Current and rate equation for resonant tunneling

John H. Davies,* Selman Hershfield,† Per Hyldgaard, and John W. Wilkins
Department of Physics, The Ohio State University, Columbus, Ohio 43210-1168
(Received 17 June 1992)

We have used nonequilibrium quantum statistical mechanics to derive an exact expression for the current and a rate equation for a simple model of a resonant-tunneling diode including scattering within the resonant state. The rate equation is identical to the classical rate equation for resonant tunneling provided that two conditions are met. First, the driving frequency must be slow compared with the Fermi energy of incoming electrons; the lifetime of the resonance does *not* set a scale. Second, the resonance must be narrow compared with the range of incoming energies and on a scale set by the variation in energy of the tunneling rates of the individual barriers. Our derivation shows that the rate equation holds both in the coherent limit, in which transport can be described by elementary wave mechanics alone, and in the limit where scattering within the resonant state makes a classical "sequential" picture more appropriate; all information about coherence is lost in the averaging required to deduce the rate equation. The current through the resonant state is independent of scattering provided that the rate equation holds. We also examine the current when the rate equation does not hold, and show how strongly inelastic processes enter the quantum kinetic description through scattering-out and scattering-in processes at different energies.

I. INTRODUCTION

Resonant tunneling through a double-barrier structure was one of the first proposed applications of semiconductor heterostructures, and has been extensively investigated since its observation.[1] A resonant state is trapped in the well between the two barriers, which isolate it from the leads on the left and right. The band profile through a typical resonant-tunneling diode is shown in Fig. 1. This is the simplest structure; more complicated ones are also used, such as a double barrier within the base of a heterojunction bipolar transistor, but the fundamental description of resonant tunneling remains unchanged. Most current passes through a narrow range of energies around the resonant state. The energy ε_0 of the resonant state lies above the Fermi energy when no bias is applied. It is pulled down into the range of energies of incoming electrons on the left as the bias in increased, and current begins to flow. The current increases with the bias until the resonant state passes through the bottom of the band in the left lead ϕ_L, at which point the current falls rapidly to a low value. This gives a region of negative differential resistance, the important feature of the device.

The modeling of resonant-tunneling structures has posed a challenge to theorists, because this is one of the simplest devices that relies on quantum mechanics and whose operation cannot be described in terms purely of classical particles. Although it is not too difficult to calculate the characteristics of a perfect device without additional scattering, at least if self-consistency is neglected, the treatment of inelastic scattering remains ill-understood except within highly simplified models.

Two elementary methods have been used. The first is the "coherent" picture, based on elementary wave mechanics. Schrödinger's equation is solved for a potential like that in Fig. 1, numerically if necessary, which gives the transmission coefficient as a function of energy. Integrating this over the density of incoming states gives the current.[2] Luryi[3] questioned the basis of this method, and argued that an electron spends so long in the resonant state that it will almost certainly be scattered. He proposed that tunneling should instead be regarded as a "sequential" process, with an electron tunneling into the well and later tunneling out with no phase coherence between these two processes. This picture lends itself naturally to classical rate equations, and the rates can be calculated using a transfer Hamiltonian.[4–6] It was suggested initially that the two pictures led to different

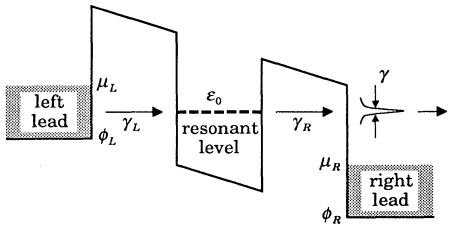


FIG. 1. Schematic band profile through a double-barrier resonant-tunneling diode under typical operating conditions. The bottoms of the bands on the left and right are ϕ_L and ϕ_R. The chemical potentials on the left and right, μ_L and μ_R, differ by the (large) applied bias. The tunneling rates through the individual barriers are γ_L and γ_R. The full width γ of the resonant state, centered on ε_0, is flooded with electrons from the left.

results but it was soon shown[4,5] that the current in a steady state is the same in both pictures.

Several authors[7–9] have modified the coherent picture to include scattering within the resonance in a way analogous to a lossy optical cavity. They found that the peak in the transmission function was broadened and reduced in height, but its area and therefore the total current were unaffected. In another approach, the resonant state is coupled to a reservoir to provide dephasing.[10,11] More sophisticated methods have also been used to model resonant-tunneling diodes, and treat scattering in better detail. A one-dimensional model in which a single electron in the resonant state interacts with optic phonons can be solved exactly,[12–16] and shows that a second peak due to phonon-assisted tunneling appears in the $I(V)$ characteristic. Further calculations using tight-binding models are described in Refs. 17–19. The current has been calculated for more realistic models of practical devices using simpler approximations[20–23] and is in general agreement with experiment.[24,25] Attempts have also been made to use the Wigner function to describe transport,[26,27] but this turns out to be remarkably difficult to apply in practice. Nonequilibrium Green's functions have proved more successful,[28–33] and we shall use this formalism.

Our aim in this paper is to explore a simplified but more general model of a resonant-tunneling diode using quantum kinetic theory, outlined in Sec. II. We include scattering only within the resonant state itself—a single site for a one-dimensional model, or a plane of sites for a three-dimensional model. An exact expression is derived for the current, which holds in the presence of arbitrary scattering or interactions within the resonant state. In Sec. III we examine the approximations needed to derive the "classical" rate equation from the quantum kinetic equation, and assess its range of validity. The rate equation holds under a wide range of conditions, and our derivation contains only two approximations. The first is to assume that the system is "slowly varying" in time. The frequency scale is typically set by the Fermi energy, so this is not a stringent condition. In particular, the lifetime of the resonance does not enter. The second approximation, more serious, is to assume that the resonance is narrow, both compared with the range of incident energies and so that the tunneling rates through the individual barriers can be treated as constant across the resonance. This is certainly violated if scattering by optic phonons is significant, for example, which puts new structure into the resonance. Another source of difficulty would be a random potential, which changes the component of the electron's wave vector parallel to the current, and could therefore have a large effect on its tunneling rate.[34] However, inelastic scattering that gives only a small change in energy and wave vector does not invalidate the assumption of constant tunneling rates, and therefore does not affect the rate equation.

We next consider in Sec. IV the form of the current in a steady state and provide a physical interpretation of the exact expressions, whose appearance is rather complicated. The current is independent of scattering provided that the rate equation remains valid. Scattering enters through scattering-in and scattering-out rates, as we demonstrate in Sec. V with a one-dimensional system with scattering by optic phonons. This shows how a large part of the current due to scattering can be absorbed into a term whose appearance is suggestive of purely elastic tunneling, and which gives the current expected from the rate equation. However, the rate equation is not valid in general and there is an additional correction term to the current.[28] This has also been derived in the context of a generalized Landauer formula that includes the effect of interactions.[32]

Our calculations use nonequilibrium Green's functions and the methods developed by Langreth.[35] This may seem like a "sledgehammer to crack a nut," but has the advantage that we can obtain general results without needing an explicit form of the scattering. The physical interpretation of the equations is also clear in this formalism, and we relate the results to the simple model of incoherence given by Jonson and Grincwajg[8] in Sec. V. The spin of the electron plays no role in the calculations described here and its label is suppressed, but it would need to be included in any detailed calculations of the effect of scattering. All currents and charge densities should be doubled to account for the two directions of spin. We have set $\hbar = 1$ unless otherwise noted.

We first summarize the results of the coherent and sequential pictures.

A. Coherent picture

The central quantity of the coherent picture is the transmission coefficient, which has a Lorentzian form for energies close to the resonance at ε_0,

$$T_{\rm coh}(\omega) = T_{\rm coh}^{\rm pk}\left[1+\left(\frac{\omega-\varepsilon_0}{\gamma_{\rm coh}/2}\right)^2\right]^{-1}. \tag{1.1}$$

The peak transmission is

$$T_{\rm coh}^{\rm pk} = \frac{4T_LT_R}{(T_L+T_R)^2} = \frac{4\gamma_L\gamma_R}{(\gamma_L+\gamma_R)^2}, \tag{1.2}$$

where $T_{L,R}$ are the transmission coefficients of the left and right barriers, and are proportional to the corresponding tunneling rates $\gamma_{L,R}$; the exact relation will be given later for the model studied here. The full width at half maximum for coherent tunneling is simply the sum of the two tunneling rates $\gamma_{\rm coh} = \gamma_L + \gamma_R$. The particle current for a one-dimensional system and a single spin is given by

$$I_{\rm coh} = \int \frac{d\omega}{2\pi} T_{\rm coh}(\omega), \tag{1.3}$$

where the range of integration is set by the range of incoming energies or the difference in Fermi energies. Assuming that the peak of $T_{\rm coh}(\omega)$ is well within the range of energies, as in Fig. 1, and using the approximate Lorentzian form (1.1), gives

$$I_{\rm coh} = \tfrac{1}{4}T_{\rm coh}^{\rm pk}\gamma_{\rm coh} = \frac{\gamma_L\gamma_R}{\gamma_L+\gamma_R}. \tag{1.4}$$

The current depends on the area under $T_{\rm coh}(\omega)$, not

just its peak value. In three dimensions there is also a sum over the transverse wave vector, as in the Tsu-Esaki formula.[2]

B. Sequential picture

We shall take the sequential model to be defined by the following rate equation for $n_0(T)$, the total number of electrons in the resonant state at time T:

$$\frac{dn_0}{dT} = \frac{n_L - n_0}{\tau_L} + \frac{n_R - n_0}{\tau_R}. \tag{1.5}$$

The time constants $\tau_{L,R}$ are the reciprocals of the tunneling rates $\gamma_{L,R}$. The resonant state would fill up to $n_0 = n_L$ if it were connected only to the left lead; usually $n_R = 0$ as in Fig. 1. For a one-dimensional system, $n_L = 1$ if ε_0 lies within the range of incoming energies on the left. In this case, the solution of Eq. (1.5) in a steady state is

$$n_0 = \frac{\tau_R}{\tau_L + \tau_R}, \quad I_{\text{seq}} = \frac{1}{\tau_L + \tau_R} = \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} = I_{\text{coh}}. \tag{1.6}$$

Thus the current is the same in the coherent and sequential models.[4–6]

In a three-dimensional system, the resonant state is a two-dimensional electron gas. Figure 1 shows that the effective Fermi energy for the resonant state connected to the left lead only is $(\mu_L - \varepsilon_0)$, so

$$n_L = \frac{m}{2\pi\hbar^2}(\mu_L - \varepsilon_0), \tag{1.7}$$

where $m/2\pi\hbar^2$ is the constant two-dimensional density of states for a single spin. Again the current agrees with the coherent model.

II. MODEL

We use a simple model of a resonant-tunneling system that has been employed by several authors.[15,36,37] This is a tight-binding model with two chains of sites forming the left and right leads, coupled by weaker matrix elements to a central site that carries the resonant state. Current flows in the z direction, and this is generalized to a three-dimensional system by replacing the single sites with xy planes of sites, as illustrated in Fig. 2.

