Nonlinear $I-V$ characteristics of a mesoscopic conductor

Baigeng Wang and Jian Wang
Department of Physics, The University of Hong Kong, Pokfulam Road, Hong Kong, China

Hong Guo
Center for the Physics of Materials and Department of Physics, McGill University, Montreal, PQ H3A 2T8, Canada

(Received 20 April 1999; accepted for publication 31 July 1999)

We present a general theoretical formulation, based on nonequilibrium Green's functions, for nonlinear dc transport in multiprobe mesoscopic conductors. The theory is gauge invariant and is useful for the predictions of current–voltage characteristics and the nonequilibrium charge pileups inside the conductor. We have provided a detailed comparison between the gauge invariant scattering matrix theory and our theory. We have also given several examples where the $I-V$ curve can be obtained analytically. The effects of exchange and correlation have been considered explicitly. © *1999 American Institute of Physics.* [S0021-8979(99)06121-6]

I. INTRODUCTION

Many practical electronical devices, such as diodes and transistors, operate on nonlinear current–voltage ($I-V$) characteristics: $I_\alpha = I_\alpha(\{V_\beta\})$, where subscripts α and β denote leads which connect the device to the outside world. Classically, one can predict the $I-V$ curves by solving the coupled equations of classical electron motion such as the Boltzman equation, and the Poisson equation for the electrostatic potential of the conductor, subjecting to the boundary conditions that at the asymptotic region of the lead β, the external bias voltage is fixed at V_β. For coherent quantum conductors in the mesoscopic regime, one still must solve the coupled equations but the electrons are now quantum entities. Clearly the prediction of $I-V$ curves becomes much more difficult in the quantum situation. As a consequence, most theoretical analysis of quantum transport in coherent quantum devices do not predict $I-V$ curves: they focus on the linear dc conductance $G_{\alpha\beta}$ which can be calculated from a variety of theoretical methods.[1,2] This is then compared with experiments which extract $G_{\alpha\beta}$ from the measured $I-V$ curves at a vanishing bias voltage: $I_\alpha = \Sigma_\beta G_{\alpha\beta} V_\beta$.

Because linear conductance $G_{\alpha\beta}$ does not give the whole picture concerning a nonlinear device, it is very important to, theoretically, go beyond the linear transport regime and predict the whole $I-V$ curve. At nonlinear situations, one must worry about a fundamental physics requirement: the gauge invariant condition which dictates that the predicted electric current should be the same when potential everywhere is shifted by a constant amount.[3] Büttiker and co-workers[3,4] have developed a scattering matrix theory (SMT) which satisfies the gauge invariant condition and predicted the second-order nonlinear conductance. When the scattering matrix takes a specially simple[4] form, e.g., the Breit–Wigner form, the full nonlinear $I-V$ characteristics is also obtained.[4] The key idea,[3,4] in order to maintain gauge invariance, is to include the internal potential landscape into the analysis which is a result of long range Coulomb interactions.[5] For example, Ref. 4 treated internal potential to linear order in bias voltage, thus it can predict nonlinear conductance up to the second order. Recently, this SMT has been extended to predict higher order weakly nonlinear conductances[6] and connections to the framework of response theory has been formalized.[7] In both SMT[4,6] and the response theory,[7] one calculates the nonlinear conductance perturbatively order by order in voltage. Such theories make sense for weakly nonlinear situations where the external bias is finite but small. Hence practically, it is very difficult to compute $I-V$ curves for more general situations.

From the nonequilibrium Green's functions (NEGF),[8–15] Ref. 16 provided an analysis of $I-V$ curve in the wide-band limit, where gauge invariance was satisfied by including the internal potential phenomenologically through the use of a capacitive charging model. The charging model is, however, not fully nonlinear since the internal potential is treated linearly in it. It is thus important and attractive to further develop the NEGF to the fully nonlinear regime for the purpose of predicting $I-V$ characteristics of multiprobe coherent quantum conductors. It is the purpose of this work to provide such a development: we have formulated a general gauge invariant nonlinear dc theory based on NEGF by treating the nonlinear internal potential from first principles. Our theory also goes beyond the wide-band limit and can directly predict the $I-V$ curves and the nonequilibrium charge pileups[3,17] inside the conductor. The former can be expanded to obtain weakly nonlinear conductances which we shall compare with results obtainable from SMT at lower orders;[18] while the latter gives a voltage-dependent nonlinear capacitance coefficient which is experimentally measurable. Our theory provides a solid base for further numerical predictions of $I-V$ curves for complicated coherent quantum device geometries. We will provide the theoretical formalism in Sec. II. In Sec. III, we will give detailed comparison between the result of our theory and that of SMT. We will calculate the $I-V$ curve for a number of examples. The summary will be given in Sec. IV.

II. GAUGE INVARIANT FORMULATION

Let's consider a quantum coherent multiprobe conductor with the Hamiltonian

$$H=\sum_{k\alpha}\epsilon_{k\alpha}c^{\dagger}_{k\alpha}c_{k\alpha}+H_{cen}\{d_n,d^{\dagger}_n\} + \sum_{k\alpha,n}(T_{k\alpha,n}c^{\dagger}_{k\alpha}d_n+\text{c.c.}), \quad (1)$$

where $\epsilon_{k\alpha}=\epsilon^0_k+qV_\alpha$. The first term of Eq. (1) describes the probes where dc signal is applied far from the conductor; the second term is the general Hamiltonian for the scattering region which is a polynomial in $\{d^{\dagger}_n,d_n\}$ that commutes with the electron number operator[16] $N=\Sigma_n d^{\dagger}_n d_n$; the last term gives the coupling between probes and the scattering region with the coupling matrix $T_{k\alpha,n}$. Here, $c^{\dagger}_{k\alpha}$ ($c_{k\alpha}$) is the creation (annihilation) operator of electrons inside the α probe. Similarly, $d^{\dagger}_n$ (d_n) is the operator for the scattering region.

