c_i) = c_i \sum_{n=1}^{n_{\max}^{(i)}} D_n^{(i)}(E) e^{-E/\tau_i}, \quad (15a)$$

$$\rho_{\text{toy}}^{(o)}(E, \tau_o, c_o) = c_o \sum_{n=1}^{n_{\max}^{(o)}} D_n^{(o)}(E) e^{-E/\tau_o}, \quad (15b)$$

where the c_i 's are overall normalization constants that effectively replace the chemical potentials as independent variables. Assuming that the junction is an ideal electron filter, we fix the number of occupied input-lead subbands $n_{\max}^{(i)}$ and output-lead subbands $n_{\max}^{(o)}$. For example, we can assume that two input-lead subbands are occupied, and that the junction removes all electrons from the second subband and transmits

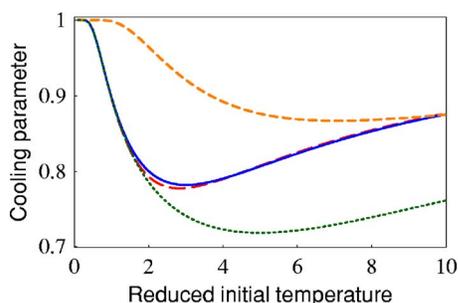


FIG. 2. (Color online) The cooling parameter η as a function of the dimensionless input-lead temperature τ_i , for the toy model of Eq. (16a). The solid curve corresponds to a two-band model in which the idealized device transmits all electrons from the first input-lead subband and removes all electrons from the second. The long-dashed curve, which lies almost on top of the solid curve, shows the analytic approximation of Eq. (16b). The other two curves correspond to numerical solution of a three-band model: for the short-dashed curve the device removes all electrons from the third input-lead subband. For the dotted curve, the device removes all electrons from the second and third subbands. In our dimensionless units, the threshold energy of the second subband is 3 and that of the third subband is 8.

all electrons from the first subband: $n_{\max}^{(i)} = 2$ and $n_{\max}^{(o)} = 1$.

In this model, the number-flux and energy-flux integrals can be evaluated analytically. Equating the incoming and outgoing fluxes, we can eliminate the c_i 's to obtain a transcendental equation that relates the cooling parameter η to the input-lead temperature τ .

$$\eta = \frac{1}{1 + 3e^{-(3/\eta\tau)} / (1 + e^{-(3/\eta\tau)}) \eta\tau}. \quad (16a)$$

We can expand this equation in a Taylor series about $\eta=1$ to get the approximate solution

$$\eta(\tau) \approx 1 - \frac{3(e^{3/\tau}\tau + \tau + 3)}{(1 + e^{3/\tau})(e^{3/\tau}\tau^2 + \tau^2 + 3\tau + 9)}. \quad (16b)$$

In our dimensionless units the threshold of the second subband is at an energy of 3. We find that maximum cooling occurs when $\tau \approx 3$, where there are thermally excited electrons in the second band. At very large temperatures ($\tau \gg 3$), the populations in the two input-lead subbands are nearly equal, and losing electrons from the second subband corresponds to losing half of the total number of electrons without appreciable cooling. Figure 2 shows these results along with those for calculations in which *three* input-lead subbands are occupied and electrons are lost only from the second and third subbands or only from the third subband. This figure makes clear that cooling is possible within this toy model. Furthermore, these results provide useful bounds for considering cooling from real devices, to which we now turn.

To achieve optimum cooling, higher-subband electrons should scatter into the sidearm, and lower-subband electrons should scatter into the output lead. In a real device not all higher-subband electrons will scatter into the sidearm, and not all the lower-subband electrons will scatter into the output lead. The probabilities for electron scattering, as quanti-

fied in transmission coefficients, depend on system properties such as the geometry and on scattering potentials. We consider “perfect devices” that have no impurities, so electrons are scattered only by the boundaries of the device. We further assume that the potential energy of the electrons in the leads is zero; this simplifying assumption is *not* essential to either our formalism or the cooling effect. We can alter the scattering of electrons by changing the ratio of the width of the sidearm to that of the input lead (Fig. 1) or by changing the geometry altogether.

A. T-junction cooling devices

To calculate the cooling parameter η we need to know the population of electrons in each subband. To determine this quantity we must in turn know transmission coefficients from a state in the input lead to states in the output lead. To calculate these transmission coefficients we use a generalization of R -matrix theory that we summarize in the Appendix. To determine the results reported here, we calculated cooling parameters η using transmission coefficients for a T-junction device for various ratios of the width w_s of the sidearm to the width w_i of the input lead, keeping the width w_o of the output lead equal to w_i .

1. T-junction device with $w_s = w_i = 1.0$

We first consider a T-junction device in which all leads have the same width: $w_i = w_s = w_o$. Figure 3 (top) shows “state-to-lead” transmission coefficients for scattering into the output lead of electrons in different subbands of the input lead. These coefficients are sums over all energetically accessible (open) subbands n_o of the output lead of state-to-state transmission coefficients $\mathcal{T}_{n_o n_i}^{\ell, i}$ [Eq. (8)] from a given state n_i of the input lead. This figure illustrates the loss of higher-subband electrons from the initial electron distribution.

