

Voltage drop in mesoscopic systems: A numerical study using a quantum kinetic equation

Michael J. McLennan,* Yong Lee, and Supriyo Datta

School of Electrical Engineering, Purdue University, West Lafayette, Indiana 47907

(Received 15 October 1990)

In this paper, we present a numerical method for evaluating the full Wigner function throughout a device by solving a steady-state quantum kinetic equation in two dimensions, in the linear-response regime. This method has two advantages over conventional treatments of mesoscopic devices. First, dissipative processes can be included within the device, thus allowing a smooth transition from the quantum to the semiclassical regime. Second, the contacts are treated in the same manner as in semiclassical device analysis. A short phase-breaking time can be used in the contact regions so that oscillations in the electron density due to interference effects die out quickly; this is particularly useful when obtaining self-consistent solutions with the Poisson equation. Any quantity of interest, such as electron density or current density per unit energy, can be computed throughout the entire device. We will first show that under low-bias, low-temperature conditions, the diagonal elements of the Wigner function can be used to define a local electrochemical potential (μ) that lends insight into the internal transport physics. We show that separate electrochemical potentials μ_L and μ_R for left- and right-moving electrons show unphysical behavior when defined in a local sense. But sensible results are obtained when these potentials are defined in an average sense over regions the size of a de Broglie wavelength. We then examine the difficulties associated with measuring μ , with numerical examples. Next, we use the local electrochemical potential profile to clarify the nature of the spreading resistance associated with the narrowing of a current lead. Finally, we show that the electrostatic potential (ϕ) can be viewed as a convolution of μ with a screening function and present example computations of ϕ .

I. INTRODUCTION

Mesoscopic devices are usually composed of a small "interesting" region connected through wide leads to an external source. It is common to neglect the details of the connection to the external source, which hopefully contributes only a small series resistance. Instead, it is assumed that the external source imposes the following *boundary condition* on the structure: The region to the left of A' is maintained at a constant electrochemical potential μ_1 , while the region to the right of B' is maintained at μ_2 (Fig. 1). It can be shown that the current I is related to the total transmission T through the structure¹⁻⁷

$$I = \frac{e^2}{h} T(\mu_1 - \mu_2), \quad (1.1)$$

where

$$T \equiv \text{Tr}\{\underline{t}^\dagger \underline{t}\}, \quad (1.2)$$

and \underline{t} is the transmission matrix between the leads (from A' to B'). Since the voltage drop between A' and B' is $(\mu_1 - \mu_2)$, the conductance of the overall structure is

$$G \equiv \frac{I}{\mu_1 - \mu_2} = \frac{e^2}{h} T. \quad (1.3)$$

For a device having perfect transmission, one might expect the conductance to be infinite, since there is no scattering. Equation (1.3), however, predicts a finite conductance. Imry⁶ has interpreted this finite conductance

as arising from the contact regions between the device and the wide leads. To compute the conductance of the device alone, we need to determine what fraction of the voltage applied between A' and B' is usually dropped across the device between points A and B . This question was originally raised by Landauer in his pioneering paper⁴ and has led to many different Landauer-type conductance formulas,⁸⁻¹⁰ depending on how the voltage drop across the device is measured.

Engquist and Anderson¹¹ suggested that an electrochemical potential could be measured at various points within a structure by attaching additional reservoirs and allowing them to equilibrate with the sample. Büttiker

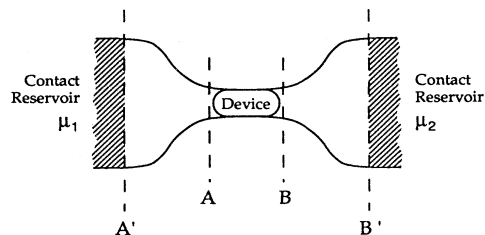


FIG. 1. A typical quantum device has a small interesting region connected to an external source through wide leads. The external source is assumed to maintain the region to the left of A' at a constant electrochemical potential μ_1 and the region to the right of B' at μ_2 .

recognized¹² that such voltage probes could be treated on an equal footing with current probes in the Landauer model, and he generalized the expression for the current (1.1) to account for multiple leads,

$$I_i = \frac{e^2}{h} \sum_j T_{ij} (\mu_i - \mu_j), \quad (1.4)$$

where T_{ij} represents the total transmission from lead j to lead i . This equation has proven to be remarkably successful at explaining experimental four-terminal measurements, such as the symmetry properties of magnetoresistance and the integer quantum Hall effect. Recent reviews of this work can be found in Refs. 13–15. Usually the coefficients T_{ij} are computed from the Schrödinger equation, assuming coherent transport within the device. However, Büttiker has shown¹⁶ that the presence of phase-breaking scatterers can be simulated by including additional voltage probes in the analysis.

Indeed the equation that we use for most of this paper looks like a generalization of (1.4) to a continuous distribution of voltage probes:

$$I(\mathbf{r}) = \frac{e^2}{h} \int d\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}') [\mu(\mathbf{r}') - \mu(\mathbf{r})]. \quad (1.5)$$

However, a direct generalization of (1.4) is purely phenomenological and does not tell us how to evaluate the function $T_0(\mathbf{r}, \mathbf{r}')$. Moreover, it raises subtle questions about whether a local electrochemical potential $\mu(\mathbf{r})$ can be meaningfully defined within a device. For this reason we start from a general quantum kinetic equation and through various approximations reduce it to the form shown in (1.5). The kinetic equation allows us to compute the energy distribution of carriers directly and thus verify the conditions under which this distribution has a Fermi-Dirac form, so that a local electrochemical potential $\mu(\mathbf{r})$ can be defined unambiguously. Moreover, we can compute the full Wigner function and thus obtain any quantity of interest such as the electron density or the current density anywhere within the device.

Our starting point is the formulation of quantum transport developed by Kadanoff and Baym and Keldysh.¹⁷ This approach was originally applied primarily to homogeneous systems, and several excellent reviews are available.^{18,19} It has also been applied to tunneling devices and to current fluctuations in mesoscopic devices.²⁰ The fundamental quantity in this formulation is the Wigner function $-iG^<(\mathbf{r}; \mathbf{k}; E; t)$, which can be computed by solving a quantum kinetic equation. Any observable, such as the electron density or the current density, can then be obtained by taking appropriate moments of this function. From this standpoint, the Wigner function plays a role analogous to the semiclassical distribution function. This approach has two advantages over conventional treatments of mesoscopic devices.

(1) Arbitrary dissipative processes can be included within the device. This allows one to study the transition from quantum to semiclassical behavior.

(2) The contacts are treated in exactly the same manner

as in classical device analysis. Instead of solving the drift-diffusion equation, we solve a quantum kinetic equation subject to the same boundary conditions at the contacts. A short phase-breaking time can be used in the contact regions so that oscillations in the electron density due to interference effects die out quickly; this is particularly useful when obtaining self-consistent solutions with the Poisson equation.

Most of the recent work on quantum kinetic equations has focused on high-field transport in homogeneous materials, although there have been efforts to extend this work to heterostructures,²¹ and numerical results for resonant-tunneling diodes have been reported.^{22–24} To our knowledge, however, quantum kinetic equations have not been applied to mesoscopic structures where the two-dimensional current flow leads to interesting phenomena that are absent when the flow is essentially one-dimensional.

A numerical solution of the quantum kinetic equation in two dimensions is made difficult by the large number of independent variables: \mathbf{r} , \mathbf{k} , E , and t . Assuming two components each for \mathbf{r} and \mathbf{k} , the number of independent variables is six. In this paper, we use a simplification scheme proposed earlier²⁵ to reduce this number to two, thus making the problem manageable. These simplifications are briefly summarized in Sec. II. By restricting our analysis to steady state, we eliminate t . By assuming a special model for phase-breaking scattering processes, as discussed in Sec. II C, we eliminate \mathbf{k} . If we further restrict ourselves to low-bias voltages at low temperatures, the energy distribution of the Wigner function can be characterized by a Fermi-Dirac form with a local electrochemical potential $\mu(\mathbf{r})$, as discussed in Sec. III A. This eliminates E as an independent variable, leaving only the two spatial dimensions \mathbf{r} .

With these assumptions, the quantum kinetic equation reduces to an integral equation for $\mu(\mathbf{r})$:

$$\mu(\mathbf{r}) = \frac{\int d^3\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}') \mu(\mathbf{r}')}{\int d^3\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}')}. \quad (1.6)$$

Equation (1.6) is really the same as (1.5) with the external current $I(\mathbf{r})$ set to zero, which is true everywhere except in the contacts; to get from (1.6) to (1.5), a careful discussion of the contacts is required, as explained in Sec. III C. Equation (1.6) has a simple physical interpretation. It states that the electrochemical potential at any point is a weighted average of the potentials at surrounding points. The weighting function $T_0(\mathbf{r}, \mathbf{r}')$ is related to the probability that an electron that suffers a phase-breaking scattering at \mathbf{r}' will suffer its next phase-breaking event at \mathbf{r} . This function is computed from the Green function of the Schrödinger equation modified to include an optical potential derived from the self-energy. Contacts are simulated using *open* boundary conditions.

Once the weighting function $T_0(\mathbf{r}, \mathbf{r}')$ is obtained, we can solve the integral equation [Eq. (1.6)] subject to the boundary condition that the electrochemical potential

take on specified constant values in the contact regions. This is the same boundary condition that one uses in classical device analysis. The only difference is that in classical device analysis one solves the drift-diffusion equation $\nabla \cdot (\sigma \nabla \mu) = 0$, while in the quantum regime we solve the integral equation, Eq. (1.6). In fact, it has been shown that (1.6) reduces to the diffusion equation in homogeneous samples.²⁵

In a typical simulation, a device is discretized into approximately 1000 nodes. To compute $\mu(\mathbf{r})$ throughout, the device takes approximately 1 h on a Sun 4 workstation [once $\mu(\mathbf{r})$ is obtained all other quantities of interest such as electron density and current density are readily obtained]. Clearly, this is far from the limitations of modern computing, and as such, this method could prove to be a powerful and practical tool for device analysis in the quantum regime. The details of the numerical method are described in the Appendix.

It is interesting to note that the solutions to Eq. (1.6) look fairly classical for short scattering times. For example, Fig. 2 shows the potential profile and current flow pattern in a three-moded wire with two impurities. One of the impurities is exactly centered in the wire, while the other is off to one side. Overall, the potential μ drops almost linearly due to the uniform background of phase-breaking scatterers, which, in our model, not only randomize the phase but also the momentum. Around the impurities the potential drops in a localized manner. Indeed, for the off-center impurity, the potential drops sharply on the side of the wire near the impurity, but falls

smoothly along the other. The current redistribution around the impurities is also quite in keeping with our intuition. Note, however, that we have used a fairly short phase-breaking time ($\tau_\phi = 0.2$ ps) in this particular example. With longer times, the potential shows oscillatory behavior, while the current flow develops circulating patterns due to interference effects.²⁶

It is easy to see from Eq. (1.6) why the electrochemical potential drops sharply around an obstacle. An obstacle effectively isolates the left from the right so that the weighting function $T_0(\mathbf{r}, \mathbf{r}')$ is large when the points \mathbf{r} and \mathbf{r}' are both to the left or both to the right of the obstacle. Consequently, the potential to the left of an obstacle is determined largely by the left contact and to the right of an obstacle by the right contact. This leads to an abrupt potential drop as we cross the obstacle. It is important to note that the potential drop need not be accompanied by an Ohmic power loss. This is apparent since a static obstacle may not have the internal degrees of freedom necessary to dissipate energy. Thus, the question of where the voltage drop is may have a different answer from the question of where the power is lost. In this paper we do not address the latter question.

Actually, the question of where is the voltage drop requires two answers: one in terms of the electrostatic potential $\phi(\mathbf{r})$, and the other in terms of the electrochemical potential $\mu(\mathbf{r})$.⁴ The electrostatic potential ϕ is a well-defined concept that has a clear meaning, even at the microscopic level. The electrochemical potential, on the other hand, is usually viewed as a macroscopic concept, and it is not obvious that this potential can be meaningfully defined in a microscopic sense.

In this paper, we define a local electrochemical potential by considering the energy distribution of the electron density $n(\mathbf{r}; E)$, which is related to the diagonal element of the correlation function $G^<(\mathbf{r}, \mathbf{r}; E)$. The electron density $n(\mathbf{r}; E)$ can be expressed as the product of the local density of states $N_0(\mathbf{r}; E)$ and an occupation factor $f(\mathbf{r}; E)$. At equilibrium, this occupation factor is simply the Fermi-Dirac factor with a constant electrochemical potential μ_0 . We will show in Sec. III that, at low temperatures and low bias, $f(\mathbf{r}; E)$ is well described by a Fermi-Dirac function with a local electrochemical potential $\mu(\mathbf{r})$. Under a large bias, $f(\mathbf{r}; E)$ is significantly distorted and cannot be characterized by a single parameter $\mu(\mathbf{r})$. It seems, however, that such hot-carrier effects would also lead to nonlinear response, since the degree of heating (and hence the conductance) becomes bias dependent. We therefore believe that the local electrochemical potential is a valid and useful concept in the linear-response regime at low temperatures, where carrier heating should be negligible.

Intuitively, we can view the local electrochemical potential as a measure of the electron energy at a point. We can also view it as a measure of the excess electron density;⁷ however, part of this excess charge is neutralized by screening charges, as described in Sec. VI. The resulting electrostatic potential is given by the electrochemical potential convolved with a screening function. In a highly conductive medium the screening length is very short so that the two potentials are nearly identical. But in less

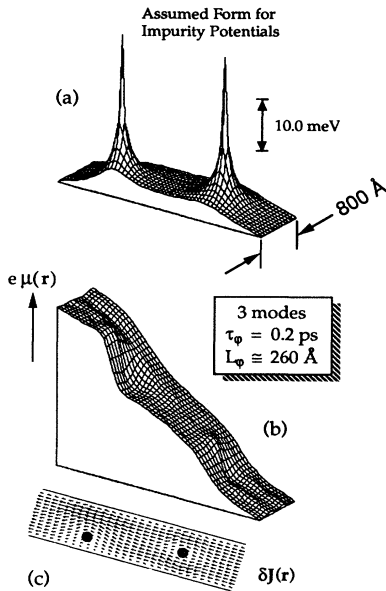


FIG. 2. Potential drop and current flow pattern in a wire with three modes having two localized impurities ($\tau_\phi = 0.2$ ps). (a) Assumed impurity potential; (b) potential variation; (c) current flow pattern.

conductive structures, the electrostatic potential varies slowly compared to the electrochemical potential, and the local space charge is proportional to the difference between the two potentials.

It is often appealing to define separate electrochemical potentials μ_L and μ_R for left-moving and right-moving carriers. Using the present approach, we can compute the left-moving and right-moving electron densities $n_L(\mathbf{r}; E)$ and $n_R(\mathbf{r}; E)$ individually from the Wigner function. However, these quantities are not positive definite, unlike the total electron density $n = n_L + n_R$, so that the local electrochemical potentials μ_L and μ_R may show unphysical behavior, as we will demonstrate with a numerical example of Sec. III. We believe this problem arises because the concept of left-moving and right-moving electrons at a particular point is in violation of the uncertainty principle. We will also show that this unphysical behavior can be avoided by defining these potentials in an average sense over regions the size of a de Broglie wavelength. The results obtained after averaging behave as we might intuitively expect.^{27,28}

We would like to stress that the local electrochemical potential $\mu(\mathbf{r})$ defined by us characterizes the *energy distribution* of the electrons and has nothing to do with the *momentum distribution*. A common source of misunderstanding is to assume that if a local electrochemical potential exists, then there must be equal numbers of electrons moving to the left and to the right. This is not true in general even in equilibrium. For example, in the quantum Hall regime, even at equilibrium, all electrons move to the left near one edge and to the right near the other edge. However, we believe that the energy distribution of electrons is still characterized by an electrochemical potential throughout the structure. It will also be noted that while \mathbf{k} is conjugate to \mathbf{r} , E is the conjugate to t . We can thus talk about the energy distribution (but not the momentum distribution) at a point without violating the uncertainty principle.

The local electrochemical potential μ is difficult to measure for two reasons, which we examine in Sec. IV. First, measurement probes are usually *invasive*, that is, they disturb the electrochemical potential within the device. It should, however, be possible to design weakly coupled noninvasive probes that do not cause any significant perturbation. But even with such a noninvasive probe, the potential that is finally measured may depend on the details of the probe-to-sample coupling. We present two numerical examples—one where this is not the case and one where it is. We first consider a narrow wire with a tunneling barrier where the intrinsic electrochemical potential oscillates rapidly. We find that noninvasive probes render a surprisingly accurate measurement of the oscillations regardless of their geometry. We then consider the “quenching” of the Hall resistance observed in narrow wires at low magnetic fields.^{29–35} By directly computing the electrochemical potential in a narrow wire with attached probes, we show that the usual Hall voltage is present near the voltage probes, and yet the probes fail to measure it properly. Noninvasive probes thus do not guarantee a faithful measurement as stressed by Büttiker.³⁶

Although the electrochemical potential may be difficult to measure precisely, it is still a useful concept that can lend insight into the transport physics. For example, in Sec. V we compute the electrochemical potential for a ballistic device (Fig. 1). We find that the potential drop across the “interesting” region (between points A and B) is nearly zero, as we might expect, since there is no scattering within the device. There are sharp drops at A and B where the current funnels into and out of the narrow region; this is due to the spreading resistance described by Imry.⁶ In addition, we find drops at A' and B' where the *boundary conditions are applied*. This may seem surprising since there are no apparent obstacles to current flow at A' and B' . To explain the drops at A' and B' , we note that the assumption of a constant electrochemical potential is accurate only for an infinitely wide lead with zero series resistance, as emphasized by Landauer. As a corollary to this argument, we can state that imposing the constant potential boundary conditions at A' and B' is *physically equivalent* to widening the leads to infinity at these points. Consequently, there are spreading resistances at A' and B' similar to those at A and B . We also examine the contact resistance at A and B and show that it is reduced when the width of the leads is reduced, as argued by Landauer. However, we find that the full contact resistance is recovered if the width of the leads is much larger than the phase-breaking length L_ϕ . We believe that this is because a lead that is much wider than L_ϕ is effectively infinite in width.

The equivalence between a constant potential boundary condition and an infinitely wide lead is very important from the standpoint of numerical simulation. Actual experimental structures typically have leads that are infinitely wide for all practical purposes. Such wide leads are very difficult to simulate directly. However, if we are not interested in a detailed description of current flow in the lead, then we can neglect any widening altogether and simply impose a constant μ boundary condition when solving the kinetic equation, Eq. (1.5).

Localized drops in the electrochemical potential occur around obstacles because electrons pileup on one side and drain off on the other.⁷ The resulting dipole, together with the induced screening charge, acts as a source term in the Poisson equation giving rise to an electrostatic potential ϕ . In Sec. VI we show that the electrostatic potential ϕ can be viewed as a convolution of the electrochemical potential μ with a screening function and compute ϕ for a few examples.

II. GENERAL APPROACH

Any device analysis—semiclassical or quantum—involves the simultaneous solution of a transport equation and the Poisson equation. In Sec. II A, we outline the steps involved in obtaining such a solution. These steps are identical for both the semiclassical and the quantum regime; the only difference lies in the particular transport equation that is solved. In the semiclassical regime, we solve either the Boltzmann equation or the drift-diffusion equation. In the quantum regime, we solve a quantum kinetic equation, such as the one developed by

Kadanoff and Baym and Keldysh.¹⁷ A very brief overview of this quantum kinetic equation is presented in Sec. II B. In Sec. II C we review the simple model that we assume for phase-breaking scatterers in order to simplify the kinetic equation. A further simplification is achieved by restricting to near-equilibrium solutions at low bias as described in Sec. II D. Finally, in Sec. II E we summarize the simplified equations and describe the solution technique.

A. Overview of device analysis

Step 1 in any device analysis is to obtain the equilibrium band diagram (Fig. 3). Throughout this paper, we restrict our attention to only one type of carrier, namely electrons in the conduction band. To obtain the equilibrium conduction-band profile, we solve the Poisson equation for the electrostatic potential ϕ ,

$$\nabla^2\phi(\mathbf{r}) = \frac{e}{\epsilon} [N_D^+(\mathbf{r}) - n_{\text{eq}}(\mathbf{r})], \quad (2.1)$$

self-consistently with the equilibrium electron density $n_{\text{eq}}(\mathbf{r})$,

$$n_{\text{eq}}(\mathbf{r}) = \int dE N_0(\mathbf{r}; E) f_0(E), \quad (2.2)$$

where $N_D^+(\mathbf{r})$ is the density of ionized donors, and $N_0(\mathbf{r}; E)$ is the local electronic density of states, which depends implicitly on ϕ . One way of solving these two equations is to assume some initial form for ϕ , and then iterate between them until the solution converges. Although this procedure is conceptually simple, it is computationally intensive, since Eq. (2.2) requires us to compute the density of states at all energies. To simplify our calculations, we neglect this first step, and simply assume some form for $\phi(\mathbf{r})$ in equilibrium. However, this step could be important in describing certain phenomenon such as the influence of electron-electron interactions on

weak localization.

Step 2 is to solve the transport equation subject to the boundary conditions imposed at the contacts. We assume that these contacts remain in local equilibrium and that some external source shifts the electrochemical potential μ of one contact relative to another. By solving the transport equation in the region between contacts, we determine both the electron density and the current density within the device. Note that we are treating an open system where particles move in and out of the device freely. This is appropriate for transport problems. For problems involving optical or magnetic properties a canonical ensemble with a fixed number of particles may be more appropriate.

If electrons were neutral particles, this would complete our analysis. However, since electrons are charged particles, the charge imbalances arising under bias, together with the associated screening charges, lead to a correction $\delta\phi$ to the electrostatic potential. Step 3 is to solve the Poisson equation for this correction $\delta\phi$ to the electrostatic potential. In general, this correction could influence the solution of the transport problem (step 2). So we should iterate between steps 2 and 3 until convergence is achieved. Self-consistent calculations of this sort have been performed in the analysis of resonant-tunneling diodes in the high-bias regime.^{22,23}

This paper is primarily concerned with the *linear-response regime*. It is believed that in this regime it is not necessary to iterate between steps 2 and 3. This is because the correction $\delta\phi$ to the electrostatic potential does not lead to any first-order change in the terminal currents. It could, however, lead to a first-order change in the detailed current flow pattern within the device, as discussed in Sec. III B. In much of this paper (Secs. III–V), we concentrate on step 2 and discuss solutions to the transport equation. It is only in Sec. VI that we carry out step 3, to illustrate the nature of the correction $\delta\phi$; however, we neglect any influence that $\delta\phi$ might have on step 2, leaving it for future investigations.

B. Quantum kinetic equation

Less than 30 years ago, Kadanoff and Baym and Keldysh¹⁷ developed a general framework for quantum transport which we use as our starting point. The central quantity in their formulation is the correlation function $G^<$,

$$-i\hbar G^<(\mathbf{r}_1, \mathbf{r}_2; t_1, t_2) = \langle \psi^\dagger(\mathbf{r}_2, t_2) \psi(\mathbf{r}_1, t_1) \rangle, \quad (2.3)$$

where ψ^\dagger and ψ are electron field operators. It is often convenient to transform to center of mass and relative coordinates, and then Fourier transform with respect to the relative coordinates,

$$\frac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2) \rightarrow \mathbf{r}, \quad \mathbf{r}_1 - \mathbf{r}_2 \rightarrow \mathbf{k}, \quad (2.4a)$$

$$\frac{1}{2}(t_1 + t_2) \rightarrow t, \quad t_1 - t_2 \rightarrow E, \quad (2.4b)$$

and express the resulting function as $G^<(\mathbf{r}; \mathbf{k}; E; t)$. We can make an immediate simplification if we restrict ourselves to steady state, so that we are left with only three

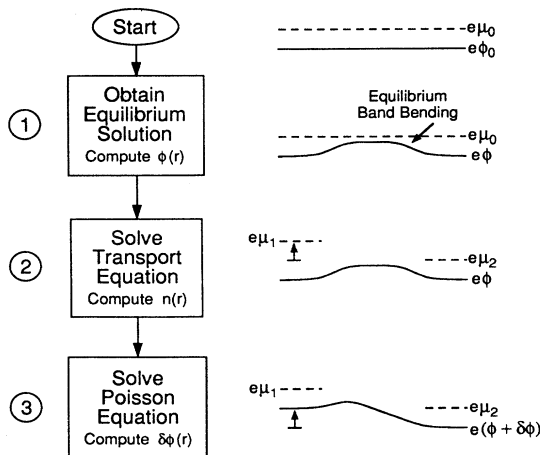


FIG. 3. Overview of the procedure for device analysis.

coordinates $(\mathbf{r}; \mathbf{k}; E)$.

The Wigner function $-iG^<(\mathbf{r}; \mathbf{k}; E)$ is the quantum-mechanical analog of the semiclassical distribution function. Any observable can be computed by taking moments of this function.^{18,19} For instance, the electron density and the current density can be expressed as

$$n(\mathbf{r}; E) = \frac{-i}{2\pi} \int \frac{d^3\mathbf{k}}{(2\pi)^3} G^<(\mathbf{r}; \mathbf{k}; E), \quad (2.5)$$

$$\mathbf{J}(\mathbf{r}; E) = \frac{-i}{2\pi} \int \frac{d^3\mathbf{k}}{(2\pi)^3} e \left[\frac{\hbar\mathbf{k} - e\mathbf{A}(\mathbf{r})}{m} \right] G^<(\mathbf{r}; \mathbf{k}; E). \quad (2.6)$$

These quantities must then be integrated over energy to obtain the total electron density $n(\mathbf{r})$ and total current density $\mathbf{J}(\mathbf{r})$ at each position,

$$n(\mathbf{r}) = \int dE n(\mathbf{r}; E), \quad (2.7)$$

$$\mathbf{J}(\mathbf{r}) = \int dE \mathbf{J}(\mathbf{r}; E). \quad (2.8)$$

The $(\mathbf{r}; \mathbf{k})$ representation is useful in bringing out the analogy between the Wigner function $-iG^<$ and the semiclassical distribution function. However, we could just as well express the electron density (2.5) and the current density (2.6) in the $(\mathbf{r}, \mathbf{r}')$ representation. By making use of the Fourier-transform relationship (2.4a), we obtain

$$n(\mathbf{r}; E) = \frac{-i}{2\pi} G^<(\mathbf{r}, \mathbf{r}; E), \quad (2.9)$$

$$\begin{aligned} \mathbf{J}(\mathbf{r}; E) &= \frac{-e\hbar}{4\pi m} (\nabla - \nabla') G^<(\mathbf{r}, \mathbf{r}'; E) \Big|_{\mathbf{r}=\mathbf{r}}, \\ &+ \frac{ie^2}{2\pi m} \mathbf{A}(\mathbf{r}) G^<(\mathbf{r}, \mathbf{r}; E). \end{aligned} \quad (2.10)$$

To compute the correlation function $G^<$, we solve the quantum kinetic equation,^{25,37}

$$\begin{aligned} G^<(\mathbf{r}_1, \mathbf{r}_2; E) &= \int d^3\mathbf{r}' \int d^3\mathbf{r}'' G^R(\mathbf{r}_1, \mathbf{r}'; E) \\ &\times G^{R*}(\mathbf{r}_2, \mathbf{r}''; E) \Sigma^<(\mathbf{r}', \mathbf{r}''; E), \end{aligned} \quad (2.11)$$

where G^R is the Green function of the Schrödinger equation modified to include a nonlocal optical potential Σ^R ,

$$\begin{aligned} [E - H_0(\mathbf{r}_1)] G^R(\mathbf{r}_1, \mathbf{r}_2; E) - \int d^3\mathbf{r}' \Sigma^R(\mathbf{r}_1, \mathbf{r}'; E) \\ \times G^R(\mathbf{r}', \mathbf{r}_2; E) = \delta(\mathbf{r}_1 - \mathbf{r}_2). \end{aligned} \quad (2.12)$$

Physically, these equations can be understood as follows (Fig. 4). $\Sigma^<$ represents the in-scattering rate from other energies, while $G^R(\mathbf{r}_1, \mathbf{r}_2; E)$ represents the wave function at \mathbf{r}_1 if an electron is injected at \mathbf{r}_2 with energy E by scattering from another energy. The kinetic equation (2.11) states that the overall correlation $G^<(\mathbf{r}_1, \mathbf{r}_2; E)$ between points \mathbf{r}_1 and \mathbf{r}_2 is given by the product of three factors: the initial correlation $\Sigma^<(\mathbf{r}', \mathbf{r}''; E)$ of the in-scattered electrons, the correlation $G^R(\mathbf{r}_1, \mathbf{r}'; E)$ due to propagation from \mathbf{r}' to \mathbf{r}_1 , and the correlation

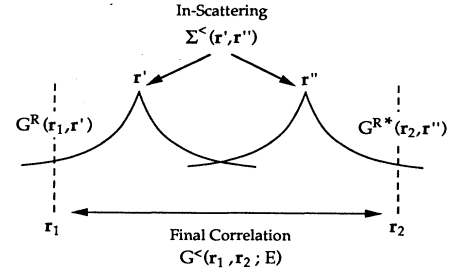


FIG. 4. A physical interpretation of the quantum kinetic equation (2.11).