We first split the system into unperturbed and perturbed parts. The unperturbed system H_0 comprises three uncoupled sections. There is a left lead (planes $m \leq -1$), a central plane ($m = 0$) that will provide the resonant state, and a right lead (planes $m \geq 1$). Each of these subsystems is separately in thermal equilibrium at $t = -\infty$, and the chemical potentials of the leads differ by the applied bias. The chemical potential of the central plane is unimportant, as it will be washed out as soon as this plane is attached to the leads. The leads are perfect, meaning that there is no scattering within them and that they are translationally invariant along their length. We shall later assume that the left and right leads are identical. The sites in the central plane have energy ε_0, which will form the resonant state when coupled to the leads, with coupling between the sites in the plane (which is usually the same as the intraplane coupling in the leads).

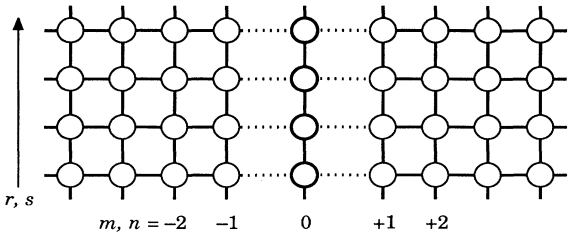


FIG. 2. Model of a resonant-tunneling diode. The resonant state occupies the central plane $m = 0$. The dotted lines represent the weak matrix elements that couple it to the leads on either side, $m = \pm 1$. Scattering is restricted to sites within the central plane. The planes are labeled by m, n and sites within each plane are labeled by r, s.

We need to be able to drive the system in a time-dependent way in order to get a transport equation. The simplest way to do this is to let the energy ε_0 of the resonant state vary with time, keeping the Fermi energies fixed. In practice the bias across the device is changed; this moves the Fermi energies and potentials in the two leads $\mu_{L,R}$ and $\phi_{L,R}$ as well as the energy of the resonant state. However, μ_R is irrelevant under typical operating conditions (Fig. 2) because it is so low that no electrons enter the resonant state from the right. The only effect that remains is the relative motion of ε_0 with respect to μ_L, which we have therefore captured. This simplification would fail at low bias, and all three energies would have to be considered.

Two perturbations are then turned on adiabatically. The first is a coupling between the end planes of the leads and the central plane,

$$\hat{V} = \sum_s (V_{10} c_{1,s}^+ c_{0,s} + V_{01} c_{0,s}^+ c_{1,s} + V_{\bar{1}0} c_{\bar{1},s}^+ c_{0,s} + V_{0\bar{1}} c_{0,s}^+ c_{\bar{1},s}), \tag{2.1}$$

where the matrix elements are independent of s, the label for the sites in the planes. There is a larger gap between planes $m = 0$ and ± 1 in Fig. 2 to emphasize that the matrix elements coupling these planes must be much weaker than those inside the leads, or there will not be a well-defined resonant state. This technique of treating the matrix elements for tunneling within perturbation theory was introduced by Caroli *et al.*[38,39] and Combescot[40] in their pioneering work on the theory of tunneling within the Keldysh formalism.

The second perturbation is due to interactions within the central plane, and gives rise to scattering and a self-energy in the Green's functions. The form of the interactions is unimportant—they could be due to disorder, electron-electron interaction, or electron-phonon scattering—but they must be local to the central plane. This is a crucial simplification of the model. In contrast to the general questions that we address in this paper, a specific calculation of the effect of phonons on reso-

nant tunneling using nonequilibrium Green's functions has been presented,[30] and there are further very recent studies.[29,31]

Formalism

We use a theory of nonequilibrium quantum statistical mechanics using Green's functions, based on the work of Kadanoff and Baym[41] and Keldysh.[42] The techniques and results summarized in this section are due to Langreth.[35]

The desired statistical information is carried by the Green's functions

$$g^{<}_{mn,rs}(t_1,t_2) = \langle c^{+}_{n,s}(t_2)c_{m,r}(t_1)\rangle\,, \tag{2.2a}$$

$$g^{>}_{mn,rs}(t_1,t_2) = \langle c_{m,r}(t_1)c^{+}_{n,s}(t_2)\rangle\,. \tag{2.2b}$$

These can be interpreted as densities of occupied ($g^{<}$) or empty ($g^{>}$) states. The subscripts identify the sites, as shown on Fig. 2: the first pair labels the planes; the second pair labels sites within the planes, and will frequently be suppressed. The superscript identifies the type of function and may contain a 0 to denote the unperturbed system. We shall also need advanced and retarded functions, defined conventionally; for example,

$$g^{r}_{mn,rs}(t_1,t_2) = -i\Theta\,(t_1-t_2)\,\langle\{c_{m,r}(t_1),c^{+}_{n,s}(t_2)\}\rangle\,. \tag{2.3}$$

These describe the propagation of an extra electron added to the system, and do not contain the statistical information present in $g^{<,>}$. They can be combined to give the "real part,"

$$\bar{g} = \tfrac{1}{2}\,(g^{r}+g^{a})\,. \tag{2.4}$$

The Green's functions are related through the spectral function a,

$$g^{<}+g^{>} = a = i\,(g^{r}-g^{a})\,. \tag{2.5}$$

The commutation relation of the operators leads to the important identity at equal times,

$$a_{mn,rs}(t,t) = \delta_{mn}\delta_{rs} = \int\frac{d\omega}{2\pi}\tilde{a}_{mn,rs}(\omega), \tag{2.6}$$

which takes a more familiar form as a sum rule on the Fourier transform $\tilde{a}$. The local density of states is the diagonal part of $\tilde{a}$ divided by 2π, and Eq. (2.6) then states that the integral of this local density of states is unity on each site. We shall make extensive use of this identity.

The Green's functions can be Fourier transformed with respect to t_1-t_2 in equilibrium or a steady state. In equilibrium,

$$\tilde{g}^{<}_{mn,rs}(\omega) = \tilde{a}_{mn,rs}(\omega)f(\omega), \tag{2.7}$$

where $f(\omega)$ is a Fermi function, and all the spatial dependence is in the spectral function. A kinetic equation is needed to find $g^{<}$ when the system is driven away from equilibrium, and can be derived from Dyson's equations. Those for the retarded function have their usual forms,

$$g^{r} = g^{0r} + g^{0r}\hat{V}g^{r} + g^{0r}\sigma^{r}g^{r}, \tag{2.8a}$$

$$g^{r} = g^{0r} + g^{r}\hat{V}g^{0r} + g^{r}\sigma^{r}g^{0r}. \tag{2.8b}$$

The two parts of the perturbation, the coupling $\hat{V}$ and the self-energy from the scattering σ^{r}, are kept separate because their effects are very different. The Dyson equations for $g^{<}$ are more complicated:

$$\begin{aligned} g^{<} &= g^{0<} + g^{0r}\hat{V}g^{<} + g^{0<}\hat{V}g^{a} \\ &\quad + g^{0r}\sigma^{r}g^{<} + g^{0r}\sigma^{<}g^{a} + g^{0<}\sigma^{a}g^{a}, \end{aligned} \tag{2.9a}$$

$$\begin{aligned} g^{<} &= g^{0<} + g^{r}\hat{V}g^{0<} + g^{<}\hat{V}g^{0a} \\ &\quad + g^{r}\sigma^{r}g^{0<} + g^{r}\sigma^{<}g^{0a} + g^{<}\sigma^{a}g^{0a}. \end{aligned} \tag{2.9b}$$

There are implied integrations over time and summations over sites between the terms. Fortunately the properties of the model simplify evaluation of these equations: $\hat{V}$ only couples plane -1 to 0 and 0 to $+1$, and the self-energies exist only within the central plane, 0. As well as the conventional functions $\sigma^{r,a}$ there are self-energies $\sigma^{<}$ and $\sigma^{>}$ which have the meaning of "scattering-in" and "scattering-out" rates. The self-energies are related by

$$\sigma^{<}_{00} + \sigma^{>}_{00} = \gamma_{00} = i\,(\sigma^{r}_{00} - \sigma^{a}_{00})\,, \tag{2.10}$$

where γ is the "scattering rate," in analogy to Eq. (2.5) relating the spectral function to the Green's functions.

A kinetic equation for individual sites in the central plane can be derived from Eqs. (2.9a) and (2.9b),[35] and could be used for studying the distribution of electrons in the resonant state. However, it provides far more detail than we need here because we are only interested in the *total* current flowing through the resonant state, which will be derived more simply from the continuity equation below.

This completes our survey of the formalism, which we shall now use to derive the current and rate equation for resonant tunneling.

III. DERIVATION OF THE RATE EQUATION

Our aim in this section is to derive the rate equation (1.5) for sequential tunneling starting from the quantum-mechanical Hamiltonian and to establish the conditions under which it is accurate. We start from the exact equation for the conservation of charge in the resonant state, plane $m=0$,

$$\frac{dn_0(T)}{dT} = I_L(T) - I_R(T), \tag{3.1}$$

where $I_L \equiv I_{0\bar{1}}$ is the current flowing into the resonant state from plane $m=-1$ on the left and $I_R \equiv I_{10}$ is the current flowing from the resonant state to plane $m=1$. The total number of electrons in the resonant state n_0 is given in terms of a Green's function by

$$n_0(T) = \sum_s g^{<}_{00,ss}(T,T) \equiv \mathrm{Tr}\,g^{<}_{00}(T,T), \tag{3.2}$$

where the trace is over the sites in plane 0. We now need

to find expressions for the currents in terms of Green's functions.