Also plotted in the middle panel in this figure are the state-to-state transmission coefficients. We see that it is indeed the case that states in higher subbands are more likely to scatter down the sidearm than are states in lower subbands. We have found this result to be generally true for all geometries we have considered.

The cooling parameter η for this case is shown in the lower panel in Fig. 3 as a function of the dimensionless “reduced” initial temperature T_i with the initial chemical potential $\mu_i = 0$. (Note that we measure the energy in terms of \mathcal{E}_1^i , and all the energies are measured from \mathcal{E}_1^i . So $\mu_i = 0$ means that the external potential of the system is such that the Fermi energy is $E_F = \hbar^2 \pi^2 / 2m^* w_i^2$.) For this geometry the cooling parameter $\eta > 1$ for all initial temperatures. That is, this device geometry *heats* electrons—hardly the desired effect.

This case is important because it demonstrates that even if high-energy electrons are lost due to scattering, a compensatory loss of low-energy electrons may produce an overall heating effect. Loss of low-energy electrons opens gaps in the electron distribution at low energies. Higher-energy electrons can then relax into these newly accessible low-energy

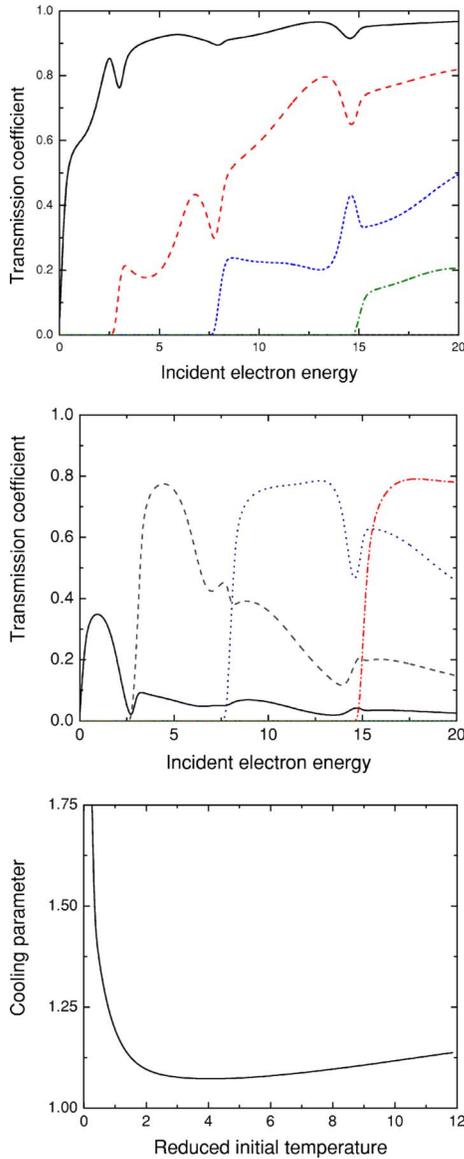


FIG. 3. (Color online) Upper panel: State-to-output-lead transmission coefficients for electrons in a T-junction device with $w_i = w_s = w_o$. The curves correspond to different subbands of the incident electrons: $n_i=1$ (solid curve), 2 (dashed), 3 (short dash), 4 (dash-dot), 5 (dotted). The horizontal axis is the energy of the incoming electron measured in terms of the first-subband energy of the input lead, $\mathcal{E}_1^i = \hbar^2 \pi^2 / 2m^* w_i^2$, from a zero of energy at \mathcal{E}_1^i . Middle panel: Electrons are scattered into the *sidearm* from subbands of the input lead $n_i=1$ (solid curve), 2 (long dash), 3 (medium dash) 4 (short dash), and 5 (dotted). This case corresponds to transmission into the output lead in Fig. 1. As the energy passes through a threshold that opens a sidearm channel, the transmission coefficient to that channel increases rapidly, then generally decreases with further increase in energy. This behavior is consistent with a semiclassical picture of electrons with large transverse momentum and small longitudinal momentum preferentially “squirt” down the sidearm. Lower panel: The cooling parameter η versus the input-lead “reduced temperature” for the coefficients shown in the upper panel for initial chemical potential $\mu_i=0$. The “reduced initial temperature” (horizontal axis) is the dimensionless quantity $k_B T_i / \mathcal{E}_1^i$.

states, with the resulting energy difference liberated as thermal energy. If this happens, then the loss of low-energy electrons will result in overall heating. The Fermi-Dirac case differs from the Boltzmann case in our toy model: in the Boltzmann case the upper subbands are *always* depopulated as $T \rightarrow 0$, while in the Fermi-Dirac case the upper subbands can be populated even at $T=0$ for large chemical potentials μ . Even a small dip in the scattered-electron distribution at low energies can significantly affect the final temperature. To *cool* electrons therefore it is not sufficient to merely scatter higher-energy electrons. We must scatter thermally excited electrons *but not significantly scatter electrons in lower-energy subbands*.

2. Alternative T-junction geometries

To determine whether a T-junction device can cool electrons at all, we now consider several widths w_s of the sidearm in Fig. 1. For each geometry we determine the initial temperature T_i and chemical potential μ_i that *minimize* the cooling parameter η . Table I shows these data and the corresponding final temperature T_o and chemical potential μ_o optimized for maximum cooling.