The electric current can be written in terms of Green functions in the usual manner[19,20] ($\hbar=1$),

$$J_\alpha=-iq\int(dE/2\pi)Tr(\Gamma_\alpha(E-qV_\alpha) \times\{[G^r(E,U)-G^a(E,U)]f(E-qV_\alpha)+G^<(E)\}), \quad (2)$$

where $G^r(E,U)$ is the retarded Green's function which depends on U, the electrostatic potential builds up inside our conductor. Although Eq. (2) is derived in momentum space, physical quantities such as current do not depend on representation. Hence, one can calculate the Green's function and current using Eq. (2) either in momentum space or in real space. To get analytic solution, one can assume the wide-band limit and work in momentum space. On the other hand, in the numerical calculations, it is more convenient to work in real space where a tight-binding form of the Hamiltonian is used. In the Hartree approximation, the retarded Green's function in real space is given by[2]

$$G^r(E,U)=\frac{1}{E-H-qU-\Sigma^r}, \quad (3)$$

where $\Sigma^r\equiv\Sigma_\alpha\Sigma^r_\alpha(E-qV_\alpha)$ is the self-energy[21] and $\Gamma_\alpha(E)=-2\,\text{Im}[\Sigma^r_\alpha(E)]$ is the linewidth function. Within the density functional theory,[22] we can further include the exchange and correlation effect,

$$G^r(E,U)=\frac{1}{E-H-qU-V_{xc}-\Sigma^r}, \quad (4)$$

where $V_{xc}=\delta E_{xc}/\delta\rho$ is the potential due to the exchange and correlation energy E_{xc} and ρ is the charge density. It is worth to emphasize that a most important departure of our theory from the previous NEGF analysis[19,20,16] is that we explicitly include the internal potential landscape $U(\mathbf{r})$ into the Green's functions self-consistently. This is the crucial step in the development of a gauge invariant nonlinear dc theory. At Hartree level $U(\mathbf{r})$ is determined by the self-consistent Poisson equation,

$$\nabla^2U(x)=-4\pi\rho=4\pi iq\int(dE/2\pi)[G^<(E,U)]_{xx}, \quad (5)$$

where $G^<$ is the lesser Green's function in real space and x labels the position. Within Hartree approximation, $G^<$ is related to the retarded and advanced Green's functions G^r and G^a,

$$G^<(E,U)=G^r\sum_\beta i\Gamma_\beta(E-qV_\beta)f(E-qV_\beta)G^a. \quad (6)$$

Equation (5) is, in general, a nonlinear equation because $G^{r,a}$ depends on $U(\mathbf{r})$ [see Eq. (3)]. Using Eq. (6) we reduce the current into the following form;

$$J_\alpha=-q\sum_\beta\int(dE/2\pi)Tr(\Gamma_\alpha G^r\Gamma_\beta G^a)(f_\alpha-f_\beta), \quad (7)$$

where we have used the notation $\Gamma_\alpha\equiv\Gamma_\alpha(E-qV_\alpha)$, $\Gamma=\Sigma_\alpha\Gamma_\alpha$, and $f_\beta\equiv f(E-qV_\beta)$. To make connection with the SMT, we introduce the screened transmission function

$$A_{\alpha\beta}=Tr[\Gamma_\alpha G^r(\Gamma\delta_{\alpha\beta}-\Gamma_\beta)G^a]/(4\pi^2), \quad (8)$$

we then arrive at the familiar form of the current in SMT,[4]

$$J_\alpha=-2\pi q\sum_\beta\int dEf_\beta A_{\alpha\beta}. \quad (9)$$

Equations (7), (3), and (5) completely determines the nonlinear $I-V$ characteristics of an arbitrary multiprobe conductor, they form the basic equations of our theory. The self-consistent nature of the problem is clear: one must solve the quantum scattering problem (the Green's functions) in conjunction with the Poisson equation. It is easy to prove that the current expression Eq. (7) is gauge invariant: shifting the potential everywhere by a constant V, $U\rightarrow U+V$ and $V_\alpha\rightarrow V_\alpha+V$, J_α from Eq. (7) remains the same. Note in Eq. (7), the quantity Γ depends on voltage and without such a voltage dependence, the gauge invariance cannot be satisfied. On a technical side, Eqs. (7), (3), and (5) also form a basis for numerical predictions of $I-V$ curves. For instance, one can compute the various Green's functions G and the coupling matrix Γ for multiprobe conductors using tight-binding models;[2] and the Poisson equation can be solved using very powerful numerical techniques.[23]

III. RESULTS

The main thrust of the previous section (and of this work) is the solution of the gauge invariance problem for dc nonlinear transport in general terms of bias voltages. In the rest of the article, we shall derive analytical expressions for a number of examples which can be solved in closed form. At low bias where SMT[3] and linear response[7] are applicable for weakly nonlinear situation, we show that our general formula reduces and becomes compatible with them. But for higher bias where these previous theories are not applicable, our theory becomes an unique approach for analyzing nonlinear dc quantum transport. Hence, we derive the general current–voltage characteristics for the entire range of nonlinearity for

a tunneling device, and prove that results obtained with or without gauge invariance can differ substantially not only quantitatively but also qualitatively.

A. Weakly nonlinear regime

For weak nonlinearity, we can series expand all quantities in terms of the small external bias voltage[3] and obtain results order by order: this is precisely the approach adapted in SMT[3] and response theory.[7] In this subsection, we confirm that our nonlinear theory indeed reduces to these previous approaches at the weakly nonlinear regime where they are applicable. In particular, we shall derive analytical expressions for the local density of states (LDOS) and the second-order weakly nonlinear dc conductance, which are the two interesting quantities for weakly nonlinear regime.

In both SMT[3] and response theory,[7] LDOS plays a very important role. From our NEGF theory, LDOS can be easily derived from the right-hand side of Eq. (5), which is the charge density, with the help of Eq. (6). Here, we shall present the explicit expression at the lowest order[7] expansion in the external voltage. Hence, we seek the solution of $U(\mathbf{r})$ in the following form:

$$U = U_{\rm eq} + \sum_{\alpha} u_{\alpha} V_{\alpha} + \frac{1}{2} \sum_{\alpha\beta} u_{\alpha\beta} V_{\alpha} V_{\beta} + \cdots, \tag{10}$$

where $U_{\rm eq}$ is the equilibrium potential and $u_{\alpha}(\mathbf{r})$, $u_{\alpha\beta..}(\mathbf{r})$ are the characteristic potentials.[3,7,6] It can be shown that the characteristic potential satisfy the following sum rules:[3,7,6]

$$\sum_{\alpha} u_{\alpha} = 1 \tag{11}$$

and

$$\sum_{\gamma \epsilon \beta} u_{\alpha\{\beta\}_l} = 0. \tag{12}$$

Here, the subscript $\{\beta\}_l$ is a short notation of l indices γ, δ, η,.... Expanding $G^<$ of Eq. (5) in power series of V_{α}, we can derive the equations for all the characteristic potentials. In particular, the expansions are facilitated by the Dyson equation to the appropriate order (in the absence of the exchange and correlation effect):

$$G^r = G_0^r + G_0^r \Big(qU - qU_{\rm eq} + \sum_{\alpha} [\Sigma_{\alpha}^r(E - qV_{\alpha}) - \Sigma_{\alpha}^r(E)] \Big) G_0^r + \cdots \tag{13}$$

with G_0^r the equilibrium retarded Green's function, i.e., when $U = U_{\rm eq}$. At the lowest order, we thus obtain

$$\begin{aligned} -\nabla^2 u_{\alpha}(\mathbf{r}) = 2q^2 & \int dE f (G_0^r u_{\alpha} G_0^r \Gamma G_0^a + \text{c.c.})_{rr} \\ & - 2q^2 \int dE f (G_0^r \partial_E \Sigma_{\alpha}^r G_0^r \Gamma G_0^a + \text{c.c.})_{rr} \\ & - 2q^2 \int dE [G_0^r (\Gamma_{\alpha} \partial_E f + \partial_E \Gamma_{\alpha} f) G_0^a]_{rr}, \end{aligned} \tag{14}$$

where Γ_{α} and f no longer depend on voltage after the expansion; and c.c. denotes complex conjugate.

