Time-dependent transport in interacting and noninteracting resonant-tunneling systems

Antti-Pekka Jauho
Mikroelektronik Centret, Technical University of Denmark, DK-2800 Lyngby, Denmark, and Nordita, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

Ned S. Wingreen
NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540

Yigal Meir
Department of Physics, University of California at Santa Barbara, Santa Barbara, California 93106
(Received 5 April 1994)

We consider a mesoscopic region coupled to two leads under the influence of external time-dependent voltages. The time dependence is coupled to source and drain contacts, the gates controlling the tunnel-barrier heights, or to the gates that define the mesoscopic region. We derive, with the Keldysh nonequilibrium-Green-function technique, a formal expression for the fully nonlinear, time-dependent current through the system. The analysis admits arbitrary interactions in the mesoscopic region, but the leads are treated as noninteracting. For proportionate coupling to the leads, the time-averaged current is simply the integral between the chemical potentials of the time-averaged density of states, weighted by the coupling to the leads, in close analogy to the time-independent result of Meir and Wingreen [Phys. Rev. Lett. **68**, 2512 (1992)]. Analytical and numerical results for the exactly solvable noninteracting resonant-tunneling system are presented. Due to the coherence between the leads and the resonant site, the current does not follow the driving signal adiabatically: a "ringing" current is found as a response to a voltage pulse, and a complex time dependence results in the case of harmonic driving voltages. We also establish a connection to recent linear-response calculations, and to earlier studies of electron-phonon scattering effects in resonant tunneling.

I. INTRODUCTION

The hallmark of mesoscopic phenomena is the phase coherence of the charge carriers, which is maintained over a significant part of the transport process. The interference effects resulting from this phase coherence are reflected in a number of experimentally measurable properties. For example, phase coherence is central to the Aharonov-Bohm effect,[1] universal conductance fluctuations,[1] and weak localization,[2] and can be affected by external controls such as temperature or magnetic field. The study of stationary mesoscopic physics is now a mature field, and in this work we focus on an alternative way of affecting the phase coherence: external *time-dependent* perturbations. The interplay of external time dependence and phase coherence can be phenomenologically understood as follows. If the single-particle energies acquire a time dependence, then the wave functions have an extra phase factor, $\psi \sim \exp[-i\int^t dt'\epsilon(t')]$. For a uniform system such an overall phase factor is of no consequence. However, if the external time dependence is different in different parts of the system, and the particles can move between these regions (without being "dephased" by inelastic collisions), the phase difference becomes important.

The interest in time-dependent mesoscopic phenomena stems from recent progress in several experimental techniques.[3] Time dependence is a central ingredient in many different experiments, of which we mention the following:

(i) *Single-electron pumps and turnstiles.* Here time-modified gate signals move electrons one by one through a quantum dot, leading to a current which is proportional to the frequency of the external signal. These structures have considerable importance as current standards. The Coulombic repulsion of the carriers in the central region is crucial to the operational principle of these devices, and underlines the fact that extra care must be paid to interactions when considering time-dependent transport in mesoscopic systems.

(ii) ac *response and transients in resonant-tunneling devices.* Resonant-tunneling devices have a number of applications as high-frequency amplifiers or detectors. For the device engineer a natural approach would be to model these circuit elements with resistors, capacitances, and inductors. The question then arises as to what, if any, are the appropriate "quantum" capacitances and inductances one should ascribe to these devices. Answering this question requires the use of time-dependent quantum-transport theory.

(iii) *Interaction with laser fields.* Ultrashort laser pulses allow the study of short-time dynamics of charge carriers. Here again, coherence and time dependence

0163-1829/94/50(8)/5528(17)/$06.00

combine with the necessity of treating interactions.

A rigorous discussion of transport in an interacting mesoscopic system requires a formalism that is capable of including explicitly the interactions. Obvious candidates for such a theoretical tool are various techniques based on Green functions. Since many problems of interest involve systems far from equilibrium, we cannot use linear-response methods, such as those based on the Kubo formula, but must use an approach capable of addressing the full nonequilibrium situation. The nonequilibrum-Green-function techniques, as developed about thirty years ago by Kadanoff and Baym,[4] and by Keldysh,[5] have during the recent years gained increasing attention in the analysis of transport phenomena in mesoscopic semiconductors systems.[6] In particular, the *steady-state* situation has been addressed by a large number of papers.[7–12] Among the central results obtained in these papers is that under certain conditions (to be discussed below) a Landauer-type conductance formula[13] can be derived. This is quite appealing in view of the wide spread success of conductance formulas in the analysis of transport in mesoscopic systems.