$G^{R*}(\mathbf{r}_2, \mathbf{r}''; E)$ due to propagation from \mathbf{r}'' to \mathbf{r}_2 .

In principle, we could use the kinetic equation (2.11) as the transport equation in step 2 of the device analysis (Fig. 3). For two-dimensional structures, however, this is extremely difficult due to the large number (five) of independent variables. Following Ref. 25, we use a simple model for phase-breaking scatters, which reduces the self-energy to a local potential; this is briefly outlined in the following section. With this simplification, the number of independent variables in the kinetic equation is reduced to three.

C. Simple model for phase-breaking scattering

We assume that the propagation of electrons is described by a one-electron effective-mass Hamiltonian of the form

$$H_0 = \frac{[\mathbf{p} - e\mathbf{A}(\mathbf{r})]^2}{2m} + e\phi(\mathbf{r}) + V(\mathbf{r}), \quad (2.13)$$

where the scalar potential $\phi(\mathbf{r})$ is obtained from the Poisson equation, as discussed in Sec. II A; it includes band bending due to space charge and external bias. Electron-electron interactions beyond this Hartree term are neglected. Any band discontinuities due to heterojunctions, as well as all sources of elastic scattering such as impurities, defects, and boundaries, are included in the potential $V(\mathbf{r})$. This part H_0 of the total Hamiltonian is treated exactly.

Phase-breaking scattering occurs when electrons interact with a bath of independent oscillators labeled by an index m ,

$$H_R = \sum_m \hbar\omega_m (a_m^\dagger a_m + \frac{1}{2}), \quad (2.14)$$

where a_m^\dagger and a_m are the creation and annihilation operators for oscillator m . We assume that each oscillator interacts with electrons through a δ potential,

$$H' = \sum_m U \delta(\mathbf{r} - \mathbf{r}_m) (a_m^\dagger - a_m), \quad (2.15)$$

where the interaction strength U is assumed to be constant, although the spatial density of oscillators may vary. This simple model would well represent point-size

phase-breaking scatterers, such as magnetic impurities. It does not represent true phonons, which have nonzero spatial extent, but it might still describe much of the physics of dissipative transport. In any case, it provides a tractable microscopic model whose predictions can be compared with experiment.

The interaction part H' of the Hamiltonian is treated approximately (taking only one-phonon processes into account) through the self-energy functions $\Sigma^<$ and Σ^R , which are local functions in this model, and can be written in the form²⁵

$$\Sigma^<(\mathbf{r}, \mathbf{r}'; E) = \frac{i\hbar}{\tau_p(\mathbf{r}; E)} \delta(\mathbf{r} - \mathbf{r}'), \quad (2.16a)$$

$$\Sigma^R(\mathbf{r}, \mathbf{r}'; E) = \frac{-i\hbar}{2\tau_\phi(\mathbf{r}; E)} \delta(\mathbf{r} - \mathbf{r}'). \quad (2.16b)$$

Actually, Σ^R also has a real part,²⁵ which we are neglecting in this paper. Using expressions (2.16a) and (2.16b), one of the integrals in the quantum kinetic equation (2.11) and the Schrödinger equation (2.12) can be performed trivially:

$$\begin{aligned} G^<(\mathbf{r}_1, \mathbf{r}_2; E) &= \int d^3\mathbf{r}' G^R(\mathbf{r}_1, \mathbf{r}'; E) G^{R*}(\mathbf{r}_2, \mathbf{r}'; E) \\ &\quad \times \frac{i\hbar}{\tau_p(\mathbf{r}'; E)}, \quad (2.17) \\ \left[E - H_0(\mathbf{r}_1) + \frac{i\hbar}{2\tau_\phi(\mathbf{r}_1; E)} \right] G^R(\mathbf{r}_1, \mathbf{r}_2; E) &= \delta(\mathbf{r}_1 - \mathbf{r}_2). \quad (2.18) \end{aligned}$$

If we now specialize (2.17) to just the diagonal elements $G^<(\mathbf{r}, \mathbf{r}; E)$, we obtain

$$n(\mathbf{r}; E) = \frac{\hbar}{2\pi} \int d^3\mathbf{r}' |G^R(\mathbf{r}, \mathbf{r}'; E)|^2 \frac{1}{\tau_p(\mathbf{r}'; E)}, \quad (2.19)$$

noting from Eq. (2.9) that the diagonal elements $G^<(\mathbf{r}, \mathbf{r}; E)$ are simply related to the electron density per unit energy $n(\mathbf{r}; E)$. The function $\tau_p(\mathbf{r}; E)$ depends only on the diagonal elements $n(\mathbf{r}; E)$ and not on the off-diagonal elements $G^<(\mathbf{r}, \mathbf{r}'; E)$, so that Eq. (2.19) only involves the diagonal elements. Once we have solved it to compute $n(\mathbf{r}; E)$ [and hence $\tau_p(\mathbf{r}; E)$, which is a functional of $n(\mathbf{r}; E)$], we can substitute τ_p into Eq. (2.17), perform an integral, and obtain the full correlation function $G^<(\mathbf{r}_1, \mathbf{r}_2; E)$. Any quantity of interest can then be obtained from this function.

D. Energy decoupling

In the preceding section, we adopted a simple model for phase-breaking scatterers, which reduced the self-energies $\Sigma^<$ and Σ^R to local functions, and reduced the independent variables in the resulting kinetic equation to \mathbf{r} and E . We can make a further simplification if we confine our attention to low-bias voltages, and low temperatures. In this regime, we can neglect the vertical flow of carriers from one energy to another and thereby decouple different energies.

At equilibrium, it can be shown²⁵ that the following

solution satisfies the kinetic equation (2.19) for the diagonal elements $n(\mathbf{r}; E)$:

$$n(\mathbf{r}; E) = f_0(E) N_0(\mathbf{r}; E), \quad (2.20)$$

where $N_0(\mathbf{r}; E)$ is the local density of states, and $f_0(E)$ is the Fermi-Dirac factor with an equilibrium electrochemical potential μ_0 ,

$$f_0(E) \equiv \frac{1}{e^{(E - e\mu_0)/k_B T} + 1}. \quad (2.21)$$

This result is reassuring, since this form of the equilibrium solution is known from more general considerations. Furthermore, it suggests that the solution for $n(\mathbf{r}; E)$ can in general be expressed as the product of some occupation function $f(\mathbf{r}; E)$ and the local density of states $N_0(\mathbf{r}; E)$,

$$n(\mathbf{r}; E) \equiv f(\mathbf{r}; E) N_0(\mathbf{r}; E). \quad (2.22)$$

In equilibrium, $f(\mathbf{r}; E)$ reduces to the Fermi-Dirac factor $f_0(E)$; away from equilibrium, $f(\mathbf{r}; E)$ may be significantly distorted. We define $p(\mathbf{r}; E)$ as the density of holes in the conduction band,

$$p(\mathbf{r}; E) = [1 - f(\mathbf{r}; E)] N_0(\mathbf{r}; E). \quad (2.23)$$

This density of holes, together with the density of electrons, accounts for the density of states in the conduction band,

$$n(\mathbf{r}; E) + p(\mathbf{r}; E) = N_0(\mathbf{r}; E). \quad (2.24)$$

Next, consider the expressions for the hole lifetime τ_p and the electron lifetime τ_n ,²⁵

$$\frac{1}{\tau_p(\mathbf{r}; E)} = \frac{2\pi}{\hbar} \int d(\hbar\omega) F(\mathbf{r}; \hbar\omega) n(\mathbf{r}; E - \hbar\omega), \quad (2.25a)$$

$$\frac{1}{\tau_n(\mathbf{r}; E)} = \frac{2\pi}{\hbar} \int d(\hbar\omega) F(\mathbf{r}; \hbar\omega) p(\mathbf{r}; E + \hbar\omega), \quad (2.25b)$$

where $F(\mathbf{r}; \hbar\omega)$ describes the spectrum of the oscillators causing phase-breaking scattering,

$$F(\mathbf{r}; \hbar\omega) = \begin{cases} U^2 J_0(\mathbf{r}; \hbar\omega) N(\hbar\omega), & \hbar\omega > 0, \\ U^2 J_0(\mathbf{r}; |\hbar\omega|) [N(|\hbar\omega|) + 1], & \hbar\omega < 0 \end{cases} \quad (2.26)$$

and $J_0(\mathbf{r}; \hbar\omega)$ is the density of oscillators per unit volume, per unit energy; $N(\hbar\omega)$ is the Bose-Einstein factor describing the average number of “phonons” with frequency ω in each oscillator. The phase-breaking time τ_ϕ is the parallel combination of the electron lifetime τ_n and the hole lifetime τ_p ,

$$\frac{1}{\tau_\phi(\mathbf{r}; E)} = \frac{1}{\tau_n(\mathbf{r}; E)} + \frac{1}{\tau_p(\mathbf{r}; E)}, \quad (2.27)$$

The divergence of the current density per unit energy $\mathbf{J}(\mathbf{r}; E)$ is proportional to the difference between the outscattering rate for holes and that for electrons,

$$\nabla \cdot \mathbf{J}(\mathbf{r}; E) = e \left[\frac{p(\mathbf{r}; E)}{\tau_p(\mathbf{r}; E)} - \frac{n(\mathbf{r}; E)}{\tau_n(\mathbf{r}; E)} \right]. \quad (2.28)$$

The left-hand side represents the vertical flow of carriers from one energy to another. This quantity can be set equal to zero if (1) the phase-breaking processes are purely elastic, or if (2) the entire energy range over which transport occurs can be considered a single channel (because of the uniformity of propagation characteristics). In either case, we obtain

$$\frac{n(\mathbf{r};E)}{\tau_n(\mathbf{r};E)} = \frac{p(\mathbf{r};E)}{\tau_p(\mathbf{r};E)}. \quad (2.29)$$

We will have changed nothing if we simply add n/τ_p to both sides,

$$\frac{n(\mathbf{r};E)}{\tau_\phi(\mathbf{r};E)} = \frac{N_0(\mathbf{r};E)}{\tau_p(\mathbf{r};E)}. \quad (2.30)$$

This expression gives a particularly simple relation between τ_p and τ_ϕ ,

$$\frac{1}{\tau_p(\mathbf{r};E)} = \frac{f(\mathbf{r};E)}{\tau_\phi(\mathbf{r};E)}. \quad (2.31)$$

If we use this relationship to replace $1/\tau_p$ in the kinetic equation (2.19), we obtain an integral equation for $f(\mathbf{r};E)$,

$$f(\mathbf{r};E) = \int d^3\mathbf{r}' \frac{\hbar |G^R(\mathbf{r}, \mathbf{r}'; E)|^2}{2\pi\tau_\phi(\mathbf{r}'; E)N_0(\mathbf{r}; E)} f(\mathbf{r}'; E). \quad (2.32)$$

It may not be apparent what we have gained by replacing $1/\tau_p$ in the kinetic equation (2.19). In Eq. (2.19), the solutions for different energies are coupled together, since τ_p at an energy E depends on $n(\mathbf{r}; E')$ at all other energies E' [see Eq. (2.24a)]. On the other hand, τ_ϕ is nearly independent of $n(\mathbf{r}; E)$, which we can see by expanding its definition (2.28),

$$\frac{1}{\tau_\phi(\mathbf{r}; E)} = \frac{2\pi}{\hbar} \int d(\hbar\omega) F(\mathbf{r}; \hbar\omega) [n(\mathbf{r}; E - \hbar\omega) + p(\mathbf{r}; E + \hbar\omega)]. \quad (2.33)$$

If the spectral function $F(\mathbf{r}; \hbar\omega)$ is relatively peaked near $\hbar\omega=0$, then the factor within the parentheses is approximately $n+p=N_0$, independent of the occupation of states.²⁵ The same is true for nondegenerate systems with $n \ll p \simeq N_0$.³⁷ In these cases, $\tau_\phi(\mathbf{r}; E)$ is independent of the solution $f(\mathbf{r}; E)$ of the kinetic equation (2.32). In general, however, we should iterate between Eqs. (2.32) and (2.33) to obtain self-consistent solutions for $f(\mathbf{r}; E)$ and $\tau_\phi(\mathbf{r}; E)$. Since τ_ϕ is insensitive to changes in f , this process should require only a few iterations to achieve convergence.

If we neglect this dependence of $\tau_\phi(\mathbf{r}; E)$ on $f(\mathbf{r}; E)$, then the energies are completely decoupled in the kinetic equation (2.32). Thus, each energy can be treated independently, and the solution $f(\mathbf{r}; E)$ at different energies E can be computed one at a time, for all points \mathbf{r} . Note that the essential condition for achieving this energy decoupling is the neglect of vertical flows [$\nabla \cdot \mathbf{J}(\mathbf{r}; E) \simeq 0$]

in going from Eqs. (2.28) to (2.29). In our previous work,²⁵ this decoupling was achieved by assuming that the occupation factor $f(\mathbf{r}; E)$ is given by a Fermi-Dirac function with a local electrochemical potential $\mu(\mathbf{r})$. This is a sufficient condition for $\nabla \cdot \mathbf{J}(\mathbf{r}; E)$ to be zero but not a necessary condition. For example, if the phase-breaking processes are purely elastic, then $\nabla \cdot \mathbf{J}(\mathbf{r}; E) = 0$ regardless of the shape of the occupation factor $f(\mathbf{r}, E)$.

E. Solving the simplified kinetic equation

In the preceding sections, we have discussed the various simplifying assumptions that reduce the quantum kinetic equation (2.11) to a more manageable form (2.32) that can be used in step 2 of the device analysis (Fig. 3). We will now describe how we can numerically solve the simplified kinetic equation (2.32) for a particular device. A block diagram of the solution procedure is shown in Fig. 5. We begin by assuming the equilibrium form for the solution $f(\mathbf{r}; E)$, so that we can obtain an initial guess for the phase-breaking time $\tau_\phi(\mathbf{r}; E)$. We then compute the Green function $G^R(\mathbf{r}, \mathbf{r}'; E)$ and the density of states $N_0(\mathbf{r}; E)$. These quantities determine the kernel of the kinetic equation (2.32). Within each contact region, the occupation factor $f(\mathbf{r}; E)$ is assumed to have a Fermi-Dirac form with a fixed electrochemical potential. In the intermediate device regions, $f(\mathbf{r}; E)$ is computed by solving

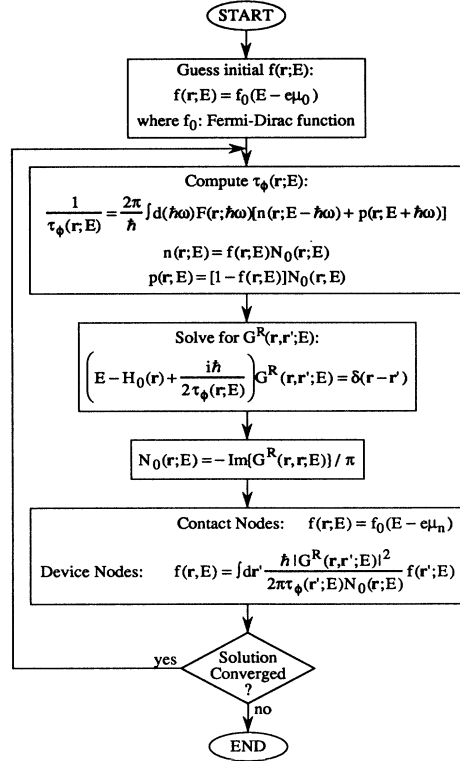


FIG. 5. Overview of the procedure for solving the simplified kinetic equation.

the kinetic equation (2.32). As we have already noted, $\tau_\phi(\mathbf{r}; E)$ depends weakly on the solution for $f(\mathbf{r}; E)$. Strictly speaking, the solution process should be repeated until both $f(\mathbf{r}; E)$ and $\tau_\phi(\mathbf{r}; E)$ form a self-consistent solution. In practice, the change in $\tau_\phi(\mathbf{r}; E)$ may be small enough that an iterative solution may not be necessary.

It will be noted that in our model the local phase-breaking time $\tau_\phi(\mathbf{r}; E)$ depends on the local density of states $N_0(\mathbf{r}; E)$. Since the local density of states often oscillates by orders of magnitude in mesoscopic structures, the phase-breaking time too should show large oscillations. However, realistic scattering processes are usually not local, and it is likely that the phase-breaking time at a point is determined by the density of states averaged over the surrounding volume rather than by the density of states at the point. Large spatial fluctuations in τ_ϕ are thus not expected in realistic structures unless the phase-breaking scatterers are truly local (such as magnetic boundaries). In this paper we will neglect these issues altogether and treat $\tau_\phi(\mathbf{r})$ as a specified input parameter independent of $N_0(\mathbf{r}; E)$ or $f(\mathbf{r}; E)$. The iteration indicated in Fig. 5 is thus unnecessary. We leave a more detailed treatment of the phase-breaking time to future work.

Once we have computed the occupation factor $f(\mathbf{r}; E)$, we can reconstruct the correlation function within the device. Using the simple approximation for $1/\tau_p$ (2.31) in the kinetic equation (2.17), we obtain

$$G^<(\mathbf{r}_1, \mathbf{r}_2; E) = \int d^3\mathbf{r}' f(\mathbf{r}'; E) G_\delta(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}'; E), \quad (2.34)$$

where

$$G_\delta(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}'; E) \equiv i\hbar \frac{G^R(\mathbf{r}_1, \mathbf{r}'; E) G^{R*}(\mathbf{r}_2, \mathbf{r}'; E)}{\tau_\phi(\mathbf{r}'; E)}, \quad (2.35)$$

With this correlation function, we can compute any quantity of interest. In particular, we can compute the electron density $n(\mathbf{r}; E)$,

$$n(\mathbf{r}; E) = \int d^3\mathbf{r}' f(\mathbf{r}'; E) n_\delta(\mathbf{r}, \mathbf{r}'; E), \quad (2.36)$$

and the current density $\mathbf{J}(\mathbf{r}; E)$,

$$\mathbf{J}(\mathbf{r}; E) = \int d^3\mathbf{r}' f(\mathbf{r}'; E) \mathbf{J}_\delta(\mathbf{r}, \mathbf{r}'; E), \quad (2.37)$$

where

$$n_\delta(\mathbf{r}, \mathbf{r}'; E) \equiv \frac{-i}{2\pi} G_\delta(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}'; E) \Big|_{\mathbf{r}_1=\mathbf{r}_2=\mathbf{r}} \quad (2.38a)$$

$$= \frac{-i}{2\pi} \frac{|G^R(\mathbf{r}, \mathbf{r}'; E)|^2}{\tau_\phi(\mathbf{r}'; E)}, \quad (2.38b)$$

$$\mathbf{J}_\delta(\mathbf{r}, \mathbf{r}'; E) \equiv \frac{-ie}{2\pi m} \left[\frac{\hbar}{2i} (\nabla_1 - \nabla_2) - e \mathbf{A}(\mathbf{r}_1) \right] G_\delta(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}'; E) \Big|_{\mathbf{r}_1=\mathbf{r}_2=\mathbf{r}} \quad (2.39a)$$

$$= \frac{e\hbar^2}{2\pi m} \frac{1}{\tau_\phi(\mathbf{r}'; E)} \left[\text{Im}[G^{R*}(\mathbf{r}, \mathbf{r}'; E) \nabla_1 G^R(\mathbf{r}, \mathbf{r}'; E)] - \frac{e}{\hbar} \mathbf{A}(\mathbf{r}) |G^R(\mathbf{r}, \mathbf{r}'; E)|^2 \right]. \quad (2.39b)$$

It is interesting to note that, by using the kinetic Eq. (2.32), we can simplify the expression for the electron density (2.36) to $n(\mathbf{r}; E) = N_0(\mathbf{r}; E) f(\mathbf{r}; E)$, as we would expect. Furthermore, if we evaluate the divergence of the current $\nabla \cdot \mathbf{J}$ from Eq. (2.37) and set it equal to zero, we obtain the kinetic Eq. (2.32). To show this, it is necessary to make use of an expression for the divergence of \mathbf{J}_δ [see Eq. (C5) of Ref. 25(b), noting that \mathbf{J}_δ as defined here has an extra factor of $\tau_\phi(\mathbf{r}'; E)$],

$$\frac{1}{e} \nabla \cdot \mathbf{J}_\delta(\mathbf{r}, \mathbf{r}'; E) = \frac{2\pi}{\hbar} \frac{N_0(\mathbf{r}; E)}{\tau_\phi(\mathbf{r}; E)} \delta(\mathbf{r} - \mathbf{r}') - \frac{|G^R(\mathbf{r}, \mathbf{r}'; E)|^2}{\tau_\phi(\mathbf{r}; E) \tau_\phi(\mathbf{r}'; E)}. \quad (2.40)$$

In Secs. III B and III C we will simplify these equations to linear response. As we will see, the equations describing linear response look just like the above equations but with local occupation factor $f(\mathbf{r}; E)$ replaced by the local electrochemical potential $\mu(\mathbf{r})$.

III. DEFINING THE ELECTROCHEMICAL POTENTIAL

In Sec. III A we show that under low-bias conditions at low temperatures the energy distribution of carriers can be characterized everywhere by a local electrochemical potential $\mu(\mathbf{r})$. In Sec. III B we use this property to obtain a linear-response kinetic equation in terms of $\mu(\mathbf{r})$. We then derive an alternative expression for the terminal current and show that this kinetic equation can be viewed as a continuous probe version of the Büttiker formula (Sec. III C). Finally in Sec. III D we present a numerical example to illustrate the conceptual difficulties associated with defining separate local potentials μ_L and μ_R for left- and right-moving electrons. We show that these difficulties can be overcome by defining these potentials in an average sense over regions the size of a de Broglie wavelength.

We would like to stress here that the local electrochemical potential $\mu(\mathbf{r})$ defined by us characterizes the *energy distribution* of the electrons at a point \mathbf{r} . It has

nothing to do with the *momentum distribution*. Thus the existence of a local potential does not imply that the number of electrons moving to the left equals that moving to the right. Such an assumption could be wrong *even at equilibrium*. For example, in the quantum Hall regime, at equilibrium, all electrons move to the left near one edge and to the right near the other edge.

A. When can carrier heating be neglected?

It is usually assumed that a certain degree of inelastic scattering is necessary to maintain a local electrochemical potential. We will show in this section that even in the absence of inelastic scattering, the occupation factor $f(\mathbf{r};E)$ has a Fermi-Dirac form (2.23) with a local electrochemical potential $\mu(\mathbf{r})$, if $eV_A < k_B T < \Gamma_c$, where V_A is the applied bias and Γ_c is the correlation energy over which the propagation characteristics of the electrons can be assumed constant.

We start from the simplified kinetic equation (2.32) which is derived neglecting all vertical flow due to inelastic processes. We have numerically solved (2.32) for a single-moded wire with a tunneling barrier in the middle (Fig. 6). Under a small bias, we find that the solution $f(\mathbf{r};E)$ in the middle of the barrier closely approximates a Fermi-Dirac factor. For a larger bias, however, the solution deviates significantly from the Fermi-Dirac form. This distortion of the occupation factor $f(\mathbf{r};E)$ can be viewed as “carrier heating.” To lowest order, it could be described by introducing a local electron temperature, indicating the degree to which the occupation factor is smeared out in energy.

We can understand this behavior in a qualitative sense as follows. The occupation factor at any point is computed as a weighted average of the occupation factors at surrounding points [see Eq. (2.32)],

$$f(\mathbf{r};E) = \int d^3\mathbf{r}' K(\mathbf{r},\mathbf{r}';E) f(\mathbf{r}';E). \quad (3.1a)$$

It can be shown²⁵ that the kernel $K(\mathbf{r},\mathbf{r}';E)$ has the following normalization property:

$$\int d^3\mathbf{r}' K(\mathbf{r},\mathbf{r}';E) = 1. \quad (3.1b)$$

Let us assume that there is hardly any phase-breaking scattering within the device so that the integral in (3.1) only needs to include the two contacts. Within each contact the distribution function has a Fermi-Dirac form with a constant electrochemical potential, so that

$$f(\mathbf{r};E) \simeq f_1(E) \int_{\mathbf{r}' \in \text{contact 1}} d^3\mathbf{r}' K(\mathbf{r},\mathbf{r}';E) + f_2(E) \int_{\mathbf{r}' \in \text{contact 2}} d^3\mathbf{r}' K(\mathbf{r},\mathbf{r}';E). \quad (3.2)$$

The distribution function $f(\mathbf{r};E)$ at a particular point \mathbf{r} within the device is then a weighted average of the distribution functions $f_1(E)$ and $f_2(E)$ at the two contacts,

$$f(\mathbf{r};E) \simeq K_1(\mathbf{r};E) f_1(E) + K_2(\mathbf{r};E) f_2(E) = \frac{k_1(\mathbf{r};E)}{e^{(E-e\mu_1)/k_B T} + 1} + \frac{K_2(\mathbf{r};E)}{e^{(E-e\mu_2)/k_B T} + 1}, \quad (3.3)$$

where $K_1(\mathbf{r};E) + K_2(\mathbf{r};E) = 1$, due to the normalization

property (3.1b). Even in the presence of phase-breaking processes it is possible to express $f(\mathbf{r};E)$ in the form indicated in (3.3) using the iterative solution procedure described on p. 8040 in Ref. 25(b).

This equation states that the solution $f(\mathbf{r};E)$ at the interior node is an average of the distributions $f_1(E)$ and $f_2(E)$ imposed at the contacts. We can conceive of two situations where this average will not have a Fermi-Dirac form. First, the weighting coefficients $K_1(\mathbf{r};E)$ and $K_2(\mathbf{r};E)$ may vary rapidly in energy, contributing extra structure to the energy dependence of $f(\mathbf{r};E)$, as shown in Fig. 7(a). Second, the applied bias may be too large relative to the thermal energy spread of electrons, causing the distribution $f(\mathbf{r};E)$ to flatten out for energies between μ_1 and μ_2 , as shown in Fig. 7(b).

From these general considerations, it appears that carrier heating can be avoided if two conditions are met: (1) The kernel of the kinetic equation (2.32), $|G^R|^2/2\pi\tau_\phi N_0$, must vary slowly within the energy range of interest ($eV_A, k_B T < \Gamma_c$), and (2) the applied bias must be much smaller than the thermal energy spread of the electrons ($eV_A < k_B T$). Note that, as the temperature tends to zero, the applied bias must also tend to zero, to avoid carrier heating. In the zero-temperature limit, any bias (no matter how small) is enough to distort the occupation factor $f(\mathbf{r};E)$.

Of course, the presence of inelastic scattering will tend to restore the occupation factor $f(\mathbf{r};E)$ to an equilibrium

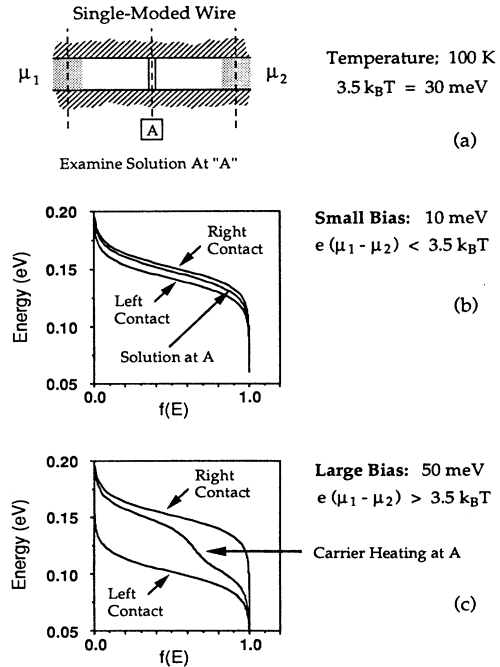


FIG. 6. (a) A single-moded wire with a tunneling barrier. The solution $f(\mathbf{r};E)$ is computed in the middle of the barrier. (b) Results for a small (10 mV) bias. (c) Results for a large (50 mV) bias.