A. Current from left-hand contact

The operator for current crossing a single bond s from plane $m = -1$ to plane $m = 0$ is

$$\hat{I}_{0\bar{1},s} = -i(V_{0\bar{1}}c^{+}_{0,s}c_{\bar{1},s} - V_{\bar{1}0}c^{+}_{\bar{1},s}c_{0,s}). \tag{3.3}$$

Summing over all bonds and taking the expectation value gives

$$I_L(T) = -i\mathrm{Tr}\left[V_{0\bar{1}}g^{<}_{\bar{1}0}(T,T) - V_{\bar{1}0}g^{<}_{0\bar{1}}(T,T)\right]. \tag{3.4}$$

The next step is to eliminate the off-diagonal Green's functions. Using the Dyson equation (2.9a) gives

$$g^{<}_{\bar{1}0} = g^{0<}_{\bar{1}\bar{1}}V_{\bar{1}0}g^{a}_{00} + g^{0r}_{\bar{1}\bar{1}}V_{\bar{1}0}g^{<}_{00}. \tag{3.5}$$

All other terms vanish. This simple result relies on specific properties of the model: the unperturbed system comprises three disconnected parts; the tunneling only couples plane -1 to 0 and 0 to $+1$; and the interactions and self-energy are restricted to plane 0. The other form (2.9b) of the Dyson equation gives a more complicated result. An important feature is that the interactions only appear in Eq. (3.5) implicitly through g_{00}; there is no explicit self-energy. Using (2.9b) for $g^{<}_{0\bar{1}}$ in the same way gives

$$\begin{aligned} I_L(T) = -i\mathrm{Tr}(\, & V_{0\bar{1}}g^{0<}_{\bar{1}\bar{1}}V_{\bar{1}0}g^{a}_{00} + V_{0\bar{1}}g^{0r}_{\bar{1}\bar{1}}V_{\bar{1}0}g^{<}_{00} \\ & - V_{\bar{1}0}g^{<}_{00}V_{\bar{1}0}g^{0a}_{\bar{1}\bar{1}} - V_{\bar{1}0}g^{r}_{00}V_{\bar{1}0}g^{0<}_{\bar{1}\bar{1}}). \end{aligned} \tag{3.6}$$

The current is now expressed entirely in terms of two sets of diagonal Green's functions: the *unperturbed* Green's function for the end plane of the uncoupled left-hand lead $g^{0}_{\bar{1}\bar{1}}$ and the *full* Green's function for the resonant state coupled to the leads g_{00}. The appearance of the unperturbed function $g^{0}_{\bar{1}\bar{1}}$ for the lead, rather than the full function $g_{\bar{1}\bar{1}}$, corresponds to the assumption in the Tsu-Esaki formula[2] that the incoming electrons have an unperturbed Fermi function with a chemical potential set by the lead, and has emerged from our exact formulation. To simplify the notation, define $g_L = g^{0}_{\bar{1}\bar{1}}$, $g_0 = g_{00}$, and put $V_{\bar{1}0} = V_{0\bar{1}} = V_L$.

There is an implied integration over an intermediate time t' in Eq. (3.6); for example the first term expands to

$$-i\mathrm{Tr}\int_{-\infty}^{\infty} V_L g^{<}_L(T,t')V_L g^{a}_0(t',T)dt'. \tag{3.7}$$

The advanced and retarded functions ensure that only past times $t' \le T$ contribute. The relations (2.5) between the Green's functions can be used to rewrite (3.6) in an explicitly retarded form with only $<>$ functions:

$$\begin{aligned} I_L(T) = \mathrm{Tr}\int_{-\infty}^{T} (\, & V_L g^{<}_L V_L g^{>}_0 - V_L g^{>}_L V_L g^{<}_0 \\ & - V_L g^{<}_0 V_L g^{>}_L + V_L g^{>}_0 V_L g^{<}_L)dt'. \end{aligned} \tag{3.8}$$

This is easily interpreted. The first and fourth terms describe the transfer of electrons from filled states in the contact ($g^{<}_L$) to empty states in the resonant state ($g^{>}_0$); the second and third describe the transfer of electrons from the resonant state to the contact and are therefore negative. Unfortunately this equation is less easy to manipulate because of the finite limit on the integral, and we return to Eq. (3.6).

The next step is to perform a "gradient expansion" of the current,[43] assuming that it is a slowly varying function of time. This approximation is only needed in deriving the time-dependent rate equation, and is exact (but superfluous) for the current in a steady state. The accuracy of the gradient expansion will be checked in Sec. III C. The current contains terms of the form

$$u(t_1,t_2) = \int_{-\infty}^{\infty} dt'\, v(t_1,t')\, w(t',t_2). \tag{3.9}$$

Define "sum and difference" times by $\bar{T} = \frac{1}{2}(t_1 + t_2)$ and $\bar{t} = (t_1 - t_2)$. The idea is that functions vary rapidly in the difference $\bar{t}$ but only slowly in the average time $\bar{T}$. Finally, take a Fourier transform from $\bar{t}$ to ω, and $u(t_1,t_2)$ becomes $\tilde{u}(\omega,\bar{T})$ in the new variables. The gradient expansion[43] replaces Eq. (3.9) by

$$\tilde{u}(\omega,\bar{T}) = \tilde{v}(\omega,\bar{T})\tilde{w}(\omega,\bar{T}) + \frac{i}{2}\left[\frac{\partial\tilde{v}}{\partial\omega}\frac{\partial\tilde{w}}{\partial\bar{T}} - \frac{\partial\tilde{v}}{\partial\bar{T}}\frac{\partial\tilde{w}}{\partial\omega}\right] + \cdots. \tag{3.10}$$

We want (3.6) at equal times $t_1 = t_2 = \bar{T} = T$, which means $\bar{t} = 0$ or an integral over all ω. The first term becomes

$$\begin{aligned} & -i\mathrm{Tr}\int_{-\infty}^{\infty} V_L g^{<}_L(T,t')V_L g^{a}_0(t',T)dt' \\ & = -i\mathrm{Tr}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}V_L\tilde{g}^{<}_L(\omega)V_L\tilde{g}^{a}_0(\omega,T) \\ & \quad + \tfrac{1}{2}\mathrm{Tr}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\left[V_L\frac{\partial\tilde{g}^{<}_L}{\partial\omega}V_L\frac{\partial\tilde{g}^{a}_0}{\partial T}\right. \\ & \qquad \left. - V_L\frac{\partial\tilde{g}^{<}_L}{\partial T}V_L\frac{\partial\tilde{g}^{a}_0}{\partial\omega}\right] + \cdots. \end{aligned} \tag{3.11}$$

The second term in square brackets disappears because g_L is an equilibrium function and so does not depend on T. Absorb the integral over ω into the trace for brevity. Similarly, the fourth term of (3.6) gives

$$\begin{aligned} & i\mathrm{Tr}\int_{-\infty}^{\infty} V_L g^{r}_0(T,t')V_L g^{<}_L(t',T)dt' \\ & = i\mathrm{Tr}\left[V_L\tilde{g}^{r}_0(\omega,T)V_L\tilde{g}^{<}_L(\omega)\right] \\ & \quad + \tfrac{1}{2}\mathrm{Tr}\left[V_L\frac{\partial\tilde{g}^{r}_0}{\partial T}V_L\frac{\partial\tilde{g}^{<}_L}{\partial\omega}\right] + \cdots. \end{aligned} \tag{3.12}$$

There are now only multiplications in the temporal variables, so the Green's functions can be freely cycled under the trace. Using the properties (2.5) and (2.4) of the Green's functions allows Eqs. (3.11) and (3.12) to be combined, giving

$$\mathrm{Tr}\left[V_L\tilde{g}_L^<(\omega)V_L\tilde{a}_0(\omega,T)\right]+\mathrm{Tr}\left[V_L\frac{\partial\tilde{g}_L^<}{\partial\omega}V_L\frac{\partial\bar{g}_0}{\partial T}\right]+\cdots. \tag{3.13}$$

The middle two terms of (3.6) can be treated in the same way, and the sum is

$$I_L(T)=\mathrm{Tr}\left[V_L\tilde{g}_L^<(\omega)V_L\tilde{a}_0(\omega,T)-V_L\tilde{a}_L(\omega,T)V_L\tilde{g}_0^<(\omega)\right]+\mathrm{Tr}\left[V_L\frac{\partial\tilde{g}_L^<}{\partial\omega}V_L\frac{\partial\bar{g}_0}{\partial T}+V_L\frac{\partial\bar{g}_L}{\partial\omega}V_L\frac{\partial\tilde{g}_0^<}{\partial T}\right]+\cdots. \tag{3.14}$$

We shall ignore all but the leading terms for now, and reconsider the corrections in Sec. III C.

Now, $g_L^<$ is an equilibrium function and is given by

$$\tilde{g}_{L,sr}^<(\omega)=\tilde{a}_{L,sr}(\omega)f_L(\omega), \tag{3.15}$$

where $f_L(\omega)$ is the Fermi function at equilibrium for the left-hand lead; it has no spatial dependence, which is carried by the spectral function. Define a tunneling rate from the resonant state into the left-hand lead by

$$\gamma_{L,sr}(\omega)=V_L\tilde{a}_{L,sr}(\omega)V_L. \tag{3.16}$$

Note that this rate is the standard one, which is twice the value used in the Kondo problem. Choose a transverse basis set p to describe the lead in which the spectral function is diagonal in the indices within the plane, s and r. Then the relation between the diagonal spectral function and the local density of states allows the rate to be rewritten as $\gamma_{L,p}(\omega)=2\pi|V_L|^2N_{L,p}(\omega)$, exactly as expected from Fermi's golden rule. Note that $N_{L,p}(\omega)=\tilde{a}_{L,p}(\omega)/2\pi$ is the local density of states in the *end* plane of a semi-infinite system, and is substantially different from the density of states in the bulk. For example, if we take a one-dimensional system with a cosine band $\varepsilon(k)=-W\cos(ka)$, the density of states per site in the bulk is

$$N(\omega)=\frac{1}{\pi W}[1-(\omega/W)^2]^{-1/2}, \tag{3.17}$$

while on the end site it is

$$N_L(\omega)=\frac{2}{\pi W}[1-(\omega/W)^2]^{1/2}. \tag{3.18}$$

The important difference is that $N_L(\omega)$, and therefore the tunneling rate, vanishes towards the extremities of the band because the velocity goes to zero. The leading term in the current is finally

$$I_L(T)\approx\mathrm{Tr}\left\{\gamma_L(\omega)\left[\tilde{a}_0(\omega,T)f_L(\omega)-\tilde{g}_0^<(\omega,T)\right]\right\}. \tag{3.19}$$

This clearly vanishes in equilibrium, when all the Fermi functions are identical and the relation (2.7) can be used. The current into the right-hand lead can be calculated in the same way.

B. Rate equation

The currents can now be substituted into the continuity equation (3.1). This becomes, showing the integral over ω explicitly,

$$\frac{d}{dT}n_0(T)=\mathrm{Tr}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\{\gamma_L(\omega)\left[\tilde{a}_0(\omega,T)f_L(\omega)-\tilde{g}_0^<(\omega,T)\right]+\gamma_R(\omega)\left[\tilde{a}_0(\omega,T)f_R(\omega)-\tilde{g}_0^<(\omega,T)\right]\}. \tag{3.20}$$

The trace implies sums of the form $\sum_{s,r}\gamma_{L,sr}\tilde{a}_{0,rs}$. The sum must be expressed in terms of the local density of states, the diagonal elements of the spectral function, in order to derive an equation with the same form as Eq. (1.5). There is a basis set p that simultaneously diagonalizes both γ_L and γ_R if the left and right leads have the same cross section; we shall take this to be the case. Transverse plane waves would work for the usual case of a large device, for example.

We now come to the second assumption in the derivation. The integral over ω is dominated by the narrow peak in $\tilde{a}_0$ and $\tilde{g}_0^<$ due to the resonant state. Assume that the tunneling rates $\gamma_{L,R}$ through the individual barriers are constant across the width of the resonance, evaluate them at the peak and pull them out of the integral. Further, assume that these values of $\gamma_{L,R}$ are the same for each state p. This is usually exact in the absence of scattering within the resonance because the potential in which the electrons move is additive, with a double barrier along z plus a confining potential (or nothing) in the x-y plane. Careful analysis of this assumption may be needed if scattering within the resonance is strong. For example, Meshkov[34] considered resonant tunneling from a two-dimensional electron gas and the effect of a static random potential, which can cause a large change in the wave vector of an electron. This does not change its total energy, but may change substantially the z component of its wave vector and therefore the rate of tunneling. The effect of such scattering on three-dimensional devices has also been examined.[44,45] The rate equation would not be valid if such scattering were significant.