Considerably fewer studies have been reported where an explicit time dependence is an essential feature. We are aware of an early paper in surface physics,[14] but only in the recent past have groups working in mesoscopic physics addressed this problem.[15–20,38] The work reported in this paper continues along these lines: we give the full details and expand on our short communication.[17]

Our main formal result from the nonequilibrium-Green-function approach is a general expression for the time-dependent current flowing from noninteracting leads to an interacting region. As we will discuss in Sec. II, the time dependence enters through the self-consistent parameters defining the model. We show that under certain restrictions, to be specified below, a Landauer-like formula can be obtained for the *time-averaged* current. To illustrate the utility of our approach we give results for an exactly solvable noninteracting case, which displays an interesting, and experimentally measurable, nonadiabatic behavior. We also establish a link between the present formulation and recently published results for linear-response and electron-phonon interactions, obtained by other techniques.

The paper is organized as follows. We examine in Sec. II the range of experimental parameters in which we expect our theoretical formulation to be valid. In Sec. III we briefly review the physics behind the nonequilibrium-Green-function technique of Keldysh, and Baym and Kadanoff, which is our main theoretical tool, and then introduce the specific model Hamiltonians used in this work. We derive the central formal results for the time-dependent current in Sec. IV. We also derive, under special restrictions, a Landauer-like formula for the average current. In Sec. V, we apply the general formulas to an explicitly solvable resonant-tunneling model. Both analytical and numerical results are presented. We also show that the linear ac-response results of Fu and Dudley[21] are contained as a special case of the exact results of this section. In Sec. VI, we illustrate the utility of our formulation by presenting a much simplified derivation of Wingreen *et al.*[22] results on resonant tunneling in the presence of electron-phonon interactions. Appendix A summarizes some of the central technical properties of the Keldysh technique: we state the definitions, give the basic equations, and provide the analytic continuation rules employed below. In Appendices B and C, we present proofs for certain statements made in the main text, and, finally, in Appendix D we describe some transformations which facilitate numerical evaluation of the time-dependent current.

II. APPLICABILITY TO EXPERIMENTS

A central question one must address is: under which conditions are the nonequilibrium techniques, applied successfully to the steady-state problem, transferrable to time-dependent situations, such as the experiments mentioned above?

The time-dependent problem has to be formulated carefully, particularly with respect to the leads. It is essential to a Landauer type of approach, that the electrons in the leads be noninteracting. In practice, however, the electrons in the leads near the mesoscopic region contribute to the self-consistent potential. We approach this problem by dividing the transport physics in two steps:[23] (i) the self-consistent determination of charge pileup and depletion in the contacts, the resulting barrier heights, and single-particle energies in the interacting region, and (ii) transport in a system defined by these self-consistent parameters. Step (i) requires a capacitance calculation for each specific geometry,[23] and we do not address it in this paper. Instead, we assume the results of (i) as time-dependent input parameters and give a full treatment of the transport through the mesoscopic region (ii). In practice, the interactions in the leads are absorbed into a time-dependent potential and from then on the electrons in the leads are treated as noninteracting. This means that when relating our results to actual experiments some care must be exercised. Specifically, we calculate only the current flowing into the mesoscopic region, while the total time-dependent current measured in the contacts includes contributions from charge flowing in and out of accumulation and depletion regions in the leads. In the *time-averaged* (dc) current, however, these capacitive contributions vanish and the corresponding time-averaged theoretical formulas, such as Eq. (27), are directly relevant to experiment. It should be noted, though, that these capacitive currents may influence the effective time-dependent parameters in step (i) above.

Let us next estimate the frequency limits that restrict the validity of our approach. Two criteria must be satisfied. First, the driving frequency must be sufficiently slow that the applied bias is dropped entirely across the tunneling structure. When a bias is applied to a sample, the electric field in the leads can only be screened if the driving frequency is smaller than the plasma frequency, which is tens of THz in typical doped semiconductor samples. For signals slower than this, the bias is established entirely across the tunneling structure by accumulation and depletion of charge near the barriers. The unscreened Coulomb interaction between net excess