To illustrate these data and the effect of changing the geometry in this way, we show in Fig. 4 the variation of η with initial temperature T_i (for initial chemical potential $\mu_i=0$) for different sidearm widths. These results show that we achieve cooling ($\eta < 1$) for some geometries and heating ($\eta > 1$) for others. Only the transmission coefficients depend on the device geometry, so it is through these quantum-mechanical scattering probabilities that we can control the extent to which a device can cool electrons.

We shall now consider in detail a device with $w_s/w_i = 0.4$. Figure 5 shows state-to-lead transmission coefficients for such a device. Corresponding transmission coefficients for scattering into the sidearm are shown in the middle panel of Fig. 3. (Note that, in regard to the sidearm data, as the incident energy of an electron passes through the activation threshold for a subband, the probability that the electron will scatter down the sidearm increases rapidly. Eventually this probability becomes larger than the scattering probability into any lower subband of the sidearm.) At low energies the transmission probability is nearly unity, so for this geometry no low-energy electrons are lost from the initial distribution. Were we to adjust the Fermi energy of this device so only the lowest two subbands were occupied, we would see cooling.

In Fig. 6 we illustrate the dependence of the cooling effect on the initial chemical potential μ_i . This figure shows the cooling parameter η for $w_s/w_i=0.4$ as a function of the initial temperature T_i for $\mu_i=0.0, 3.0, \text{ and } 6.0$. Electron cooling is maximized for $\mu_i=0$, the edge of the lowest subband. At this chemical potential all electrons scattered into the sidearm are in the thermally active region of the Fermi distribution. Cooling is also obtained for $\mu_i=3.0$, the edge of the second subband. At larger values of μ_i , however, this device heats electrons.

This example shows that a T junction *can* cool electrons. We would prefer, however, a device that produces more cooling than 4%. Investigation of other T-junction geometries could not produce significantly more cooling for any value of

TABLE I. Cooling parameters η of T junctions with different sidearm widths w_s . Also shown are the input-lead chemical potential μ_i , the output-lead chemical potential μ_o , and the output-lead temperature T_o optimized for maximum cooling.

w_s	μ_i	T_i	T_o	μ_o	η
1.0	0.0	4.06	4.36	-3.27	1.07
	3.0	5.22	5.80	-1.72	1.11
	6.0	6.19	7.18	-0.21	1.16
0.9	0.0	4.65	4.94	-3.35	1.06
	3.0	5.49	6.00	-1.28	1.09
	6.0	6.25	7.10	0.57	1.14
0.75	0.0	3.40	3.52	-1.80	1.03
	3.0	4.77	5.10	-0.01	1.07
	6.0	6.07	6.73	1.66	1.11
6.0	0.0	1.25	1.24	-0.33	0.99
	3.0	5.19	5.40	0.65	1.04
	6.0	6.38	6.83	2.67	1.07
0.5	0.0	1.43	1.37	-0.19	0.96
	3.0	1.11	1.14	2.48	1.02
	6.0	5.96	6.34	3.34	1.06
0.4	0.0	2.10	2.02	-0.14	0.96
	3.0	1.93	1.90	2.62	0.98
	6.0	4.80	4.96	4.44	1.03

w_s/w_i . So we next investigate the addition of a second sidearm. This change produces the “plus junction” illustrated in Fig. 7.

B. A “plus-junction” cooling device

Since we achieved cooling in a T-junction device with $w_s=0.4$, we shall consider a plus junction with the same lead ratios: $w_i=w_o$ and $w_s/w_o=0.4$. (The widths of the two sidearms in Fig. 7 are the same.)

In Fig. 8 we show the results of calculations for a “plus-junction” device. In panel (a) we plot state-to-lead transmission coefficients for the first three subbands of such a device with $w_s/w_i=0.40$; in panel (b) we show the corresponding cooling parameter η as a function of the initial temperature τ_i for different choices of initial chemical potential μ_i . In panels (c) and (d) we show the corresponding results for $w_s/w_i=0.32$. We first note that, in general, adding a second sidearm improves the cooling by increasing the transmission coefficients into the sidearm. In panels (a) and (c) the subband thresholds for the input lead are marked by vertical solid lines, and the thresholds for the sidearms are marked by vertical dashed lines. The transmission coefficient for the second subband shows an abrupt decrease as the energy passes the threshold for the first subband in the sidearm. This feature is large enough to affect transmission into the *second* sidearm subband at its threshold, $\epsilon=3$. We also see a dip in the first subband transmission coefficient for the first subband but this dip is comparatively narrow.

These features affect the cooling parameter as shown in panel (b). For $\mu_i=0$ and $\mu_i=3$ there is a narrow range of temperatures τ_i at which the device cools. With increasing initial temperature, electrons populate states in the first subband with very small transmission probabilities. These electrons are *not* transmitted, and the resulting hole in the population of the first subband leads to *heating*—so the cooling parameter has a relative minimum. At high temperatures, electrons populate the lowest three bands, and cooling results from loss of electrons from the third subband. At the higher electrochemical potential $\mu_i=8$, which equals the threshold of the third subband, cooling does not occur until the temperature τ_i becomes rather large. But at sufficiently high temperatures, the substantial loss of electrons from the third subband leads to appreciable cooling.