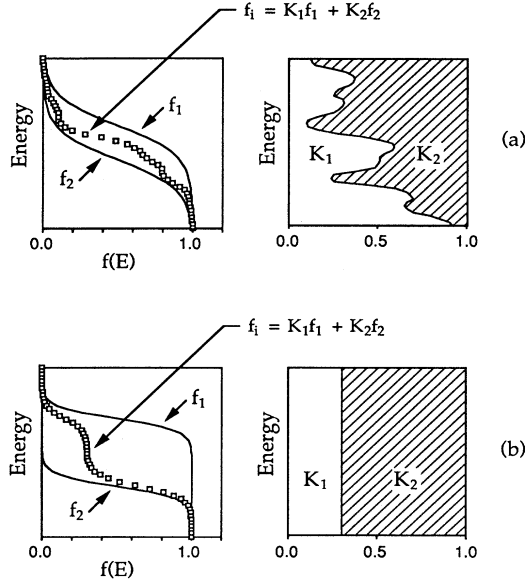


FIG. 7. A simple example to illustrate the conditions leading to carrier heating. The solution $f_i(E)$ will not have a Fermi-Dirac form if (a) the coefficients $K_1(E)$ and $K_2(E)$ vary rapidly with energy, or if (b) the applied bias is large.

form. Because of this, the relevant bias in condition (2) above is really that which is dropped over an inelastic-scattering length. We emphasize, however, that it is *not* necessary to have inelastic scattering within the device to obtain a Fermi-Dirac form for the occupation factor. It is sufficient to have inelastic scattering in the contact regions, so that the boundary distributions have a Fermi-Dirac form; the internal solution will then follow suit if both of the above conditions are met.

An interesting situation arises if the bias is low but the temperature is relatively high so that $eV_A < \Gamma_c < k_B T$ and condition 1 is not met. In this case the coefficients K_1 and K_2 vary across the energy range $k_B T$ over which transport occurs. One could then define a local energy-dependent electrochemical potential, $\mu(\mathbf{r}; E) \simeq K_1(\mathbf{r}, E)\mu_1 + K_2(\mathbf{r}; E)\mu_2$. However, there may be significant vertical flow (linked to the thermoelectric effects) that invalidate the approximation discussed in Sec. II D. In this paper, we will assume that $eV_A < k_B T < \Gamma_c$ so that K_1 and K_2 are constants over the energy range $k_B T$, leading to a unique energy-independent local electrochemical potential $\mu(\mathbf{r}) \simeq K_1(\mathbf{r})\mu_1 + K_2(\mathbf{r})\mu_2$.

B. Simplification to linear response

If the applied bias is kept sufficiently small, we have shown that the occupation factor $f(\mathbf{r}; E)$ will indeed have a Fermi-Dirac form. Knowing this, it seems unnecessary to compute the actual solution for $f(\mathbf{r}; E)$ at each energy. Instead, we simply characterize the occupation factor at

each point \mathbf{r} by a single number, the local electrochemical potential $\mu(\mathbf{r})$,

$$f(\mathbf{r}; E) \equiv \frac{1}{e^{[E - e\mu(\mathbf{r})]/k_B T} + 1}. \quad (3.4)$$

We can reformulate our analysis in terms of this new potential. For a small applied bias, we can expand the distribution function in a Taylor series about $\mu = \mu_0$,

$$f(\mathbf{r}; E) = f_0(E) + \left[-\frac{\partial f_0}{\partial E} \right] e[\mu(\mathbf{r}) - \mu_0] + \dots \quad (3.5)$$

For linear response, we keep only terms to first order in $\delta\mu \equiv \mu(\mathbf{r}) - \mu_0$. After substituting this expansion (3.5) into the kinetic equation (2.32), we obtain a linear-response kinetic equation,

$$\mu(\mathbf{r}) = \frac{\int d^3\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}') \mu(\mathbf{r}')}{\int d^3\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}')}, \quad (3.6)$$

where

$$\begin{aligned} T_0(\mathbf{r}, \mathbf{r}') &= \int dE \left[-\frac{\partial f_0}{\partial E} \right] \frac{\hbar^2 |G^R(\mathbf{r}, \mathbf{r}'; E)|^2}{\tau_\phi(\mathbf{r}; E) \tau_\phi(\mathbf{r}'; E)} \\ &\simeq \frac{\hbar^2 |G^R(\mathbf{r}, \mathbf{r}'; \mu_0)|^2}{\tau_\phi(\mathbf{r}; \mu_0) \tau_\phi(\mathbf{r}'; \mu_0)}, \end{aligned} \quad (3.7a)$$

$$\left[E - H_0 + \frac{i\hbar}{2\tau_\phi(\mathbf{r}; E)} \right] G^R(\mathbf{r}, \mathbf{r}'; E) = \delta(\mathbf{r} - \mathbf{r}'). \quad (3.7b)$$

In writing (3.6) we have made use of the identity [see Eq. (3.4) of Ref. 25(b)]

$$\int d^3\mathbf{r}' \frac{|G^R(\mathbf{r}, \mathbf{r}'; E)|^2}{\tau_\phi(\mathbf{r}'; E)} = \frac{2\pi}{\hbar} N_0(\mathbf{r}; E). \quad (3.8)$$

In this linear-response form, the kinetic equation (3.6) is easily solved. We begin by dividing our structure, as before, into contact regions (where we will apply a boundary condition) and device regions (where we will compute the solution). In the contact regions, the electrochemical potential $\mu(\mathbf{r})$ is fixed to a constant value. By shifting the potential of one contact relative to another, we can apply a bias to the structure. In the remaining “device” regions, we compute the potential $\mu(\mathbf{r})$ by solving the linear-response kinetic equation (3.6).

Once the solution for the electrochemical potential $\mu(\mathbf{r})$ has been obtained, we can use it to reconstruct the occupation factor (3.5). The correlation function $G^<(\mathbf{r}_1, \mathbf{r}_2; E)$ can then be computed,

$$\begin{aligned} G^<(\mathbf{r}_1, \mathbf{r}_2; E) &= \int d^3\mathbf{r}' \left[f_0(E) + \left[-\frac{\partial f_0}{\partial E} \right] e[\mu(\mathbf{r}') - \mu_0] \right] \\ &\quad \times G_\delta(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}'; E), \end{aligned} \quad (3.9)$$

where G_δ was defined in Sec. II E [Eq. (2.35)]. Any quantity of interest can be obtained from this function.

Note that the occupation factor $f(\mathbf{r}; E)$ can be separated into two parts: an equilibrium contribution $f_0(E)$, and a Fermi-surface term $(-\partial f_0/\partial E)\delta\mu$. Since the correlation function $G^<$ has this structure, any quantity computed from it can also be separated in this manner. Consider, for instance, the current density $\mathbf{J}(\mathbf{r}; E)$,

$$\mathbf{J}(\mathbf{r}; E) = \mathbf{J}_{\text{eq}}(\mathbf{r}; E) + \delta\mathbf{J}(\mathbf{r}; E), \quad (3.10)$$

where

$$\mathbf{J}_{\text{eq}}(\mathbf{r}; E) = \int d^3\mathbf{r}' f_0(E) \mathbf{J}_\delta(\mathbf{r}, \mathbf{r}'; E), \quad (3.11)$$

$$\begin{aligned} \delta\mathbf{J}(\mathbf{r}; E) &= \left[-\frac{\partial f_0}{\partial E} \right] \int d^3\mathbf{r}' e^{[\mu(\mathbf{r}') - \mu_0]} \mathbf{J}_\delta(\mathbf{r}, \mathbf{r}'; E) \\ &\simeq e \int d^3\mathbf{r}' [\mu(\mathbf{r}') - \mu_0] \mathbf{J}_\delta(\mathbf{r}, \mathbf{r}'; \mu_0). \end{aligned} \quad (3.12)$$

\mathbf{J}_δ was defined in Sec. II E [Eq. (2.39)]. The equilibrium component $\mathbf{J}_{\text{eq}}(\mathbf{r}; E)$ is nonzero over a wide range of energies. It is much simpler to compute only the *change* $\delta\mathbf{J}$ arising under bias, since this change is limited to the Fermi energy at low temperatures.

The equilibrium current density \mathbf{J}_{eq} is zero if there are no magnetic fields. However, even at equilibrium, circulating currents could exist in the presence of a magnetic field. But these currents will not contribute to the terminal currents, since we can have no net current flow in equilibrium. Thus, our expression for the current density (3.12) is adequate for computing the terminal currents; we simply integrate this current density over the surface at each contact,

$$I_n = \int_{\text{contact } n} \delta\mathbf{J}(\mathbf{r}) \cdot d\mathbf{S}_n. \quad (3.13)$$

One subtle point remains in our discussion of the linear response. In computing the current density $\delta\mathbf{J}(\mathbf{r})$ we have kept the term

$$e \int d^3\mathbf{r}' \delta\mu(\mathbf{r}') \mathbf{J}_\delta(\mathbf{r}, \mathbf{r}'; E), \quad (3.14a)$$

which is first order in $\delta\mu$. An electrostatic potential $\delta\phi$ of the same order as $\delta\mu$ will arise due to screening charges, as discussed in Sec. II A; this is treated in greater detail in Sec. VI. The correction $\delta\phi$ will influence the Green function $G^R(\mathbf{r}, \mathbf{r}'; E)$, and will therefore influence \mathbf{J}_δ . The term

$$e \int d^3\mathbf{r}' \delta\mu(\mathbf{r}') \delta\mathbf{J}_\delta(\mathbf{r}, \mathbf{r}'; E) \quad (3.14b)$$

is clearly second order ($\sim \delta\mu \delta\phi$) and therefore negligible. However, the equilibrium component contributes a term of the form

$$\delta\mathbf{J}_{\text{eq}}(\mathbf{r}; E) \equiv \int d^3\mathbf{r}' f_0(E) \delta\mathbf{J}_\delta(\mathbf{r}, \mathbf{r}'; E), \quad (3.14c)$$

which is first order in $\delta\phi$. Strictly speaking, our expression (3.12) for the current density is not the entire linear response; we should add $\delta\mathbf{J}_{\text{eq}}$ to it. As we have mentioned, in the presence of a magnetic field, there can be circulating currents in equilibrium. The change in the pattern of circulating currents due to $\delta\phi$ is given by $\delta\mathbf{J}_{\text{eq}}$; however, this additional term does not contribute to the terminal currents I_n , and we will neglect it in this paper.

We summarize the linear-response solution procedure in Fig. 8, and compare it to the procedure we would adopt in solving the drift-diffusion equation for a classical device. Indeed, we can write the linear-response transport equation (3.6) in the form

$$T_{\text{op}}\mu(\mathbf{r}) = 0, \quad (3.15)$$

where T_{op} is the appropriate integral operator obtained from the linear-response kinetic equation (3.6). This could be viewed as a quantum analog of the drift-diffusion equation $\nabla \cdot (\sigma \nabla \mu) = 0$. The solution procedure is identical, as shown in Fig. 8.

C. Connection to the Büttiker formula

At this point, we are ready to compute and study solutions in specific devices. Before we do so, however, we will establish a connection between our kinetic equation (3.6) and the Büttiker formula (1.4). We accomplish this by obtaining an alternative expression for the terminal currents. When we solve the diffusion equation (or its quantum-mechanical counterpart) using the procedure shown in Fig. 8, the divergence of the current density is nonzero near the device-contact boundary. To see this, consider a simple one-dimensional resistor [Fig. 9(a)]. The potential μ obtained from the diffusion equation is constant in the contacts and varies linearly within the device. The current density $\mathbf{J} = -\sigma \nabla \mu$ is zero in the contacts, and constant within the device. This implies that the divergence $\nabla \cdot \mathbf{J}$ is nonzero at the device-contact interfaces. By integrating the diverging current

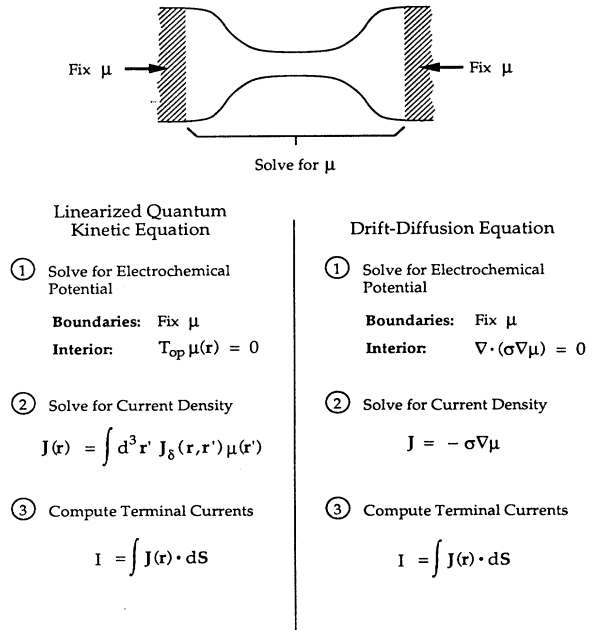


FIG. 8. The linear-response solution procedure is compared with a conventional drift-diffusion analysis.

$$I(\mathbf{r}) \equiv -\nabla \cdot \mathbf{J} \quad (3.16)$$

over the contact region, we can obtain the terminal current I_n ,

$$I_n = \int_{\text{contact } n} d^3\mathbf{r} I(\mathbf{r}). \quad (3.17)$$

This should yield precisely the same result that we would obtain by integrating the current density over the surface S_n of the contact [Eq. (3.13)].

We make use of this idea to derive an alternative expression for the terminal current in the quantum case. To obtain an expression for $I(\mathbf{r})$, we take the divergence of the current density (3.12), and simplify it using our earlier expression (2.40) for $\nabla \cdot \mathbf{J}_\delta$. We then obtain, using (3.7c),

$$I(\mathbf{r}) = \frac{e^2}{h} \int d^3\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}') [\mu(\mathbf{r}) - \mu(\mathbf{r}')]. \quad (3.18)$$

In the device regions, the current $I(\mathbf{r})=0$, so that Eq. (3.18) reduces to our kinetic equation (3.6). We solve this equation for the potential $\mu(\mathbf{r})$ within the device using the procedure described earlier (step 1 of Fig. 8). We then compute the diverging current $I(\mathbf{r})$ in the contact region, and integrate it over the volume of each contact region to obtain the corresponding terminal current [Eq. (3.17)]. In our numerical calculations, we find that both methods of computing the terminal currents [Eqs. (3.13) and (3.17)] yield the same result to within five or six decimal

places. Figure 9(b) shows an actual numerical example. Unlike the diffusion equation, the diverging current $J(\mathbf{r})$ for the quantum equation extends a length L_ϕ in the contact region, as the current density $\mathbf{J}(\mathbf{r})$ decays to zero. This is because the operator T_{op} in the quantum equation couples together all nodes within a phase-breaking length L_ϕ , while the ∇^2 operator in the classical equation couples together only nearest-neighbor nodes.

Note the striking similarity between Eq. (3.18) for $I(\mathbf{r})$ and the Büttiker formula, repeated here for convenience,

$$I_i = \frac{e^2}{h} \sum_j T_{ij} (\mu_i - \mu_j). \quad (3.19)$$

It appears that we have generalized the Büttiker formula to a continuous distribution of probes. Indeed, Büttiker has shown¹⁶ that additional probes can be added to a device to simulate the effect of phase-breaking scatterers. It appears that we have simply taken his suggestion to the continuum limit. A phenomenological extension of this kind has actually been used to study the effects of distributed phase-breaking scattering.³⁸ However, our kinetic equation (3.18) has been rigorously derived from a model Hamiltonian,²⁵ and thus places Büttiker's "probe" model for phase-breaking scatterers on a firm quantitative footing.

In this continuous-probe interpretation, the kernel $T_0(\mathbf{r}, \mathbf{r}')$ of our kinetic equation (3.18) plays the role of a transmission probability density. It answers the question: "How likely is it that an electron injected into the sample at position \mathbf{r}' will reach a position \mathbf{r} without scattering into another probe along the way?" Of course, the majority of electrons will have scattered out within a phase-breaking length L_ϕ of the injection point, so that in general this function decays as $\exp(-\rho/L_\phi)$, where $\rho = |\mathbf{r} - \mathbf{r}'|$. As Eq. (3.6) shows, the electrochemical potential at any point is a weighted average of the values of surrounding points. Because the weighting coefficients $\sim T_0$ decay as $\exp(-\rho/L_\phi)$, nonlocal quantum effects only arise on the length scale defined by L_ϕ . If we neglect phase-breaking processes within the device, then $T_0(\mathbf{r}, \mathbf{r}')$ is nonzero only if points \mathbf{r} and \mathbf{r}' both lie inside some contact. The overall transmission T_{ij} from contact j to contact i [see Eq. (3.19)] is obtained by summing $T_0(\mathbf{r}, \mathbf{r}')$ over all $\mathbf{r}' \in \text{contact } i$ and all $\mathbf{r} \in \text{contact } j$.

D. Can we define μ_L and μ_R ?

We have defined the electrochemical potential $\mu(\mathbf{r})$ in terms of the total electron density per unit energy $n(\mathbf{r}; E)$ [see Eqs. (2.22) and (2.23)]. It is often intuitively appealing to define separate electrochemical potentials $\mu_L(\mathbf{r})$ and $\mu_R(\mathbf{r})$ for the left-moving and right-moving electrons.^{27,28} However, the left-moving and right-moving electron densities $n_L(\mathbf{r}; E)$ and $n_R(\mathbf{r}; E)$ are not positive-definite unlike the total electron density $n = n_L + n_R$. Consequently the local potentials $\mu_L(\mathbf{r})$ and $\mu_R(\mathbf{r})$ may show unphysical behavior. We believe this problem arises because the concept of left-moving and right-moving electrons at a particular point is in violation of the uncertainty principle. The problem can be avoided

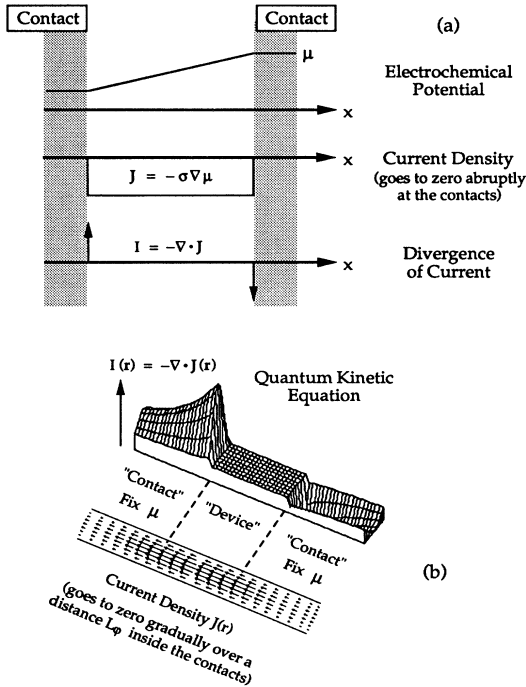


FIG. 9. (a) In a drift-diffusion analysis, the continuity equation is violated at the boundaries. (b) In our quantum analysis, this effect extends over a phase-breaking length L_ϕ .

by defining these potentials in an average sense over regions the size of a de Broglie wavelength. The resulting potentials $\tilde{\mu}_L$ and $\tilde{\mu}_R$ behave as we might intuitively expect.^{27,28} Our objective in this section is to illustrate the above statements with a numerical example.

We consider a single-moded wire with a tunneling barrier in the middle [Fig. 10(a)]. The potential $\mu(r)$ obtained by solving Eq. (3.6) [or Eq. (3.18)] oscillates on either side of the barrier [Fig. 10(b)]. These oscillations decay away from the barrier, an indication of the phase-breaking scattering present in our model. Note that the oscillations in μ are not simply a consequence of the standing waves in the electron density that one expects near a barrier; these would arise even if the potential μ were constant, due to oscillations in the local density of states $N_0(r;E)$. As we have stated, the potential μ at any point is a weighted average (3.6) of the values at surrounding points. We can understand the oscillations, therefore, by investigating the weighting coefficients $T_0(r,r')$, as shown in Fig. 10(c). At a point L , where the electrochemical potential reaches a peak, the weighting coefficients favor the reservoir on the left-hand side of the wire. This is understandable, since the reservoir on the

right-hand side is shielded by the tunneling barrier. If we move a half-wavelength nearer to the left-hand reservoir, however, the coupling to that reservoir actually *decreases*. At this point, left-moving and right-moving waves interfere destructively, so that the magnitude of the Green function $G^R(r,r')$ is small; this magnitude continues to decay as electrons travel toward the left-hand reservoir. Thus, at the point R , the weighting coefficients favor the right-hand reservoir, and the electrochemical potential dips to a minimum. These results are in good qualitative agreement with earlier work,³⁸⁻⁴⁰ as we might expect, since our kinetic equation (3.18) can be viewed as a continuous-probe version of the Büttiker formula.

It will be noted that the oscillations in $\mu(r)$ are always limited between μ_1 and μ_2 , that is, $\mu_1 > \mu > \mu_2$. However, when we compute μ_L and μ_R find that the oscillations extend beyond μ_1 and μ_2 , which is unphysical (Fig. 11). The potentials μ_L and μ_R are calculated as follows. We compute the electron densities n_L and n_R , for left-moving and right-moving electrons, from the Wigner function $-iG^<$,

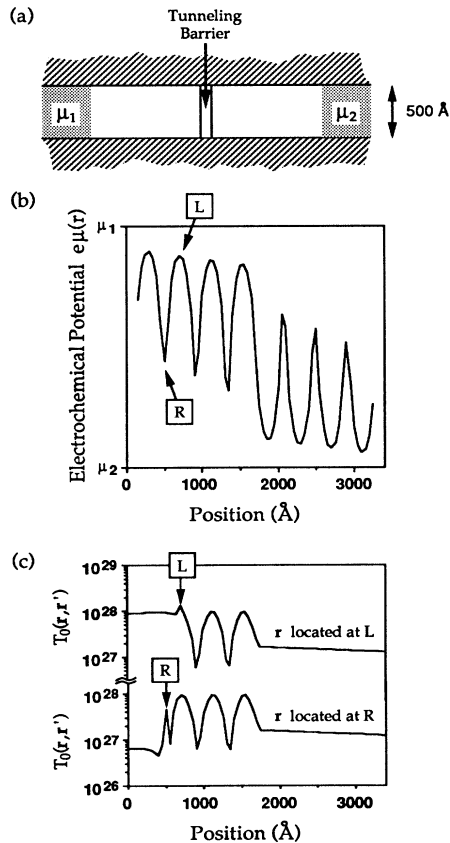


FIG. 10. (a) Single-moded wire with a tunneling barrier. (b) Solution for the electrochemical potential μ along the center of the wire. (c) The kernel $T_0(r,r')$ as a function of r' along the wire, for r located at the points L and R indicated in (b).

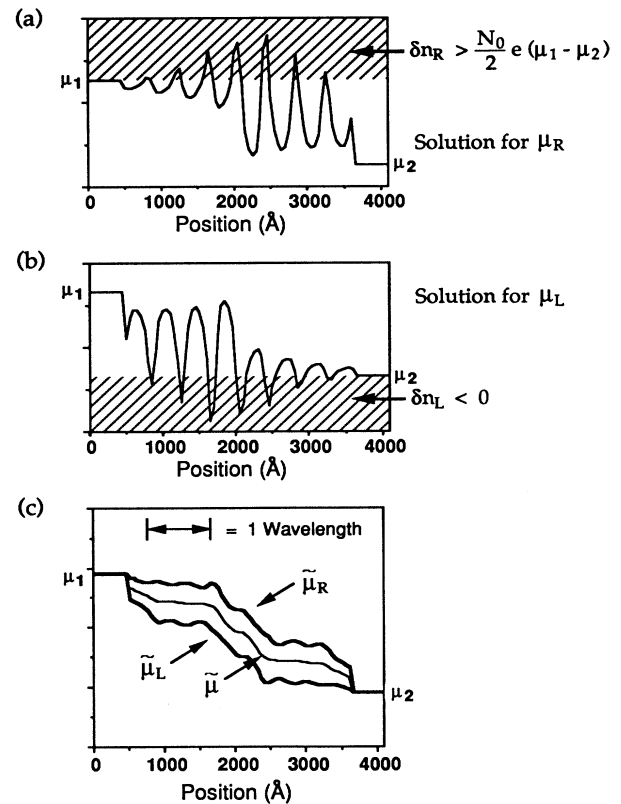


FIG. 11. Electrochemical potentials in a single-moded wire with a tunneling barrier (see Fig. 10). Solutions for (a) μ_R and (b) μ_L show unphysical behavior. However, (c) solutions for the potentials μ_L , $\tilde{\mu}_R$, and $\tilde{\mu}$ (averaged over a wavelength) show sensible behavior.

$$n_L(\mathbf{r}; E) = \frac{-i}{2\pi} \int \frac{d^2\mathbf{k}}{(2\pi)^2} \int_{k_x < 0} \frac{dk_x}{2\pi} G^<(\mathbf{r}; \mathbf{k}; E), \quad (3.20a)$$

$$n_R(\mathbf{r}; E) = \frac{-i}{2\pi} \int \frac{d^2\mathbf{k}}{(2\pi)^2} \int_{k_x > 0} \frac{dk_x}{2\pi} G^<(\mathbf{r}; \mathbf{k}; E). \quad (3.20b)$$

We can reconstruct the Wigner function $-iG^<$ using Eq. (3.9), once we have obtained our solution for the local electrochemical potential $\mu(\mathbf{r})$. As we discussed earlier, the Wigner function consists of two terms: an equilibrium quantity and a Fermi-surface term. We evaluate only

$$\Delta(\mathbf{r}) = \frac{-e\hbar}{2\pi^2} \lim_{\xi \rightarrow 0} \int_0^\infty d\rho_x \frac{\rho_x}{\rho_x^2 + \xi^2} \int d^3\mathbf{r}' \frac{\mu(\mathbf{r}')}{\tau_\phi(\mathbf{r}'; \mu_0)} \text{Im} \left[G^{R*} \left[\mathbf{r} - \frac{\rho_x}{2} \hat{x}, \mathbf{r}'; \mu_0 \right] G^R \left[\mathbf{r} + \frac{\rho_x}{2} \hat{x}, \mathbf{r}'; \mu_0 \right] \right]. \quad (3.22)$$

Using these electron densities, we can define electrochemical potentials separately for left-moving and right-moving electrons (assuming that there are no magnetic fields, so that half the total density of states N_0 is associated with each direction),

$$\mu_L(\mathbf{r}) \equiv \mu(\mathbf{r}) - \frac{\Delta(\mathbf{r})}{eN_0(\mathbf{r})/2}, \quad (3.23a)$$

$$\mu_R(\mathbf{r}) \equiv \mu(\mathbf{r}) + \frac{\Delta(\mathbf{r})}{eN_0(\mathbf{r})/2}. \quad (3.23b)$$

The potentials μ_L and μ_R for the narrow wire with a tunneling barrier are displayed in Figs. 11(a) and 11(b). In keeping with our earlier result (Fig. 10), the presence of a tunneling barrier causes both potentials to oscillate. Excursions into the shaded areas, however, indicate that these potentials have unphysical properties: For left-moving electrons, the excursions represent a *negative* electron density δn_L , while for right-moving electrons, they represent an electron density δn_R in excess of the density of states. This behavior is difficult to reconcile with our semiclassical intuition.

Such strange, quantum-mechanical behavior might be expected. In a sense, we have attempted to localize electrons in both position and momentum, thereby violating the Heisenberg uncertainty principle. If we were to average the electron densities δn_L and δn_R over a wavelength, however, we should obtain sensible results. Indeed, the electrochemical potentials defined in terms of the average electron densities,

$$\bar{\mu}_L(\mathbf{r}) \equiv \frac{\langle \delta n_L(\mathbf{r}) \rangle}{\langle N_0(\mathbf{r}) \rangle / 2}, \quad (3.24a)$$

$$\bar{\mu}_R(\mathbf{r}) \equiv \frac{\langle \delta n_R(\mathbf{r}) \rangle}{\langle N_0(\mathbf{r}) \rangle / 2}, \quad (3.24b)$$

are well-behaved quantities, as shown in Fig. 11(c). The idea of averaging over a wavelength is not novel. Other authors have recognized that the off-diagonal elements of the Wigner function have undesirable properties and have developed Gaussian smoothing techniques (such as the use of Husimi functions) to obtain sensible results.⁴²

the Fermi-surface term, so that our results reflect the changes δn_L and δn_R arising under bias. Using the definition for G_δ (2.35) and the Fourier-transform relationship (2.4a), it can be shown that

$$\delta n_L(\mathbf{r}) = \frac{1}{2} \delta n(\mathbf{r}) - \Delta(\mathbf{r}), \quad (3.21a)$$

$$\delta n_R(\mathbf{r}) = \frac{1}{2} \delta n(\mathbf{r}) + \Delta(\mathbf{r}), \quad (3.21b)$$

where

In the above discussion we assumed that the density of states for left- and right-moving electrons are equal, so that each equals $N_0(\mathbf{r}; E)/2$. This is not true in a magnetic field. For example, in the quantum Hall regime the entire density of states on one side of the sample is associated with left-moving electrons and that on the other side with right-moving electrons; thus, μ is identical with μ_L on one side and with μ_R on the other. In general, we could define the left- and right-moving density of states as follows. The density of states $N_0(\mathbf{r}; E)$ is related to the diagonal elements of the retarded Green function,⁴¹

$$N_0(\mathbf{r}; E) = -\text{Im}[G^R(\mathbf{r}, \mathbf{r}; E)]/\pi. \quad (3.25)$$

Using the Fourier-transform relationship (2.4a), we can rewrite $N_0(\mathbf{r}; E)$ as

$$N_0(\mathbf{r}; E) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \left[-\frac{1}{\pi} \text{Im}[G^R(\mathbf{r}; \mathbf{k}; E)] \right]. \quad (3.26)$$

It seems that we can obtain the left-moving density of states $N_L(\mathbf{r}; E)$ by restricting the integral in (3.26) to $k_x < 0$ and the right-moving density of states $N_R(\mathbf{r}; E)$ by restricting to $k_x > 0$. These quantities, however, may not be positive-definite.