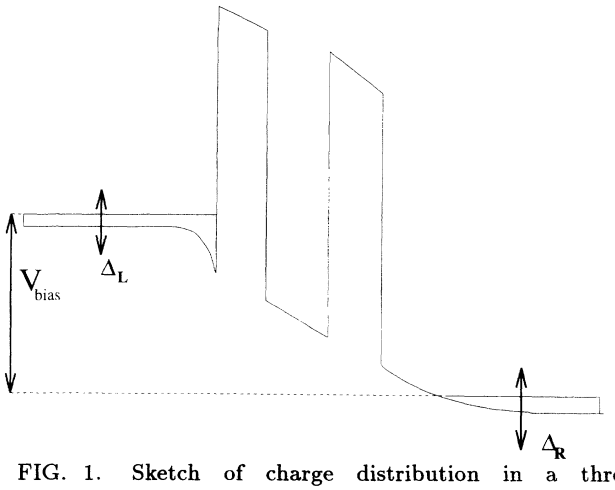

FIG. 1. Sketch of charge distribution in a three dimensional resonant-tunneling device under dc bias $V_{\rm bias} = \mu_L - \mu_R$ with a time modulation of amplitude $\Delta_{L/R}$ superposed on the leads. As argued in the text, only a tiny fraction of charge carriers participates in setting up the voltage drop across the structure.

charge is quite strong, and hence the bias across a tunneling structure is caused by a relatively small excess of charge in accumulation and depletion layers. The formation of these layers then causes a rigid shift [see Eq. (2) below] of the bottom of the conduction band deeper in the leads, which is the origin of the rigid shift of energy levels in our treatment of a time-dependent bias.

The second frequency limit on our approach is that the buildup of electrons required for the formation of the accumulation and depletion layers must not significantly disrupt the coherent transport of electrons incident from the leads. One way to quantify this is to ask—what is the probability that an electron incident from the leads participates in the buildup of charge associated with a time-dependent bias? This probability will be the ratio of the net current density flowing into the accumulation region to the total incident flux of electrons. For a three-dimensional double-barrier resonant-tunneling structure (see Fig. 1) the ac charging the accumulation layer is $I_{\rm acc}^{\rm rms} = 2\pi\nu C V^{\rm rms}/A$, where ν is the driving frequency, C is the capacitance, $V^{\rm rms}$ is the applied bias, and A is the area. In comparison, the total incident flux is $I_{\rm inc} = 3/8\, env_{\rm F}$. Using the parameters appropriate for a typical experiment (we use that of Brown *et al.*[24]), we find that up to 10 THz the probability of an electron participating in the charge buildup is only 1%. Summarizing, these estimates indicate that our approach should be accurate up to frequencies of tens of THz, which are large by present experimental standards, and consequently the analysis presented in what follows should be valid for most experimental situations.

III. THEORETICAL TOOLS AND THE MODEL

A. Baym-Kadanoff-Keldysh nonequilibrium techniques

Here we wish to outline the physical background behind the Keldysh formulation, and in particular its connection to tunneling physics. Readers interested in technical details should consult any of the many available review articles, such as Refs. 25–27. The basic difference between construction of equilibrium and nonequilibrium perturbation schemes is that in nonequilibrium one cannot assume that the system returns to its ground state (or a thermodynamic equilibrium state at finite temperatures) as $t \to +\infty$. Irreversible effects break the symmetry between $t = -\infty$ and $t = +\infty$, and this symmetry is heavily exploited in the derivation of the equilibrium perturbation expansion. In nonequilibrium situations one can circumvent this problem by allowing the system to evolve from $-\infty$ to the moment of interest (for definiteness, let us call this instant t_0), and then continue the time evolvement from $t = t_0$ back to $t = -\infty$.[28] (When dealing with quantities that depend on two time variables, such as Green functions, the time evolution must be continued to the later time.) The advantage of this procedure is that all expectation values are defined with respect to a well defined state, i.e., the state in which the system was prepared in the remote past. The price is that one must treat the two time branches on an equal footing (See Fig. 2).

A typical object of interest would be a two time Green function (see Appendix A); the two times can be located on either of the two branches of the complex time path (e.g., τ and τ' in Fig. 2). One is thus led to consider 2×2 Green-function matrices, and the various terms in the perturbation theory can be evaluated by matrix multiplication. Since the internal time integrations run over the complex time path, a method of bookkeeping for the time labels is required, and there are various ways of doing this. In the present work we employ a version of the Keldysh technique.