The first term on the right-hand side of Eq. (14), which depends on internal potential u_{α}, describes the induced charge density in the conductor. It can be simplified using the fact $i\Gamma = (G_0^r)^{-1} - (G_0^a)^{-1}$, hence it becomes $4\pi q^2 \Sigma_{r'} \Pi_{rr'} u_{\alpha}(r')$, where Π is the Lindhard function[24,7] defined as $\Pi_{rr'} = -i \int (dE/2\pi) f (G_{0rr'}^r G_{0r'r}^r - G_{0rr'}^a G_{0r'r}^a)$. The second and third term of Eq. (14) which do not depend on characteristic potential correspond to the charge density due to external injection. They are the local partial density of states (LPDOS) $dn_{\alpha}(\mathbf{r})/dE$ called injectivity in the language of the scattering matrix,[3]

$$\begin{aligned} dn_{\alpha}(x)/dE = & - \int (dE/2\pi) [G_0^r \Gamma_{\alpha} \partial_E f G_0^a] \\ & - \int (dE/2\pi) f [G_0^r \partial_E \Gamma_{\alpha} G_0^a + G_0^r \partial_E \Sigma_{\alpha}^r G_0^r \Gamma G_0^a \\ & + G_0^r \Gamma G_0^a \partial_E \Sigma_{\alpha}^r G_0^a] \\ = & - \int (dE/2\pi) f [G_0^r (G_0^r \Gamma_{\alpha} + \Gamma_{\alpha} G_0^a) G_0^a] \\ & + \int (dE/2\pi) f [G_0^r (\partial_E \Sigma^r G_0^r \Gamma_{\alpha} + \Gamma_{\alpha} G_0^a \partial_E \Sigma^a \\ & - \partial_E \Sigma_{\alpha}^r G_0^r \Gamma - \Gamma G_0^a \partial_E \Sigma_{\alpha}^a) G_0^a]. \end{aligned} \tag{15}$$

Comparing this result with that derived by SMT,[5] the SMT result[5] corresponds to the first term on the right-hand side of Eq. (15). Hence, the local partial density of states obtained from our general theory is slightly different from that defined in SMT.[5] However, it can be proven that the difference, e.g., the second integral of Eq. (15), becomes negligible in the limit of large scattering volume. The proof follows the approach detailed in our earlier work Ref. 25.

We obtain LDOS $dn(\mathbf{r})/dE$ from Eq. (15):

$$\begin{aligned} dn(\mathbf{r})/dE = & \sum_{\alpha} dn_{\alpha}/dE \\ = & - \int (dE/2\pi) f [G_0^r (G_0^r \Gamma + \Gamma G_0^a) G_0^a] \\ = & \, i \int (dE/2\pi) f (G_0^r G_0^r - G_0^a G_0^a). \end{aligned} \tag{16}$$

This is exactly the same as the LDOS we obtained from the response theory.[7] The agreement is actually not surprising: because we are dealing with dc transport where there is time reversal symmetry, the result of NEGF should be the same as

that of the linear response. Our result of LDOS exactly satisfies the general relationship:[24,3] $\Sigma_{r'}\Pi_{r,r'}=dn(\mathbf{r})/dE$.

As a second comparison to the results obtainable from SMT and response theory, we now derive a formula for the second-order nonlinear conductance $G_{\alpha\beta\gamma}$ from our theory of Sec. II. In weakly nonlinear regime, the electric current can be expanded in a series form in terms of external bias voltages,

$$J_\alpha=\sum_\beta G_{\alpha\beta}V_\beta+\sum_{\beta\gamma} G_{\alpha\beta\gamma}V_\beta V_\gamma + \sum_{\beta\gamma\delta} G_{\alpha\beta\gamma\delta}V_\beta V_\gamma V_\delta+\cdots \tag{17}$$

To compute the second-order nonlinear conductance $G_{\alpha\beta\gamma}$, it is enough to calculate u_α for the internal potential. Expanding the general expression for the current Eq. (7) to the second order in voltage, it is straightforward to obtain the second-order nonlinear conductance:

$$\begin{aligned}G_{\alpha\beta\gamma}=&-q^3\int (dE/2\pi)Tr[\partial_E\Gamma_\alpha G_0^r(\Gamma\delta_{\alpha\gamma}-\Gamma_\gamma)G_0^a\partial_E f\delta_{\alpha\beta}\\ &-\Gamma_\alpha G_0^r(u_\beta-\partial_E\Sigma_\beta^r)G_0^r(\Gamma\delta_{\alpha\gamma}-\Gamma_\gamma)G_0^a\partial_E f\\ &+\Gamma_\alpha G_0^r\partial_E\Gamma_\gamma G_0^a\partial_E f\delta_{\alpha\beta}-\Gamma_\alpha G_0^r\partial_E\Gamma_\gamma G_0^a\partial_E f\delta_{\beta\gamma}\\ &+(1/2)\Gamma_\alpha G_0^r(\Gamma\delta_{\alpha\gamma}-\Gamma_\gamma)G_0^a\partial_E^2 f\delta_{\beta\gamma}\\ &-\Gamma_\alpha G_0^r(\Gamma\delta_{\alpha\gamma}-\Gamma_\gamma)G_0^a(u_\beta-\partial_E\Sigma_\beta^a)G_0^a\partial_E f].\end{aligned} \tag{18}$$

This result is gauge invariant as one can explicitly verify that it satisfies the gauge invariant condition[4] $\Sigma_\beta(G_{\alpha\beta\gamma}+G_{\alpha\gamma\beta})=0$. This result agrees with that derived from SMT[4] if we neglect the terms involving $\partial_E\Gamma$ and $\partial_E\Sigma$. In that case, we obtain

$$\begin{aligned}G_{\alpha\beta\gamma}=&q^3\int (dE/2\pi)Tr[(G_0^a\Gamma_\alpha G_0^r u_\beta G_0^r\\ &+G_0^a u_\beta G_0^a\Gamma_\alpha G_0^r-1/2G_0^a\Gamma_\alpha G_0^r G_0^r\delta_{\beta\gamma}\\ &-1/2G_0^a G_0^a\Gamma_\alpha G_0^r\delta_{\beta\gamma})(\Gamma\delta_{\alpha\gamma}-\Gamma_\gamma)]\partial_E f\\ =&2\pi q^2\int dE(-\partial_E f)[1/2q\partial_E A_{\alpha\beta}\delta_{\beta\gamma}+\partial_{V_\beta}A_{\alpha\gamma}],\end{aligned} \tag{19}$$

which agrees exactly with the result in Ref. 4. The second-order nonlinear conductance has been investigated numerically for several systems using the scattering approach.[26] For detailed discussion of the numerical technique needed for the calculation, see Ref. 26.