IV. MEASURING THE ELECTROCHEMICAL POTENTIAL

In Sec. III we have seen how under low-bias conditions the electron energy distribution at each point can be characterized by a local electrochemical potential $\mu(\mathbf{r})$. Our objective in this section is to explore the difficulties associated with measuring the local potential in mesoscopic samples by connecting a voltage probe. The difficulty is twofold. First, voltage probes are usually invasive; that is, they significantly alter the very potential distribution they are supposed to measure. This is discussed in Sec. IV A. However, this is a problem that can possibly be avoided by designing weakly coupled noninvasive probes. A more serious conceptual issue emphasized by Büttiker³⁶ is that even with noninvasive probes, the measured potential could be affected by the

probe geometry. We present a numerical example in Sec. IV B showing how the oscillating potential near a barrier (see Fig. 10) can be measured fairly accurately using noninvasive probes with different geometries. Finally, in Sec. IV C we present a counterexample. We show that the intrinsic Hall potential in a narrow quantum wire is present around the voltage probes and yet the probes fail to measure it.

A. Invasive voltage probes

We can get some insight into the problems of measuring the electrochemical potential by considering the kinetic equation [Eq. (3.6)], which is rewritten here for convenience,

$$\mu(\mathbf{r}) = \frac{\int d^3\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}') \mu(\mathbf{r}')}{\int d^3\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}')}. \quad (4.1)$$

In this form this equation has a simple physical interpretation. It simply states that the electrochemical potential at any point is a weighted average of those at surrounding points. The weighting coefficient $T_0(\mathbf{r}, \mathbf{r}')$ is related to the probability that an electron suffering a phase-breaking scattering at \mathbf{r}' will suffer its next phase-breaking event at \mathbf{r} .

Measuring the local potential in large macroscopic samples is straightforward. This is because the potential varies slowly compared to the phase-breaking length, which is equal to the spatial range $|\mathbf{r} - \mathbf{r}'|$ over which the weighting function $T_0(\mathbf{r}, \mathbf{r}')$ is significant. Consequently, an external probe only samples a small portion of the device over which the potential is nearly constant. But in mesoscopic samples the potential varies rapidly over the range of the weighting function $T_0(\mathbf{r}, \mathbf{r}')$ leading to significant nonlocal effects.

Recent experiments have made it apparent that one cannot simply attach a probe to a small sample and measure the voltage at a given point. An example of this can be found in the literature on conductance fluctuations, a topic that has received widespread attention in recent years.^{43,44} In the presence of a magnetic field, the conductance of disordered wires exhibits noiselike, but repeatable, fluctuations.⁴⁵ Many of the early experiments were performed with a large spacing ($L > L_\phi$) between the voltage probes. From a semiclassical viewpoint, we might expect the fluctuations to decrease as the voltage probes are brought together, so that at some point, when the probes are sufficiently close, the measured voltage would be zero. Experimentally, however, the voltage fluctuations are found to level off at a constant value as the separation between the probes is decreased.⁴⁶

Of course, in these experiments, there is no clear division between the probes and the sample. The probe is simply a lead made of the same material, attached to the side of the sample under test. For this reason, it is difficult to localize precisely where the voltage is being measured. Indeed, Büttiker has shown that even a perfectly ordered wire will exhibit voltage fluctuations if impurities are introduced into the voltage probes.⁴⁷ In effect, an electron “sees” every part of the device within a

phase-breaking length L_ϕ . This includes regions within the voltage probes, even though these regions are outside of the classical current path.

From these experiments, it is clear that voltage probes must be treated as an integral part of a device; their presence or absence, even their form, can affect the outcome of an experiment. Using the theory described in Sec. III, we are in a rather unique position to view the measurement process in microscopic detail. Consider, for example, the electrochemical potential $\mu(\mathbf{r})$ in a homogeneous wire both with and without a voltage probe, as shown in Fig. 12. Our voltage probe is modeled as a lead made of the same material, extending out to infinity. At some point within the lead, the electrochemical potential settles out to a constant value—the value that would be “measured” at the contact in a real experiment. Without the probe, the potential drops linearly due to the uniform background of phase-breaking scatterers. When we attach the voltage probe, however, the potential oscillates within a phase-breaking length of the junction, almost as if we had inserted a tunneling barrier into the wire (cf. Fig. 10). By attaching a probe, we have actually introduced a large scattering center into the wire, thereby disturbing the potential we were trying to measure.

B. Noninvasive voltage probes

Strongly coupled probes, such as those discussed above, are easily fabricated; but they are by no means the

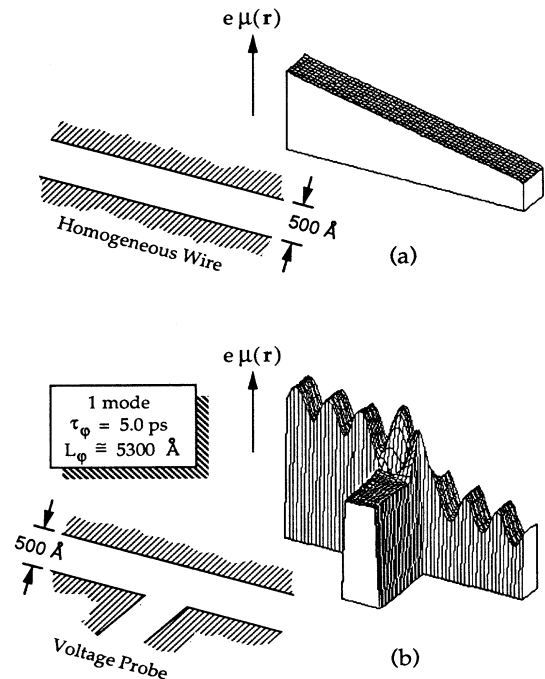


FIG. 12. (a) In a homogeneous wire, the electrochemical potential μ drops linearly. (b) The presence of a voltage probe disturbs the solution for μ .

only way of performing a measurement. A scanning tunneling microscope, for instance, can be modified so that it measures both the surface topography and the electrochemical potential.^{66,67} Since the tip is separated from the sample by a tunneling barrier, a measurement of this type should be minimally invasive. Results obtained in polycrystalline Au₆₀Pd₄₀ films⁶⁷ are striking: In the presence of bias, the electrochemical potential remains approximately constant across each grain, and drops sharply at the grain boundaries. This is precisely the behavior we might expect, since the grain boundaries offer the greatest resistance to current flow. In the absence of bias, of course, the potential shows no variation. Because these experiments provide an internal view of the sample, they represent a powerful tool for the exploration of transport physics.

In our simulations, we can make a voltage probe less invasive by including a tunneling barrier at the mouth of the probe. We will use fairly long and shallow barriers, having an extent of ~ 150 Å and an apparent height of ~ 3 meV. A similar probe could be realized experimentally by a point contact operating near pinchoff.⁶⁸ To illustrate the effect that the probe-to-sample coupling has on the measurement, we investigate two probes with slightly different tunneling barriers, shown in Fig. 13(a). One of these barriers has a uniform thickness, while the other does not. For a suitable test structure, we return to the single-moded wire with a tunneling barrier studied in the preceding section. Interference effects cause the electrochemical potential to oscillate on either side of the barrier (Fig. 10). We position each of these probes at various points along the wire, and compare the measured potentials to the actual solution in an unprobed sample, as shown in Fig. 13.

Both probes measure the electrochemical potential far better than we might have expected. It is evident from Eq. (4.1) that the potential μ_P at a point P in the probe is a weighted average of the potentials $\mu(\mathbf{r}')$ at surrounding points,

$$\mu_P = \int d\mathbf{r}' K_P(\mathbf{r}') \mu(\mathbf{r}') . \quad (4.2)$$

The weighting function $K_P(\mathbf{r}')$ extends over a range L_ϕ , which in this example is 5300 Å. Since L_ϕ is much larger than the period of the oscillations, we would not expect the measured potential to oscillate. We would expect the probe to float to some average potential, rather than to the potential μ_C at the point C in front of the probe. Note, however, that μ_C itself is an average of the surrounding potentials,

$$\mu_C = \int d\mathbf{r}' K_C(\mathbf{r}') \mu(\mathbf{r}') . \quad (4.3)$$

If the weighting functions $K_P(\mathbf{r}')$ and $K_C(\mathbf{r}')$ are similar, then we would expect

$$\mu_P \approx \mu_C . \quad (4.4)$$

The results in Fig. 13(b) suggest that this is indeed the case. We have studied a large variety of noninvasive probes and have found that they all seem to provide good qualitative measurements.

However, the measured potentials are quantitatively different for the two probes. It appears that the probe with a uniform barrier measures the potential more faithfully. At first glance, it seems that the probe with the nonuniform barrier may have its focal point on the left side of the probe, so that the values measured by this probe might simply be off center. But a simple shift of the data points is not enough to account for the measurement error at all points. Presumably, the two probes couple to the sample in a slightly different manner, due to the different shapes of the tunneling barriers. As a result, the different weighting coefficients $K_P(\mathbf{r}')$ for the two different probes yield slightly different measurements μ_P for the potential. Similar effects have been reported in an analysis of scanned probe measurements,⁵¹ indicating that in general, potential measurements should be interpreted with care.

C. Quenching of the Hall resistance

Finally we consider the well-known problem of the quenching of the Hall resistance in mesoscopic wires. We

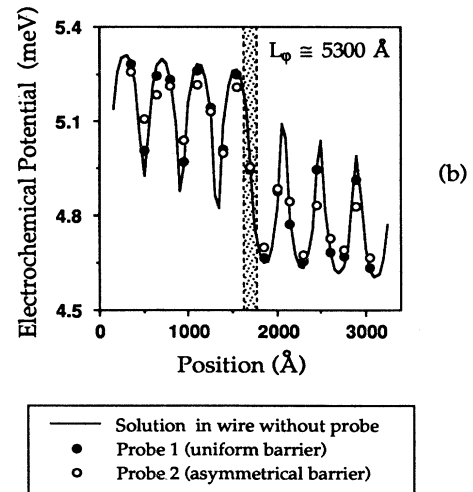
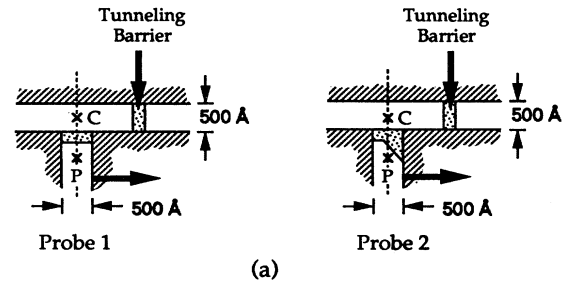


FIG. 13. (a) Two slightly different but weakly coupled probes are investigated; the position of each probe is scanned along the edge of the wire. (b) The two probes give slightly different measurements, but both measurements are in good agreement with the intrinsic potential represented by the solid line.

will show that the intrinsic Hall potential is present near the voltage probe, and yet the probes do not measure it. It thus seems that noninvasive probes may not guarantee a faithful measurement, despite the surprisingly good results shown in Sec. IV B (Fig. 13).

Semiclassically, the Hall resistance is expected to be linear in the magnetic field, $\rho_{xy} = B/en$. From this relationship, the slope of the Hall resistance (plotted against magnetic field) can be used to determine the electron density in a particular sample. Experiments performed on narrow wires, however, show that ρ_{xy} is nearly zero or “quenched” at low magnetic fields.^{29–31} Theoretical calculations^{48,49} also show this effect. Using the present approach, we computed the Hall resistance as a function of magnetic field for the geometry shown in Fig. 14(a), with two different phase-breaking times τ_ϕ . Quenching is apparent in the sample with the longer τ_ϕ , while the usual linear dependence is recovered with a shorter τ_ϕ [Fig. 14(b)]. Note, however, that the phase-breaking scatterers

in our model randomize momentum as well as phase. Because of this, it is not clear whether the quenching is destroyed by the additional phase breaking, or the additional momentum randomization. As we change τ_ϕ , we go from a quantum ballistic to a semiclassical diffusive regime. Unfortunately, we cannot investigate the semiclassical ballistic regime, which corresponds to the billiard-ball model of Beenakker and van Houten.⁴⁹ One way to do this is to evaluate the weighting function $T_0(\mathbf{r}, \mathbf{r}')$ from semiclassical dynamics; however, this is outside the scope of this paper.

It should be mentioned that in obtaining the quenched characteristic shown in Fig. 14(b), the equilibrium Fermi energy μ_0 had to be selected carefully. We present the Hall resistance versus μ_0 at a fixed magnetic field ($B = 0.2T$) in Fig. 14(c), for two different phase-breaking times $\tau_\phi = 1.0$ and 0.1 ps; the difference between these curves indicates the degree of quenching, which evidently oscillates as a function of μ_0 . Such selective quenching

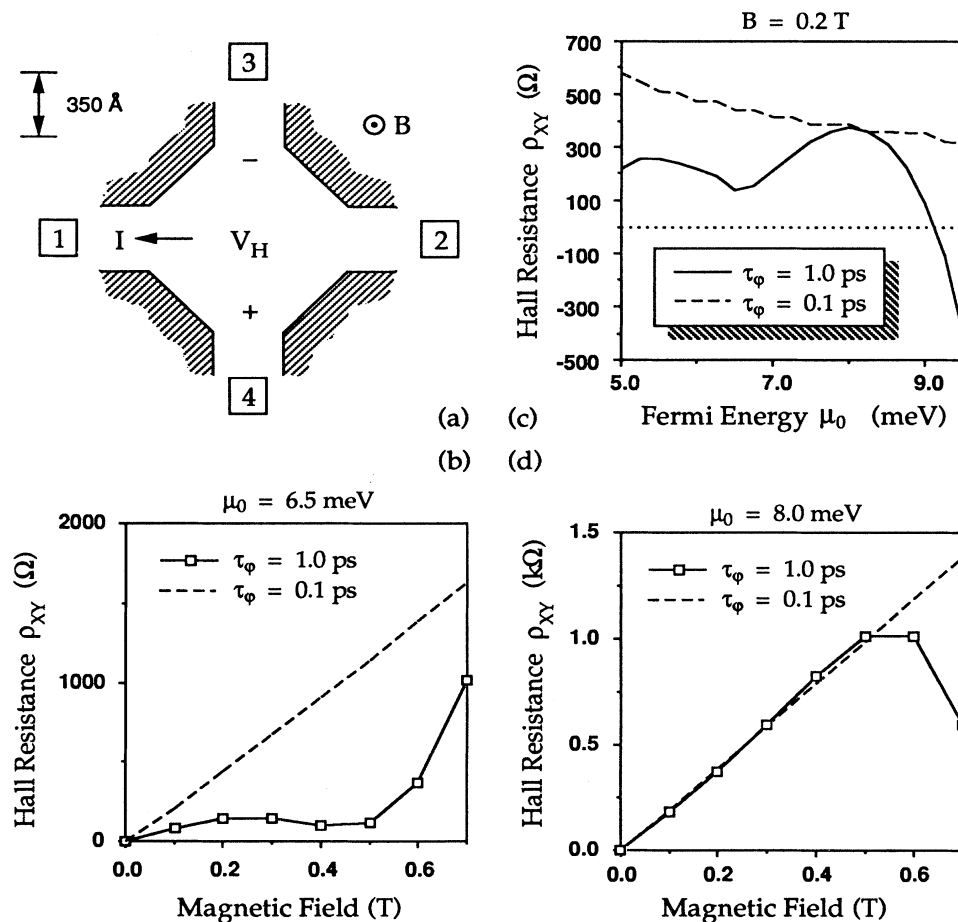


FIG. 14. (a) The measurement junction for Hall resistance calculations; all leads are single moded. (b) The Hall resistance shows quenching when the phase-breaking time τ_ϕ is long. (c) Quenching of the Hall resistance changes as a function of the equilibrium Fermi energy μ_0 . (d) At a particular energy $\mu_0 = 8.0$ meV, the Hall resistance shows the onset of quenching at higher magnetic fields.

behavior has been observed both experimentally³¹ and theoretically.^{48,50} It is also possible for quenching to manifest itself at higher magnetic fields, although it may be absent at low fields [Fig. 14(d)]. This delicate dependence on both energy and magnetic field reflects the complex behavior of scattering within the junction, which has been discussed by a number of authors.^{35,48–50}

An important question regarding the quenching is whether it is an intrinsic property of narrow wires or an artifact of the measurement probes. By computing the transverse electrostatic potential in narrow wires without probes, it has been shown^{33,34} that the Hall resistance is not intrinsically quenched. Using the present approach, we can include the presence of voltage probes and examine the electrochemical potential $\mu(r)$ throughout the structure. As we will shown in Sec. VI, this potential $\mu(r)$ can be used to compute the electrostatic potential $\phi(r)$. The two potentials share the same qualitative behavior, although it is the electrochemical potential $\mu(r)$ that is actually measured by the probes.

In Fig. 15(a), we present the solution for the electrochemical potential $\mu(r)$ throughout the structure. The usual Hall buildup is evident in regions of the wire away

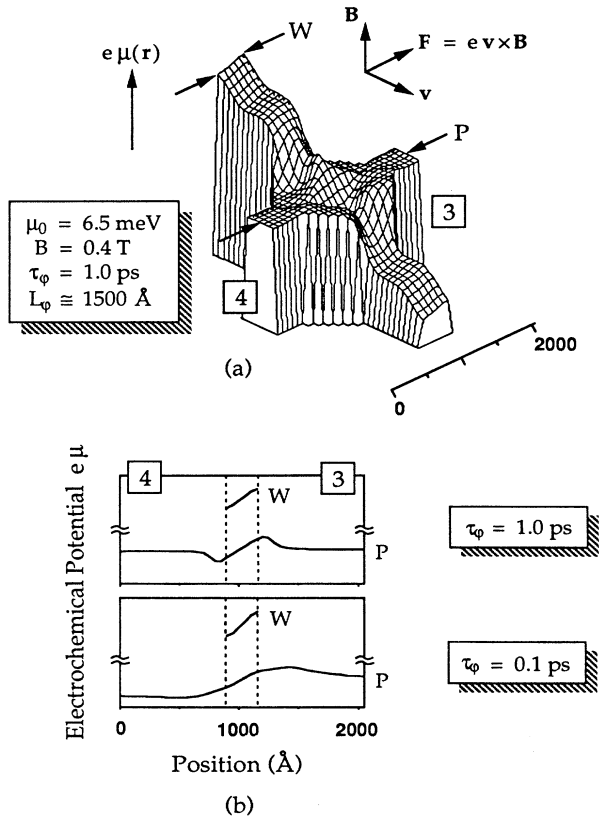


FIG. 15. (a) The electrochemical potential μ within the junction. (b) Cross sections of μ taken at the points indicated in (a) for two different phase-breaking times $\tau_\phi = 1.0$ and 0.1 ps; these cross sections have been normalized to reflect equal currents for the two values of τ_ϕ .

from the junction. We can see this buildup more clearly by examining a cross section of the wire marked W , shown in Fig. 15(b) for two different phase-breaking times τ_ϕ ; these curves have been scaled so that the current is the same in both cases. Within the junction, however, the Hall buildup is disturbed. Since this disturbance is antisymmetric with respect to the center of each probe, we can consider the cross section along this line [marked P in Fig. 15(a)] to be representative of the average potential within the junction. We compare this potential (along P) to that which arises outside of the junction (along W). We focus our attention first on the central region of the wire—the region that would still exist if the probes were removed. In this region, the Hall buildup within the junction corresponds quite well to that away from the junction, even for the sample with the longer phase-breaking time. Thus, in an average sense, the usual Hall buildup exists within the junction, although the probes fail to measure it properly when the phase-breaking time τ_ϕ is long.

As we have mentioned earlier, the electrochemical potential $\mu(r)$ at any point r is a weighted average of the values at surrounding points [Eq. (4.1)]. We can gain a little insight by examining the weighting coefficients $T_0(r, r')$. These coefficients are presented in Fig. 16 for a point r in probe 3, for two different phase-breaking times

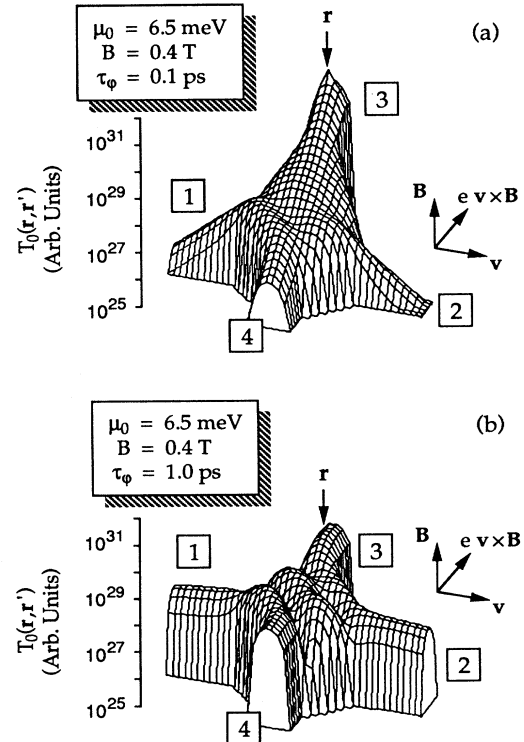


FIG. 16. The weighting coefficients $T_0(r, r')$ as a function of r' for a point r within probe 3, for two different phase-breaking times: (a) $\tau_\phi = 0.1$ ps and (b) $\tau_\phi = 1.0$ ps.

τ_ϕ . We would expect that electrons traveling from contact 1 to contact 2 would be deflected by the Lorentz force into probe 3. As a result, the weighting coefficients for probe 3 should favor contact 1. When the phase-breaking time τ_ϕ is short, this is indeed the case. When τ_ϕ is long, however, the two contacts are weighted almost equally, and the probe floats to a lower potential [see cross sections in Fig. 15(b)]. Because probes 3 and 4 “see” each other so strongly, it is difficult for them to register a potential difference.

The quenching of the Hall effect is also reflected in the pattern of current density within the junction. Usually, the Hall buildup gives rise to an electric field which counteracts the Lorentz force, so that the current flow pattern is virtually undisturbed in a magnetic field. In Sec. VI, we will actually compute such corrections to the electrostatic potential. For the present, however, we are effectively treating electrons as neutral particles. Even so, we expect the current density to remain undisturbed. This is because a Hall buildup of neutral particles will establish a diffusion current that counteracts the Lorentz force. When the junction exhibits quenching, however, one (or both) of the probes will float to an improper potential, and the current that would normally counteract the Lorentz force may change. As a result, we expect to see the current flow pattern change from its $B=0$ configuration. This is precisely what we observe in the calculations of the current density. In Fig. 17, we plot the current density both with and without a magnetic field, for the longer phase-breaking time $\tau_\phi=1$ ps. In the

presence of a magnetic field the current is diverted toward probe 3, indicating that the potential difference between probes 3 and 4 is too small to counteract the Lorentz force. We have checked that for a shorter $\tau_\phi=0.1$ ps, the current flow pattern hardly changes when a magnetic field is applied as we might expect since there is no quenching.

V. POTENTIAL DROP AT A CONSTRICTION

Recent experiments with point contacts^{52,53} have shown that the two-probe conductance measured between two reservoirs is approximately equal to $(2e^2/h)M$, if the device in Fig. 1 is a ballistic channel with M modes. Theoretical calculations based on the two-probe Landauer formula (1.3) also lead to this result.^{54,55,35} An obvious question to ask is where the voltage drop associated with this two-probe resistance $(h/2e^2M)$ occurs. We do not expect any drop in the electrochemical potential across the narrow device region, since it is ballistic. Neither do we expect any significant drop in the wide regions, since they are highly conductive. Imry⁶ has argued that the potential drop is due to the *spreading resistance* associated with the transition region where the current funnels into or out of the narrow channel. This resistance is known as the Sharvin resistance in inelastic point-contact spectroscopy.⁵⁶

In Secs. III and IV we discussed how the local electrochemical potential can be defined and measured. Our objective in this chapter is to examine the spatial variation of the *electrochemical* potential in several structures in order to clarify the nature of the spreading resistance. We will first discuss the contact resistance that occurs when we connect any conductor to an ideal reservoir (Sec. V A). We will then examine the geometry shown in Fig. 1 where a narrow device flares out into wider leads at the ends (Sec. V B). The purpose is to shed light on the question of how wide a lead needs to be in order to function as an ideal reservoir.

At the outset we would once again like to remind the reader that the kinetic equation we use for our calculations simply states that the electrochemical potential at any point \mathbf{r} is a weighted average of the potentials at surrounding points \mathbf{r}' ; the weighting function $T_0(\mathbf{r}, \mathbf{r}')$ is related to the probability that an electron suffering a phase-breaking scattering at \mathbf{r}' suffers the next one at \mathbf{r} [see Eq. (4.1)]. As we have mentioned earlier, this simple physical picture is very useful in understanding the numerical results.

A. Contact resistance

Consider first a single-moded wire, 800 \AA long, with $\mu=\mu_1$ at one end and $\mu=\mu_2$ at the other. Subject to these boundary conditions we solve the kinetic equation, Eq. (4.1), to obtain the potential distribution $\mu(\mathbf{r})$ in the wire. The external current is then obtained from Eq. (3.6) by integrating $I(\mathbf{r})$ over one of the contacts. The total resistance is computed from the ratio of the applied potential $(\mu_1-\mu_2)$ to the total current.

Figure 18 shows the potential profile for a wire with a

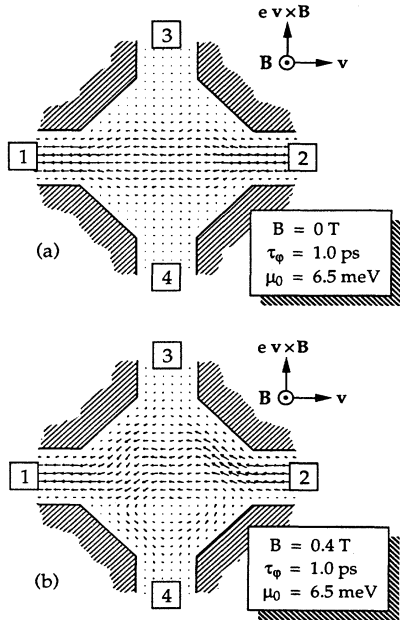


FIG. 17. Current density δJ within the junction both with and without a magnetic field, for a long phase-breaking time $\tau_\phi=1.0$ ps. The diversion of current toward probe 3 is a symptom of the quenched Hall resistance.

long $\tau_\phi (=20 \text{ ps})$ and a short $\tau_\phi (=0.5 \text{ ps})$. For the ballistic wire there is no potential drop across the wire as we might expect. However, there are sharp drops at the ends which may come as a surprise since there is no apparent discontinuity or obstacle to flow at these points. It is easy to see why Eq. (4.1) leads to this form for the potential distribution. The integrals in (4.1) include the left contact (C_1), the device (D), and the right contact (C_2). Since there is hardly any scattering within the device, we can neglect the integral over the device and rewrite (4.1) as [cf. Eq. (3.3)]

$$\mu(\mathbf{r}) = K_1(\mathbf{r})\mu_1 + K_2(\mathbf{r})\mu_2, \quad (5.1a)$$

where

$$K_1(\mathbf{r}) = \frac{\int_{\mathbf{r}' \in C_1} d^3\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}')}{\int_{\mathbf{r}' \in C_1 + C_2} d^3\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}')}, \quad (5.1b)$$

$$K_2(\mathbf{r}) = \frac{\int_{\mathbf{r}' \in C_2} d^3\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}')}{\int_{\mathbf{r}' \in C_1 + C_2} d^3\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}')}. \quad (5.1c)$$

Clearly $K_1 + K_2 = 1$. Thus the potential inside the device is a weighted average of μ_1 and μ_2 . $K_1(\mathbf{r})$ is the fraction of electrons at \mathbf{r} that suffered their last scattering in the left contact, while $K_2(\mathbf{r})$ is the fraction that suffered their last scattering in the right contact. In a ballistic device the Green function $G^R(\mathbf{r}, \mathbf{r}')$ [and hence $T_0(\mathbf{r}, \mathbf{r}')$] hardly changes as we move \mathbf{r} from left to right. Consequently $K_1(\mathbf{r})$ and $K_2(\mathbf{r})$ are independent of \mathbf{r} and equal to 0.5 throughout the device. Hence [from (5.1a)] $\mu = 0.5$ everywhere inside the device.