Panels (c) and (d) show the results of an attempt to improve cooling by narrowing the sidearm. Panel (c) shows that narrowing the sidearm *increases* the threshold for each of its subbands. This change results in less scattering out of the first subband. Furthermore, there is less scattering out of the second below threshold. The third subband is not strongly affected. The consequences of these changes are evident in the cooling parameter in panel (d). Narrowing the sidearm improves cooling, because the decrease in the number of electrons lost from the first subband suppresses the bump in the cooling parameter for the larger sidearm-width in panel (b). The low-temperature behavior of the cooling curve highlights the sensitivity of the effect to the loss of low-energy electrons. Figure 3(c) shows that there is a small notch near $E=0$ in the transmission coefficient into the lowest subband. The loss of these very-low-energy electrons means that the cooling curves for $\mu=3$ and $\mu=8$ will diverge as $T\rightarrow 0$, since higher-energy states will always be populated even at $T=0$, and when the electrons relax to fill this notch their final temperature will *not* be zero. On the other hand, the cooling curve for $\mu=0$ diverges until the temperature is so low that there are no electrons to relax into the lost states, resulting in a well-behaved cooling parameter in the $T\rightarrow 0$ limit.

With increasing temperature, an increasing number of electrons are lost from the second and third subbands, increasing the cooling effect. Thus cooling at an initial electrochemical potential of $\mu_i=8$ in the narrower sidearm is better than at any temperature in the wider sidearm [panel (b)], because in the narrower sidearm fewer electrons are lost from the first subband.

These figures illustrate the dependence of the cooling parameter on the detailed dependence of the transmission coefficients on energy. These coefficients, in turn, depend on device geometry. For wide sidearms, more electrons may be lost from the second subband—but only at the cost of loss from the *first* subband, which produces heating. Some choices of sidearm widths we explored produced cooling curves with a relative minimum; in such cases increasing the input temperature degrades device performance, because higher-energy electrons fall into gaps in the population of lower subbands. In all cases, however, the trend at high temperatures ($5 < \tau < 15$) is to increase cooling; eventually the behavior of devices that lose a substantial fraction of electrons from the third subband approaches the classical model

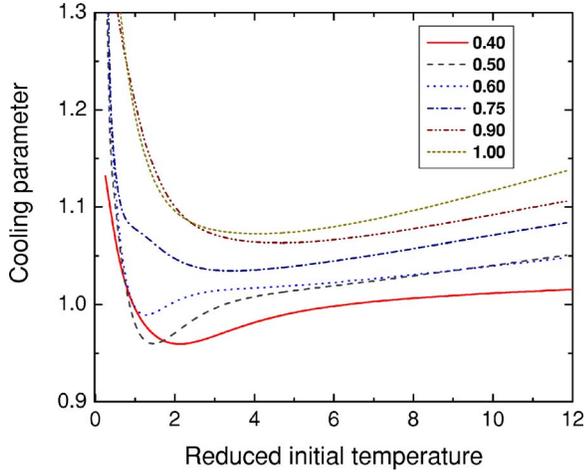


FIG. 4. (Color online) The cooling parameter η as a function of the reduced initial temperature with initial chemical potential $\mu_i = 0$ for different values of w_s/w_i . The cooling parameter decreases with increasing magnitude of $w_s/w_i = 1.0$ (solid line), 0.9 (long dash), 0.75 (medium dash), 0.6 (short dash), 0.5 (dash-dot), and 0.4 (dotted). The reduced initial temperature is defined in the caption to Fig. 3.

of Eq. (16a). Apparent improvement in cooling at high temperatures must be tempered by the realization that the model breaks down if the electron-phonon scattering length is too short compared to the size of the device. We discuss this point in more detail below.

One could try to further optimize the geometry of a device by, for example, allowing the width of the sidearms to increase with increasing distance from the junction with the

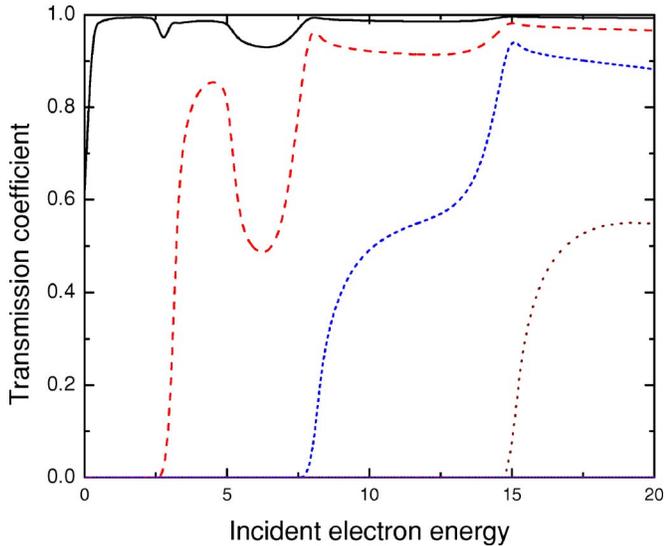


FIG. 5. (Color online) State-to-lead transmission coefficients for electrons in a T-junction device with $w_i = w_o$ and $w_s = 0.4w_i$. Electrons are scattered into the output lead from subbands of the input lead $n_i = 1$ (solid curve), 2 (long dash), 3 (medium dash), 4 (short dash), and 5 (dotted). The horizontal axis is the energy of the incoming electron measured in terms of the first-subband energy of the input lead, $\mathcal{E}_1^i = \hbar^2 \pi^2 / 2m^* w_i^2$, from a zero of energy at \mathcal{E}_1^i . See also the data in Table I.