Making these simplifications, we find

$$\frac{d}{dT}n_0(T)=\gamma_L\sum_p\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\left[\tilde{a}_{0,pp}(\omega,T)f_L(\omega)-\tilde{g}_{0,pp}^<(\omega,T)\right]+\gamma_R\sum_p\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\left[\tilde{a}_{0,pp}(\omega,T)f_R(\omega)-\tilde{g}_{0,pp}^<(\omega,T)\right]. \tag{3.21}$$

Now,

$$\sum_p\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\tilde{a}_{0,pp}(\omega,T)f_L(\omega)=n_L(T), \tag{3.22}$$

the number of electrons that would build up in the resonance if it were filled to the equilibrium distribution of

the left-hand lead $f_L(\omega)$. This is very close to the definition of $n_L(T)$ in the rate equation (1.5) for the classical model, except that it takes account of the width of the resonance and is therefore more generally applicable. Similarly,

$$\sum_p \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \tilde{g}^<_{0,pp}(\omega, T) = n_0(T), \tag{3.23}$$

the number of electrons in the resonant state. The lifetime τ_L is the reciprocal of the rate γ_L, and Eq. (3.21) therefore reduces immediately to

$$\frac{d}{dT} n_0(T) = \frac{n_L(T) - n_0(T)}{\tau_L} + \frac{n_R(T) - n_0(T)}{\tau_R}. \tag{3.24}$$

This is exactly the rate equation for sequential tunneling, although with the more general definitions of n_L and n_R.

Resonant-tunneling diodes are usually operated under high bias, as shown in Fig. 1. For most values of transverse momentum the peak of the spectral function is entirely within or outside the range of incident energies. In this case we can use the sum rule (2.6) on the spectral function, that its integral over all frequencies is unity, to reduce Eq. (3.22) to

$$n_L(T) \approx \sum_p f_L(\omega_p), \tag{3.25}$$

where ω_p is the energy at which $\tilde{a}_{0,pp}(\omega, T)$ is peaked. This reduces to the simple form of Eq. (1.7) if there are two transverse dimensions, and we have returned to the classical rate equation of the sequential picture. Scattering only enters the rate equation (3.24) through the spectral function in n_L and n_R, and integrating over $\tilde{a}_0$ has removed its last vestige. This is the form in which the rate equation is usually used. Equation (3.24) is more accurate than (1.5), because it includes the case where the resonance might only be partially occupied — if it were near a chemical potential or the edge of a band, for example. Moreover, the resonance may become broad if structure is introduced by phonons, and (1.5) may become inaccurate.

We have now derived the "classical" rate equation under general conditions: it is valid in the case of purely coherent tunneling as well as when there is scattering within the resonant plane. In fact the rate equation is so "averaged" a picture of a resonant-tunneling diode that all information about coherence in tunneling has been lost. The only assumption made about the form of the scattering is that it keeps the spectral function narrow enough for γ_L and γ_R to be treated as constant across the width of the resonance. We shall now examine the accuracy of this and the only other approximation, the gradient expansion in time.

C. Accuracy of the rate equation

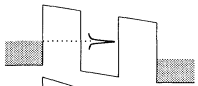
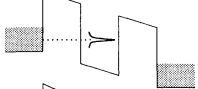
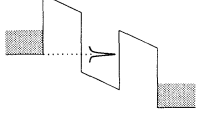
Two important approximations were made in the derivation of the rate equation (3.24): the gradient expansion in time, which means that its validity is restricted to "low" frequencies, and the assumption of a narrow resonance. The accuracy of these must now be considered, and the results are listed in Table I.

1. Gradient expansion

We shall verify the accuracy of the gradient expansion used above by checking the magnitude of the next term in the series, although this does not take account of the possibility that the expansion might fail due to nonperturbative terms; it is essentially a small-signal analysis, with the structure driven at a frequency ω_{ext}. The dependence on time was introduced by driving the energy of the resonant state ε_0 rather than that of the leads, so the only functions that depend on time are those in the central plane of sites. The first-order corrections to the lowest-order gradient expansion, Eq. (3.14), have a derivative with respect to ω multiplying one with respect to T. Start by considering the second pair of terms in Eq. (3.14). Dividing the correction by the main contribution gives

$$\left(\frac{1}{\tilde{a}_L}\frac{\partial \tilde{g}_L(\omega)}{\partial \omega}\right)\left(\frac{1}{\tilde{g}^<_0}\frac{\partial \tilde{g}^<_0(\omega, T)}{\partial T}\right). \tag{3.26}$$

TABLE I. Conditions under which the classical rate equation is accurate, showing the limits set by the gradient expansion and the assumption of a narrow resonance. The system is driven externally at frequency ω_{ext}, and Γ characterizes the total width of the resonance. It is assumed that there is no large-angle scattering, which would invalidate the rate equation.

	Bias condition	Gradient expansion valid	Narrow resonance approximation valid
	$\varepsilon_0 \approx \mu_L$	$\hbar\omega_{\text{ext}} \ll k_B T$	$\Gamma \ll k_B T$
	$\phi_L < \varepsilon_0 < \mu_L$	$\hbar\omega_{\text{ext}} \ll (\mu_L - \phi_L)$	$\Gamma \ll (\mu_L - \phi_L)$
	$\varepsilon_0 \approx \phi_L$	invalid	invalid

The real part $\bar{g}_L(\omega)$ contains no information about the occupation of the states, and therefore generally varies only on the scale of the bandwidth. The exception is near the edge of the band $\varepsilon_0 \approx \phi_L$, where $\bar{g}_L$ is singular (its slope is discontinuous in one dimension, and the singularity is weaker in higher dimensions). This can be neglected provided that the chemical potential is well away from the edge of the band. This correction will therefore be negligible unless $\tilde{g}_0^<(\omega, T)$ varies with T at a frequency comparable with the bandwidth, in which case it makes little sense to talk about a resonant state. This term leads to "renormalization of the driving force"[35] because it multiplies $\partial\tilde{g}_0^</\partial T$ and could therefore be taken over onto the left-hand side of the rate equation (3.24).

The other term [from Eq. (3.13)] is

$$\left(\frac{1}{g_L^<}\frac{\partial\tilde{g}_L^<(\omega)}{\partial\omega}\right)\left(\frac{1}{\tilde{a}_0}\frac{\partial\bar{g}_0(\omega,T)}{\partial T}\right). \tag{3.27}$$

In this case $\tilde{g}_L^<(\omega)$ contains the occupation function $f_L(\omega)$ as well as the spectral function for the lead. It will therefore vary on a scale k_BT near μ_L. Thus the bandwidth again sets the maximum frequency for which the gradient approximation is valid, unless ε_0 is close to μ_L in which case k_BT sets the scale. Numerically, $k_BT/\hbar \approx 100\,\mathrm{GHz}$ at 4 K, so this is not stringent even at this low temperature, and the bandwidth usually gives a much higher frequency. A similar time scale occurs in the theory of transport in bulk semiconductors in high electric fields, where it has been identified with the Landau quasiparticle formation time.[46]

The vital feature is that no derivatives of g_0 with respect to *frequency* appear (to this order). This would have been a disaster, because it would have introduced the width of the resonance as an energy scale, and the gradient expansion would only have held on time scales long compared with τ_L and τ_R. A rough value for the width of the resonance is 10 μeV, which would have set a maximum frequency of only 2 GHz. It turns out instead that the limit is set by the energy scales of the leads or k_BT, which are larger and give a much wider range of application for the gradient approximation.

2. Narrow resonance

We assumed that the resonance was narrow compared with the range of incoming energies $(\mu_L - \phi_L)$ to simplify the rate equation from Eq. (3.20) to (3.24). Also, we assumed that the tunneling rates γ_L and γ_R could be taken as constant across its width. There are three contributions to the total width γ_{tot} of the resonance (to be discussed more fully in Sec. V): the two tunneling rates $\gamma_{L,R}$ and the inelastic scattering rate γ_0. If the inelastic scattering simply broadens the resonance, without introducing extra structure, the assumption of a constant width will be satisfactory provided that $\gamma_{\mathrm{tot}}\,(d\ln\gamma_{L,R}/d\omega) \ll 1$. This means that the resonance must be narrow compared with the bandwidth, an obvious condition. This criterion also occurs in bulk transport, where it measures the importance of collisional broadening.[47]

Often the inelastic scattering does not simply broaden the resonance, but introduces characteristic structure; optic phonons provide an important example. The spectral function can be spread over a wide range, which we characterize with a single quantity Γ. It is possible for Γ to be bigger than $(\mu_L - \phi_L)$, and the rate equation will not be accurate. These results are summarized in Table I.

In the next section we shall derive an exact expression for the current in a steady state and verify the conditions under which this agrees with the prediction of the rate equation.

IV. CURRENT IN A STEADY STATE

The rate equation (3.24) is readily solved in a steady state; the number of electrons is given by

$$n_0 = \frac{\gamma_L n_L + \gamma_R n_R}{\gamma_L + \gamma_R}, \tag{4.1}$$

and the particle current is

$$I = \frac{\gamma_L\gamma_R}{\gamma_L + \gamma_R}\,(n_L - n_R)\,. \tag{4.2}$$

Inelastic scattering within the resonant state has no effect on the current if the densities n_L and n_R are given by the simple form (3.25). This has been found as an exact result in a model of resonant tunneling in the presence of optic phonons that assumed infinitely wide bands in the leads.[13] Under what conditions does this result hold in a more general model?