It thus appears that for a ballistic device the entire resistance is lumped at the two ends, right where the boundary conditions are applied [Fig. 18(c)]. This can be understood if we remember that the constant potential boundary condition that we have applied is accurate only if these regions are infinitely wide so that the series resistance is zero.⁵⁷ As a natural corollary to this argument, we can state that imposing a constant potential boundary condition is *physically equivalent* to assuming that the wire suddenly becomes infinitely wide in the contact regions. Consequently, we expect contact resistances at these points, just as we get a spreading resistance when

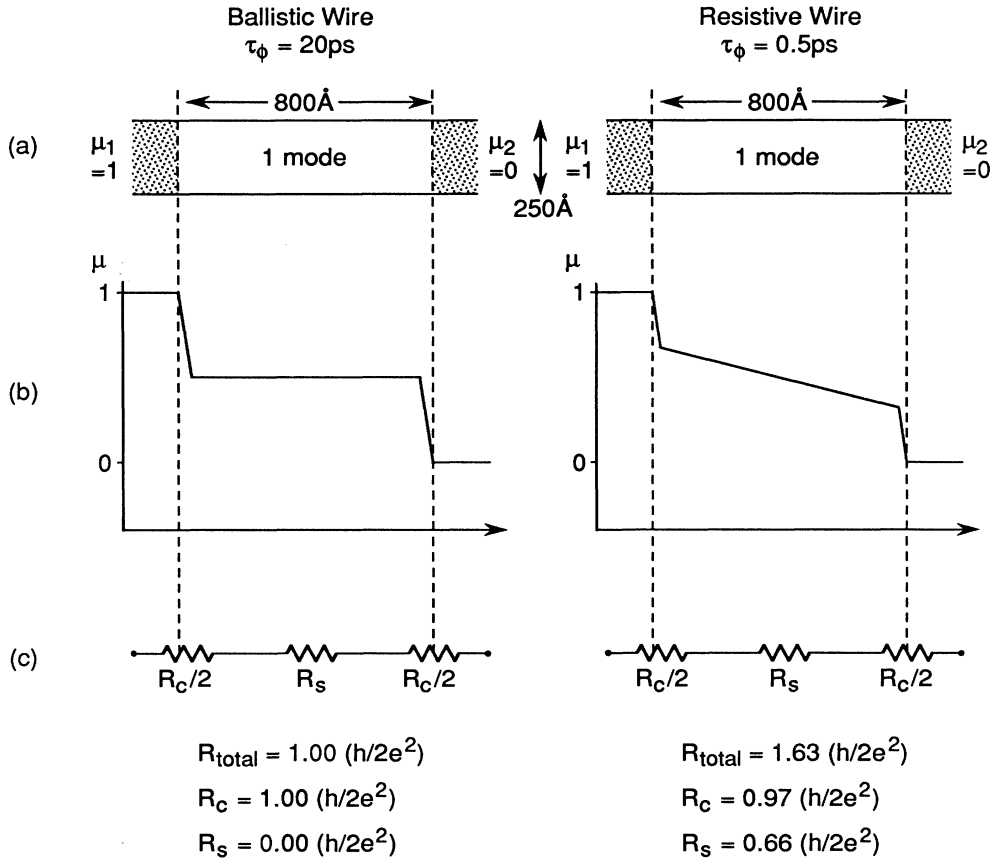


FIG. 18. Potential drop in a single-moded wire with two different values of τ_ϕ . (a) Schematic diagram; (b) potential variation along the center; (c) equivalent circuit.

the width of a conductor changes.⁶

The resistive wire in Fig. 18 having $\tau_\phi=0.5$ ps shows nearly the same contact resistance localized at the ends with an additional series resistance distributed throughout the wire as shown in the equivalent circuit [Fig. 18(c)]. Figure 19 shows similar results for a wire with four modes. From these examples we can write the contact resistance R_C for a wire with M modes approximately as

$$R_C \approx \frac{1}{M}(h/2e^2) . \tag{5.2}$$

This is the resistance that appears at the ends when the wire is suddenly widened to infinity or, equivalently, when a constant potential boundary condition is applied at the ends. For a ballistic wire R_C represents the total resistance. But if there is scattering inside the device, then an additional resistance appears in series with R_C .

It will be noted that the total resistance is related to the transmission T through the device [Eq. (1.3)],

$$R_{\text{total}} = R_C + R_S = \frac{1}{T}(h/2e^2) . \tag{5.3}$$

From (5.2) and (5.3) we can solve for R_S , the resistance of the channel alone,

$$R_S = \frac{1 - (T/M)}{T}(h/2e^2) . \tag{5.4}$$

For a single-moded device, (5.4) yields the familiar Landauer formula

$$R_S = \frac{1 - T}{T}(h/2e^2) . \tag{5.5}$$

B. How ideal is a reservoir?

Landauer has shown that if the contact has a finite number of modes W , then the contact resistance is reduced from the value we expect from (5.2). This can be understood as follows. Consider the structure shown in Fig. 20(a) with $\tau_\phi=20$ ps. Since there is hardly any scattering within the device, the potential at any point can be obtained from Eq. (5.1). In the central narrow region $K_1=K_2=0.5$ (from symmetry) so that the potential is 0.5. Now, if we consider a point located in the wide region to the left we can estimate K_1 and K_2 from a simple semiclassical argument as follows. Consider electrons entering the device from the left contact. These electrons first traverse the wide region to the left. A fraction N/W is then transmitted onto the right through the narrow region, and the rest $(1 - N/W)$ traverse the left-hand wide

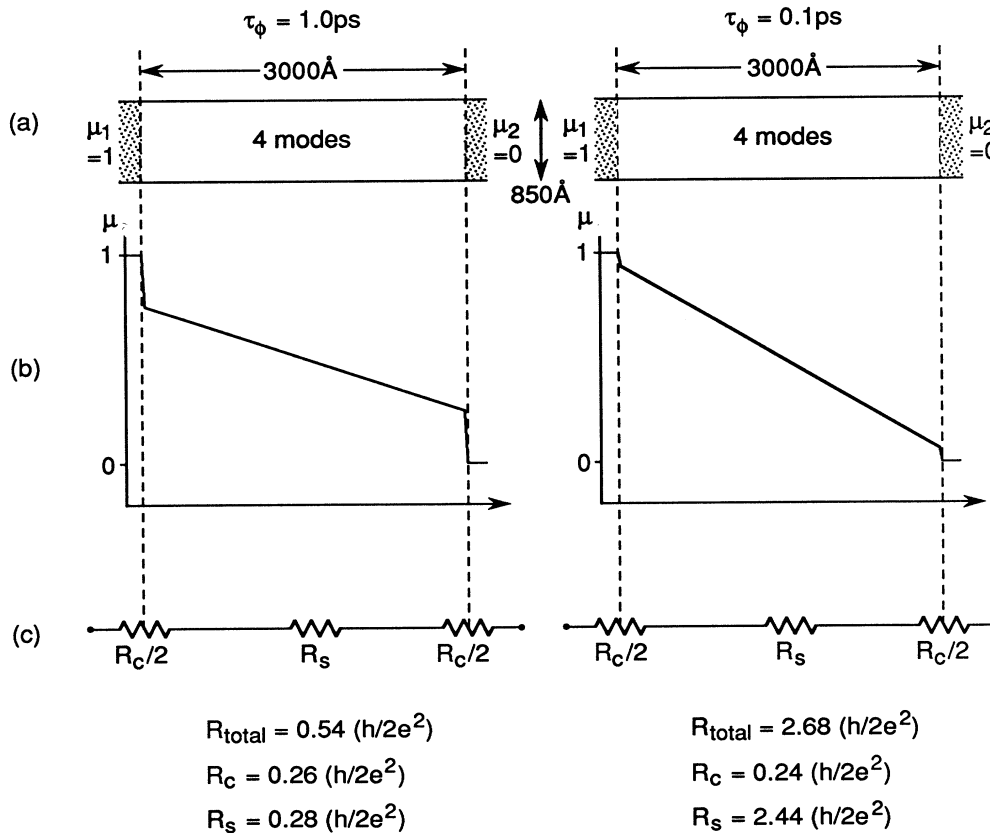


FIG. 19. Same as Fig. 18 but for a wire with four modes.

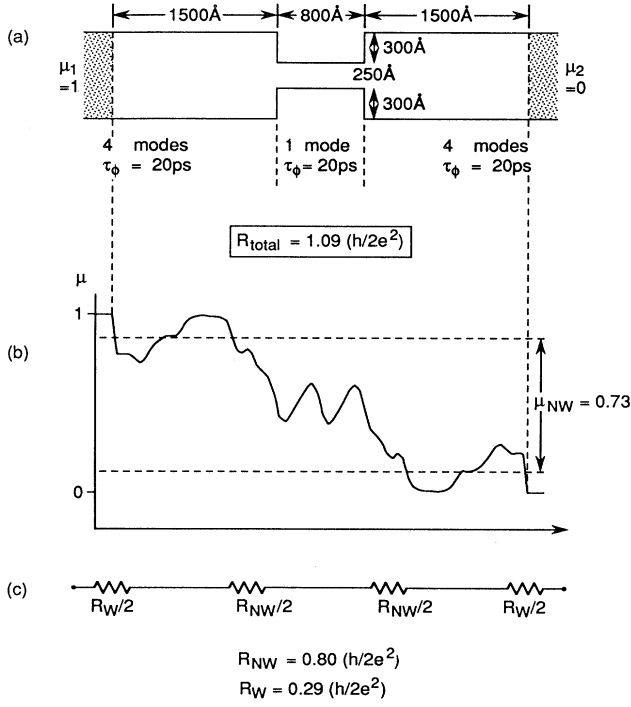


FIG. 20. Potential drop in a wide-narrow-wide (WNW) structure with abrupt transitions. (a) Schematic diagram; (b) potential variation along the center; (c) equivalent circuit.

region back into the left contact. A similar argument applies to electrons coming in from the right contact. Thus, if we look at any point in the wide region to the left, we find $(1 + 1 - N/W)$ electrons that came from the left contact and N/W electrons that came from the right contact. Hence, $K_1 = 1 - (N/2W)$ and $K_2 = N/2W$. We thus obtain the potential μ_{WL} in the left-hand wide region from (5.1):

$$\mu_{WL} = 1 - \frac{N}{2W}. \quad (5.6a)$$

Similarly the potential μ_{WR} in the right-hand wide region is given by

$$\mu_{WR} = \frac{N}{2W}. \quad (5.6b)$$

Also, the current I is given by

$$I = \frac{2e^2}{h} N. \quad (5.6c)$$

From Eqs. (5.6) we compute the contact resistance between the wide and narrow regions,

$$R_{NW} = \frac{\mu_{WL} - \mu_{WR}}{I} = \left[\frac{1}{N} - \frac{1}{W} \right] (h/2e^2). \quad (5.7a)$$

We also have a contact resistance $R_W/2$ between the wide regions and the contacts (where the boundary conditions on μ are applied) as shown in Fig. 20,

$$R_W = 2 \frac{\mu_{WR}}{I} = 2 \frac{1 - \mu_{WL}}{I} = \frac{1}{W} (h/2e^2). \quad (5.7b)$$

Adding (5.7a) and (5.7b),

$$R_{NW} + R_W = \frac{1}{N} (h/2e^2). \quad (5.7c)$$

Note that if the wide regions are made infinitely wide, the boundary resistance $R_W/2$ tends to zero, and the contact resistance R_{NW} tends to the ideal value of $(1/N)(h/2e^2)$. If a lead acting as a reservoir is not sufficiently wide, the contact resistance R_{NW} between the lead and the narrow device is less than the ideal value as evident from (5.7a). This agrees with the result derived by Landauer (apart from a factor of $2/\pi$) from a very different approach.⁵⁷

To check the above results we first consider the structure shown in Fig. 20(a). Figure 20(b) shows the computed potential profile. There are Friedel-like⁵⁸ oscillations in the potential due to interference effects, just as we saw earlier for a tunneling barrier (Fig. 10). Ignoring the oscillations we can estimate the fraction of the potential μ_{NW} that drops across the narrow region and compute the contact resistance R_{NW} between the narrow region with one mode ($N=1$) and the wide region with four modes ($W=4$). We obtain, as shown in Fig 20(b),

$$R_{NW} \simeq 0.80 (h/2e^2), \quad (5.8a)$$

$$R_W \simeq 0.29 (h/2e^2). \quad (5.8b)$$

Both R_{NW} and R_W are somewhat larger than what we expect from (5.7a) and (5.7b) with $N=1$, $W=4$. It should be noted that the individual values of R_{NW} and R_W are

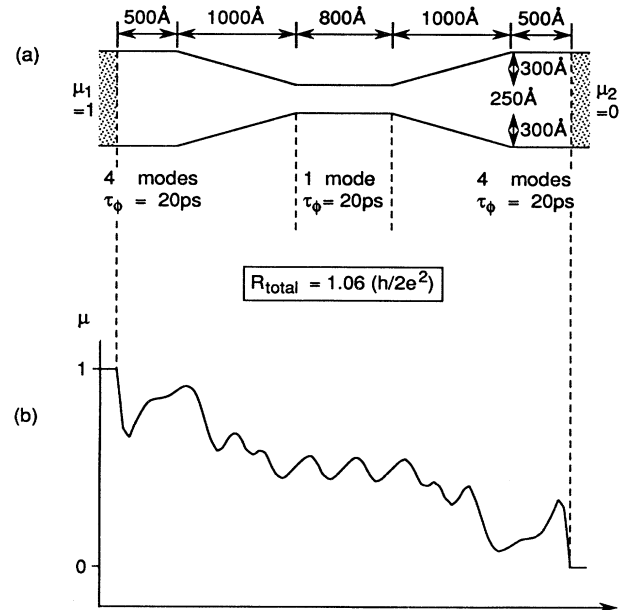


FIG. 21. Same as Fig. 20 for a wide-narrow-wide (WNW) structure with graded transitions.

only approximate because of the oscillatory nature of the potential. Only the sum $R_{\text{total}} = R_{NW} + R_W$ is known accurately. The total resistance $R_{\text{total}} = 1.09(h/2e^2)$ is reasonably close to the estimate in (5.7c). It is reduced slightly to $1.06(h/2e^2)$ when we grade the junction between the wide and narrow regions (Fig. 21).

The examples in Figs. 20 and 21 both involve nearly ballistic structures ($\tau_\phi = 20$ ps). We will now look at two examples where the narrow regions are ballistic ($\tau_\phi = 20$ ps), but the wide regions have significant amounts of scattering ($\tau_\phi = 1$ and 0.1 ps). This helps damp out the oscillations, making the potential drops across the different regions less ambiguous. Unfortunately, it also introduces a distributed series resistance across the wide regions. However, this series resistance can be estimated from the potential drop computed earlier for a straight wire [Figs. 19(a) and 19(b)] and subtracted out. We can then estimate the contact resistance R_{NW} between the narrow and wide regions as shown in Figs. 22(b) and 23(b),

$$R_{NW} \simeq 0.74(h/2e^2), \quad \tau_\phi(\text{wide}) = 1 \text{ ps}, \quad (5.9a)$$

$$R_{NW} \simeq 1.05(h/2e^2), \quad \tau_\phi(\text{wide}) = 0.1 \text{ ps}. \quad (5.9b)$$

With $\tau_\phi(\text{wide}) = 1$ ps there are oscillations in the potential [Fig. 22(b)] making it difficult to estimate R_{NW} and R_W

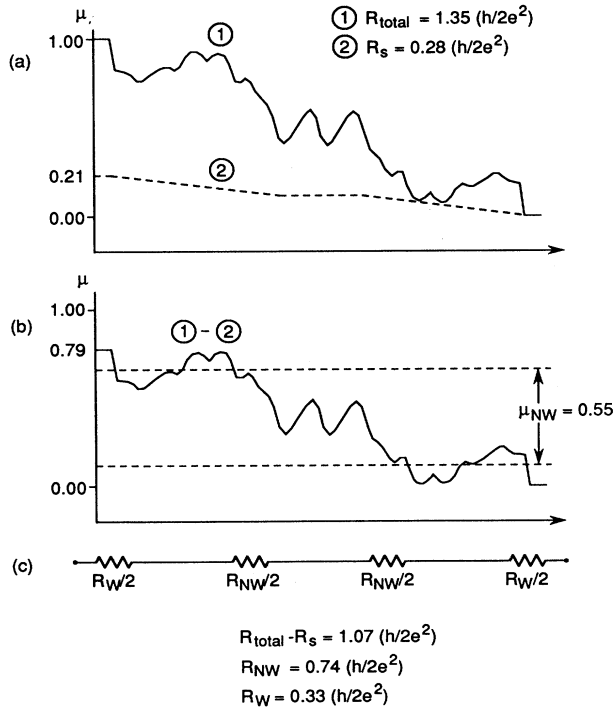


FIG. 22. (a) Potential drop in the same structure as in Fig. 20 but with $\tau_\phi = 1$ ps in the wide region. The dashed curve shows the potential drop due to the series resistance of the wide regions, as deduced from Fig. 19(a). (b) Potential drop after subtracting the series resistance. (c) Equivalent circuit.

individually. We had this same difficulty earlier with $\tau_\phi = 20$ ps (Fig. 20). But with $\tau_\phi(\text{wide}) = 0.1$ ps the oscillations are completely damped, making the estimates for R_{NW} and R_W quite unambiguous [Fig. 23(b)]. It is interesting to note how flat the potential is in the wide regions after subtracting the series resistance, considering that this series resistance was obtained from a *different* structure [Fig. 19(b)].

The contact resistance R_{NW} for $\tau_\phi = 1$ ps is in agreement with our estimate in (5.8). However, for $\tau_\phi = 0.1$ ps, R_{NW} is significantly larger. We can understand this by noting that the phase-breaking length L_ϕ in this case is $\simeq 250$ Å, which is more than a factor of 3 smaller than the width of the wide regions. Since the wide region is much wider than L_ϕ , it is *effectively infinite in width*. Consequently we expect R_{NW} to increase from $0.75(h/2e^2)$ to its full value of $1.00(h/2e^2)$. This qualitative change in the nature of the wide region is also reflected in the current flow pattern as we reduce $\tau_\phi(\text{wide})$ from 1 to 0.1 ps (Fig. 24). With $\tau_\phi(\text{wide}) = 1$ ps there are obvious interference effects in the current flow pattern in the wide region, which should be sensitive to the width of the structure. But with $\tau_\phi(\text{wide}) = 0.1$ ps the current flow pattern is just what one expects intuitively in the classical diffusive regime and should remain unaffected by the width.

Although the numerical results are only approximate, they serve to illustrate the localized nature of the contact resistance. We conclude that a fairly localized drop

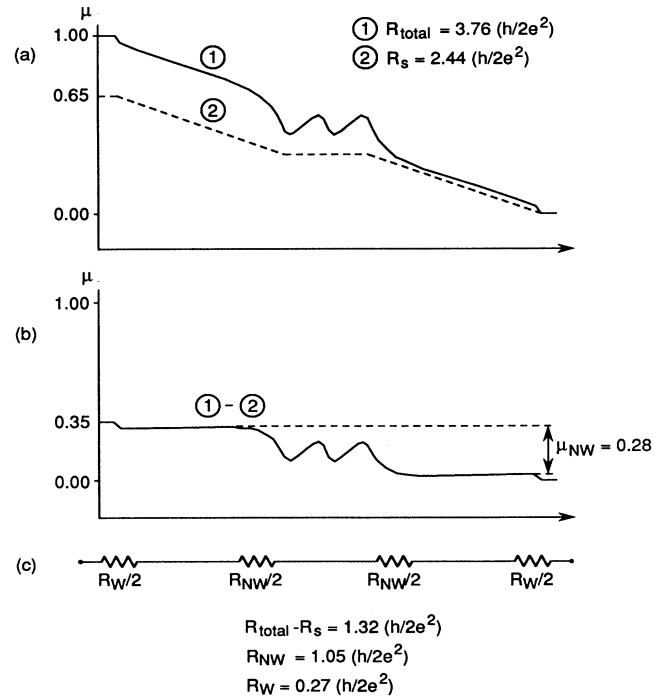


FIG. 23. Same as Fig. 22 but with $\tau_\phi = 0.1$ ps in the wide region. The series resistance is obtained from Fig. 19(b).

occurs when the number of current-carrying channels is abruptly constricted from W to N ($W > N$). This drop can be described by a contact resistance equal to $(1/N)(h/4e^2)$ if the wide region is wider than L_ϕ . However, if the wide region is narrower than L_ϕ , then the contact resistance is reduced by the factor $(1 - N/W)$.

There is one caveat we would like to sound before concluding this section. In many of the figures we have shown equivalent circuits with resistances in series. The resistances are drawn to reflect the spatial drops in the electrochemical potential. However, these resistances do *not* necessarily give us the spatial distribution of the dissipated power. From our examples it is apparent that a drop in the electrochemical potential occurs whenever there is an obstacle to current flow such as an impurity (Fig. 2) or a constriction [Fig. 24(a)]. However, such an obstacle may not have the internal degrees of freedom necessary for power dissipation. Thus the IR drop is not necessarily associated with an I^2R loss, though we believe that the I^2R loss should occur within an inelastic-

scattering length of the obstacle. The spatial distribution of the dissipated power will depend on the spectrum of the phase-breaking scatterers. In all our calculations in this paper we have not specified the nature of the phase-breaking scatterers that give rise to τ_ϕ ; indeed, these scatterers could be magnetic impurities giving rise to purely elastic scattering. In that case all the power dissipation would occur in the contacts where one must have inelastic scattering in order to enforce the assumed Fermi-Dirac boundary conditions. We believe that the question of where the power is lost can be answered by solving the full kinetic equation [Eq. (2.19)] and computing the divergence of the heat current. We leave this for future investigations.

VI. THE ELECTROSTATIC POTENTIAL

To this point, we have neglected the charge of an electron, treating it as a neutral particle. We now recognize that any charge imbalances within the sample will give rise to an electrostatic potential $\phi(\mathbf{r})$ which accounts for electron-electron interactions in the Hartree approximation. This potential is computed from the Poisson equation,

$$\nabla^2 \phi(\mathbf{r}) = \frac{e}{\epsilon} [N_D^+(\mathbf{r}) - n(\mathbf{r})], \quad (6.1)$$

where $N_D^+(\mathbf{r})$ is the density of ionized donors, and $n(\mathbf{r})$ is the electron density. Assuming that we have already computed an equilibrium solution for $\phi(\mathbf{r})$, we may consider only the changes $\delta\phi(\mathbf{r})$ arising under bias,

$$\nabla^2 \delta\phi(\mathbf{r}) = -\frac{e}{\epsilon} \delta n(\mathbf{r}). \quad (6.2)$$

In general, these changes could be significant enough to influence the solution of the transport problem. In the linear-response regime, however, we can neglect these effects to first order, as explained in Sec. II A.

To evaluate $\delta n(\mathbf{r})$, we note that the equilibrium electron density is given by

$$n_{\text{eq}}(\mathbf{r}) = \int dE N_0(\mathbf{r}; E) f_0(E), \quad (6.3)$$

where $f_0(E)$ is the Fermi-Dirac factor with a constant electrochemical potential μ_0 . Under a small bias, $f_0(E)$ changes to $f(\mathbf{r}; E)$ with a local potential $\mu(\mathbf{r})$, suggesting that

$$n(\mathbf{r}) = \int dE N_0(\mathbf{r}; E) f(\mathbf{r}; E). \quad (6.4)$$

However, we should note that the electrostatic potential change $\delta\phi(\mathbf{r})$ also causes a first-order change in the local density of states N_0 , and we should take this into account. In the Thomas-Fermi picture, we can simply assume that $N_0(\mathbf{r}; E) \rightarrow N_0(\mathbf{r}; E - e\delta\phi)$. Consequently, the electron density under bias (6.4) can be determined from

$$\begin{aligned} n(\mathbf{r}) &= \int dE N_0(\mathbf{r}; E - e\delta\phi) f(\mathbf{r}; E) \\ &= \int dE N_0(\mathbf{r}; E) f(\mathbf{r}; E + e\delta\phi). \end{aligned} \quad (6.5)$$

Subtracting this from the equilibrium electron density (6.3), we find

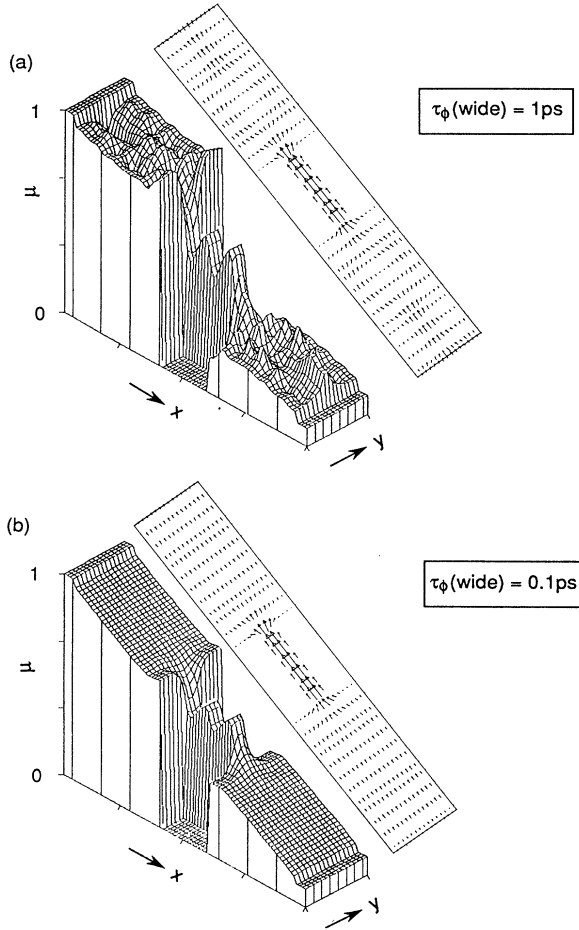


FIG. 24. Potential profile and current flow pattern for the wide-narrow-wide structure in Fig. 20 with $\tau_\phi = 20$ ps in the narrow region and (a) $\tau_\phi(\text{wide}) = 1$ ps and (b) 0.1 ps.

$$\begin{aligned}\delta n(\mathbf{r}) &= n(\mathbf{r}) - n_{\text{eq}}(\mathbf{r}) \\ &= \int dE N_0(\mathbf{r}; E) [f(\mathbf{r}; E + e\delta\phi) - f_0(E)].\end{aligned}\quad (6.6)$$

Using a Taylor series expansion for $f(\mathbf{r}; E + e\delta\phi)$ up to the first derivative, we obtain

$$\delta n(\mathbf{r}) \simeq \int dE \left[-\frac{\partial f_0}{\partial E} \right] N_0(\mathbf{r}; E) e [\delta\mu(\mathbf{r}) - \delta\phi(\mathbf{r})], \quad (6.7)$$

where $\delta\mu(\mathbf{r}) \equiv \mu(\mathbf{r}) - \mu_0$. At low temperatures, we can neglect the thermal spread of electron energies ($-\partial f_0/\partial E$), and write

$$\delta n(\mathbf{r}) = e N_0(\mathbf{r}; \mu_0) [\delta\mu(\mathbf{r}) - \delta\phi(\mathbf{r})]. \quad (6.8)$$

Thus, the excess electron density can be viewed as the charge pileup due to the transport process [$\propto \delta\mu(\mathbf{r})$] minus the screening charge [$\propto \delta\phi(\mathbf{r})$].