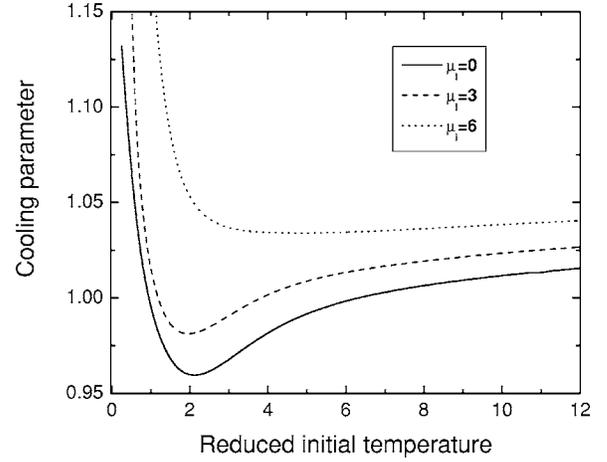


FIG. 6. The cooling parameter η as a function of the reduced initial temperature for a T-junction device with $w_i = w_o$ and $w_s = 0.4w_o$ for three initial chemical potentials, $\mu_i = 0.0$ (solid curve), 3.0 (dashed), and 6.0 (dotted). These data are based on the transmission coefficients shown in Fig. 5. Maximum cooling is obtained for $\mu_i = 0$. The reduced initial temperature is defined in the caption to Fig. 3.

central region. Alternatively, one could round the sharp corners at each junction into smooth curves. However, our initial explorations of such alterations did not produce substantially more cooling than obtained with the far simpler plus junction in Fig. 7. The essential feature that makes a plus junction more effective than any T junction is the presence of more than one sidearm. For $w_s/w_i = 0.4$ we obtain almost 8% cooling ($\eta \approx 0.92$) compared to 4% ($\eta \approx 0.96$) in the T junction. It is also important that the threshold for scattering into these sidearms lies above the threshold for the second subband of the input lead. Further improvements might be gained from cascading stages of the plus junction.

IV. EXPERIMENTAL CONSIDERATIONS

We now consider a plus-junction device using realistic experimental parameters for InSb and GaAs. All material properties of an actual device depend on the Fermi energy E_F of the system. This quantity is inversely proportional to the effective mass m^* of the electrons in the material and is

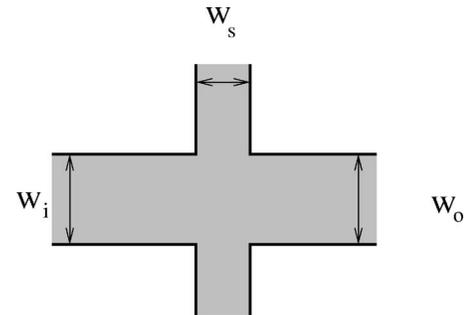


FIG. 7. Schematic of a plus-junction device. While the widths of the input and output leads are equal ($w_i = w_o$), the width w_s of the two identical sidearms may be smaller or larger than w_i .