Our earlier result for the current, Eq. (3.14), which is exact in a steady state, gives

$$\begin{aligned} I_L &= \sum_p \int \frac{d\omega}{2\pi} I_{L,p}(\omega) \\ &= \sum_p \int \frac{d\omega}{2\pi}\gamma_{L,p}(\omega)\left[a_{0,pp}(\omega)f_L(\omega) - g^<_{0,pp}(\omega)\right], \end{aligned} \tag{4.3a}$$

$$\begin{aligned} I_R &= \sum_p \int \frac{d\omega}{2\pi} I_{R,p}(\omega) \\ &= -\sum_p \int \frac{d\omega}{2\pi}\gamma_{R,p}(\omega)\left[a_{0,pp}(\omega)f_R(\omega) - g^<_{0,pp}(\omega)\right]. \end{aligned} \tag{4.3b}$$

We have dropped the tildes, as all functions will now depend on energy rather than time. Conservation of current requires that $I_L = I_R$ in a steady state. To increase the symmetry of the expressions that follow, we define the measured current to be $I = \frac{1}{2}(I_L + I_R)$. This expression would also hold for the time-dependent current in a symmetric structure according to the Ramo-Shockley theorem.[48,49]

It is far from obvious that $I_L = I_R$ because of the different γ's and occupation functions. To make the expressions more symmetric, insert

$$1 = \frac{2\gamma_{R,p}}{\gamma_{L,p} + \gamma_{R,p}} + \frac{\gamma_{L,p} - \gamma_{R,p}}{\gamma_{L,p} + \gamma_{R,p}} \tag{4.4}$$

into Eq. (4.3a), and the expression obtained by inter-

changing L and R into Eq. (4.3b). The physical motivation for this trick will be discussed in Sec. V. The total current $I = \frac{1}{2}(I_L + I_R)$ can then be split into two terms, $I = I_a + I_c$, given by[28] FL

$$I_a = \sum_p \int \frac{d\omega}{2\pi} \frac{\gamma_{L,p}(\omega)\gamma_{R,p}(\omega)}{\gamma_{L,p}(\omega) + \gamma_{R,p}(\omega)} \times a_{0,pp}(\omega)[f_L(\omega) - f_R(\omega)], \tag{4.5}$$

$$I_c = \tfrac{1}{2} \sum_p \int \frac{d\omega}{2\pi} \frac{\gamma_{L,p} - \gamma_{R,p}}{\gamma_{L,p} + \gamma_{R,p}} [I_{L,p}(\omega) - I_{R,p}(\omega)]. \tag{4.6}$$

The first of these, I_a, is usually the main contribution and corresponds to the current from the rate equation if the sum rule on the spectral function can be used; scattering only appears implicitly through the form of the spectral function $a_{0,pp}(\omega)$. However, there are situations when the "correction" term I_c provides the *only* current, although it vanishes in linear response.[28] It arises because scattering changes the distribution function of electrons in the resonant state, as we shall show in Sec. V D. We shall now examine the two terms in more detail.

Consider a narrow resonance again, so that the tunneling rates $\gamma_{L,R}$ may be pulled out of the integrals. The correction term then vanishes with the aid of the continuity equation:

$$I_c \approx \frac{1}{2}\frac{\gamma_L - \gamma_R}{\gamma_L + \gamma_R} \sum_p \int \frac{d\omega}{2\pi} [I_{L,p}(\omega) - I_{R,p}(\omega)] = \frac{1}{2}\frac{\gamma_L - \gamma_R}{\gamma_L + \gamma_R} (I_L - I_R) = 0. \tag{4.7}$$

The current I_a reduces to that from the rate equation, Eq. (4.2), if the sum rule on the spectral function is used.

Some theoretical work on resonant tunneling in the presence of phonons[13] has used infinitely wide bands, in which case γ_L and γ_R are rigorously constant. The total current is therefore strictly independent of inelastic scattering provided that the whole resonance is flooded with incoming electrons, in which case the sum rule on the spectral function can be employed. In this case the current is limited by the rate at which electrons can get on (γ_L) and off (γ_R) the resonant state, and by the total number of states in the resonance (integral over the spectral function, which gives unity for each transverse state). This simple picture holds regardless of whether transport is purely coherent, with no inelastic processes, or whether there is strong scattering by phonons, provided that the width of the resonance remains small compared with the Fermi energy of the incoming electrons. This condition is the same as the requirement that the electron bounce back and forth in the well several times before it is scattered, and is an obvious criterion for the existence of a resonance.

The correction term I_c must be included if γ_L and γ_R depend on energy. Futhermore, since the tunneling rates depend on energy, I_a no longer reduces to an integral over the spectral function alone and its contribution to the current changes in the presence of interactions, even at large bias. An example of a system where this becomes important is resonant tunneling with narrow bands in the leads.[15] The bandwidth in these systems can be comparable to the energy of the optic phonons.

In the next section we shall pursue this analysis further for a simpler one-dimensional model and develop a physical motivation for the trick in Eq. (4.4).

V. PHYSICAL INTERPRETATION

Our aim in this section is to bring out the physical content of the exact expressions for the current derived in the preceding section, and to demonstrate the distinct ways in which coherent and incoherent (scattering) processes enter the current. The preceding section has shown that the current is independent of the degree of incoherence (scattering) provided that the bias is large and that the correction I_c can be neglected. Scattering enters into the equations only indirectly, through the form of the spectral function. We shall show below that the current can be split into "coherent" and "incoherent" contributions, where the incoherent part depends explicitly on scattering processes. However, these two terms can be recombined to show again that the total current is almost independent of scattering.

We also address the question of whether it is possible to write the current for an *interacting* system in the form

$$I = \int \frac{d\omega}{2\pi} T(\omega)\,[f_L(\omega) - f_R(\omega)], \tag{5.1}$$

perhaps with the generalized transmission coefficient $T(\omega)$ split into coherent and incoherent contributions? This would be a straightforward extension of Eq. (1.3) which holds for a noninteracting (coherent) system, and would be very much in the spirit of a number of treatments of inelastic or dephasing scattering.[7–9]

While Eq. (5.1) is attractive because of its simplicity, it is far from obvious that there should exist such an equation for an interacting system. In Eq. (5.1) the current appears as a sum of *independent* tunneling events for single electrons, whose transmission probability depends on energy. In an interacting system one would expect that an electron going through the sample alters the probability for a subsequent electron to transverse the sample. Thus it would seem that the tunneling events should not be independent and should not be described by simple transmission coefficients.

We have already addressed this question somewhat in an earlier paper,[28] where we concluded that there was no simple way to interpret the current in terms of an effective transmission coefficient in the most general case when the tunneling rates through the individual barriers γ_L and γ_R depend on energy. We examine this question in more detail here. As in our earlier treatments we specialize to the one-dimensional case to simplify the algebra, although the results carry over with trivial modification to a three-dimensional system which is translationally invariant in the transverse plane.

We start in Sec. V A by reviewing the approach of Jon-

son and Grincwajg[8] to inelastic scattering in a resonant-tunneling diode. Their approach yields a current of the form of Eq. (5.1) and it will be our guideline throughout this section. In Sec. V B we manipulate the exact expression for the current into two pieces, which we call the coherent and incoherent parts. The incoherent part comprises those electrons that have been scattered, with scattering-in and scattering-out rates $\sigma^<$ and $\sigma^>$, while the coherent part contains no such factors. The coherent part is written exactly in the form of Eq. (1.3), with the same coherent transmission probability found by Jonson and Grincwajg. The incoherent contribution must also be included, and we have to use an approximation consistent with their picture to get a form like Eq. (5.1) for the *total* current. Part of the incoherent current has now been omitted. Current conservation requires the missing piece to be zero when the rates γ_L and γ_R are independent of energy, but it does not vanish in general and may even carry the total current under some conditions. We illustrate this in Sec. V F by considering the case of resonant tunneling in the presence of an optical phonon.

Thus, our principal results are as follows.

(1) The current may not in general be written in the form of Eq. (5.1) as an integral over a simple effective transmission coefficient.[28]

(2) Only in the special case when the tunneling rates γ_L and γ_R are independent of energy is it possible to write the current as in Eq. (5.1), and to interpret it within the framework of Jonson and Grincwajg.

(3) The conservation law for charge provides the underlying reason that allows the current to be written in this form in the special case of constant γ_L and γ_R.

We start by reviewing the heuristic picture of incoherence in resonant tunneling.

A. Heuristic picture

Several authors[7–9] have treated inelastic scattering in a resonant-tunneling diode; we shall follow the work of Jonson and Grincwajg,[8] which is summarized below in our notation. They considered a resonant-tunneling diode under typical operating conditions as in Fig. 1, when the resonance is flooded with electrons from the left and the electrons coming from the right make no contribution to the current. The transmission coefficient $T(\omega)$ for a system without scattering is calculated by summing the waves produced as an electron bounces back and forth between the two barriers. This leads to a Lorentzian peak of height $T_{\rm pk}^{\rm coh}$ [Eq. (1.2)] and full width $\gamma_{\rm coh} = \gamma_L + \gamma_R$. Scattering is then introduced through an exponential decay of the waves, as would be found in an optical Fabry-Pérot étalon with absorption. The width of $T(\omega)$ increases to $\gamma_{\rm tot}$ while the peak transmission falls to $T_{\rm pk}^{\rm inc} = T_{\rm pk}^{\rm coh}\,(\gamma_{\rm coh}/\gamma_{\rm tot})^2$. The area under the peak is therefore reduced by $\gamma_{\rm coh}/\gamma_{\rm tot}$, and the current carried by the coherent fraction of the electrons falls by this factor. Electrons cannot be absorbed in the same way as photons, and those that have been scattered out of the coherent channel must be accounted for. Jonson and Grincwajg argue that a fraction $\gamma_R/(\gamma_L+\gamma_R)$ of these travel to the right, and contribute to the transmitted current, while the remaining fraction return to the left. Including these processes and integrating over the full width of the resonance restores the transmitted current to the value that it had without incoherence.

B. Current

The current from the left-hand lead to the resonant state for a one-dimensional system in a steady state, Eq. (4.3a), can be rewritten for a single transverse state as

$$I_L = \int \frac{d\omega}{2\pi}\left[V_L g_L^<(\omega) V_L g_0^>(\omega) - V_L g_L^>(\omega) V_L g_0^<(\omega)\right], \tag{5.2}$$

using the properties (2.5) of the Green's functions. The two terms describe the transfer of electrons into and out of the resonant state, as discussed for Eq. (3.8). To expand this, we need an explicit expression for $g_0^{<,>}$. In general, $g^<$ satisfies the integral equation[50,51]

$$g^< = [1 + g^r(\sigma^r + \hat{V})]g^{0<}[(\hat{V} + \sigma^a)g^a + 1] + g^r\sigma^< g^a. \tag{5.3}$$

We shall only want this for $g_0^<$. Usually the last term is the important one: $\sigma^<$ gives the rate at which electrons were scattered in to the state at some time in the past, g^r gives the probability amplitude that they survive to the present time, and g^a completes the square to give a probability density for $g^<$.[35] This is the only term that survives in bulk transport.[47] The first term can be rewritten, using the equations of motion for the advanced and retarded Green's functions, as $[g^r(g^0)^{-1}]g^{0<}[(g^0)^{-1}g^a]$, where the operator $(g^0)^{-1} = [i(d/dt) - H_0]$. Turning the first operator $(g^0)^{-1}$ around gives $(g^0)^{-1}g^{0<} = 0$. The first term in Eq. (5.3) therefore vanishes except for any boundary terms introduced by reversing the operator. The boundary term in time contains the memory of the initial state, before the interactions and tunneling were turned on. Usually this decays in time and can be neglected; the exception is if the resonance lies at an energy such that it can never make contact with the states in either lead, which we ignore. The spatial boundary terms, by contrast, are essential in an open system; effectively they provide another scattering-in rate, due to electrons tunneling from the leads. The result is

$$g_0^< = g_0^r\left(V_L g_L^< V_L + V_R g_R^< V_R + \sigma_0^<\right) g_0^a \tag{5.4}$$

$$= g_0^r\left(\gamma_L f_L + \gamma_R f_R + \sigma_0^<\right) g_0^a. \tag{5.5}$$