Finally, we comment that by expanding Eq. (5) to higher order in terms of voltage, it is straightforward to show that the nonlinear characteristic potentials $u_{\alpha\beta\ldots}$ satisfy a Poisson equation similar to Eq. (5) with different source terms $dn_{\alpha\beta\ldots}/dE$ which correspond to nonlinear LDOS. The higher order nonlinear coefficient $G_{\alpha\beta\ldots}$ can be obtained in similar fashion as we have done here for the second-order coefficient.

B. *I–V* curve in the wideband limit at zero temperature

In the last subsection, we examined the limit of weak nonlinearity. However, the main advance we have obtained from the gauge-invariant NEGF formalism developed in Sec. II is to be able to predict the full nonlinear current–voltage ($I-V$) curves. Several analysis will be presented in this and the next subsections for the $I-V$ curves.

In the commonly used wide-band limit,[19] the coupling matrix Γ is independent of energy which drastically simplifies the algebra. The wide-band limit corresponds to cases where the probes have no feature, thus the internal potential $U(\mathbf{r})$ becomes just a space-independent constant U_0 (the value of U_0 depends on the voltages $\{V_\alpha\}$). In wide-band limit the steady state Green's function takes a very simple form, $G_0^r=1/(E-E_0+i\Gamma/2)$, thus, the integral in Eq. (7) can be done exactly. We obtain

$$J_\alpha=-\frac{q}{\pi\Gamma}\sum_\beta (\Gamma\delta_{\alpha\beta}-\Gamma_\beta)\Gamma_\alpha \times\arctan\left(\frac{\Delta E-qU_0+qV_\beta}{\Gamma/2}\right), \tag{20}$$

where $\Delta E=E_F-E_0$ and the constant U_0 is determined by the charge conservation condition $\int dETr[G^<(E,V)]=\int dETr[G^<(E,0)]$, i.e.,

$$\sum_\beta \Gamma_\beta\arctan\left(\frac{\Delta E-qU_0+qV_\beta}{\Gamma/2}\right)=\Gamma\arctan\left(\frac{\Delta E}{\Gamma/2}\right). \tag{21}$$

To obtain this equation, the quasineutrality approximation[5] is assumed which neglects the charge polarization in the system in addition to the use of total charge neutrality. The gauge-invariant condition in Eq. (20) is clearly satisfied: raising both V_β and U_0 by the same amount does not alter the current. Equations (20) and (21) have been obtained before[4] from SMT where a Breit–Wigner form of the scattering matrix is assumed. Hence, we may conclude that in this sense the wide-band limit in NEGF approach is equivalent to the Breit–Wigner approximation in the scattering matrix approach. It is, however, different from that derived in Ref. 16 where a linear charging model is used for the internal potential buildup. It is not difficult to confirm that the result of Ref. 16 is recovered if we solve for U_0 in Eq. (21) to the first order in voltage V, i.e., we compute the internal potential U_0 by neglecting the contributions of higher order characteristic potentials $u_{\alpha\beta}\ldots$. In this limit, we obtain $U_0=\Sigma_\alpha(\Gamma_\alpha/\Gamma)V_\alpha$. Substitute this into Eq. (20) we arrive at the result of Ref. 16. This exercise also allows us to identify the phenomenological parameter C_i of Ref. 16 to be Γ_i, and it indicates that the linear charging model for the internal potential is not complete for the full nonlinear $I-V$ curve predictions.

Next, let's derive the full nonlinear $I-V$ curve for a quantum dot with two resonant levels. For two resonant levels in a quantum dot, the retarded Green's function[27] can be derived to have the following expression:

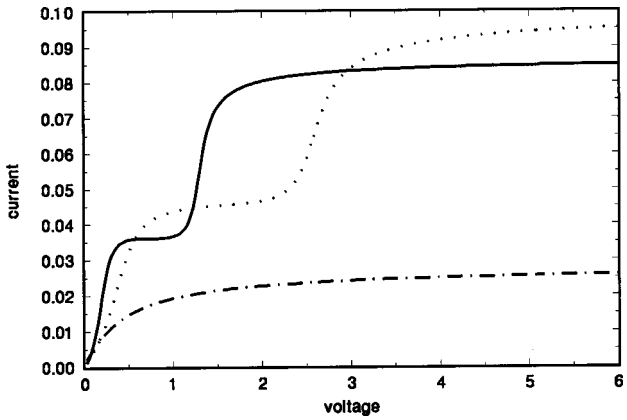

FIG. 1. Current vs the voltage for a resonant tunneling structure with two resonant levels for self-consistent solution and non-self-consistent solution with two different voltage gauges. Solid line: non-self-consistent solution for $V_1=V$ and $V_2=0$; dotted line: non-self-consistent solution for $V_1=V/2$ and $V_2=-V/2$; dot-dashed line: self-consistent solution. The system parameters are $\Gamma_1=\Gamma_2=0.1$, $E_1=0.3$, $E_2=1.4$, and $E_F=0.11$.

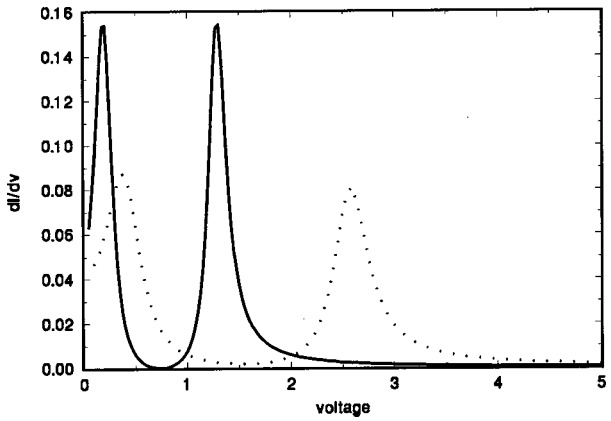

FIG. 2. Corresponding differential conductance for the non-self-consistent solutions of Fig. 1. Solid line: non-self-consistent solution for $V_1=V$ and $V_2=0$; dotted line: non-self-consistent solution for $V_1=V/2$ and $V_2=-V/2$.

$$G^r=\frac{1}{[1/(E-E_1-qU)+1/(E-E_2-qU)]^{-1}+i\Gamma/2}, \tag{22}$$

where E_1 and E_2 are energies of the two resonant levels. From Eq. (7), we obtain the current is

$$J_\alpha=\frac{q}{2\pi\Gamma}\sum_\beta\ (\Gamma\delta_{\alpha\beta}-\Gamma_\beta)\Gamma_\alpha \times \mathrm{Im}\left[\ln(a_\beta^2-b^2)+\frac{i\Gamma}{2b}\ln\left(\frac{a_\beta+b}{a_\beta-b}\right)\right], \tag{23}$$

where $a_\beta=E_F+qV_\beta-qU_0-(E_1+E_2)/2+i\Gamma/2$ and $b^2=(E_1-E_2)^2/4-\Gamma^2/4$. In the quasineutrality approximation, we derive that the internal potential U_0 is determined by following equation:

$$\sum_\beta\ \Gamma_\beta\,\mathrm{Im}\left[\ln(a_\beta^2-b^2)+\frac{i\Gamma}{2b}\ln\left(\frac{a_\beta+b}{a_\beta-b}\right)\right] =\Gamma\ \mathrm{Im}\left[\ln(a^2-b^2)+\frac{i\Gamma}{2b}\ln\left(\frac{a+b}{a-b}\right)\right], \tag{24}$$

where $a=E_F-(E_1+E_2)/2+i\Gamma/2$.