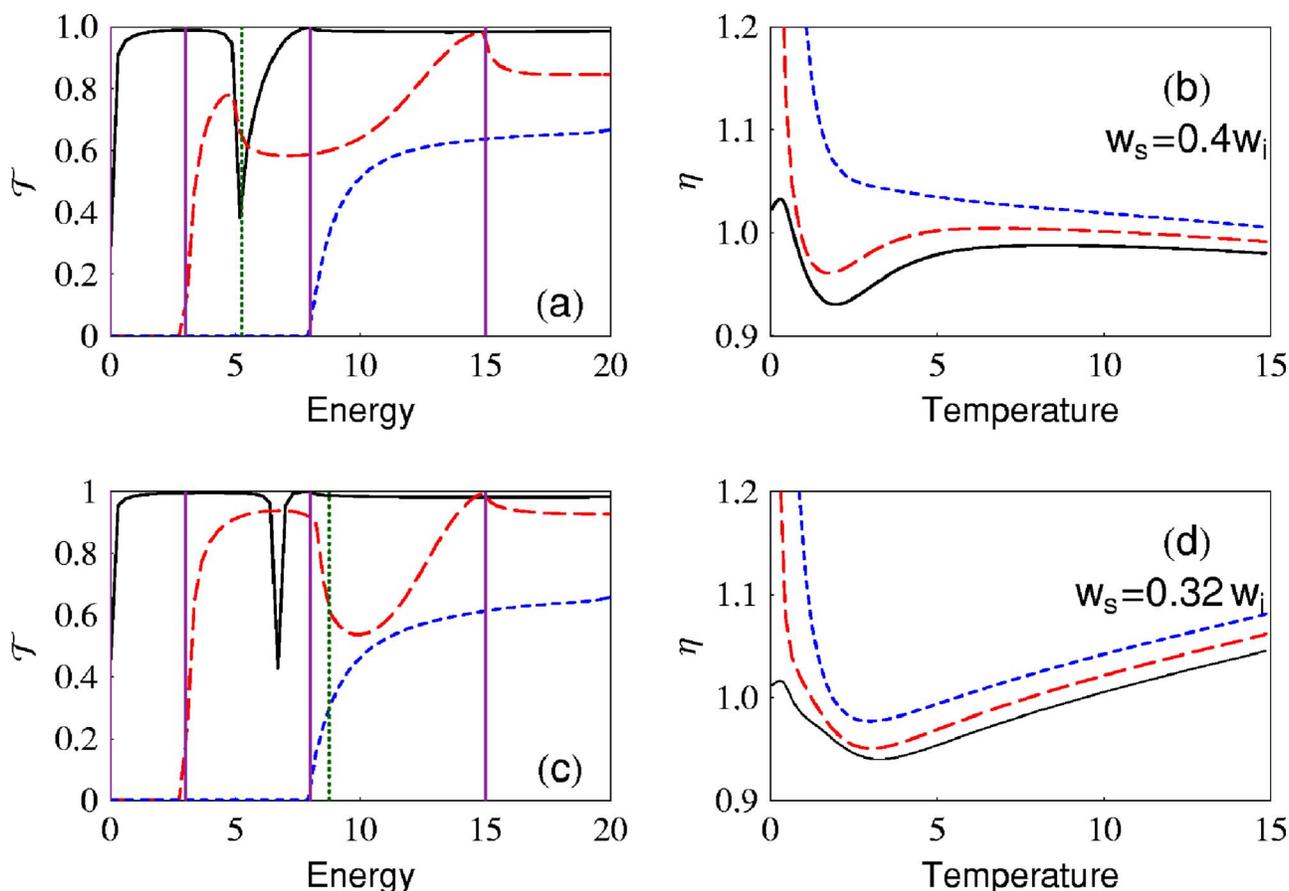


FIG. 8. (Color online) (a) State-to-output-lead transmission coefficients for electrons in a plus-junction device with $w_s/w_i=0.4$. The curves correspond to different subbands of the incident electrons: $n_i=1$ (solid), 2 (long dash), and 3 (short dash). The horizontal axis is the energy of the incoming electron measured in terms of the first-subband energy of the input lead, $\hbar^2 \pi^2 / 2m^* w_i^2$, from a zero at this energy. (b) Cooling parameter η , calculated from the coefficients in (a), as a function of the reduced initial temperature τ_i . The reduced initial temperature is defined in the caption to Fig. 3. Plotted here are curves for initial chemical potentials of $\mu_i=0$ (solid), $\mu_i=3$ (long dashed) and $\mu_i=8$ (short dash). (c) Transmission coefficients as in (a) but with $w_s/w_i=0.32$. (d) Cooling parameters as in (b) but with $w_s/w_i=0.32$. In panels (a) and (c) the subband thresholds for the input lead are marked by vertical solid lines, and the thresholds for the sidearms are marked by vertical dashed lines.

determined by the electron density in the reservoir. For our device we chose for the initial chemical potential (which is approximately equal to the Fermi energy) $\mu_i=0$. Since we have scaled the energy by \mathcal{E}_1^i and chosen \mathcal{E}_1^i as the zero of energy, setting the initial chemical potential equal to zero means that

$$\mu_i = E_F = \frac{\hbar^2 \pi^2}{2m^* w_i^2}, \quad (17)$$

where w_i is the width of the input lead of the device in Fig. 7. Since the Fermi energy and the subband energies depend on the effective mass in the same fashion, the width of the input lead is independent of the material. For a sample with electron density n , we have

$$\frac{\pi \hbar^2}{m^*} n = \frac{\hbar^2 \pi^2}{2m^* w_i^2}, \quad (18)$$

which gives

$$w_i = \sqrt{\frac{\pi}{2n}}. \quad (19)$$

If, for example, $n=1.0 \times 10^{11} \text{ cm}^{-2}$, then $w_i=39.6 \text{ nm}$, quite a small value. We can increase this value by decreasing the electron density.

For a plus junction, we were able to maximize cooling by setting the initial temperature to $\tau_i \sim 2$ in dimensionless units. In dimensional units, this optimum temperature is

$$k_B T = T_{\text{opt}} \mathcal{E}_1^i. \quad (20a)$$

Using Eq. (17), we obtain

$$k_B T = T_{\text{opt}} E_F. \quad (20b)$$

For a sample with $n=1.0 \times 10^{11} \text{ cm}^{-2}$, the initial temperature for maximum cooling is $T_i \sim 82 \text{ K}$ for GaAs and $T_i \sim 399 \text{ K}$ for InSb. Room temperature (300 K) corresponds to $T \sim 1.5$ for InSb. *One can therefore obtain substantial cooling due to quantum effects in a room-temperature device.* A cooling parameter of $\eta \sim 0.95$ implies that the elec-

tron population is cooled by 15 K. The resulting cooled electrons could be used for photodetection of optical frequencies corresponding to thermal energies near room temperature.