In the context of tunneling problems the time-independent Keldysh formalism works as follows. In the remote past the contacts (i.e., the left and right lead) and the central region are decoupled, and each region is in thermal equilibrium. The equilibrium distribution functions for the three regions are characterized by their respective chemical potentials; these do not have to coincide nor are the differences between the chemical potentials necessarily small. The couplings between the different regions are then established and treated as perturbations via the standard techniques of perturbation theory, albeit on the two-branch time contour. It is important to notice that the couplings do not have to be small, e.g., with respect level spacings or $k_{\rm B}T$, and typically must be treated to all orders.

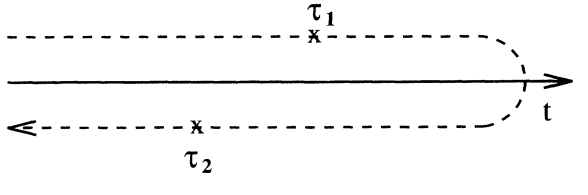

FIG. 2. The complex-time contour on which nonequilibrium-Green-function theory is constructed. In the contour sense, the time τ_1 is earlier than τ_2 even though its real-time projection appears larger.

The time-dependent case can be treated similarly. Before the couplings between the various regions are turned on, the single-particle energies acquire rigid time-dependent shifts, which, in the case of the noninteracting contacts, translate into extra phase factors for the propagators (but not in changes in occupations). The perturbation theory with respect to the couplings has the same diagrammatic structure as in the stationary case. The calculations, of course, become more complicated because of the broken time-translational invariance.

B. Model Hamiltonian

We split the total Hamiltonian in three pieces: $H = H_c + H_T + H_{\rm cen}$, where H_c describes the contacts, H_T is the tunneling coupling between contacts and the interacting region, and $H_{\rm cen}$ models the interacting central region, respectively. Below we discuss each of these terms.

1. Contacts, H_c

Guided by the typical experimental geometry in which the leads rapidly broaden into metallic contacts, we view electrons in the leads as noninteracting except for an overall self-consistent potential. Physically, applying a time-dependent bias between the source and drain contacts corresponds to accumulating or depleting charge to form a dipole around the central region. The resulting electrostatic-potential difference means that the single-particle energies become time dependent: $\epsilon^0_{k\alpha} \to \epsilon_{k\alpha}(t) = \epsilon^0_{k\alpha} + \Delta_\alpha(t)$ [here α labels the channel in the left (L) or right (R) lead]. The occupation of each state $k\alpha$, however, remains unchanged. The occupation, for each contact, is determined by an equilibrium distribution function established in the distant past, before the time-dependence or tunneling matrix elements are turned on. Thus, the contact Hamiltonian is

$$H_c = \sum_{k,\alpha\in L,R} \epsilon_{k\alpha}(t)\mathbf{c}^\dagger_{k\alpha}\mathbf{c}_{k\alpha}\,, \tag{1}$$

and the exact time-dependent Green functions in the leads for the uncoupled system are

$$\begin{aligned} g^<_{k\alpha}(t,t') &\equiv i\langle \mathbf{c}^\dagger_{k\alpha}(t')\mathbf{c}_{k\alpha}(t)\rangle \\ &= if(\epsilon^0_{k\alpha})\exp\left[-i\int_{t'}^{t} dt_1 \epsilon_{k\alpha}(t_1)\right] \\ g^{r,a}_{k\alpha}(t,t') &\equiv \mp i\theta(\pm t \mp t')\langle\{\mathbf{c}_{k\alpha}(t),\mathbf{c}^\dagger_{k\alpha}(t')\}\rangle \\ &= \mp i\theta(\pm t \mp t')\exp\left[-i\int_{t'}^{t} dt_1 \epsilon_{k\alpha}(t_1)\right]. \end{aligned} \tag{2}$$

One should note that our model for $g^<$ differs from the choice made in the recent study of Chen and Ting.[15] The difference does not affect calculations carried out to linear response in the ac drive, but is significant in nonlinear response. Specifically, Chen and Ting allow the electrochemical potential in the distribution function f to vary with time: $\mu_L - \mu_R = e[V + U(t)]$, where $U(t)$ is the ac signal. This assumption implies that the *total number* of electrons in the contacts varies with time. This behavior is inconsistent with what happens in real devices: it is only the relatively small number of electrons in the accumulation-depletion layers that is time dependent. In addition to the unphysical charge pileup in the contacts, the model of Chen and Ting leads to an instantaneous loss of phase coherence in the contacts, and hence does not display any of the interesting interference phenomena predicted by our phase-conserving model.