There is a similar equation for $g_0^>$; adding them gives an integral equation for the spectral function,

$$a_0 = g_0^r\left(\gamma_L + \gamma_R + \gamma_0\right) g_0^a. \tag{5.6}$$

Substituting (5.4) into (5.2) gives

$$I_L = \int \frac{d\omega}{2\pi}\{\ [V_L g_L^< V_L g_0^r V_L g_L^> V_L g_0^a + V_L g_L^< V_L g_0^r V_R g_R^> V_R g_0^a + V_L g_L^< V_L g_0^r \sigma_0^> g_0^a]$$
$$- [V_L g_L^> V_L g_0^r V_L g_L^< V_L g_0^a + V_L g_L^> V_L g_0^r V_R g_R^< V_R g_0^a + V_L g_L^> V_L g_0^r \sigma_0^< g_0^a]\}. \quad (5.7)$$

The terms with $g^{<,>}$ both on the left cancel. The current I_R can be treated in the same way and the two averaged to give a symmetric expression as before. Again, split the current into two terms: all factors containing a scattering rate $\sigma_0^<$ or $\sigma_0^>$ go into the incoherent current $I_{\rm inc}$, while the remainder give the coherent current $I_{\rm coh}$. Thus $I = \frac{1}{2}(I_L + I_R) = I_{\rm coh} + I_{\rm inc}$, where

$$I_{\rm coh} = \tfrac{1}{2}\int \frac{d\omega}{2\pi}\{\ [V_L g_L^< V_L g_0^r V_R g_R^> V_R g_0^a$$
$$- V_L g_L^> V_L g_0^r V_R g_R^< V_R g_0^a]$$
$$- [V_R g_R^< V_R g_0^r V_L g_L^> V_L g_0^a$$
$$- V_R g_R^> V_R g_0^r V_L g_L^< V_L g_0^a]\}, \quad (5.8)$$

$$I_{\rm inc} = \tfrac{1}{2}\int \frac{d\omega}{2\pi}\{\ [V_L g_L^< V_L g_0^r \sigma_0^> g_0^a - V_L g_L^> V_L g_0^r \sigma_0^< g_0^a]$$
$$- [V_R g_R^< V_R g_0^r \sigma_0^> g_0^a - V_R g_R^> V_R g_0^r \sigma_0^< g_0^a]\}. \quad (5.9)$$

These equations have a simple interpretation, despite their formidable appearance.

C. Coherent current

In this section we first show why $I_{\rm coh}$, defined by Eq. (5.8), may be interpreted as the coherent contribution to the current. Consider its fourth term first. If the equation were expressed in time rather than frequency, $g_L^<$ would have the earliest times because it is sandwiched between a retarded and an advanced function. This is the density of occupied states on the end site of the left lead. The matrix element V_L transfers the electron onto the resonant state, where g_0^r propagates it forward in time. Scattering causes g_0^r to decay as a function of time, reducing the current carried by the coherent channel; there may also be intermode scattering in a multimode system. Thus the coherent current is influenced by scattering processes, although they do not appear explicitly in its definition. The matrix element V_R transfers the electron into the right-hand lead, where $g_R^>$ is the density of empty states that the electron can enter. The other part of this term, with g_0^a, completes the square[35] to give a probability density. This term therefore transfers electrons from the left to the right, with a decay due to electrons that are scattered to other energies. Similarly, the first term describes the propagation of holes from right to left; it is clearly identical to the fourth if its components are reordered, and the pair of terms cancel the $\frac{1}{2}$ in front. The remaining two terms describe the propagation of electrons from right to left and therefore enter with a minus sign.

Grouping the terms as just described, reordering them, and using the expansion (2.7) of the equilibrium functions reduces Eq. (5.8) to

$$I_{\rm coh} = \int \frac{d\omega}{2\pi} g_0^r g_0^a \gamma_L \gamma_R (f_L - f_R). \quad (5.10)$$

Using Eq. (5.6) to replace $g_0^r g_0^a$ gives

$$I_{\rm coh} = \int \frac{d\omega}{2\pi} \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R + \gamma_0} a_0 (f_L - f_R). \quad (5.11)$$

Assume as usual that the tunneling and scattering rates are constant, and that $f_L = 1$ and $f_R = 0$ over the width of the resonance. Then everything can be taken out of the integral except for the spectral function, and the sum rule gives

$$I_{\rm coh} = \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R + \gamma_0} = \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} \frac{\gamma_{\rm coh}}{\gamma_{\rm tot}}, \quad (5.12)$$

where $\gamma_{\rm coh} = \gamma_L + \gamma_R$ and $\gamma_{\rm tot} = \gamma_L + \gamma_R + \gamma_0$. Inelastic scattering therefore reduces the coherent fraction of the current by a factor of $\gamma_{\rm coh}/\gamma_{\rm tot}$ as found by Jonson and Grincwajg.[8]

We next examine the incoherent current, and determine the conditions under which this supplements the coherent current back to its value before scattering was introduced.

D. Incoherent current

The "incoherent" current $I_{\rm inc}$, Eq. (5.9), which comprises electrons that have suffered real scattering events, can be interpreted along the same lines as the coherent current. The self-energies (scattering rates) are always the earliest functions in time. Start with the last term again. Electrons are scattered into the resonant state with energy ω at a rate given by $\sigma_0^<$. They are transferred to the right lead by V_R, and $g_R^>$ gives the density of empty final states. Again, g_0^a completes the square. This is a positive contribution to the current. The first term describes the same process but for holes moving to the left, or can be viewed as electrons that tunnel from the left lead to the resonant state, but are scattered out by $\sigma_0^>$ before they can tunnel off to the right lead. Thus the terms with $\sigma_0^>$ describe the incoherent current due to electrons that enter the resonance with energy ω and are scattered to some other energy before they can leave, and $\sigma_0^<$ gives the current due to electrons leaving the resonance with energy ω after being scattered from some other energy; each electron therefore enters twice.

These terms can be regrouped and simplified using similar ideas to those of Jonson and Grincwajg. They argued that one should calculate the total scattering rate, and break it into forward and backward contributions according to the tunneling rates through the barriers. We need to make some slight modifications to this, because the forward and backward currents also depend on the occupation of the final states, a factor that could be ignored in Jonson and Grincwajg's single-particle treatment but

must be included here.

In equilibrium, the scattering rates are related by the analog of Eq. (2.7),

$$\sigma_0^<(\omega) = \gamma_0(\omega) f(\omega), \quad \sigma_0^>(\omega) = \gamma_0(\omega)[1 - f(\omega)], \tag{5.13}$$

where $f = f_L = f_R$ is the common distribution function at equilibrium. Can a modified distribution function be found that obeys this relation away from equilibrium? Although Eq. (2.7) for $g^<(\omega)$ no longer holds, one could instead turn it around into the definition of a nonequilibrium distribution function,

$$f_0(\omega) = \frac{g_0^<(\omega)}{a_0(\omega)} = \frac{\gamma_L f_L + \gamma_R f_R + \sigma_0^<}{\gamma_L + \gamma_R + \gamma_0}. \tag{5.14}$$

Unfortunately substituting f_0 for f in Eq. (5.13) is not exact away from equilibrium, and f_0 contains the quantity that we are trying to find, $\sigma_0^<$. A simpler choice is to drop the terms containing interactions from Eq. (5.14) which leaves[28]

$$f^{\text{eff}}(\omega) = \frac{\gamma_L(\omega) f_L(\omega) + \gamma_R(\omega) f_R(\omega)}{\gamma_L(\omega) + \gamma_R(\omega)}. \tag{5.15}$$

This would be the correct distribution function in the resonant state if there were no interactions; it contains steplike features at μ_L and μ_R from the distributions of incoming electrons, which are smoothed by scattering in f_0.[28] Now use this approximate distribution function to write

$$\sigma_0^<(\omega) = \gamma_0(\omega) f^{\text{eff}}(\omega) + \left[\sigma_0^<(\omega) - \gamma_0(\omega) f^{\text{eff}}(\omega)\right], \tag{5.16}$$

and a corresponding expression containing $(1 - f^{\text{eff}})$ for $\sigma_0^>$. It is hoped that $\gamma_0 f^{\text{eff}}$ gives the main contribution to $\sigma_0^<$, and the rest is a small correction. In the range of energies of interest, $f_L \approx 1$ and $f_R \approx 0$, so

$$f^{\text{eff}}(\omega) \approx \frac{\gamma_L(\omega)}{\gamma_L(\omega) + \gamma_R(\omega)}, \qquad \sigma_0^<(\omega) \approx \frac{\gamma_L(\omega)}{\gamma_L(\omega) + \gamma_R(\omega)} \gamma_0(\omega). \tag{5.17}$$

These contain exactly the same factors postulated by Jonson and Grincwajg, and which we used earlier in Eq. (4.4). Using the approximate form $\sigma_0^< \approx \gamma_0 f^{\text{eff}}$, the first term of I_{inc} (Eq. 5.9) becomes

$$\tfrac{1}{2} \int \frac{d\omega}{2\pi} V_L g_L^< V_L g_0^r \sigma_0^> g_0^a = \tfrac{1}{2} \int \frac{d\omega}{2\pi} \gamma_L f_L g_0^r g_0^a \gamma_0 \frac{\gamma_L(1 - f_L) + \gamma_R(1 - f_R)}{\gamma_L + \gamma_R}. \tag{5.18}$$

Similarly, the next term gives

$$-\tfrac{1}{2} \int \frac{d\omega}{2\pi} V_L g_L^> V_L g_0^r \sigma_0^< g_0^a = -\tfrac{1}{2} \int \frac{d\omega}{2\pi} \gamma_L(1 - f_L) g_0^r g_0^a \gamma_0 \frac{\gamma_L f_L + \gamma_R f_R}{\gamma_L + \gamma_R}. \tag{5.19}$$

Adding the two and substituting for $g_0^r g_0^a$ gives

$$\tfrac{1}{2} \int \frac{d\omega}{2\pi} \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R + \gamma_0} \frac{\gamma_0}{\gamma_L + \gamma_R} a_0 (f_L - f_R). \tag{5.20}$$

The third and fourth terms of Eq. (5.9) give an equal contribution, removing the $\frac{1}{2}$ from Eq. (5.20). The sum can now be added to the coherent current I_{coh} from Eq. (5.11). The scattering rate γ_0 cancels in the result:

$$\int \frac{d\omega}{2\pi} \frac{\gamma_L \gamma_R}{\gamma_L + \gamma_R} a_0 (f_L - f_R) = I_a. \tag{5.21}$$

We have recovered the result of Sec. IV for the main part of the current I_a, which is given by an integral over the local density of states (spectral function) on the resonant site. This density of states can be changed strongly by interactions as in, for example, the Kondo effect,[28] but the total current is almost independent of interactions if the tunneling rates can be treated as constant, and the sum rule on the spectral function used.