Having derived an expression for the electron density $\delta n(\mathbf{r})$, we can now express the Poisson equation (6.2) as

$$\nabla^2 \delta\phi(\mathbf{r}) = -\frac{e^2 N_0(\mathbf{r}; \mu_0)}{\epsilon} [\delta\mu(\mathbf{r}) - \delta\phi(\mathbf{r})]. \quad (6.9)$$

One way of solving this equation is to compute the impulse response $G_\phi(\mathbf{r}, \mathbf{r}')$,

$$\left[\nabla^2 - \frac{e^2 N_0(\mathbf{r}; \mu_0)}{\epsilon} \right] G_\phi(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \quad (6.10)$$

and then integrate over the source function,

$$\begin{aligned}\delta\phi(\mathbf{r}) &= \int d^3\mathbf{r}' G_\phi(\mathbf{r}, \mathbf{r}') \\ &\times \left[-\frac{e^2 N_0(\mathbf{r}'; \mu_0)}{\epsilon} \delta\mu(\mathbf{r}') \right].\end{aligned}\quad (6.11)$$

Note that if the density of states N_0 is a constant, then the impulse response is well known:

$$G_\phi(\mathbf{r}, \mathbf{r}') = \frac{e^{-k_s |\mathbf{r} - \mathbf{r}'|}}{4\pi |\mathbf{r} - \mathbf{r}'|}, \quad (6.12a)$$

$$k_s^2 \equiv e^2 N_0 / \epsilon. \quad (6.12b)$$

This is the usual result for the screening of a point charge in an electron gas. Since the medium is homogeneous, the impulse response G_ϕ depends only on the difference coordinate $|\mathbf{r} - \mathbf{r}'|$. In this approximation, the electrostatic potential $\delta\phi$ can be expressed as a convolution of the electrochemical potential $\delta\mu$ with a screening function,

$$\delta\phi(\mathbf{r}) = \frac{e^2 N_0}{\epsilon} \int d^3\mathbf{r}' \delta\mu(\mathbf{r}') G_\phi(\mathbf{r} - \mathbf{r}'). \quad (6.13)$$

From this relationship, we gain some insight into the nature of these two potentials: *Under no circumstances will $\delta\phi$ vary more rapidly than $\delta\mu$.* At best, $\delta\phi$ will track $\delta\mu$ in a highly conductive medium; usually, it will be more smeared out.

Of course, in our numerical calculations we can actually compute the density of states $N_0(\mathbf{r}; \mu_0)$, and thus account for the screening in a more quantitative manner.

To accomplish this, it is useful to separate $\delta\phi$ into two components: one $\delta\phi_e$ satisfying both the homogeneous equation,

$$\nabla^2 \delta\phi_e(\mathbf{r}) = 0 \quad (6.14)$$

and the boundary conditions supplied by an external source, and another $\delta\phi_i$ arising from the charges induced by screening within the sample. The electrostatic potential $\delta\phi$ is the sum of the two,⁵⁹

$$\delta\phi \equiv \delta\phi_e + \delta\phi_i. \quad (6.15)$$

The component $\delta\phi_e$ depends only on the boundary conditions for a particular device. If we assume that a device has two contacts represented by semi-infinite conductive sheets (Fig. 25), then a solution for $\delta\phi_e$ which satisfies the homogeneous Poisson equation (6.14) is

$$\delta\phi_e(x, y) = \frac{V_L - V_R}{\pi} \cos^{-1} \left[\frac{x}{d} \right] + V_R, \quad (6.16)$$

where V_L and V_R are the voltages applied to the contact plates located at $x = \pm d$.

An equation for the component $\delta\phi_i$ is obtained by substituting the definition (6.15) for $\delta\phi$ into the Poisson equation (6.9), and using the property (6.14) of $\delta\phi_e$,

$$\begin{aligned}\nabla^2 \delta\phi_i(\mathbf{r}) &= -\frac{e^2 N_0(\mathbf{r}; \mu_0)}{\epsilon} \\ &\times [\delta\mu(\mathbf{r}) - \delta\phi_e(\mathbf{r}) - \delta\phi_i(\mathbf{r})].\end{aligned}\quad (6.17)$$

Once we have solved the transport problem for the electrochemical potential $\delta\mu$, we can compute the induced potential $\delta\phi_i$ by solving this equation, and compute the electrostatic potential $\delta\phi$ by summing the components $\delta\phi_e$ and $\delta\phi_i$.

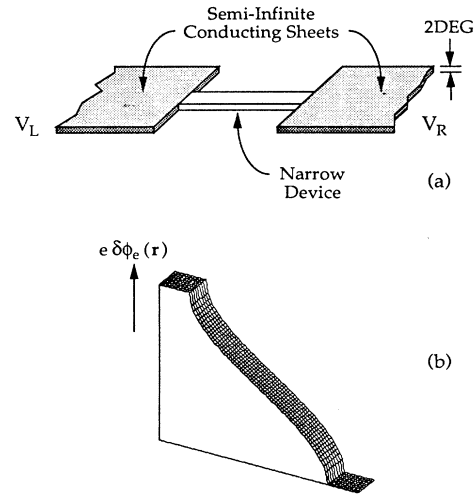


FIG. 25. (a) Contact regions at the ends of the wire are simulated as semi-infinite conductive sheets. (b) The electrostatic potential $\delta\phi_e$ in the absence of any charges within the device.

Suppose, for a moment, that our sample is a homogeneous resistor. Then the electrochemical potential $\delta\mu(\mathbf{r})$ will satisfy the continuity equation,²⁵

$$\nabla \cdot \mathbf{J} = \sigma \nabla^2 \delta\mu = 0. \quad (6.18)$$

This is precisely the same equation (6.14) satisfied by $\delta\phi_e$. Since these two potentials have the same boundary conditions, the solutions will be identical: $\delta\mu = \delta\phi_e$. Thus, there will be no charge imbalance within the sample, and the solution for the induced potential (6.17) will be $\delta\phi_i = 0$.

In a quantum device, however, $\delta\mu(\mathbf{r})$ must be computed from the transport equation (3.6), and it will differ slightly from $\delta\phi_e$. This difference will provide source charges in the Poisson equation (6.17) and induce a screening potential $\delta\phi_i$. To illustrate this behavior, we return to the single-moded wire with a tunneling barrier, studied in Sec. III. Interference effects caused by reflections from the tunneling barrier cause the solution for $\delta\mu$ to oscillate. However, the potential $\delta\phi_e$ varies smoothly between contacts (6.16), since its solution is independent of the details of the transport within the wire. The difference between these two potentials creates charge imbalances within the wire, which induce the screening potential $\delta\phi_i$ shown in Fig. 26(b). This potential tends to compensate for the variations in $\delta\mu$. On one side of the barrier, electrons pile up, so that the screening potential $\delta\phi_i$ floats up to compensate; on the other, elec-

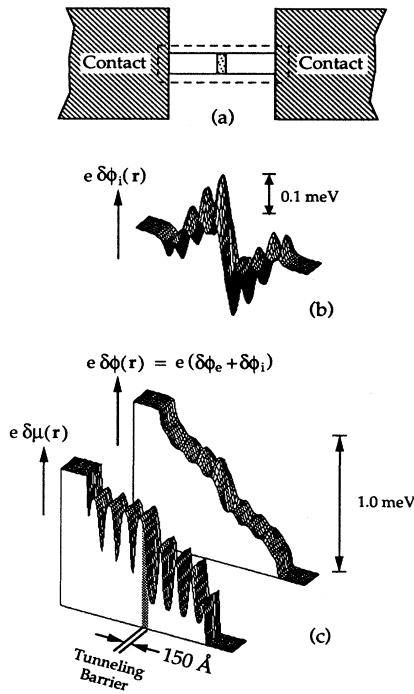


FIG. 26. (a) Geometry of a wire with a tunneling barrier, including the contact regions. (b) Differences between $\delta\mu$ and $\delta\phi_e$ induce a screening potential $\delta\phi_i$. (c) A comparison of the potentials $\delta\phi$ and $\delta\mu$ within the structure.

trons drain off, so $\delta\phi_i$ floats downward. This is precisely the residual resistivity dipole described by Landauer.^{4,7} When this induced potential $\delta\phi_i$ is added to the homogeneous solution $\delta\phi_e$, the result is a potential $\delta\phi$ which is similar to the electrochemical potential $\delta\mu$. The response of the induced potential, however, is limited by the screening length within the material, so that the variations in $\delta\phi$ are more smeared out than the variations in $\delta\mu$, as shown in Fig. 26(c).

It is interesting to compare the electrostatic potential $\delta\phi$ to the electrochemical potential $\delta\mu$ for two of the structures encountered in the preceding sections. First, we examine these potentials for the ballistic constriction studied in Sec. V. Again, the rapid variations in $\delta\mu$ appear to some extent in $\delta\phi$, as shown in Fig. 27. This is particularly true in the wide regions, where the density of states N_0 is larger, and hence the screening length is shorter. In the narrow channel, however, $\delta\phi$ is distinctly sloped, while $\delta\mu$ is nearly flat. It is conceivable that localized probes could be constructed to measure these two potentials within the narrow channel.⁵⁷ If measurements of this sort were actually performed, these two different potentials should indicate different resistances for the ballistic region.

As a final example, we compare the potentials $\delta\mu$ and $\delta\phi$ in the four-probe Hall junction studied in Sec. IV. To obtain a clearer view of the junction, we exclude the contact regions from these plots, which are presented in Fig. 28. The usual Hall buildup evident in the electrochemical potential $\delta\mu$ is tracked to some extent by $\delta\phi$. But again these two potentials clearly differ. Of course, the

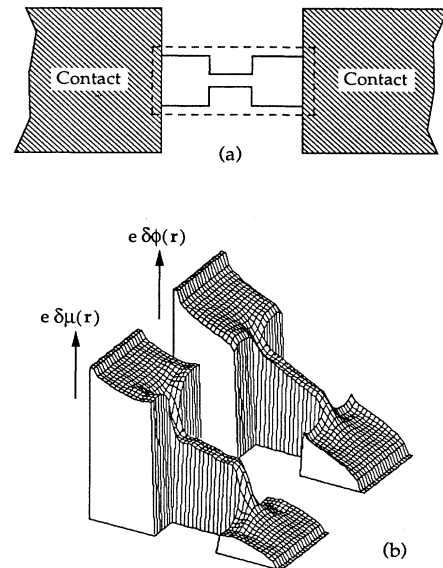


FIG. 27. (a) Geometry of a ballistic constriction, including the contact regions. There are five modes in the wide regions and one mode in the narrow region; τ_ϕ is assumed to be 0.1 ps everywhere. (b) A comparison of the potentials $\delta\mu$ and $\delta\phi$ within the structure.

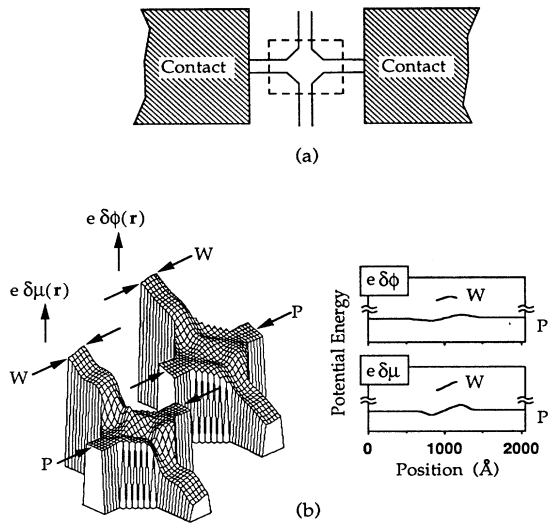


FIG. 28. (a) Geometry of the four-probe Hall junction; plots exclude the contact regions for a better view of the junction. (b) The potentials $\delta\mu$ and $\delta\phi$ within the structure; cross sections along W and P are presented for better comparison.

wires in this structure are single moded and extremely narrow. As a result, the relatively long screening length keeps the electrostatic Hall potential from building up to its proper value. Presumably, the two potentials would show better agreement in wider structures.

From this formulation, it appears that $\mu(\mathbf{r})$, rather than $\phi(\mathbf{r})$, may be a better choice as the basis for a transport theory in the linear-response regime. Once a solution for $\mu(\mathbf{r})$ has been obtained, it is a simple matter to perform the integral (6.11) required to compute $\phi(\mathbf{r})$. However, performing the reverse—computing a solution for $\phi(\mathbf{r})$ and extracting a solution for $\mu(\mathbf{r})$ —could be quite difficult. Essentially, this would require deconvolving a smeared-out function to obtain a rapidly varying quantity. From a numerical standpoint, such a procedure would be extremely prone to error.

VII. SUMMARY AND CONCLUSIONS

A. Assumptions

(1) We assume a simple model for phase-breaking scattering in which each scatterer acts independently and interacts with electrons through a δ potential. Note that these scatterers not only randomize the phase but also the momentum (Sec. II C).

(2) We neglect any vertical flow of carriers (from one energy to another) so that different energies can be decoupled (Sec. II D).

(3) We compute only the linear response, that is, the change in different quantities arising under bias. This change involves only a Fermi-surface term, so that at low temperatures we only need to consider electrons with a single energy, namely, the equilibrium Fermi energy flow. Strictly speaking, the linear-response current density

$\delta\mathbf{J}(\mathbf{r})$ includes a component $\delta\mathbf{J}_{\text{eq}}(\mathbf{r})$ that involves all energies. However, this component does not contribute to the terminal current, and we neglect it in this paper (Sec. III B).

B. Solution procedure

(1) Solve Eq. (3.7b) for $G^R(\mathbf{r}, \mathbf{r}'; \mu_0)$. The phase-breaking time $\tau_\phi(\mathbf{r}; E)$ is treated as an input parameter (Appendix, Sec. 1).

(2) Solve Eq. (3.6) for $\mu(\mathbf{r})$ with the appropriate boundary conditions. The current density $\delta\mathbf{J}(\mathbf{r})$ is then computed from Eq. (3.12). The terminal currents are obtained either from Eq. (3.13) or from Eqs. (3.17) and (3.18) (Appendix, Sec. 2).

(3) Solve Eq. (6.9) for the electrostatic potential $\delta\phi(\mathbf{r})$ (Appendix, Sec. 3).

C. Conclusions

(1) A local electrochemical potential $\mu(\mathbf{r})$ can be defined if $eV_A < k_B T < \Gamma_c$, where V_A is the applied bias and Γ_c is the correlation energy over which the propagation characteristics of electrons can be assumed constant (Sec. III A). The electrochemical potential at any point \mathbf{r} is a weighted average of the potentials at surrounding points \mathbf{r}' ; the weighting function is equal to the probability that an electron suffering a phase-breaking scattering at \mathbf{r}' suffers the next one at \mathbf{r} (Sec. III B).

(2) Separate potentials μ_L and μ_R for left- and right-moving electrons are best defined in an average sense over a de Broglie wavelength, rather than in a local sense (Sec. III D).

(3) Noninvasive probes can render a faithful measurement of the local electrochemical potential (Sec. IV B) but do not guarantee it (Sec. IV C).

(4) A fairly localized potential drop occurs when the number of current-carrying channels is abruptly constricted from W to N ($W > N$). This drop can be described by a contact resistance equal to $(1/N)(h/4e^2)$ if the wide region is wider than L_ϕ . However, if the wide region is narrower than L_ϕ , then the contact resistance is reduced by the factor $(1 - N/W)$ (Sec. V).

(5) The electrostatic potential $\delta\phi(\mathbf{r})$ can be viewed as a convolution of the electrochemical potential $\delta\mu(\mathbf{r})$ with a screening function. In very conductive samples the screening function $\sim \delta(\mathbf{r})$, so that $\delta\mu$ and $\delta\phi$ are identical, and no space charge appears. But in less conductive samples, $\delta\phi(\mathbf{r})$ varies slowly compared to $\delta\mu(\mathbf{r})$ (Sec. VI).

(6) Broadly speaking, there is a drop in the electrochemical potential wherever there is an obstacle to current flow, while the electrostatic potential drop is spread out over a screening length.

Some of the conclusions in this paper have been drawn by other researchers (notably Landauer, Büttiker, and Imry) using heuristic arguments. The main contribution of this work lies in putting these ideas on a rigorous quantitative footing using a quantum kinetic approach.

The calculations implemented in this paper represent a step toward the simulation of quantum devices at the microscopic level. Since the method has proven to be quite

tractable, it may be possible to extend the model by removing some of the simplifying assumptions. The thermal energy spread of electrons, for instance, could be taken into account by solving Eq. (2.32) instead of Eq. (3.6). The finite thickness of the two-dimensional electron gas (2DEG) could be taken into account by summing the contributions to the kernel $T_0(\mathbf{r}, \mathbf{r}')$ from a number of modes. Finally, it may be possible to go beyond the linear-response regime and also take into account spatially extended phase-breaking processes with nonlocal self-energy functions. We leave such considerations to future work.

ACKNOWLEDGMENTS

The authors would like to acknowledge many helpful discussions with Roger Lake. We also thank Rolf Landauer for his critical comments. This work was supported by the Semiconductor Research Cooperation (Contract No. 89-SJ-089) and the National Science Foundation (Grant No. ECS-83-51-036).

APPENDIX: DETAILS OF THE NUMERICAL METHOD

The purpose of this appendix is to describe a numerical method for computing the linear-response behavior of any arbitrary device. For simplicity, we restrict our attention to electrons in the conduction band of an n -type device. We assume that the temperature is extremely low (so that we can neglect the thermal spread of electron energies) and that the bias is extremely small (so that we can neglect carrier heating). Finally, we recognize that experiments are commonly performed on modulation-doped samples, in which electrons form a two-dimensional electron gas (2DEG). Hence, we confine our analysis to a two-dimensional plane of infinitesimal thickness. Of course, this plane can be patterned to form the specific geometry of any given device. Within the device, electrons obey an effective-mass Hamiltonian,

$$H_0 = \frac{1}{2m} \left[\frac{\hbar}{i} \nabla - e \mathbf{A}(\mathbf{r}) \right]^2 + V(\mathbf{r}), \quad (\text{A1})$$

where m is the electron effective mass, $\mathbf{A}(\mathbf{r})$ and $V(\mathbf{r})$ are the vector and scalar potentials, and $e = -|e|$ is the electronic charge. Because of our simplification to a two-dimensional geometry, we now regard the vector \mathbf{r} as spanning a two-dimensional space.

Strictly speaking, the solution procedure should begin with an analysis of the equilibrium state. In this analysis, the equilibrium Fermi energy μ_0 is chosen to be constant, and an initial guess of the conduction-band profile $V(\mathbf{r})$ is made. Based upon this guess, we compute a local density of states $N_0(\mathbf{r}; E)$ and use this to compute the electron density $n(\mathbf{r})$,

$$n(\mathbf{r}) = \int dE \frac{1}{e^{(E - \mu_0)/k_B T} + 1} N_0(\mathbf{r}; E). \quad (\text{A2a})$$

We then solve the Poisson equation for the electrostatic potential $\phi(\mathbf{r})$,

$$\nabla^2 \phi(\mathbf{r}) = \frac{e}{\epsilon} [N_D^+(\mathbf{r}) - n(\mathbf{r})], \quad (\text{A2b})$$

and use this to improve our guess of the conduction-band profile,

$$V(\mathbf{r}) = \Delta E_C(\mathbf{r}) + e\phi(\mathbf{r}), \quad (\text{A2c})$$

where ΔE_C accounts for any offsets in the conduction band due to changing material composition, impurities, etc. With a new guess of the conduction-band profile, we compute a new density of states $N_0(\mathbf{r}; E)$, a new electron density $n(\mathbf{r})$, and a new electrostatic potential $\phi(\mathbf{r})$. This iterative process is repeated until the equilibrium solution converges.

Obviously, the equilibrium solution entails a considerable amount of work. To keep our method tractable, we neglect this first step and simply assume some form of the equilibrium solution for $\phi(\mathbf{r})$. In all of our present calculations, we assume $\phi(\mathbf{r}) = 0$; it is possible, however, to obtain a better guess of $\phi(\mathbf{r})$ through a semiclassical analysis.

Having defined the equilibrium state, we can compute the electrochemical potential $\mu(\mathbf{r})$ arising under bias from the transport equation derived in Sec. III,

$$\mu(\mathbf{r}) = \frac{\int d^2 \mathbf{r}' T_0(\mathbf{r}, \mathbf{r}') \mu(\mathbf{r}')}{\int d^2 \mathbf{r}' T_0(\mathbf{r}, \mathbf{r}')}, \quad (\text{A3})$$

where

$$T_0(\mathbf{r}, \mathbf{r}') \equiv \frac{\hbar^2 |G^R(\mathbf{r}, \mathbf{r}'; \mu_0)|^2}{\tau_\phi(\mathbf{r}; \mu_0) \tau_\phi(\mathbf{r}'; \mu_0)}, \quad (\text{A4a})$$

$$\left[E - H_0 + \frac{i\hbar}{2\tau_\phi(\mathbf{r}; \mu_0)} \right] G^R(\mathbf{r}, \mathbf{r}'; E) = \delta(\mathbf{r} - \mathbf{r}'). \quad (\text{A4b})$$

Note that the kernel $T_0(\mathbf{r}, \mathbf{r}')$ is obtained from the Hamiltonian H_0 with the equilibrium solution for $\phi(\mathbf{r})$. As such, it is independent of bias and ensures a *linear* relationship between current and voltage.

Given a solution for the electrochemical potential $\mu(\mathbf{r})$, we can reconstruct the Wigner function and compute any quantity of interest. In particular, we can compute the change in the current density arising under bias,

$$\begin{aligned} \delta \mathbf{J}(\mathbf{r}) = & \frac{e\hbar^2}{2\pi m} \int d^2 \mathbf{r}' \frac{\mu(\mathbf{r}') - \mu_0}{\tau_\phi(\mathbf{r}'; \mu_0)} \\ & \times \text{Im} [G^{R*}(\mathbf{r}, \mathbf{r}'; \mu_0) \nabla_{\mathbf{r}'} G^R(\mathbf{r}, \mathbf{r}'; \mu_0)] \\ & - \frac{e^2 \hbar}{2\pi m} \mathbf{A}(\mathbf{r}) \int d^2 \mathbf{r}' \frac{\mu(\mathbf{r}') - \mu_0}{\tau_\phi(\mathbf{r}'; \mu_0)} |G^R(\mathbf{r}, \mathbf{r}'; \mu_0)|^2. \end{aligned} \quad (\text{A5})$$

We can also compute the terminal currents I_n for each contact n ,

$$I_n = \int_{\text{contact } n} d^2 \mathbf{r} I(\mathbf{r}), \quad (\text{A6a})$$

$$I(\mathbf{r}) = \frac{2e^2}{h} \int d^2 \mathbf{r}' T_0(\mathbf{r}, \mathbf{r}') [\mu(\mathbf{r}) - \mu(\mathbf{r}')], \quad (\text{A6b})$$

where a factor of 2 has been inserted [cf. Eq. (3.18)] to ac-

count for two spins. Of course, we could also compute the terminal currents by integrating the current density over the surface of a contact. As we have stated in Sec. III, both methods produce the same result. We prefer the formulation (A6), however, since numerically it is more robust.

If electrons were neutral particles, our picture of transport would be complete. Because they are not, however, the electrostatic potential will react to screen out any charge imbalances arising under bias. We can compute the change in the electrostatic potential $\delta\phi(\mathbf{r})$ in the presence of an applied bias by solving the Poisson equation, as shown in Sec. VI:

$$\nabla^2\delta\phi(\mathbf{r}) = \frac{e^2 N_0(\mathbf{r}; \mu_0)}{\epsilon} [\delta\mu(\mathbf{r}) - \delta\phi(\mathbf{r})]. \quad (\text{A7})$$

Although this gives us some insight into the nature of charge screening, it does not affect the rest of the analysis. In the linear-response regime, the kernel $T_0(\mathbf{r}, \mathbf{r}')$ is computed based upon the *equilibrium* potential $\phi(\mathbf{r})$; changes in $\delta\phi(\mathbf{r})$ are assumed to be small enough that they have no effect (to first order) on the terminal currents.

If our computing resources were unlimited, we could simply discretize any device (no matter how large) into a finite number of position-space nodes, and then solve the equations shown above. In any practical analysis, however, we must focus only on the interesting part of the device and draw boundaries between what is simulated and what is not. Some of these boundaries are hard potential walls which confine the electrons. Others are *open boundaries* which lead to a large contact reservoir at infinity. To avoid spurious edge effects, we must include the open character of the contact leads, a detail which can complicate the analysis considerably.

In the remainder of this appendix, we describe how a solution of the above equations is actually performed. In Sec. 1, we use a tight-binding Hamiltonian to express the Dyson equation in a matrix form. We solve this equation for the Green function $G^R(\mathbf{r}, \mathbf{r}'; E)$, which is then used to compute the kernel $T_0(\mathbf{r}, \mathbf{r}')$. In Sec. 2, we describe an iterative scheme for computing the electrochemical potential $\mu(\mathbf{r})$ from the transport equation (A3). We then describe how the current density $\delta\mathbf{J}(\mathbf{r})$ and the terminal currents I_n are computed. Finally, in Sec. 3 we describe a three-dimensional solution of the Poisson equation which is used to compute the electrostatic potential $\delta\phi(\mathbf{r})$.

1. Computing the Green function $G^R(\mathbf{r}, \mathbf{r}'; E)$

a. Tight-binding Hamiltonian

Our entire solution method hinges upon a knowledge of the equilibrium Green function $G^R(\mathbf{r}, \mathbf{r}'; E)$. To evaluate this function numerically, we must first discretize our device into a series of nodes. For simplicity, we choose our nodes to be uniformly spaced by a distance a in both directions, so that we can evaluate the derivatives in the Hamiltonian (A1) using finite differences. In the absence of a magnetic field,

$$H_0\psi(\mathbf{r}) \leftrightarrow \frac{-\hbar^2}{2m} \left[\frac{\psi_R - 2\psi + \psi_L}{a^2} + \frac{\psi_B - 2\psi + \psi_T}{a^2} \right] + V\psi, \quad (\text{A8})$$

where ψ represents the value of the wave function at a single node, and we have used the subscripts R , L , T , and B to denote the wave function at surrounding nodes (right, left, top, and bottom). We can rewrite this expression in a more familiar form,

$$H_0\psi(\mathbf{r}) \leftrightarrow \frac{-\hbar^2}{2ma^2} (\psi_R + \psi_L + \psi_T + \psi_B - 4\psi) + V\psi. \quad (\text{A9})$$

This is identical to the tight-binding Hamiltonian for a square lattice with nearest-neighbor interactions

$$H_{\text{TBM}} \equiv \sum_j (U_{jk} |j\rangle \langle k| + \epsilon_j |j\rangle \langle j|), \quad (\text{A10})$$

where

$$U_{jk} \equiv \begin{cases} U & \text{for } j, k \text{ nearest neighbors,} \\ 0 & \text{otherwise,} \end{cases} \quad (\text{A11a})$$

$$\epsilon_j \equiv -4U + V_j, \quad (\text{A11b})$$

and $U \equiv -\hbar^2/2ma^2$ is the amplitude of the elements in the hopping matrix \underline{U} .

We emphasize the relationship between these different viewpoints for two reasons. First, the finite-difference viewpoint provides us with a simple derivation for the site energies ϵ_j and the hopping elements U_{jk} . Second, the tight-binding viewpoint suggests that as our mesh spacing becomes coarse, electrons will feel the effects of the underlying tight-binding lattice. Instead of the simulating free electrons with a parabolic dispersion relationship, we will simulate tightly bound electrons with the energy dispersion

$$E(k) = 2U[\cos(ka) - 1]. \quad (\text{A12})$$

Although the physics may be somewhat different, it will not be completely wrong. We could not have drawn this conclusion from the finite-difference viewpoint alone.

In the presence of a magnetic field, the hopping elements derived above simply acquire phase factors,

$$U_{jk} \rightarrow U_{jk} e^{-ie \int \mathbf{A} \cdot d\mathbf{l} / \hbar}. \quad (\text{A13})$$

As Feynman has shown,⁶⁰ this corresponds to the so-called minimal substitution,

$$\frac{\hbar}{i} \nabla \rightarrow \frac{\hbar}{i} \nabla - e \mathbf{A}(\mathbf{r}) \quad (\text{A14})$$

in the limit as the lattice spacing a tends to zero. We choose to work in the following gauge:

$$\mathbf{A}(\mathbf{r}) = (0, A_y(\mathbf{r}), 0). \quad (\text{A15})$$

Thus, the hopping elements in the y direction are modified by a phase $\exp(\pm ie A_y a / \hbar)$, where A_y represents the average vector potential between the two nodes connected by the hopping element. A summary of

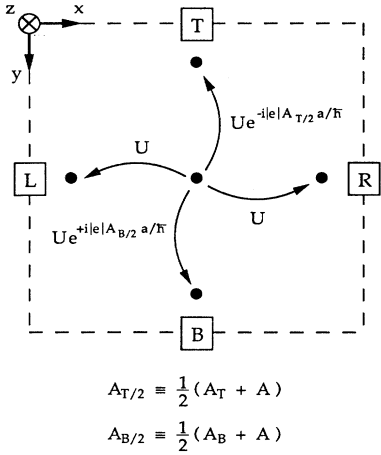


FIG. 29. Hopping elements in the right-binding model.

the hopping elements in our model is illustrated in Fig. 29.

b. Solving the Dyson equation

In the preceding section, we reduced the Hamiltonian of the Schrödinger equation to a discrete form. We can now determine the Green function $G^R(\mathbf{r}, \mathbf{r}'; E)$ by solving a matrix equation. First, we notice that if all the sites in our lattice were *unconnected*, computing the Green function would be trivial. In other words, if we could ignore the hopping elements U_{jk} in the tight-binding Hamiltonian (A10),

$$\tilde{H}_0 \equiv \sum_j \varepsilon_j |j\rangle \langle j|, \quad (\text{A16})$$

then the solution for the Green function (A4b) would simply be

$$G_0(\mathbf{r}_j, \mathbf{r}_k; E) = \frac{\delta_{jk}/a^2}{E - \varepsilon_j + i\hbar/2(\tau_\phi)_j}. \quad (\text{A17})$$

Thus, injection at a particular site k causes a response only at that site, as shown in Fig. 30. We can introduce the coupling between sites by solving the Dyson equation, written here in matrix form,

$$\underline{G}^R = \underline{G}_0 + \underline{G}_0 \underline{U} \underline{G}^R. \quad (\text{A18})$$

where \underline{U} is the matrix of hopping elements. If we solve for the Green-function matrix \underline{G}^R , the resulting equation is of the form $\underline{A} \mathbf{x} = \mathbf{b}$,

$$(\underline{I} - \underline{G}_0 \underline{U}) \underline{G}^R = \underline{G}_0. \quad (\text{A19})$$

Each column in \underline{G}_0 supplies the excitation for injecting at a particular site; each corresponding column in \underline{G}^R represents the response at all nodes in the device. Thus, the solution for \underline{G}^R can be computed one column at a time by supplying a single column of \underline{G}_0 .