2. Coupling between leads and central region, H_T

The coupling between the leads and the central (interacting) region can be modified with time dependent gate voltages, as is the case in single-electron pumps. The precise functional form of the time dependence is determined by the detailed geometry and by the self-consistent response of charge in the contacts to external driving. We assume that these parameters are known, and simply write

$$H_T = \sum_{\substack{k,\alpha\in L,R \\ n}} [V_{k\alpha,n}(t)\mathbf{c}^\dagger_{k\alpha}\mathbf{d}_n + \text{H.c.}]\,. \tag{3}$$

Here $\{\mathbf{d}^\dagger_n\}$ and $\{\mathbf{d}_n\}$ form a complete orthonormal set of single-electron creation and annihilation operators in the interacting region.

3. The central-region Hamiltonian $H_{\rm cen}$

The form chosen for $H_{\rm cen}$ in the central interacting region depends on geometry and on the physical behavior being investigated. Our results relating the current to local properties, such as densities of states and Green functions, are valid generally. To make the results more concrete, we will discuss two particular examples in detail. In the first, the central region is taken to consist of noninteracting, but time-dependent levels,

$$H_{\rm cen} = \sum_m \epsilon_m(t)\mathbf{d}^\dagger_m\mathbf{d}_m. \tag{4}$$

Here $\mathbf{d}^\dagger_m$ ($\mathbf{d}_m$) creates (destroys) an electron in state m. The choice (4) represents a simple model for time-dependent resonant tunneling. Below we shall present general results for an arbitrary number of levels, and analyze the case of a single level in detail. The latter is interesting both as an exactly solvable example, and for predictions of coherence effects in time-dependent experiments.

The second example we will discuss is resonant tunneling with electron-phonon interaction,

$$H^{\text{el-ph}}_{\rm cen} = \epsilon_0\mathbf{d}^\dagger\mathbf{d} + \mathbf{d}^\dagger\mathbf{d}\sum_{\mathbf{q}} M_{\mathbf{q}}[\mathbf{a}^\dagger_{\mathbf{q}} + \mathbf{a}_{-\mathbf{q}}]\,. \tag{5}$$

In the above, the first term represents a single site, while the second term represents the interaction of an electron on the site with phonons: $\mathbf{a}^\dagger_{\mathbf{q}}(\mathbf{a}_{\mathbf{q}})$ creates (destroys) a phonon in mode $\mathbf{q}$, and $M_{\mathbf{q}}$ is the interaction matrix element. The full Hamiltonian of the system

must also include the free-phonon contribution $H_{\rm ph} = \sum_{\mathbf{q}} \hbar\omega_{\mathbf{q}} \mathbf{a}_{\mathbf{q}}^\dagger \mathbf{a}_{\mathbf{q}}$. This example, while not exactly solvable, is helpful to show how interactions influence the current. Furthermore, we can directly compare to previous time-independent results[22] using (5) to demonstrate the power of the present formalism.

IV. TIME-DEPENDENT CURRENT AND KELDYSH GREEN FUNCTIONS

A. General expression for the current

The current from the left contact through the left barrier to the central region can be calculated from the time evolution of the occupation number operator of the left contact:

$$J_L(t) = -e\langle \dot{N}_L \rangle = -\frac{ie}{\hbar}\langle [H, N_L] \rangle \,, \tag{6}$$

where $N_L = \sum_{k,\alpha\in L} \mathbf{c}_{k\alpha}^\dagger \mathbf{c}_{k\alpha}$ and $H = H_c + H_T + H_{\rm cen}$. Since H_c and $H_{\rm cen}$ commute with N_L, one readily finds

$$J_L = \frac{ie}{\hbar} \sum_{\substack{k,\alpha\in L \\ n}} \left[V_{k\alpha,n} \langle \mathbf{c}_{k\alpha}^\dagger \mathbf{d}_n \rangle - V_{k\alpha,n}^* \langle \mathbf{d}_n^\dagger \mathbf{c}_{k\alpha} \rangle \right] . \tag{7}$$

Now define two Green functions

$$G_{n,k\alpha}^<(t,t') \equiv i\langle \mathbf{c}_{k\alpha}^\dagger(t') \mathbf{d}_n(t) \rangle, \tag{8}$$

$$G_{k\alpha,n}^<(t,t') \equiv i\langle \mathbf{d}_n^\dagger(t') \mathbf{c}_{k\alpha}(t) \rangle \,. \tag{9}$$

Using $G_{k\alpha,n}^<(t,t) = -\left[G_{n,k\alpha}^<(t,t) \right]^*$, and inserting the time labels, the current can be expressed as

$$J_L(t) = \frac{2e}{\hbar} \mathrm{Re} \left\{ \sum_{\substack{k,\alpha\in L \\ n}} V_{k\alpha,n}(t) G_{n,k\alpha}^<(t,t) \right\} . \tag{10}$$