Figure 1 plots the predicted $I-V$ curve Eq. (23) with the parameter $\Gamma_1=\Gamma_2=0.1$, $E_1=0.3$, $E_2=1.4$, and $E_F=0.11$. To compare the $I-V$ curves with and without gauge invariance, in Fig. 1 we have plotted three curves. The dot-dashed line represents the gauge-invariant solution Eq. (23). Since this solution is gauge invariant, we choose $V_1=V$ and $V_2=0$ in Fig. 1. Both the other two lines (solid and dotted) are for $U_0=0$ thus no internal potential is taken into account self-consistently. Two observations warrant to be discussed. First, the two non-self-consistent $I-V$ curves, solid line with $V_1=V$ and $V_2=0$ and dotted line with $V_1=V/2$ and $V_2=-V/2$, give different $I-V$ curves. This is clearly wrong because electric current must only depend on the bias voltage difference which is V for both curves, and not on the choice of the reference point for potential. This is a direct consequence of the flaw of a non-self-consistent theory. Second, the qualitative behavior of the current versus voltage curve is different. Both curves of the non-self-consistent analysis show quantized steps, which the self-consistent analysis with quasineutrality approximation does not give. This difference in qualitative behavior can be understood as due to the quasineutrality approximation. The electric current for the incident electrons with energy E_F will increase sharply when the chemical potential $\mu=E_F+qV$ is close to the first resonant energy level E_1. When the internal potential build-up is not included in the non-self-consistent analysis, this current saturates after crossing the first level but increases again when μ is near the second resonant level. This is actually a reasonable picture. In the self-consistent solution within quasineutrality approximation, however, the internal energy U_0 solved from quasineutrality condition Eq. (24) increases linearly $V=\kappa U_0$ with the voltage with coefficient κ close to one. Hence, as the chemical potential rises, the resonant level also increases with approximately the same amount. Thus, the second resonance level will not be reached (within the quasi-neutrality approximation) for the range of voltages of Fig. 1. Finally, In Fig. 2, we have plotted the differential conductance dI/dV of two non-self-consistent solutions. For the case $V_1=V$ and $V_2=0$ (solid line), two peaks show up near $V=0.2$ and $V=1.3$. Since the Fermi level in the equilibrium is $E_F=0.11$, those two peaks reflect the resonant behavior when the chemical potential qV_1+E_F is in line with two resonant levels $E_1=0.3$ and $E_2=1.4$. When $V_1=V/2$ and $V_2=-V/2$ (dotted line), the chemical potential is again $qV_1+E_F=qV/2+E_F$, so we found two peaks at $V=0.38$ and $V=2.6$. However, the spacing between two peaks are different for two different choice of voltage V_1 and V_2: it is therefore important to include the Coulomb interaction so that the theory is gauge invariant.

Our results strongly suggest that as far as the $I-V$ curve prediction is concerned, without interaction one violates gauge invariance, but including interaction within the quasineutrality approximation is still not enough as it misses the expected resonance levels in the $I-V$ curve, which are often observed in experimental situations.[28] Hence, it maybe necessary to go beyond the quasineutrality approximation.

C. Exchange and correlation effect

So far the electron-electron interaction which gives rise to the internal potential buildup has been treated within the Hartree approximation with the quasineutrality condition. In this subsection, we examine the effects of exchange and correlation to the nonlinear $I-V$ curves within the wide-band limit for a resonant tunneling structure. We must also go beyond the quasineutrality approximation.

Going beyond quasineutrality approximation means that we must consider the local charge distribution under the condition of overall charge neutrality. For a double-barrier tunneling structure, let's introduce capacitance coefficients C_1 and C_2 for the left and the right barrier, respectively. Then, the charge in the quantum well due to Coulomb interaction is given by[29]

$$\Delta Q = -i\int (dE/2\pi)[G^{<}(E,U_0)-G_0^{<}]$$
$$= C_1(U_0-V_1)+C_2(U_0-V_2), \tag{25}$$

where ΔQ is the total charge in the well, U_0 is the overall shift of the band bottom of the well due to the Coulomb interaction, and $G_0^{<}$ is the equilibrium lesser Green's function. For the system with only one resonant level, this equation reduces to

$$\sum_{\beta} \Gamma_{\beta} \arctan\left(\frac{E_F-H_0+qV_{\beta}}{\Gamma/2}\right) - \Gamma \arctan\left(\frac{E_F-H_0}{\Gamma/2}\right)$$
$$= [C_1(U_0-V_1)+C_2(U_0-V_2)]\pi\Gamma/q, \tag{26}$$

where the wideband limit is assumed and $H_0=E_0+qU_0+qV_{xc}$. The current is given by

$$J_{\alpha} = -\frac{q}{\pi\Gamma}\sum_{\beta} (\Gamma\delta_{\alpha\beta}-\Gamma_{\beta})\Gamma_{\alpha}\arctan\left(\frac{E_F-H_0+qV_{\beta}}{\Gamma/2}\right). \tag{27}$$

Here, we note that Eqs. (21) and (26) are equations to determine the internal potential U_0 which is needed in order that expressions for current in Eqs. (20) and (27) to be gauge invariant.

To plot the $I-V$ curve determined by Eqs. (26) and (27), we use $V_{xc}=-1.5\alpha\Delta Q^{1/3}$ in the Hartree–Fock–Slater approximation,[22,30] where $2/3\leqslant\alpha\leqslant 1$. Parametrized by the coupling constants Γ_i and the geometrical capacitance coefficients C_i, we thus can calculate the nonlinear $I-V$ curves from Eqs. (26) and (27). Figure 3 presents current as a function of voltage difference $V=V_1-V_2$. We have compared the cases with or without exchange and correlation potential V_{xc} for two different set of system parameters: symmetric barrier with $\Gamma_1=\Gamma_2=0.5$, $C_1=C_2=5.0$, $\Delta E=-1.0$, $\alpha=0.7$ (dotted line with V_{xc}, solid line without); and asymmetric barrier with $\Gamma_1=0.1$, $\Gamma_2=0.5$, $C_1=1.0$, $C_2=5.0$, $\Delta E=-1.0$, $\alpha=0.7$ (dashed line with V_{xc} and dot-dashed line without). We observe that the current for the symmetric barrier is much larger than that of the asymmetric barrier for $V>2$, but it can be smaller for smaller bias between $1<V<2$, and becomes larger again at very small $V<1$. Without the internal potential buildup taken into account, it is well known that symmetrical tunneling barriers have larger transmission coefficients hence larger current than those of asymmetrical barriers. The behavior of the $I-V$ curves in Fig. 3 at the $V\rightarrow 0$ limit is consistent with this picture. However, at larger voltages, this expectation may or may not be true, due to the nonlinear effects and the internal potential buildup. We found that for both symmetrical and asymmetrical barriers, the exchange and correlation effects are to increase the electric current. This is seen more clearly from the differential conductance dI/dV versus voltage in Fig. 4. Since the exchange and correlation term V_{xc} is to lower the bottom of the conduction band, the peak of dI/dv shifts to the small voltage as a result.