One problem is that at high temperatures the system will no longer act in a quantum-mechanical fashion. While point-contact experiments have shown features at temperatures as high as 44 K,¹⁵ these features wash out at higher temperatures. However, we do not require long-range quantum-mechanical coherence across the device. Only electrons in the scattering regions must behave quantum mechanically. Once the electrons leave the interaction region, any dephasing collisions will not affect their subsequent transport, so long as nothing heats the electron gas before its temperatures is measured.

The parameters for these devices operate in the “low-temperature” regime. We could design a device to operate in the “high-temperature regime” ($\tau \gg 1$) by using wider quantum wires. This change would decrease the characteristic energy \mathcal{E}_1^i so as to increase the reduced temperature τ for a fixed actual temperature T . But we cannot choose too large a value for the width w_o of the output lead, since we want to retain $\mu_i \sim 0$; this condition with a large value of w_o would require an initial Fermi energy (and thus an initial doping) that would be too low to be practical. For example, for InSb we could set \mathcal{E}_1^i to one-tenth the thermal energy at room temperature, which would involve approximately tripling the width of the wire. However, we would also need to decrease the doping by a factor of 10.

The devices we have considered have only one cooling stage. One could increase cooling by connecting multiple plus junctions in series. The spacing between junctions, however, must be large enough that scattering resonances between sidearms are negligible. If not, one would have to treat the device as a single large quantum-mechanical scattering target. While the presence of such resonances would not preclude cooling, they would make calculations for a chain of junction devices more difficult and sensitive to details of phase breaking.

V. CONCLUSIONS AND PROSPECTS FOR FUTURE RESEARCH

Many photodetection applications require a cold detector. We have presented results for a prototype device that demonstrates electron cooling in a single-particle picture. We have shown that, while a naive T junction can produce modest cooling, adding an additional sidearm yields a device—the plus junction—that can produce as much as 8% cooling. The abrupt discontinuities in the confining potentials in these models are not essential to cooling; what is essential is that higher-subband states, in which electrons have larger transverse momenta, scatter appreciably into the sidearms. We therefore expect electron cooling in such devices to be insensitive to details of the potential so long as the potential does not eliminate the states of the lowest subband.

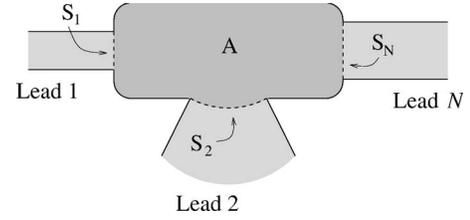


FIG. 9. Schematic of a two-dimensional device for the present scattering calculations. The surfaces S_1, S_2, \dots, S_N separate the interior region A from the N leads.

ACKNOWLEDGMENTS

This project was supported in part by the U.S. National Science Foundation under Grant No. MRSEC DMR-0080054, EPS-9720651, and PHY-0071031. One of the authors (T.J.) would like to acknowledge the University of Oklahoma Graduate College for travel and research grants.

APPENDIX: R-MATRIX THEORY FOR A TWO-DIMENSIONAL SYSTEM

We consider the two-dimensional system in Fig. 9. This system has a central region A connected to N external regions or “leads.” The leads and the interior region meet at a set of boundary surfaces we denote by S_0, S_1, \dots, S_N . We treat the boundaries between the shaded and unshaded regions as “hard walls” (infinite potential) so electron wave functions are nonzero only in the shaded regions. Since there may be more than three leads, we depart from the notation used in the body of this paper (in which the input, output, and side-arm leads were denoted by subscripts i, o , and s) and denote the input lead by a zero subscript and all other leads by positive integer subscripts. We measure all distances in units of w_0 and energies in terms of $\hbar^2/2m^*w_0^2$. We seek an analytic solution for the amplitudes of outgoing states in the leads when only one incoming state is occupied.

The time-independent Schroedinger equation for the scattering function is

$$(\hat{H} - E)|\Psi_{E,n_0}\rangle = 0, \quad (\text{A1})$$

where $|\Psi_{E,n_0}\rangle$ represents the state of an electron with kinetic energy E incident in input-lead subband n_0 . Note that $|\Psi_{E,n_0}\rangle$ is well defined in all leads. In a *finite* region, the Hamiltonian \hat{H} is not Hermitian. We can produce a Hermitian operator by adding to \hat{H} the so-called Bloch operator \hat{L}_B .¹⁰ We denote the eigenfunctions of $\hat{H} + \hat{L}_B$ in the interior region A by $|\phi_i\rangle$ and write the so-called Bloch eigenvalue equation as

$$(\hat{H} + \hat{L}_B)|\phi_i\rangle = E_i|\phi_i\rangle. \quad (\text{A2})$$

Inserting the Bloch operator into the Schroedinger equation we get

$$(\hat{H} + \hat{L}_B - E)\Psi_E = \hat{L}_B\Psi_E. \quad (\text{A3})$$

We now expand the scattering wave function $|\Psi_E\rangle$ in the set of orthonormal Bloch eigenfunctions

$$|\Psi_E\rangle = \sum_j C_j |\phi_j\rangle. \quad (\text{A4})$$

Inserting this expansion into the Schroedinger equation and using the properties of the Bloch eigenfunctions yields

$$|\Psi_{E,n_0}\rangle = \sum_j \frac{\langle \phi_j | \hat{L}_B | \Psi_{E,n_0} \rangle}{E_j - E} |\phi_j\rangle, \quad (\text{A5})$$

where E_j is the eigenvalue corresponding to the Bloch eigenfunction $|\phi_j\rangle$. This expansion is valid throughout the interior region A and on its surface (see Fig. 9).