One next needs an expression for $G_{n,k\alpha}^<(t,t')$. For the present case, with noninteracting leads, a general relation for the contour-ordered Green function $G_{n,k\alpha}(\tau,\tau')$ can be derived rather easily (either with the equation-of-motion technique, or by a direct expansion of the S matrix; the details are given in Appendix B), and the result is

$$G_{n,k\alpha}(\tau,\tau') = \sum_m \int d\tau_1 G_{nm}(\tau,\tau_1) \times V_{k\alpha,m}^*(\tau_1) g_{k\alpha}(\tau_1,\tau') \,. \tag{11}$$

Here $G_{nm}(\tau,\tau_1)$ is the contour-ordered Green function for the central region, and the τ variables are now defined on the contour of Fig. 2. Note that the time dependence of the tunneling matrix elements and single-particle energies has broken the time-translational invariance. The analytic continuation rules (A3) of Appendix A can now be applied, and we find

$$G_{n,k\alpha}^<(t,t') = \sum_m \int dt_1 V_{k\alpha,m}^*(t_1) \times [G_{nm}^r(t,t_1) g_{k\alpha}^<(t_1,t') + G_{nm}^<(t,t_1) g_{k\alpha}^a(t_1,t')] \,, \tag{12}$$

where the Green functions $g^{<,a}$ for the leads are defined in (2) above. Combining (2), (10), and (12), yields

$$J_L(t) = -\frac{2e}{\hbar} \mathrm{Im} \left\{ \sum_{\substack{k,\alpha\in L \\ n,m}} V_{k\alpha,n}(t) \int_{-\infty}^t dt_1 \times e^{i\int_{t_1}^t dt_2 \epsilon_{k\alpha}(t_2)} V_{k\alpha,m}^*(t_1) \times [G_{nm}^r(t,t_1) f_L(\epsilon_{k\alpha}) + G_{nm}^<(t,t_1)] \right\} . \tag{13}$$

The discrete sum over k in $\sum_{k\alpha}$ can be expressed in terms of densities of states in the leads: $\int d\epsilon \rho_\alpha(\epsilon)$. Then it is useful to define

$$\left[\Gamma^L(\epsilon,t_1,t) \right]_{mn} = 2\pi \sum_{\alpha\in L} \rho_\alpha(\epsilon) V_{\alpha,n}(\epsilon,t) V_{\alpha,m}^*(\epsilon,t_1) \times \exp\left[i\int_{t_1}^t dt_2 \Delta_\alpha(\epsilon,t_2) \right] , \tag{14}$$

where $V_{k\alpha,n} = V_{\alpha,n}(\epsilon_k)$. In terms of this generalized linewidth function (14), the general expression for the current is

$$J_L(t) = -\frac{2e}{\hbar} \int_{-\infty}^t dt_1 \int \frac{d\epsilon}{2\pi} \mathrm{Im}\mathrm{Tr} \left\{ e^{-i\epsilon(t_1-t)} \mathbf{\Gamma}^L(\epsilon,t_1,t) \times [\mathbf{G}^<(t,t_1) + f_L(\epsilon)\mathbf{G}^r(t,t_1)] \right\} . \tag{15}$$

Here the boldface notation indicates that the level-width function $\mathbf{\Gamma}$ and the central-region Green functions $\mathbf{G}^{<,r}$ are matrices in the central-region indeces m,n. An analogous formula applies for $J_R(t)$, the current flowing into the central region through the right barrier.

This is the central formal result of this work, and the remainder of this paper is devoted to the analysis and evaluation of Eq. (15). The current is expressed in terms of local quantities: Green functions of the central region. The first term in Eq. (15), which is proportional to the lesser function $G^<$, suggests an interpretation as the out-tunneling rate [recalling $\mathrm{Im}\, G^<(t,t) = N(t)$]. Likewise, the second term, which is proportional to the occupation in the leads and to the density of states in the central region, can be associated to the in-tunneling rate. However, one should bear in mind that all Green functions in Eq. (15) are to be calculated in the presence of tunneling. Thus, $G^<$ will depend on the occupation in the leads. Furthermore, in the presence of interactions G^r will depend on the central-region occupation. Consequently, the current can be a nonlinear function of the occupation factors. This issue has recently been discussed also by other authors.[29]