Suppose a given device is defined on a mesh of 50×50

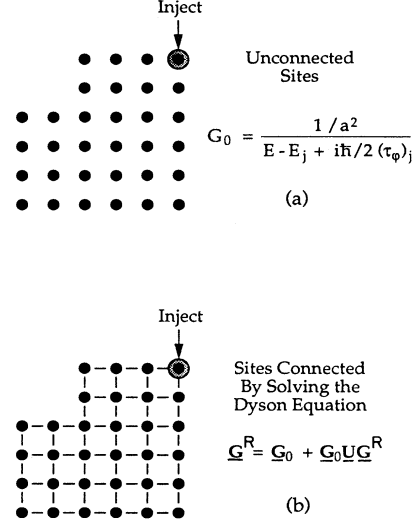


FIG. 30. (a) In an unconnected lattice, injection at a particular site causes a response only at that site. (b) Connections between the sites are incorporated by solving the Dyson Equation.

nodes; the Green function \underline{G}^R describes the response at 2500 nodes due to injection at each of the 2500 nodes. In other words, each of the matrices in Eq. (A19) would have 2500×2500 elements. Actually, the matrices \underline{G}_0 and \underline{U} are sparse and therefore require far less storage space. Nevertheless, handling this matrix equation in an efficient manner is challenging, to say the least.

Another popular technique for the calculation of Green functions is the so-called recursive technique.⁶¹ A generalization of this technique can be found in Ref. 62. In this method, the Green function is computed by adding the connections between sites one slice at a time, as shown in Fig. 31. Although the matrix equation must be solved repeatedly to establish all of the connections, the matrices involved at each step are smaller (each matrix is $N \times N$, where N is the number of sites in the slice). The final result is a Green function connecting one side of the device to the other, which can be related to the transmission matrix for the device.⁶³ When coupled with the Büttiker formula (1.4), this has proven to be an extremely powerful method for obtaining the terminal characteristics of a given device.

In our analysis, however, we require the Green function for injection at *all* nodes in the sample—not just for injection at the ends. For our purposes, therefore, it is considerably more efficient if we introduce the connections between all sites in one step. This eliminates the need to update all Green-function elements at each intermediate step of connecting a slice. Thus, although we must solve a larger matrix equation, we need only solve it once to establish the connections between all nodes.

Before we can solve the Dyson equation (A19), we need to fill in the nonzero elements of matrices \underline{U} and \underline{G}_0 . If all of the sites in the lattice are unconnected, the matrix \underline{G}_0 is diagonal and the elements are computed using Eq. (A17), as discussed above. Things become slightly more

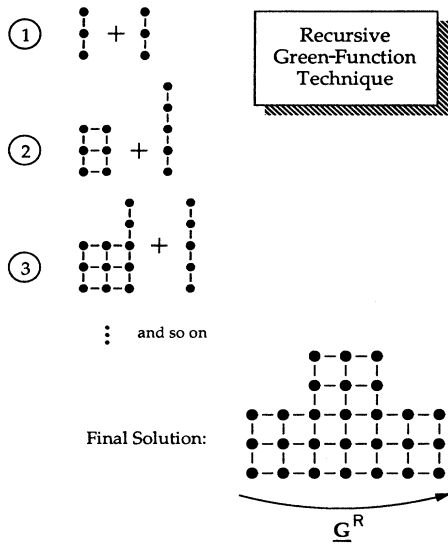


FIG. 31. In the recursive Green-function technique, the sites are connected one slice at a time across the device.

complicated, however, if we attach open boundaries to the device. Sites sharing an open boundary are connected according to the Green function for that boundary, which is derived below in Sec. 1 c. Thus, if we inject at an open-boundary site, we obtain a response at all other sites sharing the open boundary; because of this, \underline{G}_0 gains some off-diagonal elements, as shown in Fig. 32.

Determining the elements of \underline{U} is also relatively straightforward. Each row in this matrix contains the

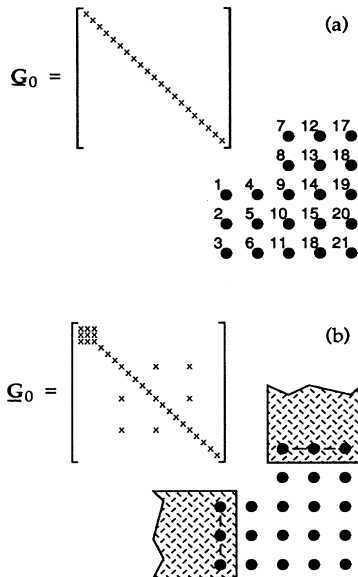


FIG. 32. (a) In a lattice of unconnected sites, \underline{G}_0 is diagonal. (b) After attaching open boundaries, \underline{G}_0 gains some off-diagonal elements.

hopping elements which connect a given site to its nearest neighbors (Fig. 29). Thus, we scan through all sites in the lattice, filling in the rows of \underline{U} one at a time. At each site, we check to see if another site exists above, below, to the right, and to the left. If any of these sites exist, we fill in the corresponding element in \underline{U} with the value shown in Fig. 29. There is one exception to this rule: If the current site and the adjacent site are both part of the same open boundary (discussed further below), then the hopping element is omitted. This is because the connections for sites along an open boundary are already included in the calculation of \underline{G}_0 .

Using these simple rules, we can compute the elements in the overall matrix $(\underline{I} - \underline{G}_0 \underline{U})$, and then solve the Dyson equation (A19) for each column of \underline{G}^R . For a large device, however, it may be impractical to actually store the entire matrix \underline{G}^R in memory. Indeed, in the previous example with 2500 sites, this would require the storage of more than six million elements. Instead, we could simply compute each column of \underline{G}^R as it is required, thereby storing only a single column (for example, 2500 elements) at a time. Although this costs us a little extra computer time, it liberates us from the realm of supercomputers with supermemory capacities.

One of the more common techniques for solving a matrix equation of the form $\underline{A} \underline{x} = \underline{b}$ is L - U decomposition. In this method, the matrix \underline{A} is decomposed into lower-triangular and upper-triangular matrices \underline{L} and \underline{U} (not to be confused with our hopping matrix \underline{U}). For a particular vector \underline{b} , the solution of $\underline{A} \underline{x} = \underline{b}$ is then performed using two back-substitution steps,

$$\underline{A} \underline{x} = (\underline{L} \underline{U}) \underline{x} = \underline{b} \rightarrow \underline{L} \underline{y} = \underline{b}, \tag{A20a}$$

$$\rightarrow \underline{U} \underline{x} = \underline{y}. \tag{A20b}$$

Since the matrices \underline{L} and \underline{U} are triangular, each back-substitution step is trivial. The advantage of this method is that the \underline{L} - \underline{U} decomposition need only be computed once, and yet the matrix equation $\underline{A} \underline{x} = \underline{b}$ can be solved quite rapidly again and again for a number of different vectors \underline{b} . For the current problem, therefore, we compute and store the L - U decomposition of the matrix $[\underline{I} - \underline{G}_0 \underline{U}]$. Whenever a particular column of \underline{G}^R is required, we simply perform two back-substitution steps using the appropriate column from \underline{G}_0 .

Since the L - U decomposition is performed only once, the process need not be optimized. A simple and memory-efficient means of accomplishing this task is Doolittle's method,⁶⁴

$$U_{kj} = A_{kj} - \sum_{p=1}^{k-1} L_{kp} U_{pj}, \quad j = k, k+1, \dots, n, \tag{A21a}$$

$$L_{jk} = \frac{1}{U_{kk}} \left[A_{jk} - \sum_{p=1}^{k-1} L_{jp} U_{pk} \right], \quad j = k+1, \dots, n, \tag{A21b}$$

where the diagonal elements of \underline{L} are assumed to be $L_{kk} = 1$. The elements of \underline{L} and \underline{U} must be determined in order as follows. The first row of \underline{U} is evaluated from left to right, followed by the first column of \underline{L} , from top to

bottom. The next row of \underline{U} is evaluated from left to right, followed by the next column of \underline{L} , from top to bottom, and so on. This method is particularly efficient because it requires no more memory than that allocated for the original matrix \underline{A} . Elements of \underline{A} can actually be replaced by elements of \underline{L} and \underline{U} as the method is carried out.

After working out a few examples, we find that the elements in \underline{L} and \underline{U} fill in toward the diagonal. In other words, if a particular element in \underline{U} is nonzero, all elements below it (down to the diagonal) will also be nonzero; if an element in \underline{L} is nonzero, all elements to the right of it (up to the diagonal) will also be nonzero. Even if these elements were zero in the original matrix \underline{A} , they will “fill in” as the L - U decomposition is performed. Such considerations are important from the standpoint of memory conservation. Since the matrices \underline{L} and \underline{U} are sparse, we need only allocate memory for the nonzero elements. To provide a sense of what these matrices look like, the matrix $(\underline{I} - \underline{G}_0 \underline{U})$ is shown for two example structures in Fig. 33. This matrix is gradually replaced in memory by the elements of \underline{L} and \underline{U} during the L - U decomposition.

When allocating the memory elements for \underline{L} and \underline{U} , the emphasis should be placed on optimizing the two back-substitution steps (A20). After all, these steps will be performed again and again as each column of \underline{G}^R is needed. For this reason, the elements of \underline{L} should be allocated first, across each row from left to right. This is the way these elements will be recalled during the first back-substitution step. The elements of \underline{U} should be allocated

next, across each row from left to right, starting at the last row of \underline{U} and working up. Again, this is the way these elements will be recalled during the second back-substitution step. The memory requirements for these matrices may be large, and this allocation scheme will avoid excessive page faults (thrashing) on machines with virtual memory.

At this point, our solution method for Green function \underline{G}^R is complete. After allocating storage space, the elements of the matrix $(\underline{I} - \underline{G}_0 \underline{U})$ for the Dyson equation (A19) are initialized. An L - U decomposition (A21) is performed, replacing the elements of $(\underline{I} - \underline{G}_0 \underline{U})$ in storage. Any column of \underline{G}^R can then be computed, given the appropriate column in \underline{G}_0 for injection at a particular site.

c. Green function in a semi-infinite wire

Our solution method for the Green function \underline{G}^R is complete except for one detail: We need an expression for the Green function \underline{G}_0 at an open boundary. This expression can be derived as follows. It is well known that any Green function can be expanded in terms of a complete set of eigenfunctions $\phi_\alpha(\mathbf{r})$,

$$G(\mathbf{r}, \mathbf{r}'; E) = \sum_{\alpha} \frac{\phi_{\alpha}(\mathbf{r}) \phi_{\alpha}^*(\mathbf{r}')}{E - E_{\alpha} + i\hbar/2\tau_{\phi}}, \quad (\text{A22})$$

where

$$H_0 \phi_{\alpha}(\mathbf{r}) = E_{\alpha} \phi_{\alpha}(\mathbf{r}). \quad (\text{A23})$$

To obtain the Green function for a semi-infinite wire, therefore, we simply perform this expansion. First, assume that the wire is purely one-dimensional; we will later add the effects of a finite width. The eigenfunctions for a wire with N sites, and thus a length $L = Na$, are known to be

$$\phi_k(x) = \frac{1}{\sqrt{2L}} (e^{ikx} - e^{-ikx}), \quad (\text{A24a})$$

$$E_k = 2U [\cos(ka) - 1], \quad (\text{A24b})$$

where $U = -\hbar^2/2ma^2$ is the amplitude of the hopping elements. Note that at $x=0$, the wave function vanishes. We now extend the number of sites N to infinity, so that the wave vectors k form a continuum. Hence, the summation in the expansion (A22) can be replaced by an integral,

$$G(\mathbf{r}_p, \mathbf{r}_q; E) = \left(\frac{1}{2L} \right) \frac{L}{\pi} \int_0^{\pi/a} dk \times \frac{(e^{ikpa} - e^{-ikpa})(e^{-ikqa} - e^{ikqa})}{E + i\delta - 2U [\cos(ka) - 1]}, \quad (\text{A25})$$

where we have defined $\delta = \hbar/2\tau_{\phi}$ to simplify the notation. Since we are only interested in the Green function along the edge of the wire, we restrict our attention to $p = q = 1$,

$$G(1, 1; E) = \frac{1}{2\pi} \int_0^{\pi/a} dk \frac{2 - e^{2ika} - e^{-2ika}}{E + i\delta + 2U - 2U \cos(ka)}. \quad (\text{A26})$$

Again, to simplify the notation, we define

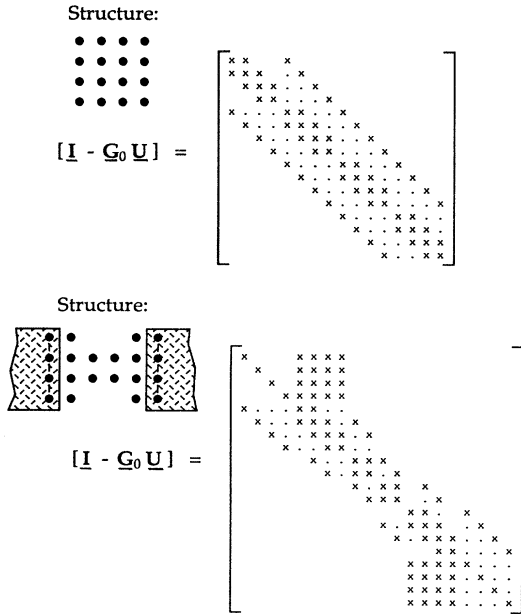


FIG. 33. Form of the matrix $(\underline{I} - \underline{G}_0 \underline{U})$ for two simple structures. An x represents nonzero elements in the original matrix, while a dot represents elements that will fill in during the L - U decomposition.

$$\chi \equiv (E + i\delta + 2U)/2U, \quad (\text{A27a})$$

$$\theta \equiv ka, \quad (\text{A27b})$$

so that the Green function (A26) can be written as

$$G(1, 1; E) = \frac{1}{4\pi Ua} \int_{-\pi}^{\pi} d\theta \frac{1 - e^{2i\theta}}{\chi - \cos\theta}. \quad (\text{A28})$$

This integral is easily performed using contour integration, with a substitution of the form $W = e^{i\theta}$. The result is (where the sign is chosen such that $|\chi^{\pm}(\chi^2 - 1)^{1/2}| < 1$)

$$G(1, 1; E) = \frac{1}{Ua} [\chi \pm (\chi^2 - 1)^{1/2}]. \quad (\text{A29})$$

Of course, this is only the result for a one-dimensional wire. We must now incorporate the effects of a finite width. Fortunately, the result is not much more complicated, since the eigenfunctions for a wire of uniform width are separable,

$$\phi_{\alpha\beta}(x_p, y_q) = \phi_{\alpha}^x(p) \phi_{\beta}^y(q), \quad (\text{A30})$$

$$\phi_{\alpha}^x(p) = \frac{1}{\sqrt{2L}} (e^{ik_{\alpha} p a} - e^{-ik_{\alpha} p a}), \quad (\text{A31a})$$

$$\phi_{\beta}^y(q) = C_{\beta} e^{ie A q a / \hbar} \sin \left[\frac{\beta \pi q}{M+1} \right], \quad (\text{A31b})$$

where M is the width of the wire in sites, and p and q are integer site indices. To be completely general, we have allowed for a constant vector potential $\mathbf{A} = (0, A, 0)$ within the wire. When we substitute these eigenfunctions in the expansion (A22), we are faced with two summations. One of these (involving ϕ_{α}^x) we have already per-

formed. The other (involving ϕ_{β}^y) we leave intact. Thus, the expression for the Green function \underline{G}_0 for sites along the edge of a semi-infinite wire of width M sites is

$$G_0(q, q') = \sum_{\beta} C_{\beta}^2 e^{ie A (q - q') / \hbar} \sin \left[\frac{\beta \pi q}{M+1} \right] \sin \left[\frac{\beta \pi q'}{M+1} \right] \times \frac{1}{Ua} [\chi_{\beta} \pm (\chi_{\beta}^2 - 1)^{1/2}], \quad (\text{A32})$$

$$\chi_{\beta} \equiv \frac{E + i\hbar/2\tau_{\phi} - E_{\beta} + 2U}{2U}, \quad (\text{A33a})$$

$$E_{\beta} \equiv 2U \left[\cos \left[\frac{\beta \pi}{M+1} \right] - 1 \right], \quad (\text{A33b})$$

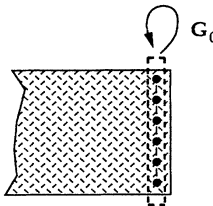
$$C_{\beta} = \left[\frac{M}{2} + \frac{1}{2} \text{Re} \left[\frac{1 - e^{i2\pi\beta M/(M+1)}}{1 - e^{-i2\pi\beta/(M+1)}} \right] \right]^{-1/2}. \quad (\text{A33c})$$

This result is summarized in Fig. 34.

d. Extending the Green function into open boundaries

To this point, we have described how the Green function $G^R(\mathbf{r}, \mathbf{r}'; E)$ is computed between any two sites \mathbf{r} and \mathbf{r}' within the sample. We emphasize that we can only inject and detect at sites actually simulated on the lattice. This is because each injection site corresponds to a particular column, and each detection site to a particular row, in the matrix \underline{G}^R . Thus, the sites within a semi-infinite lead attached to an open boundary are not actually simulated, although their presence is felt through the Green function \underline{G}_0 (Fig. 35). When solving the transport equation, it will be necessary to evaluate the Green function $G^R(\mathbf{r}, \mathbf{r}'; E)$ at sites within a semi-infinite lead. We will now describe how the solution for the Green function obtained along an open boundary can be extended into the attached lead.

Recall that the Green function can be expanded in terms of a complete set of eigenfunctions,



$$G_0(q, q') = \sum_{\beta} C_{\beta}^2 e^{ie A (q - q') / \hbar} \sin \left[\frac{\beta \pi q}{M+1} \right] \sin \left[\frac{\beta \pi q'}{M+1} \right] \frac{1}{Ua} [\chi_{\beta} \pm \sqrt{\chi_{\beta}^2 - 1}]$$

$$\chi_{\beta} = \frac{E + i\hbar/2\tau_{\phi} - E_{\beta} + 2U}{2U}$$

$$E_{\beta} = 2U \left[\cos \left[\frac{\beta \pi}{M+1} \right] - 1 \right]$$

$$C_{\beta} = \left[\frac{M}{2} + \frac{1}{2} \text{Re} \left[\frac{1 - e^{i2\pi\beta M/(M+1)}}{1 - e^{-i2\pi\beta/(M+1)}} \right] \right]^{-1/2}$$

$$U = -\frac{\hbar^2}{2ma^2}$$

FIG. 34. The Green function \underline{G}_0 for a semi-infinite wire of width M sites, in a constant vector potential $\mathbf{A} = (0, A, 0)$. Injection and detection points are along the edge of the wire.

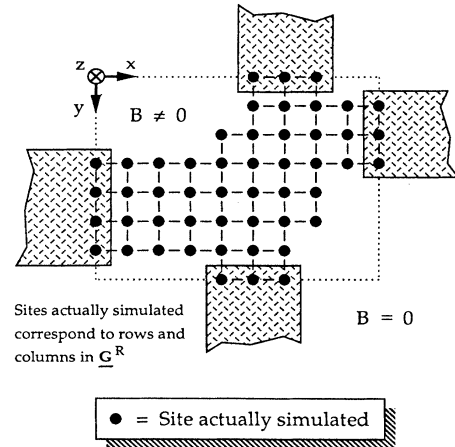


FIG. 35. Sites within the device are actually simulated, while sites within attached leads are felt through the Green function \underline{G}_0 .

$$G^R(\mathbf{r}, \mathbf{r}'; E) = \langle \mathbf{r} | G^R | \mathbf{r}' \rangle = \sum_{\alpha} \frac{\langle \mathbf{r} | \phi_{\alpha} \rangle \langle \phi_{\alpha} | \mathbf{r}' \rangle}{E - E_{\alpha} + i\delta}. \quad (\text{A34})$$

We have chosen the ket notation because of its simplicity; note that this is identical to the expression (A22) presented earlier. Suppose that we inject at some point \mathbf{r}' , and detect along an open boundary. We would like to know the amplitudes for each of the modes excited in the attached lead. If these modes are represented by a basis $|\psi_{\beta}\rangle$, then we can express this Green function as

$$\begin{aligned} G^R(\beta; \mathbf{r}'; E) &= \langle \psi_{\beta} | G^R | \mathbf{r}' \rangle \\ &= \sum_{\alpha} \frac{\langle \psi_{\beta} | \phi_{\alpha} \rangle \langle \phi_{\alpha} | \mathbf{r}' \rangle}{E - E_{\alpha} + i\delta}. \end{aligned} \quad (\text{A35})$$

Of course, we can only compute the Green function if we detect at sites (not modes) along the open boundary. But converting between the two representations is a simple matter. If we insert a complete set of states $|\mathbf{r}\rangle$, we can rewrite the Green function (A35) as

$$G^R(\beta; \mathbf{r}'; E) = \sum_{\mathbf{r}} \langle \psi_{\beta} | \mathbf{r} \rangle \left[\sum_{\alpha} \frac{\langle \mathbf{r} | \phi_{\alpha} \rangle \langle \phi_{\alpha} | \mathbf{r}' \rangle}{E - E_{\alpha} + i\delta} \right]. \quad (\text{A36})$$

This relationship is important, since the modes in a straight wire evolve in a particularly simple way. If we move a distance x into the lead, each mode merely acquires a phase factor, so that the Green function can be written as

$$G^R(x, \beta; \mathbf{r}'; E) = \sum_{\mathbf{r}} \langle \psi_{\beta} | \mathbf{r} \rangle e^{ik_{\beta}x} G^R(\mathbf{r}, \mathbf{r}'; E), \quad (\text{A37})$$

where

$$\langle \psi_{\beta} | \mathbf{r} \rangle = C_{\beta} e^{-ieAq/\hbar} \sin \left[\frac{\beta\pi q}{M+1} \right], \quad (\text{A38a})$$

$$C_{\beta} = \left[\frac{M}{2} + \frac{1}{2} \text{Re} \left[\frac{1 - e^{i2\pi\beta M/(M+1)}}{1 - e^{-i2\pi\beta/(M+1)}} \right] \right]^{-1/2}, \quad (\text{A38b})$$

for a lead of width M sites, with $q \leftrightarrow \mathbf{r}$ representing the position along the open boundary. The wave vector k_{β} satisfies

$$E = 2U(\cos k_{\beta}a - 1) + 2U \left[\cos \left[\frac{\beta\pi}{M+1} \right] - 1 \right]. \quad (\text{A38c})$$

Note that the detection points \mathbf{r} include only the sites along the open boundary. Thus, we need only evaluate $G^R(\mathbf{r}, \mathbf{r}'; E)$ at sites actually simulated on the lattice, and then use a simple expansion (A37) to extend this Green function into a semi-infinite lead.

Following similar arguments, we can extend the injection point into a different lead, so that the Green function for injection in one lead and detection in another becomes

$$\begin{aligned} G^R(x, \beta; x', \beta'; E) &= \sum_{\mathbf{r}} \sum_{\mathbf{r}'} \langle \psi_{\beta} | \mathbf{r} \rangle e^{ik_{\beta}x} \langle \mathbf{r}' | \psi_{\beta'} \rangle e^{ik_{\beta'}x'} \\ &\quad \times G^R(\mathbf{r}, \mathbf{r}'; E). \end{aligned} \quad (\text{A39})$$

Again, the coordinates \mathbf{r}' and \mathbf{r} run over the sites of the

open boundaries for the injecting and detecting leads; the distance x within any lead is zero at the open boundary and positive within the lead.

2. Solving the transport problem

a. Computing the electrochemical potential

In the preceding section, we described how the Green function $G^R(\mathbf{r}, \mathbf{r}'; E)$ can be computed within any arbitrary structure. We can use this Green function to compute the kernel $T_0(\mathbf{r}, \mathbf{r}')$ of the transport equation,

$$T_0(\mathbf{r}, \mathbf{r}') \equiv \frac{\hbar^2 |G^R(\mathbf{r}, \mathbf{r}'; \mu_0)|^2}{\tau_{\phi}(\mathbf{r}; \mu_0) \tau_{\phi}(\mathbf{r}'; \mu_0)}, \quad (\text{A40})$$

and then solve the transport equation for the electrochemical potential $\mu(\mathbf{r})$,

$$\mu(\mathbf{r}) = \frac{\int d^2\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}') \mu(\mathbf{r}')}{\int d^2\mathbf{r}' T_0(\mathbf{r}, \mathbf{r}')}. \quad (\text{A41})$$

These integrals are easy to compute numerically; we simply multiply the integrand at each site by the area a^2 surrounding that site.

In principle, we could express the transport equation (A41) as a matrix equation of the form $\underline{A}\mathbf{x} = \mathbf{b}$. Such an equation, however, would be extremely difficult to solve. In a structure with 2500 sites, the matrix \underline{A} would contain more than six million elements. Solving a matrix of this size with conventional Gaussian elimination would be unwieldy, to say the least. Instead, we use an iterative solution technique known as Gauss-Seidel iteration. First, we generate an initial guess of the solution for $\mu(\mathbf{r})$. The values for $\mu(\mathbf{r})$ are fixed to a constant value within “contact” regions, and determined from some smooth interpolation everywhere in between. We then solve the transport equation (A41) repeatedly, to improve the solution. Only a single row of the kernel $T_0(\mathbf{r}, \mathbf{r}')$ is needed at each step, so there is no need to store the entire kernel in memory.

Using the method outlined in Sec. 1, we can obtain any column of the Green function \underline{G}^R simply by supplying a single column of the matrix \underline{G}_0 , and then performing two back-substitution steps. This column describes the response at all sites due to injection at a single site; to solve the transport equation (A41), however, we need the opposite—the response at a single site due to injection at all sites. Fortunately, the transport equation can be expressed in another form which is more compatible with the Green-function solution,

$$\mu(\mathbf{r}) = \frac{\int d^2\mathbf{r}' T_0(\mathbf{r}', \mathbf{r})|_{-B} \mu(\mathbf{r}')}{\int d^2\mathbf{r}' T_0(\mathbf{r}', \mathbf{r})|_{-B}}. \quad (\text{A42})$$

In obtaining this equation, we have used a special property of the Green function in a magnetic field,

$$G^R(\mathbf{r}, \mathbf{r}'; E)|_B = G^R(\mathbf{r}', \mathbf{r}; E)|_{-B}, \quad (\text{A43a})$$

$$T_0(\mathbf{r}, \mathbf{r}')|_B = T_0(\mathbf{r}', \mathbf{r})|_{-B}. \quad (\text{A43b})$$

The notation $B \rightarrow -B$ reminds us that this transport equation (A42) must be solved using the opposite sense of the magnetic field. Thus, to compute $\mu(\mathbf{r})$ at any point, we simply inject at that point, compute a single column of the Green function \underline{G}^R , compute a single column of the kernel \underline{T}_0 , and then integrate over all sites.