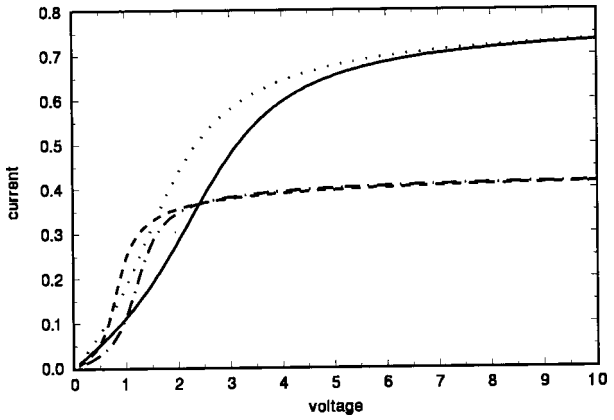

FIG. 3. Gauge invariant current vs the voltage for a resonant tunneling structure with one resonant level. Solid line (V_{xc} not included) and dotted line (V_{xc} included): $\Gamma_1=\Gamma_2=0.5$, $C_1=C_2=5.0$, $\Delta E=-1.0$, $\alpha=0.7$; dot-dashed line (V_{xc} not included) and dashed line (V_{xc} included): $\Gamma_1=0.1$, $\Gamma_2=0.5$, $C_1=1.0$, $C_2=5.0$, $\Delta E=-1.0$, $\alpha=0.7$.

When exchange and correlation effects are included in the two-level tunneling system, Eqs. (23) and (24) are modified in a similar way as Eqs. (27) and (26). Figure 5 shows the $I-V$ curve for two-level system for $\Gamma_1=\Gamma_2=0.1$, $C_1=C_2=1.0$, $E_1=0.3$, $E_2=1.4$, $E_F=0.11$, and $\alpha=0.7$. The $I-V$ curves with (dotted line) and without (solid line) V_{xc} are plotted for comparison, but both $I-V$ curves now show two steps reflecting the two resonance levels for tunneling. Hence, by going beyond the quasineutrality approximation, the gauge-invariant theory developed in Sec. II predicts a

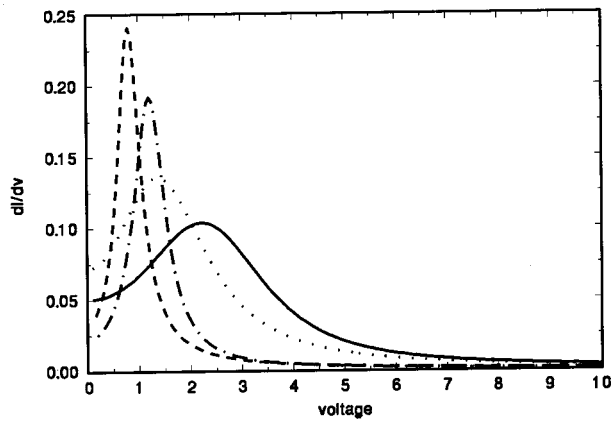

FIG. 4. Differential conductance of a resonant tunneling structure with one resonant level. The parameters are the same as that of Fig. 3.

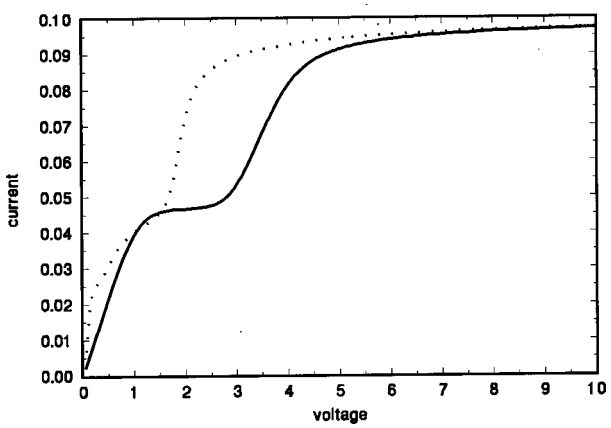

FIG. 5. Gauge invariant current vs the voltage for a resonant tunneling structure with two resonant levels. Solid line (V_{xc} not included) and dotted line (V_{xc} included): $\Gamma_1=\Gamma_2=0.1$, $C_1=C_2=1.0$, $E_1=0.3$, $E_2=1.4$, $E_F=0.11$.

''quantized'' $I-V$ curve which, as mentioned above, is physically reasonable. Again, when V_{xc} is present, the current increases.

In Fig. 6, we show the differential conductance dI/dV for the same system parameters as that of Fig. 5. When the exchange and correlation potential is included (solid line), it is surprising to observe that there are three peaks in dI/dV instead of two (dotted line without V_{xc}). In addition, the peaks of dI/dV are shifted towards smaller values of bias as compared to the dI/dV curve for $V_{xc}=0$. The entire behavior of dI/dV can be understood as the following. The internal Coulomb potential U_0 tends to move the resonant level up (to higher energy) and the exchange and correlation potential V_{xc} tends to lower it down, these two effects give compensating factors to move the resonance levels inside the quantum well. At very small voltage, $|V_{xc}|$ increases much faster than the internal potential U_0 does as the bias is increased (see inset of Fig. 6), thus the resonant level E_1 moves downwards in energy from its ''bare'' value $E_1=0.3$: a resonance peak is expected when it is lowered to the chemical potential. As the voltage increases such that the level E_1 is lowered to below the chemical potential, dI/dV decreases from the resonance peak consistent with the fact of going off resonance. When the voltage increases further, the resonant levels E_1 and E_2 start to move upward in energy due to the effect of U_0, and it will again pass the chemical potential resulting to the resonance peak near $V=0.5$. Finally, when the voltage is around $V=2.0$, the chemical potential is near the second resonant level and we see a sharp increase of dI/dV and the third peak shows up as a result. We conclude that the quasineutrality condition may need to be extended in predicting $I-V$ curves when the system is in the tunneling regime. Here, we have used a phenomenological but nonlinear capacitance charging model to include the charge polarization effects, which are seen to play an important role in predicting quantized $I-V$ curves. It is further found that the exchange-correlation potential V_{xc} can be quite important as it provides a compensating effect to the Hartree internal potential U_0.