To derive an equation for the R matrix, we now apply this expansion of the scattering state on each boundary S_i . At each such boundary we can expand the scattering function in either lead eigenfunctions or Bloch eigenfunctions in the interior region. To be specific, we introduce a local Cartesian coordinate system for each lead: x_q and y_q are the longitudinal and transverse coordinates of the q th lead, respectively. We choose $x_q=0$ on each boundary. (One can easily choose any orthonormal coordinate system, *mutatis mutandis*.) Each lead eigenfunction is then a product of a plane wave in the x_q direction and a transverse bound-state eigenfunction $\chi_n(y_q)$. The scattering wave function in the q th lead therefore becomes

$$\begin{aligned} \Psi_{E,n_0}(x_p, y_p) &= e^{-ik_{0,n_0}x_0} \chi_{0,n_0}(y_0) \delta_{p,0} \\ &+ \sum_{q,n_q=1}^N \tau_{q,n_q}(E) e^{ik_{q,n_q}x_q} \chi_{q,n_q}(y_q) \delta_{p,q}, \end{aligned} \quad (\text{A6})$$

where k_{q,n_q} and τ_{q,n_q} are the wave vector and transmission

amplitude for the channel with quantum number n_q in channel q . Also, $\chi_{n_q}(y_q)$ is the n_q th transverse eigenfunction for lead q . Finally, $\delta_{p,q}$ is the Kronecker delta function, which ensures that each wave function is defined in only one lead. If we measure energy in units of E_0 then we can express energy conservation in lead q as $E = k_{q,n_q}^2 + n_q^2 \pi^2 / w_q^2$, where w_q is the width of the q th lead (in units of w_0). We use this equation to determine the wave vector k_{q,n_q} .

After considerable algebra we get a set of linear algebraic equations that we can solve for the transmission amplitudes:

$$\begin{aligned} i \sum_{p,n_p} \tau_{p,n_p}(E) k_{p,n_p} M_{q,n_q,p,n_p}(E) - \tau_{q,n_q}(E) \\ = \delta_{q,0} \delta_{n_q,n_0} + ik_{0,n_0} M_{q,n_q,0,n_0}. \end{aligned} \quad (\text{A7})$$

In writing these equations we have defined matrix elements

$$M_{q,n_q,p,n_p} = \int_{y_p} \int_{y_q} \chi_{q,n_q}^*(y_q) R_E(y_q, y_p) \chi_{p,n_p}(y_p) dy_q dy_p. \quad (\text{A8})$$

Finally, the R matrix is given by

$$R(E, y_p, y_q) \equiv \sum_j \frac{\phi_j^*(x_q=0, y_q) \phi_j(x_p=0, y_p)}{E_j - E}. \quad (\text{A9})$$

This equation is general in that we can easily adapt it to any number of leads and to different choices of input lead.

*Electronic address: morrison@mail.nhn.ou.edu

¹ *Thermodynamics and an Introduction to Thermostatistics*, Herbert H. Callen (John Wiley and Sons, New York, 1985).

² L. G. C. Rego and G. Kirczenow, *Appl. Phys. Lett.* **75**, 2262 (1999).

³ Thushari Jayasekera, K. Mullen, and Michael A. Morrison (unpublished).

⁴ Junren Shi and X. C. Xie, *Phys. Rev. Lett.* **88**, 086401 (2002).

⁵ W. Ketterle and N. J. van Druten, in *Advances in Atomic Molecular and Optical Physics*, edited by B. Bederson and H. Walther (Academic, San Diego, 1996), Vol. 37, p. 181.

⁶ R. Landauer, *Philos. Mag.* **21**, 863 (1970).

⁷ M. Büttiker, *Phys. Rev. Lett.* **57**, 1761 (1986).

⁸ A. Douglas Stone and A. Szafer, *IBM J. Res. Dev.* **32**, 384

(1988).

⁹ E. P. Wigner and I. Eisenbud, *Phys. Rev.* **72**, 29 (1947).

¹⁰ C. Bloch, *Nucl. Phys.* **4**, 503 (1957).

¹¹ U. Wulf, J. Kucera, P. N. Racec, and E. Sigmund, *Phys. Rev. B* **58**, 16209 (1998).

¹² Thushari Jayasekera, Ph.D. thesis, University of Oklahoma, Norman.

¹³ R. Hilsch, *Rev. Sci. Instrum.* **18**, 108 (1947).

¹⁴ M. P. Silverman, *And Yet It Moves: Strange Systems and Subtle Questions in Physics* (Cambridge University Press, Cambridge, England, 1993).

¹⁵ Gregory L. Snider, Mark S. Miller, Michael J. Rooks, and Evelyn L. Hu, *Appl. Phys. Lett.* **59**, 2727 (1991).