Of course, the presence of open boundaries causes a slight complication. The integrals required for the transport equation (A42) are evaluated in a slightly different manner within each attached lead. Recall that to evaluate the Green function $G^R(\mathbf{r}, \mathbf{r}'; E)$ within the lead we must perform an expansion in terms of the eigenstates of the lead, as shown in Sec. 1*d*. We can integrate the kernel within a lead by integrating this expansion,

$$\int_{\text{lead}} d^2\mathbf{r}' T_0(\mathbf{r}', \mathbf{r}) \mu_{\text{lead}} = \mu_{\text{lead}} \frac{\hbar^2}{\tau_\phi(\mathbf{r}; \mu_0) \tau_\phi(\text{lead}; \mu_0)} \sum_{\beta} |G^R(\beta; \mathbf{r}; \mu_0)|^2 \times \frac{1}{2 \text{Im}\{k_\beta\}}, \quad (\text{A44})$$

where $G^R(\beta; \mathbf{r}; E)$ is the response at the open boundary in mode β due to injection at position \mathbf{r} ; this coefficient was derived in Sec. 1. We have assumed that the chemical potential within the lead is a constant μ_{lead} . This is certainly true if the lead is part of a contact region, where $\mu(\mathbf{r})$ is fixed as a boundary condition. This is also true if the lead is attached as a voltage probe, as long as the $\mu(\mathbf{r})$ has settled out to a constant value near the open boundary.

b. Computing the current density

Once we have obtained the solution for the electrochemical potential $\mu(\mathbf{r})$, we can compute any quantity of interest. For instance, we can compute the current density arising under bias as

$$\begin{aligned} \delta\mathbf{J}(\mathbf{r}) = & \frac{e\hbar^2}{2\pi m} \int d^2\mathbf{r}' \frac{\mu(\mathbf{r}') - \mu_0}{\tau_\phi(\mathbf{r}'; \mu_0)} \\ & \times \text{Im}[G^{R*}(\mathbf{r}, \mathbf{r}'; \mu_0) \nabla_r G^R(\mathbf{r}, \mathbf{r}'; \mu_0)] \\ & - \frac{e^2\hbar}{2\pi m} \mathbf{A}(\mathbf{r}) \int d^2\mathbf{r}' \frac{\mu(\mathbf{r}') - \mu_0}{\tau_\phi(\mathbf{r}'; \mu_0)} |G^R(\mathbf{r}, \mathbf{r}'; \mu_0)|^2. \end{aligned} \quad (\text{A45})$$

Note that the second term above depends directly on the vector potential $\mathbf{A}(\mathbf{r})$ and is therefore sensitive to our particular choice of gauge. Of course, observables such as the current density must be independent of the gauge. To account for this, there is a delicate cancellation between these two terms that maintains the property of gauge invariance. Although this may be interesting from a physics standpoint, it is a problem to be avoided in numerical solutions.

Suppose we define an ansatz

$$G^R(\mathbf{r}, \mathbf{r}'; E) \equiv e^{ie\mathbf{A}(\mathbf{r}) \cdot (\mathbf{r} - \mathbf{r}')/\hbar} \tilde{G}(\mathbf{r}, \mathbf{r}'; E) \quad (\text{A46})$$

that has the following properties:

$$|G^R(\mathbf{r}, \mathbf{r}'; E)|^2 = |\tilde{G}(\mathbf{r}, \mathbf{r}'; E)|^2, \quad (\text{A47a})$$

$$\begin{aligned} \nabla_r G^R(\mathbf{r}, \mathbf{r}'; E) \simeq & \frac{ie}{\hbar} \mathbf{A}(\mathbf{r}) G^R(\mathbf{r}, \mathbf{r}'; E) \\ & + e^{ie\mathbf{A}(\mathbf{r}) \cdot (\mathbf{r} - \mathbf{r}')/\hbar} \nabla_r \tilde{G}(\mathbf{r}, \mathbf{r}'; E). \end{aligned} \quad (\text{A47b})$$

Note that the second property (A47b) is an approximation, valid if $\mathbf{A}(\mathbf{r})$ is *slowly varying*. If we substitute this ansatz into our expression for the current density $\delta\mathbf{J}$, we are left with only a single term,

$$\begin{aligned} \delta\mathbf{J}(\mathbf{r}) = & \frac{e\hbar^2}{2\pi m} \int d^2\mathbf{r}' \frac{\mu(\mathbf{r}') - \mu_0}{\tau_\phi(\mathbf{r}'; \mu_0)} \\ & \times \text{Im}[G^{R*}(\mathbf{r}, \mathbf{r}'; \mu_0) e^{ie\mathbf{A}(\mathbf{r}) \cdot (\mathbf{r} - \mathbf{r}')/\hbar} \\ & \times \nabla_r \tilde{G}(\mathbf{r}, \mathbf{r}'; \mu_0)]. \end{aligned} \quad (\text{A48})$$

We can then eliminate \tilde{G} using our original definition (A46),

$$\begin{aligned} \delta\mathbf{J}(\mathbf{r}) = & \frac{e\hbar^2}{2\pi m} \int d^2\mathbf{r}' \frac{\mu(\mathbf{r}') - \mu_0}{\tau_\phi(\mathbf{r}'; \mu_0)} \\ & \times \text{Im}\{G^{R*}(\mathbf{r}, \mathbf{r}'; \mu_0) e^{ie\mathbf{A}(\mathbf{r}) \cdot (\mathbf{r} - \mathbf{r}')/\hbar} \\ & \times \nabla_r [e^{-ie\mathbf{A}(\mathbf{r}) \cdot (\mathbf{r} - \mathbf{r}')/\hbar} \\ & \times G^R(\mathbf{r}, \mathbf{r}'; \mu_0)]\}. \end{aligned} \quad (\text{A49})$$

Thus, we have reduced the current density to a single, slightly more complicated integral which is more stable numerically. It is easy to see why this substitution worked. In the presence of a constant vector potential, the Green function $G^R(\mathbf{r}, \mathbf{r}'; E)$ simply acquires a phase factor,

$$G^R(\mathbf{r}, \mathbf{r}'; E) \rightarrow e^{ie\mathbf{A} \cdot (\mathbf{r} - \mathbf{r}')/\hbar} G^R(\mathbf{r}, \mathbf{r}'; E), \quad (\text{A50})$$

and the exponential in our ansatz (A46) explicitly accounts for this change. If the vector potential is shown varying, however, the exponential factor accounts for most of the change, but some of the effects will appear in \tilde{G} as well.

When actually evaluating the current density, it is more convenient to inject at a single point and integrate over all detection points. In the preceding section, we modified the transport equation (A42) to be solved in this manner. Following similar arguments, we interchange the coordinates \mathbf{r} and \mathbf{r}' in our expression for the current density as well,

$$\delta\mathbf{J}(\mathbf{r}) = \frac{e\hbar^2}{2\pi m} \int d^2\mathbf{r}' \frac{\mu(\mathbf{r}') - \mu_0}{\tau_\phi(\mathbf{r}'; \mu_0)} \text{Im}\{G^{R*}(\mathbf{r}', \mathbf{r}; \mu_0)|_{-B} e^{-ieA_y(\mathbf{r})(y' - y)/\hbar} \nabla_r [e^{ieA_y(\mathbf{r})(y' - y)/\hbar} G^R(\mathbf{r}', \mathbf{r}; \mu_0)|_{-B}]\}. \quad (\text{A51})$$

In this final expression, we have specialized to our usual gauge, $\mathbf{A}=(0, A_y, 0)$.

In evaluating the current density (A51), we inject at adjacent sites and compute each component of the gradient using a finite difference. This leaves us with a function of \mathbf{r}' , which is integrated in the usual manner—the integrand at each site is multiplied by the site area a^2 , and these contributions are summed. Within a lead attached at an open boundary, however, the integrals are handled in a slightly different manner. Each of the functions G^{R*} and $e^{i\phi}\nabla e^{-i\phi}G^R$ is expanded in terms of the normal modes for the lead. The product of the modal components is then integrated, in the manner described in the preceding section.

c. Computing the terminal currents

Having computed the current density $\delta\mathbf{J}$, we could integrate the normal component over the surface of each contact to obtain the terminal currents. We could also integrate the divergence of current $I(\mathbf{r})=-\nabla\cdot\delta\mathbf{J}$ in the contact regions. In most cases, these two methods yield the same result. The current density $\delta\mathbf{J}$, however, depends upon a finite-difference approximation for the gradient (A51). If the lattice spacing a is large, this approximation is prone to error. For this reason, we adopt the latter approach to compute the terminal currents.

We can compute the divergence of current $I(\mathbf{r})$ in the contact regions from the transport equation,

$$I(\mathbf{r})=\frac{2e^2}{h}\int d^2\mathbf{r}'T_0(\mathbf{r},\mathbf{r}')[\mu(\mathbf{r})-\mu(\mathbf{r}')], \quad (\text{A52})$$

and then integrate this quantity over the contact regions to compute the terminal currents,

$$I_n=\int_{\text{contact } n} d^2\mathbf{r}I(\mathbf{r}). \quad (\text{A53})$$

Again, the presence of open boundaries complicates the analysis. Suppose a particular contact region includes an open boundary with an attached lead. We can break the integral of $I(\mathbf{r})$ into two parts: one for all contact sites on the lattice, and one for all contact sites in the attached lead,

$$I_n=\int_{\text{lattice}} d^2\mathbf{r}I(\mathbf{r})+\int_{\text{lead}} d^2\mathbf{r}I(\mathbf{r}). \quad (\text{A54})$$

We deal with the former contribution first. At each contact site, we must evaluate the current $I(\mathbf{r})$ using the transport equation (A52). In keeping with the preceding

sections, we note that the integral over \mathbf{r}' can be computed more easily if we interchange the coordinates \mathbf{r} and \mathbf{r}' . This way, we inject at a single point \mathbf{r} and integrate over all detection points \mathbf{r}' . Again, the interchange of coordinates is allowed as long as we reverse the magnetic field when evaluating the kernel,

$$I(\mathbf{r})=\frac{2e^2}{h}\int d^2\mathbf{r}'T_0(\mathbf{r}',\mathbf{r})[\mu(\mathbf{r})-\mu(\mathbf{r}')], \quad B\rightarrow-B. \quad (\text{A55})$$

At each site \mathbf{r}' , the value of the integrand is simply multiplied by the site area a^2 and the contributions are summed. Of course, if the device has other open boundaries, the integrals over \mathbf{r}' within the attached leads must be handled separately. Within each lead, the kernel is resolved into the normal modes and integrated analytically out to infinity,

$$\begin{aligned} &\int_{\text{lead}} d^2\mathbf{r}'T_0(\mathbf{r}',\mathbf{r})[\mu(\mathbf{r})-\mu_{\text{lead}}] \\ &= \hbar^2 \frac{[\mu(\mathbf{r})-\mu_{\text{lead}}]}{\tau_\phi(\mathbf{r};\mu_0)\tau_\phi(\text{lead};\mu_0)} \sum_\beta |G^R(\beta;\mathbf{r};\mu_0)|^2 \\ &\quad \times \frac{1}{2\text{Im}\{k_\beta\}}. \end{aligned} \quad (\text{A56})$$

Here, $G^R(\beta;\mathbf{r};\mu_0)$ is the response at the open boundary in mode β due to injection at position \mathbf{r} ; this coefficient was derived in Sec. 1 d.

We now handle the other contribution to the current—the integral over contact sites in an attached lead. For this contribution, we cannot afford to compute $I(\mathbf{r})$ out to infinity and integrate numerically. Instead, we use the transport equation in its original form (A52): We inject at each lattice site \mathbf{r}' and integrate over \mathbf{r} within the lead. As before, the kernel is resolved into the normal modes and integrated analytically out to infinity,

$$\begin{aligned} &\int_{\text{lead}} d^2\mathbf{r}T_0(\mathbf{r},\mathbf{r}')[\mu_{\text{lead}}-\mu(\mathbf{r}')] \\ &= \hbar^2 \frac{[\mu_{\text{lead}}-\mu(\mathbf{r}')] }{\tau_\phi(\text{lead};\mu_0)\tau_\phi(\mathbf{r}';\mu_0)} \sum_\beta |G^R(\beta;\mathbf{r}';\mu_0)|^2 \\ &\quad \times \frac{1}{2\text{Im}\{k_\beta\}}. \end{aligned} \quad (\text{A57})$$

We then sum the contributions due to all injection points \mathbf{r}' . Only one eventually remains: For injection points \mathbf{r}' within a different attached lead, we need to extend the kernel into both the injecting and detecting leads, and integrate analytically out to infinity in both directions,

$$\int_{\text{lead } 1} d^2\mathbf{r} \int_{\text{lead } 2} d^2\mathbf{r}'T_0(\mathbf{r},\mathbf{r}')(\mu_1-\mu_2)=\hbar^2 \frac{(\mu_1-\mu_2)}{\tau_\phi(1;\mu_0)\tau_\phi(2;\mu_0)} \sum_\beta \sum_{\beta'} |G^R(\beta;\beta';\mu_0)|^2 \frac{1}{2\text{Im}\{k_\beta\}} \frac{1}{2\text{Im}\{k_{\beta'}\}}. \quad (\text{A58})$$

To say that open boundaries complicate the analysis is an understatement. Although the rules above may seem a bit involved, we have simply described all special cases which arise whenever the integration sites \mathbf{r} and \mathbf{r}' fall in an attached lead. Furthermore, we have tried to optimize the integrations by interchanging injection and detection

coordinates whenever appropriate. Note that the terminal current, unlike the electrochemical potential $\mu(\mathbf{r})$ and the current density $\delta\mathbf{J}$, requires *both* polarities of magnetic field for the evaluation of the kernel. This is unfortunate, since it involves setting up and L - U decomposing the Dyson equation twice, but it is necessary.

3. Solving the Poisson equation

By solving the transport equation, we can study how electrons will diffuse from one contact to another. Near resistive obstacles, electrons will pile up on one side and drain off on the other, causing rapid variations in the local electron density. If electrons were neutral particles, this diffusion analysis would be sufficient. Since they are not, however, any charge imbalances will tend to be screened out. Such electrostatic effects can be computed by solving the Poisson equation,

$$\nabla^2\phi(\mathbf{r}) = \frac{e}{\epsilon} [N_D^+(\mathbf{r}) - n(\mathbf{r})] . \quad (\text{A59})$$

In general, however, this is a difficult problem.

We can simplify it somewhat if we assume that the equilibrium solution is known, and if we consider only the changes arising under bias,

$$\nabla^2\delta\phi(\mathbf{r}) = -\frac{e^2 N_0(\mathbf{r}; \mu_0)}{\epsilon} [\delta\mu(\mathbf{r}) - \delta\phi(\mathbf{r})] , \quad (\text{A60})$$

where $\delta\mu(\mathbf{r}) = \mu(\mathbf{r}) - \mu_0$ is the change in electrochemical potential. The change in the electrostatic potential $\delta\phi(\mathbf{r})$ can be separated into two components:

$$\delta\phi \equiv \delta\phi_e + \delta\phi_i . \quad (\text{A61})$$

The first component $\delta\phi_e$ satisfies the homogeneous equation

$$\nabla^2\delta\phi_e(\mathbf{r}) = 0 \quad (\text{A62})$$

and arises from the charges supplied at the contacts by an external source. If we assume that our device sits between two semi-infinite conducting sheets, one at a potential V_L and the other at V_R , then the solution for $\delta\phi_e$ is well known. The electrostatic potential evolves between the two contacts according to an inverse cosine relationship,⁶⁵

$$\delta\phi_e(x, y) = \frac{V_L - V_R}{\pi} \cos^{-1} \left[\frac{x}{d} \right] + V_R , \quad (\text{A63})$$

for contacts located at $x = \pm d$.

The second component $\delta\phi_i$ arises from charges induced by screening within the device. Because it depends on the distribution of $\mu(\mathbf{r})$ as well as the available density of states N_0 , this component must be computed numerically by solving the Poisson equation,

$$\nabla^2\delta\phi_i(\mathbf{r}) = -\frac{e^2 N_0(\mathbf{r}; \mu_0)}{\epsilon} [\delta\mu(\mathbf{r}) - \delta\phi_e(\mathbf{r}) - \delta\phi_i(\mathbf{r})] . \quad (\text{A64})$$

Although electrons are restricted to a two-dimensional plane within the sample, we need to solve the electrostatics problem in three dimensions. We do this by computing the response due to a single point charge within the

sample, and then integrating over the entire collection of charges. It is well known that a single charge gives rise to a potential $q/4\pi\epsilon r$ at a distance r from its location. Thus, a collection of charges ρ_i gives rise to the usual electrostatic potential,

$$\delta\phi_i(\mathbf{r}) = \int d^2\mathbf{r}' \frac{\rho_i(\mathbf{r}')}{4\pi\epsilon|\mathbf{r}-\mathbf{r}'|} , \quad (\text{A65})$$

where

$$\rho_i(\mathbf{r}) \equiv e^2 N_0(\mathbf{r}; \mu_0) [\delta\mu(\mathbf{r}) - \delta\phi_e(\mathbf{r}) - \delta\phi_i(\mathbf{r})] . \quad (\text{A66})$$

Because ρ_i depends on ϕ_i , we must perform this solution in an iterative manner. We start with an initial guess $\delta\phi_i = 0$ and compute the induced charge ρ_i . We then substitute this charge into the integral (A65) to obtain a better guess for the potential $\delta\phi_i$. Iteration continues until $\delta\phi_i$ converges to a stable solution.

Achieving convergence in this iterative approach may be somewhat difficult. Actual solutions tend to oscillate if $\delta\phi_i$ reacts too violently to the charge ρ_i at a single iteration step. A relaxation factor is helpful in slowing the response of $\delta\phi_i$,

$$\delta\phi_i^{\text{new}} \leftarrow \omega \delta\phi_i^{\text{old}} + (1 - \omega) (\delta\phi_i^{\text{new}} - \delta\phi_i^{\text{old}}) , \quad (\text{A67})$$

where ω is some fraction close to unity, and N is the iteration number. Using this technique, $\delta\phi_i$ is forced to change slowly at first, so that any initial charge imbalances are gradually screened out.

To evaluate the integral (A65) numerically, we simply sum the contributions from the charge within an area a^2 around each site,

$$\delta\phi_i(\mathbf{r}_n) = \sum_{m (\neq n)} \frac{\rho_i(\mathbf{r}_m)}{4\pi\epsilon|\mathbf{r}_n - \mathbf{r}_m|} + \int_{\text{site } n} d^2\mathbf{r}' \frac{\rho_i(\mathbf{r}')}{4\pi\epsilon|\mathbf{r}_n - \mathbf{r}'|} , \quad (\text{A68})$$

Of course, we must handle the small area around r_n separately, since the distance to this point would be zero. We can perform this integral analytically, as follows:

$$\begin{aligned} \Gamma &\equiv \frac{1}{a} \int_{-a/2}^{a/2} dx \int_{-a/2}^{a/2} dy \frac{1}{(x^2 + y^2)^{1/2}} \\ &= \frac{4}{a} \int_{-\pi/a}^{\pi/a} d\theta \int_0^{a/2 \cos\theta} dr r \frac{1}{r} \\ &= \left(\frac{1}{2} \ln |\sec\theta + \tan\theta| \right) \Big|_{-\pi/a}^{\pi/a} \\ &= 3.525 4943 48 . \end{aligned} \quad (\text{A69})$$

In terms of this constant, our final expression for $\delta\phi_i$ becomes

$$\delta\phi_i(\mathbf{r}_n) = \sum_{m (\neq n)} \frac{\rho_i(\mathbf{r}_m)}{4\pi\epsilon|\mathbf{r}_n - \mathbf{r}_m|} + \frac{\rho_i(\mathbf{r}_n)}{4\pi\epsilon} \Gamma a . \quad (\text{A70})$$

¹J. Frenkel, Phys. Rev. **36**, 1604 (1930).

²R. Tsu and L. Esaki, Appl. Phys. Lett. **22**, 562 (1973).

³C. B. Duke, *Tunneling in Solids* (Academic, New York, 1969).

⁴R. Landauer, IBM J. Res. Dev. **1**, 223 (1957).

⁵R. Landauer, Philos. Mag. **21**, 863 (1970).

⁶Y. Imry, in *Directions in Condensed Matter Physics*, edited by G. Grinstein and G. Mazenko (World Scientific, Singapore, 1986), p. 101.

- ⁷R. Landauer, IBM J. Res. Dev. **32**, 306 (1988); Z. Phys. B **68**, 217 (1987).
- ⁸M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, Phys. Rev. B **31**, 6207 (1985).
- ⁹J. K. Jain and S. A. Kivelson, Phys. Rev. B **37**, 4276 (1988).
- ¹⁰M. Yosefin and M. Kaveh, Phys. Rev. Lett. **64**, 2819 (1990).
- ¹¹H.-L. Engquist and P. W. Anderson, Phys. Rev. B **24**, 1151 (1981).
- ¹²M. Büttiker, Phys. Rev. Lett. **57**, 1761 (1986).
- ¹³C. W. J. Beenakker and H. van Houten, in *Solid State Physics*, edited by H. Ehrenreich and D. Turnbull (Academic, New York, 1991), Vol. 44.
- ¹⁴S. Datta and M. J. McLennan, Rep. Prog. Phys. (to be published).
- ¹⁵*Semiconductors and Semimetals*, edited by M. A. Reed (Academic, New York, 1990).
- ¹⁶M. Büttiker, Phys. Rev. B **33**, 3020 (1986); IBM J. Res. Dev. **32**, 63 (1988).
- ¹⁷L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Benjamin, New York, 1962); L. V. Keldysh, Zh. Eksp. Teor. Fiz. **47**, 1515, (1964) [Sov. Phys.—JETP **20**, 1018 (1965)].
- ¹⁸J. Rammer and H. Smith, Rev. Mod. Phys. **58**, 323 (1986).
- ¹⁹G. D. Mahan, Phys. Rep. **145**, 251 (1987).
- ²⁰C. Caroli, R. Combescot, P. Nozières, and D. Saint James, J. Phys. C **5**, 21 (1972); A. I. Larkin and D. E. Khmel'nitskii, Zh. Eksp. Teor. Fiz. **91**, 1815 (1986) [Sov. Phys.—JETP **64**, 1075 (1986)]; G. B. Lesovik, *ibid.* **94**, 380 (1988) [67, 1950 (1988)].
- ²¹A. P. Jauho and O. Ziep, Phys. Scr. **T25**, 329 (1989); A. P. Jauho, Solid State Electron. **32**, 1265 (1989), and references therein; G. J. Iafrate and J. B. Krieger, Phys. Rev. B **40**, 6144 (1989).
- ²²W. R. Frensley, Phys. Rev. B **36**, 1570 (1987).
- ²³N. C. Kluksdahl, A. M. Krivan, D. K. Ferry, and C. Ringhofer, Phys. Rev. B **39**, 7720 (1989).
- ²⁴F. A. Buot and K. L. Jensen, Phys. Rev. B **42**, 9429 (1990).
- ²⁵(a) S. Datta, Phys. Rev. B **40**, 5830 (1989); (b) J. Phys. Condens. Matter **2**, 8023 (1990).
- ²⁶C. S. Lent, Appl. Phys. Lett. **57**, 1678 (1990); Phys. Rev. B **43**, 4179 (1991).
- ²⁷M. C. Payne, J. Phys. Condens. Matter **1**, 4931 (1989); M. C. Payne and G. E. Engel, *ibid.* **2**, 1355 (1990).
- ²⁸R. Landauer, in *Localization, Interaction, and Transport Phenomena in Impure Metals*, edited by G. Bergmann, Y. Bruynseraede, and B. Kramer (Springer, Heidelberg, 1985).
- ²⁹M. L. Roukes, A. Scherer, S. J. Allen, Jr., H. G. Craighead, R. M. Ruthen, E. D. Beebe, and J. P. Harbison, Phys. Rev. Lett. **59**, 3011 (1987).
- ³⁰C. J. B. Ford, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, D. C. Peacock, D. A. Ritchie, J. E. F. Frost, and G. A. C. Jones, Phys. Rev. B **38**, 8518 (1988); C. J. B. Ford, S. Washburn, M. Büttiker, C. M. Knoedler, and J. M. Hong, Phys. Rev. Lett. **62**, 2724 (1989).
- ³¹A. M. Chang, T. Y. Chang, and H. U. Baranger, Phys. Rev. Lett. **63**, 996 (1989).
- ³²F. M. Peeters, Phys. Rev. Lett. **61**, 589 (1988).
- ³³G. Kirczenow, Phys. Rev. B **38**, 10958 (1988).
- ³⁴Q. Li and D. J. Thouless, Phys. Rev. Lett. **65**, 767 (1990).
- ³⁵Y. Imry, in *Nanostructure Physics and Fabrication*, edited by M. A. Reed and W. P. Kirk (Academic, San Diego, 1989), p. 379.
- ³⁶M. Büttiker, IBM J. Res. Dev. **32**, 317 (1988).
- ³⁷F. S. Khan, J. H. Davies, and J. W. Wilkins, Phys. Rev. B **36**, 2578 (1987).
- ³⁸J. L. D'Amato and H. M. Pastawski, Phys. Rev. B **41**, 7411 (1990).
- ³⁹M. Büttiker, Phys. Rev. B **40**, 3409 (1989).
- ⁴⁰P. L. Pernas, A. Martin-Rodero, and F. Flores (unpublished).
- ⁴¹S. Doniach and E. H. Sondheimer, *Green's Functions for Solid State Physicists* (Benjamin-Cummings, Reading, MA, 1974).
- ⁴²K. Takahashi, J. Phys. Soc. Jpn. **55**, 762 (1986).
- ⁴³P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985), and references cited therein.
- ⁴⁴S. Washburn, IBM J. Res. Dev. **32**, 335 (1988), and references cited therein.
- ⁴⁵P. A. Lee, A. D. Stone, and H. Fukuyama, Phys. Rev. B **35**, 1039 (1987).
- ⁴⁶A. Benoit, C. P. Umbach, R. B. Laibowitz, and R. A. Webb, Phys. Rev. Lett. **58**, 2343 (1987); W. J. Skocpol, P. M. Mankiewich, R. E. Howard, L. D. Jackel, and D. M. Tennant, *ibid.* **58**, 2347 (1987).
- ⁴⁷M. Büttiker, Phys. Rev. B **35**, 4123 (1987).
- ⁴⁸H. U. Baranger and A. D. Stone, Phys. Rev. Lett. **63**, 414 (1989); M. Büttiker, Phys. Rev. B **38**, 12724 (1988).
- ⁴⁹C. W. J. Beenakker and H. van Houten, Phys. Rev. Lett. **63**, 1857 (1989).
- ⁵⁰G. Kirczenow, Solid State Commun. **71**, 469 (1989).
- ⁵¹C. S. Chu and R. S. Sorbello, Phys. Rev. B **40**, 5950 (1989).
- ⁵²B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon, Phys. Rev. Lett. **60**, 848 (1988).
- ⁵³D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. C **21**, L209 (1988).
- ⁵⁴A. Safer and A. D. Stone, Phys. Rev. Lett. **62**, 300 (1989).
- ⁵⁵L. I. Glazman, G. B. Lesovik, D. E. Khmel'nitskii, and R. I. Shekhter, Pis'ma Zh. Eksp. Teor. Fiz. **48**, 218 (1988) [JETP Lett. **48**, 238 (1988)].
- ⁵⁶Yu. V. Sharvin, Zh. Eksp. Teor. Fiz. **48**, 984 (1965) [Sov. Phys.—JETP **21**, 655 (1965)].
- ⁵⁷R. Landauer, J. Phys. Condens. Matter **1**, 8099 (1989).
- ⁵⁸J. Friedel, Philos. Mag. Suppl. **3**, 446 (1954).
- ⁵⁹I. B. Levinson, Zh. Eksp. Teor. Fiz. **95**, 2175 (1989) [Sov. Phys.—JETP **68**, 1257 (1989)].
- ⁶⁰R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading, MA, 1966), Vol. III, pp. 21 and 22.
- ⁶¹D. J. Thouless and S. Kirkpatrick, J. Phys. C **14**, 235 (1981).
- ⁶²F. Sols, J. Appl. Phys. **66**, 3892 (1989).
- ⁶³D. S. Fisher and P. A. Lee, Phys. Rev. B **23**, 6851 (1981).
- ⁶⁴G. Dahlquist and Å. Björck, *Numerical Methods* (Prentice-Hall, Englewood Cliffs, NJ, 1974), p. 157.
- ⁶⁵S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics* (Wiley, New York, 1984), p. 336.
- ⁶⁶P. Murali and D. W. Pohl, Appl. Phys. Lett. **48**, 514 (1986).
- ⁶⁷J. R. Kirtley, S. Washburn, and M. J. Brady, Phys. Rev. Lett. **60**, 1546 (1988).
- ⁶⁸C. W. J. Beenakker, H. van Houten, and B. J. van Wees, in *Advances in Solid State Physics*, edited by U. Rössler (Pergamon/Vieweg, Braunschweig, 1989), Vol. 29.