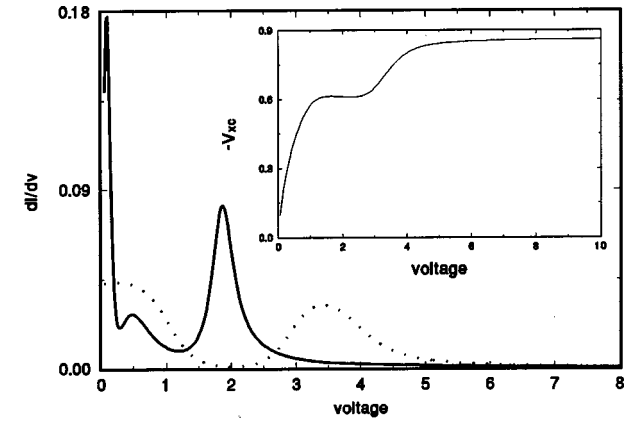

FIG. 6. Differential conductance of a resonant tunneling structure with two resonant levels. The parameters are the same as that of Fig. 5. Inset: the exchange and correlation potential V_{xc} vs voltage of a resonant tunneling structure with two resonant levels.

D. $I-V$ curve with Hubbard U term

To maintain the gauge invariance, we have so far considered the Coulomb interaction in the Hartree approximation with or without an exchange-correlation term. In this subsection, we will consider the on-site Coulomb interaction in terms of Hubbard U model with the following Hamiltonian for $H_{\rm cen}$ in Eq. (1):

$$H_{\rm cen}=\sum_{\sigma} E_0 d_\sigma^\dagger d_\sigma + U_1 n_\uparrow n_\downarrow . \tag{28}$$

Here, we assume that the quantum dot contains one energy level E_0 with Coulomb repulsion energy U_1 which accounts for the interaction between different spins. In addition to U_1 we assume that the long range Coulomb potential U between different sites gives an overall a constant shift U_0 to the bottom of conduction band. This is similar in spirit to the energy shift Δ introduced in Refs. 19 and 31. However, in our case, U_0 has to be determined self-consistently. Equation (28) has been used by many authors before,[32,9] but the gauge invariant condition has not not been considered there.

The current is still determined by Eq. (2). But now the lesser Green's function is given by[33]

$$G_\sigma^<(E,U)=-[G_\sigma^r-G_\sigma^a]\sum_\beta \Gamma_\beta f_\beta/\Gamma, \tag{29}$$

where[34–36]

$$G_\sigma^r(E,U)=\frac{\langle n_{\bar\sigma}\rangle}{E-E_0-U_1-qU+i\Gamma/2} + \frac{1-\langle n_{\bar\sigma}\rangle}{E-E_0-qU+i\Gamma/2} \tag{30}$$

and

$$\langle n_\sigma\rangle=-i\int (dE/2\pi)G_\sigma^< . \tag{31}$$

The internal Coulomb potential U can be determined in terms of the geometrical capacitances C_1 and C_2,

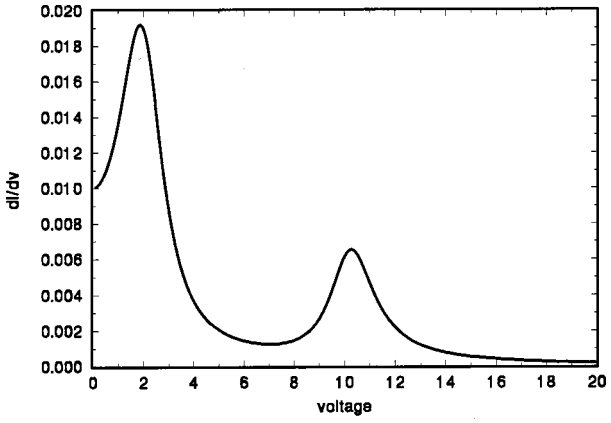

FIG. 7. Differential conductance of a resonant tunneling structure with Hubbard U term. The parameters are: $\Gamma_1=\Gamma_2=0.5$, $C_1=C_2=5.0$, $\Delta E=-1.0$, and $U_1=4.0$.

$$\int (dE/2\pi)\sum_\sigma G_\sigma^<(E,U)-\int (dE/2\pi)\sum_\sigma G_{\sigma 0}^< = C_1(U_0-V_1)+C_2(U_0-V_2) \tag{32}$$

from Eq. (30), this condition becomes

$$\sum_\beta \Gamma_\beta \sum_\sigma \langle n_{\bar\sigma}\rangle \arctan\left(\frac{\Delta E-U_1-qU+qV_\beta}{\Gamma/2}\right) + \sum_\beta \Gamma_\beta \sum_\sigma (1-\langle n_{\bar\sigma}\rangle)\arctan\left(\frac{\Delta E-qU+qV_\beta}{\Gamma/2}\right) - \Gamma\sum_\sigma \langle n_{\bar\sigma}\rangle \arctan\left(\frac{\Delta E-U_1}{\Gamma/2}\right) - \Gamma\sum_\sigma (1-\langle n_{\bar\sigma}\rangle)\arctan\left(\frac{\Delta E}{\Gamma/2}\right) = [C_1(U_0-V_1)+C_1(U_0-V_1)]2\pi\Gamma/q. \tag{33}$$

With the potential U determined this way, the current is finally written as

$$J_\alpha = -\frac{q}{\pi\Gamma}\sum_\beta (\Gamma\delta_{\alpha\beta}-\Gamma_\beta)\Gamma_\alpha \times\left[\sum_\sigma \langle n_{\bar\sigma}\rangle \arctan\left(\frac{\Delta E-U_1-qU+qV_\beta}{\Gamma/2}\right) + \sum_\sigma (1-\langle n_{\bar\sigma}\rangle)\arctan\left(\frac{\Delta E-U_1+qV_\beta}{\Gamma/2}\right)\right] \tag{34}$$

which can be calculated numerically using Eqs. (31), (33), and (34). In Fig. 7, we have plotted the differential conductance versus voltage for $\Gamma_1=\Gamma_2=0.5$, $C_1=C_2=5$, $\Delta E=-1.0$, and $U_1=4.0$. As expected, there are two peaks corresponding to two different energies E_0 and E_0+U_1. Since the Coulomb interaction U increases linearly with voltage $2U\approx V$, the separation between the two peaks becomes $2U_1$.