We next need to evaluate the remaining current that arises because f^{eff} is only an approximate distribution function. The corrections from the four terms like $(\sigma_0^< - \gamma_0 f^{\text{eff}})$ sum to

$$\tfrac{1}{2} \int \frac{d\omega}{2\pi} \frac{\gamma_L - \gamma_R}{\gamma_L + \gamma_R} \left(\sigma_0^> g_0^< - \sigma_0^< g_0^>\right) = I_c, \tag{5.22}$$

where the integral equation (5.4) and its analog for $g^>$ have been used to simplify the integrand. Although this does not look the same as the earlier form for the correction I_c in Eq. (4.6), its content is identical. To show this, subtract Eq. (5.2) for I_L from its analog for I_R, which gives

$$I_L - I_R = \int \frac{d\omega}{2\pi} [\, \left(V_L g_L^< V_L + V_R g_R^< V_R\right) g_0^> - \left(V_L g_L^> V_L + V_R g_R^> V_R\right) g_0^<]. \tag{5.23}$$

The terms in parentheses can be replaced using the integral equation (5.4), and partly cancel to leave

$$I_L - I_R = \int \frac{d\omega}{2\pi} \left(\sigma_0^> g_0^< - \sigma_0^< g_0^>\right). \tag{5.24}$$

This is an *exact* equation that must be satisfied to ensure conservation of current in a steady state.[28] It shows immediately that Eq. (5.22) must vanish if the γ's are constant, and establishes the connection with (4.7). The difference $(\sigma_0^> g_0^< - \sigma_0^< g_0^>)$ gives the *net* scattering-out rate as a function of energy. Its integral over all frequencies, Eq. (5.24), must vanish to conserve the number of electrons: an electron that is scattered out at one energy must be scattered back in at another. There is a net arrival of electrons at some energy if $(\sigma_0^> g_0^< - \sigma_0^< g_0^>)$ is negative, and a net loss if it is positive, and these must balance when integrated over all energies. Thus

the structure due to inelastic scattering is revealed by the changing sign of $(\sigma_0^> g_0^< - \sigma_0^< g_0^>)$.

The current I_c arises because interactions change the distribution function from $f^{\rm eff}$ which would hold in the absence of interactions. Often this is a small effect but it can provide the only current if elastic tunneling is impossible, as will be illustrated in Sec. V F.

We have taken a rather laborious route to derive I_a and I_c because we wished to show how the current could first be divided into $I_{\rm coh}$ and the scattered part $I_{\rm inc}$. An alternative and more direct route is to use the approximate distribution function $f^{\rm eff}$ in $g_0^<$:

$$g_0^<(\omega) = a_0(\omega) f^{\rm eff}(\omega) + [g_0^<(\omega) - a_0(\omega) f^{\rm eff}(\omega)] . \tag{5.25}$$

This can be substituted into Eqs. (4.3a) and (4.3b) for the current. The terms with $a_0 f^{\rm eff}$ combine to give I_a, while the corrections again give I_c.

E. Discussion

We have shown that the current can be split into a coherent part, from electrons that traverse the device without suffering a real scattering event, and an incoherent part. Further rearrangement gave a major contribution I_a, which is only weakly dependent on scattering because of the sum rule on the spectral function, and a correction I_c which is usually small and vanishes both in linear response and in the limit of broadbands where the tunneling rates $\gamma_{L,R}$ are constant. The current I_a containing the spectral function looks like a purely elastic term (but is *not*). Defining a transmission coefficient by

$$T(\omega) = \frac{\gamma_L(\omega)\gamma_R(\omega)}{\gamma_L(\omega) + \gamma_R(\omega)} a_0(\omega), \tag{5.26}$$

the current can be written as

$$I_a = \int \frac{d\omega}{2\pi} T(\omega) [f_L(\omega) - f_R(\omega)], \tag{5.27}$$

which is identical to the result (5.1) discussed at the beginning of Sec. V. Where have the inelastic processes gone? One might have expected a transmission coefficient that depends on two energies $T(\omega_{\rm out}, \omega_{\rm in})$, as found for example by Wingreen, Jacobsen, and Wilkins.[13] Also, the presence of one electron should affect the probability of subsequent electrons tunneling if the interaction is treated beyond the mean-field approximation leading, for example, to the Kondo effect.[28] The point is that I_a *is* changed by inelastic scattering, despite its "elastic" appearance, and further scattering processes appear explicitly in the correction I_c. This will be illustrated in the next section for a simplified treatment of scattering by optic phonons.

F. Scattering by optic phonons

As a simple example, consider weak scattering by monochromatic optic phonons of energy Ω_0 with the lattice at zero temperature, so that only emission occurs. This makes an interesting comparison with the results of Runge and Ehrenreich,[33] who performed a similar analysis for the case of elastic (alloy) scattering in a resonant-tunneling structure. The tunneling rates to left and right are equal at equilibrium, $\gamma_L(\omega) = \gamma_R(\omega)$, and increase away from the edges of the band. Scattering produces sidebands in the spectral function at energies $\varepsilon_0 \pm \Omega_0$ if states are available. Figure 3 shows the spectral function $a_0(\omega)$, the net scattering-out rate, and contributions to I_c for three situations.

In Fig. 3(a) the spectral function, including the sidebands, lies completely within the range of energies of electrons coming from the left lead. The upper sideband in the spectral function corresponds to electrons that enter the resonant site at energy $\varepsilon_0 + \Omega_0$ before emitting a phonon to fall into the main resonance. It therefore gives a positive contribution to the net scattering-out rate $\sigma_0^> g_0^< - \sigma_0^< g_0^>$. The tunneling rates satisfy $\gamma_R > \gamma_L$ at this energy, because the energy on the right is higher in the band at this bias, so this gives a negative contribution to the integrand of I_c in Eq. (5.22). Similarly, the sideband at $\varepsilon_0 - \Omega_0$ gives a positive contribution to the current because it corresponds to scattering in. The two sidebands therefore tend to cancel when the integral over all energies in I_c is performed. Also, the spectral function is contained entirely within the range of incoming energies, so that its sum rule can be used provided that the tunneling rates are only slowly varying. The current will then be nearly independent of scattering, at the value given by the classical rate equation. This corresponds to the peak in $I(V)$.

The energy of the resonant state ε_0 lies below the range of incoming energies on the left in Figs. 3(b) and 3(c), which correspond to the valley region of $I(V)$. The resonant state is not directly accessible, but $\varepsilon_0 + \Omega_0$ lies inside this range. Thus an electron from the left can emit a phonon and fall into the resonant state, from which it escapes by tunneling into the empty states on the right, giving a phonon-assisted sideband in $I(V)$. The description below illustrates how inelastic scattering enters the expressions for the current that we have derived above, I_c in particular. It is not intended as an accurate calculation of the current, which would need a more careful treatment of the spectral function than we provide here and will be described in a subsequent paper.

The scattering-in and -out rates within the self-consistent Born approximation for this simple model are

$$\sigma_0^<(\omega) = M^2 g_0^<(\omega + \Omega_0), \quad \sigma_0^>(\omega) = M^2 g_0^>(\omega - \Omega_0), \tag{5.28}$$

where M^2 is the coupling constant. The scattering-in rate is proportional to the density of occupied states at an energy shifted up by that of the phonon, which is expected as only the emission of phonons is possible. The net scattering-out rate is

$$\sigma_0^> g_0^< - \sigma_0^< g_0^> = M^2 [\, g_0^>(\omega - \Omega_0) g_0^<(\omega) - g_0^<(\omega + \Omega_0) g_0^>(\omega)]. \tag{5.29}$$

The scattering-in and scattering-out terms differ only by a displacement of Ω_0 in ω, so this approximation obvi-

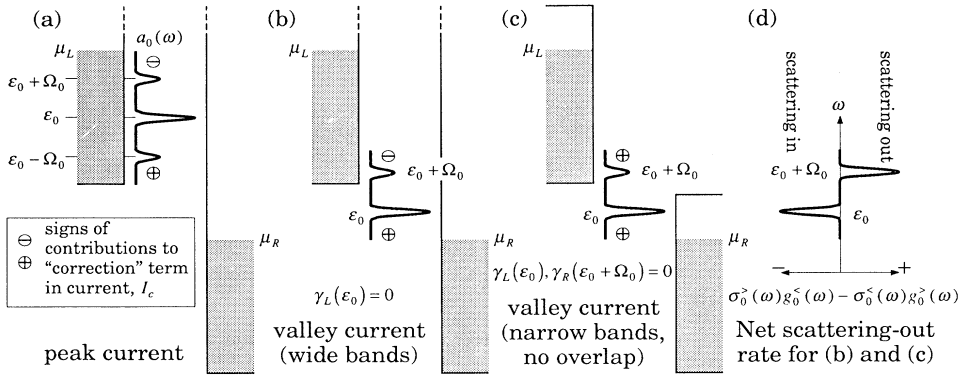


FIG. 3. One-dimensional resonant-tunneling system with scattering by optic phonons of energy Ω_0. This shows the spectral function a_0 and the signs of contributions to the correction term I_c in the current ($\oplus$ and $\ominus$). In (a) the resonance, including sidebands at $\varepsilon_0 \pm \Omega_0$ due to the emission of phonons, lies within the range of energies coming in from the left and gives the peak in $I(V)$. The lower sideband gives a negative contribution to I_c while the upper sideband gives a positive contribution, tending to cancel. The current is dominated by I_a; this contains the spectral function, whose sum rule can be used if the tunneling rates are constant to give a current that is independent of scattering as predicted by the classical rate equation. The resonance lies below the range of incoming energies in (b) and (c), the valley region of $I(V)$. Electrons from the left lead tunnel into the central plane with energy $\varepsilon_0 + \Omega_0$ and scatter out by emitting a phonon, creating the peak in the net scattering-out rate $\sigma_0^> g_0^< - \sigma_0^< g_0^>$ shown in (d). They scatter in to the resonant state at ε_0, giving the negative peak in (d), and tunnel to the right lead giving a positive contribution to I_c. In (b), where the bands are wide, the current also contains a contribution from I_a. The left and right bands are narrow and do not overlap in (c) and I_c is the total current; $I_a = 0$ and elastic tunneling is impossible.

ously obeys the condition for conserving electrons, Eq. (5.24) (but would not do so if the Born approximation were not self-consistent). We shall now analyze the scattering rates for the valley current, when the resonant state lies below the range of energies coming in from the left [Figs. 3(b)–3(d)]. The strongest structure in $a_0(\omega)$, and therefore in all the Green's functions, is the peak near the resonant energy ε_0. The contribution to each of the four functions in Eq. (5.29) is as follows.

(1) The first Green's function $g_0^>(\omega - \Omega_0)$ is peaked at $\omega = \varepsilon_0 + \Omega_0$; $g_0^<(\omega) \neq 0$ at this energy, so this peak contributes.

(2) The next function $g_0^<(\omega)$ is peaked at $\omega = \varepsilon_0$, but $g_0^>(\varepsilon_0 - \Omega_0) = 0$ because this lies below the bottom of both bands.

(3) In the second pair, $g_0^<(\omega + \Omega_0)$ is peaked at $\omega = \varepsilon_0 - \Omega_0$, where $g_0^>(\omega)$ vanishes again.

(4) Finally, $g_0^>(\omega)$ is peaked at $\omega = \varepsilon_0$, and $g_0^<(\varepsilon_0 + \Omega_0) \neq 0$.

Thus there is a peak in the scattering-in rate at $\omega = \varepsilon_0$ and one in the scattering-out rate at $\omega = \varepsilon_0 + \Omega_0$, as shown in Fig. 3(d). This is exactly what we expect: electrons tunnel into the resonant state at $\varepsilon_0 + \Omega_0$, scatter out by emitting a phonon, and scatter back in at ε_0.

In the case of the valley current in a system with wide bands shown in Fig. 3(b), $\gamma_R > \gamma_L$ at all energies so the main peak in the spectral function gives a positive contribution to I_c. The upper sideband makes a negative contribution to I_c, and tends to cancel the excessive contribution that the sideband in the spectral function makes to I_a. There will also be an elastic current.

Figure 3(c) shows the same system, again in the valley region of $I(V)$, but with such narrow bands in the leads that they no longer overlap at this bias. The contribution I_a vanishes completely in this case, not just its elastic part, because the product $\gamma_L(\omega)\gamma_R(\omega) = 0$ at all energies. All current must involve inelastic scattering and is carried by I_c under these conditions; the rate equation fails. The changing sign of $\gamma_L - \gamma_R$ means that both peaks in $a_0(\omega)$ make positive (and equal) contributions in this case.

This example has demonstrated how inelastic tunneling is treated by our formalism, and shown the extent to which these processes are included within the "spectral function" term I_a, although its appearance is reminiscent of purely elastic tunneling. It has also shown the limitations of this term, and how the correction I_c can carry the only current under extreme conditions.

VI. CONCLUSIONS

We have derived the classical rate equation for sequential tunneling from a quantum kinetic equation and shown that it is valid in both the coherent and incoherent limits: the equation contains too little information to distinguish between them. It is valid at "low" frequencies on a scale set by the Fermi energy in the leads or the scale on which the tunneling rates vary with energy, giving a broad range of validity for typical devices. The maximum frequency is not set by the width of the resonance, which would have provided a much more stringent condition. The other important condition is that the resonance must be narrow compared with the same energy scales as the frequency. This condition is much easier to violate. Scattering by optic phonons spreads the spectral function, and therefore the width of the resonance, over a wide range in energy. In a three-dimensional system, intermode scattering can have a drastic effect because

it can produce large changes in wave vector normal to the barriers, even if the total energy of the electron is unaffected.

It is clear that a fine probe is needed to determine the degree of coherence in resonant tunneling. For example, the distribution of energy for electrons in a resonant state has recently been measured optically.[52] It would be possible to calculate this distribution by computing the self-energies self-consistently in a three-dimensional system.

We have also analyzed the current in the case where the rate equation does not hold. Inelastic processes enter through scattering-in and scattering-out rates. The current can be split into two terms. The first and most important I_a involves an integral over the spectral function and has the same form that holds in purely coherent transport. Although scattering does not appear explicitly in the expression for I_a, analysis of the current into coherent and incoherent processes shows that I_a contains part of the scattered current. The remaining term in the total current I_c vanishes in the absence of scattering, if the bands are infinitely wide, or in linear response. It is usually small because its form is close to the condition for conserving current, but it carries all the current if the bias is so high that the bands in the left and right leads do not overlap and elastic transport is impossible. Strongly inelastic processes such as optic phonon scattering lead to rich structure in I_c.

Note added in proof. It has been drawn to our attention that Eq. (3.20) was derived previously by L. Y. Chen and C. S. Ting [Phys. Rev. B **41**, 8533 (1990)]. They only considered a noninteracting system and used path integrals, which appear to contain an approximation similar to our gradient expansion. We are grateful to R. Lake for bringing this paper to our notice.

ACKNOWLEDGMENTS

This work was supported primarily by the Office of Naval Research, and additionally by NSF through Grant No. PHY89-04035 to the Institute for Theoretical Physics, Santa Barbara, where part of this work was done. P.H. gratefully acknowledges the additional support of the Danish Research Academy.

*Permanent address: Department of Electronics and Electrical Engineering, Glasgow University, Glasgow, G12 8QQ, United Kingdom.

†Present address: Department of Physics, University of Florida, Gainesville, FL 32611.

[1]L. L. Chang, L. Esaki, and R. Tsu, Appl. Phys. Lett. **24**, 593 (1974).

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

[3]S. Luryi, Appl. Phys. Lett. **47**, 490 (1985).

[4]M. C. Payne, J. Phys. C **19**, 1145 (1986).

[5]T. Weil and B. Vinter, Appl. Phys. Lett. **50**, 1281 (1987).

[6]G. A. Toombs and F. W. Sheard, in *Electronic Properties of Multilayers and Low-Dimensional Semiconductor Structures*, edited by J. M. Chamberlain *et al.* (Plenum, New York, 1990), p. 257.

[7]A. D. Stone and P. A. Lee, Phys. Rev. Lett. **54**, 1196 (1985).

[8]M. Jonson and A. Grincwajg, Appl. Phys. Lett. **51**, 1729 (1987).

[9]Y. Hu, J. Phys. C **21**, L23 (1988).

[10]M. Büttiker, in *Resonant Tunneling in Semiconductors: Physics and Applications*, edited by L. L. Chang, E. E. Mendez, and C. Tejedor (Plenum, New York, 1991), p. 213.

[11]M. Büttiker, IBM J. Res. Dev. **32**, 63 (1988).

[12]L. I. Glazman and R. I. Shekhter, Solid State Commun. **66**, 65 (1988).

[13]N. S. Wingreen, K. W. Jacobsen, and J. W. Wilkins, Phys. Rev. Lett. **61**, 1396 (1988).

[14]B. Y. Gelfand, S. Schmitt-Rink, and A. F. J. Levi, Phys. Rev. Lett. **62**, 1683 (1989).

[15]P. Hyldgaard and A.-P. Jauho, J. Phys. Condens. Matter **2**, 8725 (1990).

[16]J. A. Stovneng, E. H. Hauge, P. Lipavsky, and V. Spicka, Phys. Rev. B **44**, 13 595 (1991).

[17]G. W. Wu and T. C. McGill, Phys. Rev. B **40**, 9969 (1989).

[18]B. G. R. Rudberg, Semicond. Sci. Technol. **5**, 328 (1990).

[19]X. Wu and S. E. Ulloa, Phys. Rev. B **44**, 13 148 (1991).

[20]F. Chevoir and B. Vinter, Appl. Phys. Lett. **55**, 1859 (1989).

[21]P. J. Turley and S. W. Teitsworth, Phys. Rev. B **44**, 3199 (1991).

[22]P. J. Turley and S. W. Teitsworth, Phys. Rev. B **44**, 8181 (1991).

[23]P. J. Turley and S. W. Teitsworth, Phys. Rev. B **44**, 12 959 (1991).

[24]V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Phys. Rev. B **36**, 7635 (1987).

[25]M. L. Leadbeater, E. S. Alves, L. Eaves, M. Henini, O. H. Hughes, A. Celeste, J. C. Portal, G. Hill, and M. A. Pate, Phys. Rev. B **39**, 3438 (1989).

[26]W. R. Frensley, Rev. Mod. Phys. **62**, 745 (1990).

[27]N. C. Kluksdahl, A. M. Kriman, D. K. Ferry, and C. Ringhofer, Phys. Rev. B **39**, 7720 (1989).

[28]S. Hershfield, J. H. Davies, and J. W. Wilkins, Phys. Rev. Lett. **67**, 3720 (1991).

[29]M. A. Alam, R. A. Morrisey, and A. N. Khondker, J. Appl. Phys. **71**, 3077 (1992).

[30]E. V. Anda and F. Flores, J. Phys. Condens. Matter **3**, 9087 (1991).

[31]R. Lake and S. Datta, Phys. Rev. B **45**, 6670 (1992).

[32]Y. Meir and N. S. Wingreen, Phys. Rev. Lett. **68**, 2512 (1992).

[33]E. Runge and H. Ehrenreich, Phys. Rev. B **45**, 9145 (1992).

[34]S. V. Meshkov, Zh. Eksp. Teor. Fiz. **91**, 2252 (1986) [Sov. Phys.—JETP **64**, 1337 (1986)].

[35]D. C. Langreth, in *Linear and Non-Linear Electron Transport in Solids*, edited by J. T. Devreese and E. van Doren (Plenum, New York, 1976), p. 3.

[36]L. Y. Chen and C. S. Ting, Phys. Rev. B **41**, 8533 (1990).

[37]L. I. Glazman and K. A. Matveev, Pis'ma Zh. Eksp. Teor. Fiz. **48**, 403 (1988) [JETP Lett. **48**, 445 (1988)].

[38]C. Caroli, R. Combescot, P. Nozières, and D. Saint-James, J. Phys. C **4**, 616 (1971).

[39]C. Caroli, R. Combescot, D. Lederer, P. Nozières, and D.

Saint-James, J. Phys. C **4**, 2598 (1971).
[40]R. Combescot, J. Phys. C **4**, 2611 (1971).
[41]L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Benjamin, Reading, MA, 1962).
[42]L. V. Keldysh, Zh. Eksp. Teor. Fiz. **47** 1515 (1964) [Sov. Phys.—JETP **20**, 1018 (1964)].
[43]D. C. Langreth and J. W. Wilkins, Phys. Rev. B **6**, 3189 (1972).
[44]H. A. Fertig and S. Das Sarma, Phys. Rev. B **40**, 7410 (1989).
[45]J. Leo and A. H. MacDonald, Phys. Rev. Lett. **64**, 817 (1990).
[46]P. Lipavsky, F. S. Khan, A. Kalvova, and J. W. Wilkins, Phys. Rev. B **43**, 6650 (1991).
[47]F. S. Khan, J. H. Davies, and J. W. Wilkins, Phys. Rev. B **36**, 2578 (1987).
[48]S. Ramo, Proc. IRE **27**, 584 (1939).
[49]W. Shockley, J. Appl. Phys. **9**, 639 (1938).
[50]A. Blandin, A. Nourtier, and D. W. Hone, J. Phys. (Paris) **37**, 369 (1976).
[51]P. Danielewicz, Ann. Phys. (N.Y.) **152**, 239 (1984).
[52]T. K. Woodward, D. S. Chemla, I. Bar-Joseph, H. U. Baranger, D. L. Sivco, and A. Y. Cho, Phys. Rev. B **44**, 1353 (1991).