IV. SUMMARY

In this work, we have developed a general gauge-invariant nonlinear dc transport theory based on the nonequilibrium Green's functions. The main idea of this development is to self-consistently couple the NEGF with the proper Poisson equation for the internal potential buildup inside the mesoscopic conductor. It is the consideration of the internal potential distribution which has made the NEGF theory gauge invariant. At various limiting cases, our theory predicts results consistent with those of scattering matrix theory and response theory, but our theory allows a general treatment of the full nonlinear dc transport regime which are, perhaps, impossible for the other formalisms. The present theory is natural to allow the inclusion of exchange and correlation potential within the density functional formalism. Hence, it is appropriate for transport in multiprobe conductors in the quantum coherent regime, and we have applied it to the analysis of resonant tunneling with one and two resonance levels. Our results clearly show that without self-consistent analysis the predicted current would depend on the choice of potential zero, which is wrong. For the tunneling device, our analysis also indicated the importance of including charge polarization effect. This effect can be considered using the phenomenological model involving capacitance coefficients, as done here; or it can be included through numerical solutions of a charging model as carried out in Ref. 23. Finally, we found that in general a larger current is obtained when exchange and correlation effects are included into the analysis.

Many further applications of the present formalism can be made. An important further development is to abandon the wide-band limit. In this work, we have used this limit in order to derive analytical formula, but one can go beyond this limit in numerical calculations. In the wide-band limit, the coupling matrix is independent of energy. A consequence, as we observe from the $I-V$ curves, is that the current increases monotonically with bias (no peaks in the $I-V$ curve itself). Hence, the negative differential resistance (NDR) can not be observed. To overcome this limitation, Jauho *et al.*[19] have introduced a lower energy cutoff to allow a finite occupied bandwidth of the contact. This modification allowed NDR to appear, but the current at large voltage did not agree with experimental results. Therefore, to obtain NDR quantitative correctly, we must go beyond the wide-band limit. This can be done within our formalism using the numerical method developed by McLennan *et al.*[37,2]

ACKNOWLEDGMENTS

The authors gratefully acknowledge support by a RGC grant from the SAR Government of Hong Kong under Grant No. HKU 7115/98P, and a CRCG grant from the University of Hong Kong. H. G. is supported by NSERC of Canada and FCAR of Québec. The authors thank the computer center of HKU for computational facilities.

[1] R. Landauer, IBM J. Res. Dev. **1**, 233 (1957); M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, Phys. Rev. B **31**, 6207 (1985).
[2] S. Datta, *Electronic Transport in Mesoscopic Systems*, (Cambridge University Press, New York, 1995).

[3] M. Büttiker, J. Phys.: Condens. Matter **5**, 9361 (1993).
[4] T. Christen and M. Büttiker, Europhys. Lett. **35**, 523 (1996).
[5] For a review, see M. Büttiker and T. Christen, in *Quantum Transport in Semiconductor Submicron Structures*, edited by B. Kramer (Kluwer, Dordrecht, 1996), p. 263.
[6] Z. S. Ma, J. Wang, and H. Guo, Phys. Rev. B **57**, 9108 (1998).
[7] Z. S. Ma, J. Wang, and H. Guo, Phys. Rev. B **59**, 7575 (1999).
[8] Y. Meir and N. S. Wingreen, Phys. Rev. Lett. **68**, 2512 (1992); Y. Meir, N. S. Wingreen, and P. A. Lee, *ibid.* **70**, 2601 (1993).
[9] T. K. Ng, Phys. Rev. Lett. **70**, 3635 (1993).
[10] L. Y. Chen and C. S. Ting, Phys. Rev. B **43**, 2097 (1991).
[11] S. Hershfield, J. H. Davies, and J. W. Wilkins, Phys. Rev. B **46**, 7046 (1992).
[12] C. Bruder and H. Schoeller, Phys. Rev. Lett. **72**, 1076 (1994).
[13] S. König, H. Schoeller, and G. Schön, Phys. Rev. Lett. **76**, 1715 (1996).
[14] J. C. Cuevas, A. L. Yeyati, and A. Martin-Rodero, Phys. Rev. Lett. **80**, 1066 (1998).
[15] B. G. Wang, J. Wang, and H. Guo, Phys. Rev. Lett. **82**, 398 (1999).
[16] C. A. Stafford, Phys. Rev. Lett. **77**, 2770 (1996).
[17] P. J. Price, Superlattices Microstruct. **20**, 253 (1996).
[18] In a recent paper, T. Gramespacher and M. Büttiker, Phys. Rev. B **56**, 13026 (1997), the authors discussed the relationship between scattering matrix theory and the Hamiltonian approach in which the transmission coefficient is expressed in terms of Green's function.
[19] A. P. Jauho, N. S. Wingreen, and Y. Meir, Phys. Rev. B **50**, 5528 (1994).
[20] M. P. Anantram and S. Datta, Phys. Rev. B **51**, 7632 (1995).
[21] For detailed discussion on the construction of the Green's functions and the self-energy in real space, see Ref. 2, Chap. 3.
[22] The exchange and correlation effect has been included in the ac emittance formulation in the linear regime by Büttiker. See M. Büttiker, in *Quantum Dynamics of Submicron Structures*, edited by H. A. Cerdeira *et al.* (Kluwer, The Netherlands, 1995), p. 657.
[23] J. Wang *et al.*, Phys. Rev. Lett. **80**, 4277 (1998).
[24] I. B. Levinson, Sov. Phys. JETP **68**, 1257 (1989); Zh. Eksp. Teor. Fiz. **95**, 2175 (1989).
[25] Q. R. Zheng, J. Wang, and H. Guo, Phys. Rev. B **56**, 12462 (1997); J. Wang, Q. R. Zheng, and H. Guo, *ibid.* **55**, 9770 (1997).
[26] M. K. Yip, J. Wang, and H. Guo, Z. Phys. B: Condens. Matter **104**, 463 (1997); W. D. Sheng, J. Wang, and H. Guo, J. Phys. C **10**, 5335 (1998); W. D. Sheng, Q. R. Zheng, J. Wang, and H. Guo, Phys. Rev. B **59**, 538 (1999).
[27] Q. F. Sun and Z. H. Lin (unpublished).
[28] C. T. Black, D. C. Ralph, and M. Tinkham, Phys. Rev. Lett. **76**, 688 (1996).
[29] M. H. Pedersen and M. Büttiker, Phys. Rev. B **58**, 12993 (1998); Y. M. Blanter and M. Büttiker, *ibid.* **59**, 10217 (1999).
[30] G. P. Srivastava and D. Weaire, Adv. Phys. **36**, 463 (1987).
[31] X. Q. Li and Z. B. Su, Phys. Rev. B **54**, 10807 (1996).
[32] L. I. Glazman and K. A. Matveev, JETP Lett. **48**, 445 (1988); L. Y. Chen and C. S. Ting, Phys. Rev. B **44**, 5916 (1991).
[33] T. K. Ng, Phys. Rev. Lett. **76**, 487 (1996).
[34] P. Pals and A. MacKinnon, J. Phys.: Condens. Matter **8**, 5401 (1996).
[35] H. Haug and A. P. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors* (Springer, Berlin, 1996).
[36] T. Ivanov, Europhys. Lett. **40**, 183 (1997).
[37] M. J. McLennan *et al.*, Phys. Rev. B **43**, 14333 (1991).