Wigner-function model of a resonant-tunneling semiconductor device

William R. Frensley

Central Research Laboratories, Texas Instruments Incorporated, P.O. Box 655936, Dallas, Texas 75265

A model of an open quantum system is presented in which irreversibility is introduced via boundary conditions on the single-particle Wigner distribution function. The Wigner function is calculated in a discrete approximation by solution of the Liouville equation in steady state, and the transient response is obtained by numerical integration of the Liouville equation. This model is applied to the quantum-well resonant-tunneling diode. The calculations reproduce the negative-resistance characteristic of the device, and indicate that the tunneling current approaches steady state within a few hundred femtoseconds of a sudden change in applied voltage.

I. INTRODUCTION

The progress of semiconductor fabrication technology, particularly the heteroepitaxial technology, has permitted the fabrication of structures[1] and devices[2,3] whose behavior is dominated by quantum-interference effects. By providing a mechanism to control current flow through nanometer-scale semiconductor structures, such quantum-interference effects might form the basis for a new generation of solid-state electronic technology.[4] The successful development of such a technology will require a much-more-detailed understanding of the dynamic behavior of size-quantized systems than is presently available.

A key property of any device, which has not been adequately addressed in the case of quantum tunneling devices, is the transient response to changes in the externally applied voltage. The transient response determines the ultimate switching speed and frequency response of the device. Because a device is a physical system characterized by openness and irreversibility, the analysis of its behavior, and particularly the analysis of the transient response, must be performed within the framework of statistical mechanics, rather than pure-state quantum mechanics. A preliminary report on the calculation of the transient response of a tunneling device has appeared.[5] This paper describes the calculation in detail.

In the present work the quantum-well resonant-tunneling diode[2,3,6] (RTD) has been taken to be the prototype quantum semiconductor device. This device consists of a single quantum well bounded by tunneling barriers, as shown in Fig. 1. As a bias voltage is applied to the device, the resonant state in the well is pulled down in energy with respect to the more negative electrode, and the tunneling current through this state depends on the density of occupied states in the electrode. When the resonant state is pulled below the conduction-band edge of the electrode, the tunneling current decreases. The device thus shows negative differential resistance, which is attributable to quantum interference.
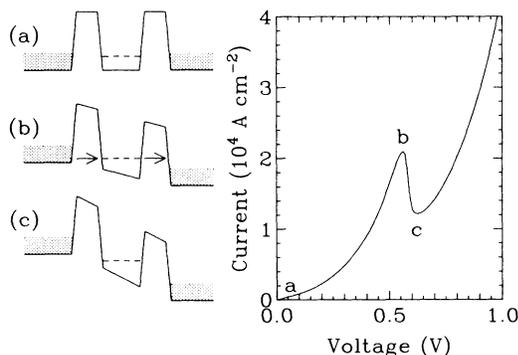

FIG. 1. Potential diagram and experimental $I(V)$ curve for a resonant-tunneling diode. The barriers are thin layers of a wider-band-gap semiconductor, typically (Al,Ga)As, and the quantum well and the regions outside the barriers are GaAs. A size-quantized state is confined in the well; its energy is indicated by the dashed line. (a) shows the structure in equilibrium. When a voltage is applied (b) electrons can resonantly tunnel out of occupied states (shaded region) through the confined state. As the voltage is increased (c) the resonant state is pulled below the occupied levels and the tunneling current decreases, leading to a negative-resistance characteristic, as the current decreases with increasing voltage. The experimental data are made available by courtesy of Reed (Ref. 6).

II. OPENNESS AND IRREVERSIBILITY OF THE DEVICE

An electron device is necessarily an open system; it is useless unless connected to an electrical circuit and able to exchange electrons with that circuit. If one wishes to study the behavior of the device apart from that of the circuit, it is convenient to represent the effects of the external circuit by ideal electron reservoirs attached to the terminals of the device. The term "open system" is used here in a more restricted sense than is common in the statistical physics literature.[7] For the present purposes, an open system is one which is connected to reservoirs of conserved particles, so that the interaction between the system and a reservoir necessarily involves a particle current through an interface (real or idealized) between the system and the reservoir. A concrete example is a semiconductor

device (the system) to which a metal wire (the reservoir) is attached through an ohmic contact (the interface). The body of literature of which Ref. 7 is representative is primarily concerned with a related problem which I will refer to as "damping." In the semiconductor-device example, this is the interaction between the electrons and phonons (regarded as a heat bath) which leads to electrical resistance.

The dynamic behavior of a device must be time irreversible. This is demonstrated by the existence of the current-voltage $I(V)$ curve that is conventionally used to characterize a device. Each point of the $I(V)$ curve represents a steady state, and all of the points except $V=0$ represent a nonequilibrium steady state. A time-reversible system cannot stably approach such a steady state.

Landauer[8] has pointed out that if one considers conduction through a sample in which only elastic scattering occurs, the resulting finite conductivity requires that energy be dissipated in the reservoirs to which the sample is connected. The connection between openness and dissipation may be readily perceived in other physical systems, perhaps most clearly in the vacuum tube (or valve) of an earlier generation of electronic technology.[9] The vacuum is certainly not a dissipative medium, but making electrical contact to it through metal electrodes results in a device which displays an $I(V)$ curve and which dissipates energy (generally in the anode). Thus openness and dissipation are related. What is perhaps less obvious is that dissipation or irreversibility is a *necessary* feature of any meaningful description of the interaction between an open system and the reservoirs to which it is connected. If the boundary conditions are reversible, then unstable solutions to the transient response are admitted. This will become apparent when we consider the effect of open-system boundary conditions on the eigenvalue spectrum of the Liouville superoperator.

The connection between open-system boundary conditions and the eigenvalue spectrum can be seen in the simpler case of the Hamiltonian. Consider the conventional demonstration of the Hermiticity of the single-particle Hamiltonian.[10] This demonstration proceeds by invoking Green's identity to transpose the Laplace operator, which leaves a surface term. This term is conventionally taken to be zero, but it can be expressed as

$$H - H^\dagger = \frac{\hbar}{i} \int_s \mathbf{j} \cdot d\mathbf{s} \ , \tag{1}$$

where $\mathbf{j}$ is the current operator and S is the boundary of the domain in which the Hamiltonian is defined. One maintains the Hermiticity of the Hamiltonian by choosing wave functions for which the surface integral vanishes: states that are well localized within the domain, or stationary scattering states for which the incoming and outgoing flux cancel. Now, in general, there must be a net change in the electron density in a device as the device passes from one steady state to another. For example, the resonant state in the RTD shows a large electron density peak in the quantum well. Therefore, during the process of establishing this state, there must be a net inward current flow to "fill up" the well. However, according to

(1), a set of basis states (on a finite spatial domain) that could describe such a process would result in a non-Hermitian Hamiltonian.

III. STATISTICAL MECHANICS

Irreversible quantum phenomena are properly treated at a statistical level.[11] That is, a level at which the state of the quantum system is represented by an operator (density matrix,[12] Wigner function,[13] or Green's function,[14] typically) rather than by a wave function. The statistical representation is required if one is to describe both the superposition of complex-valued amplitudes leading to interference effects, and the superposition of real-valued probabilities leading to incoherent phenomena. Let us briefly review the fundamental relations of quantum statistical mechanics, in order to define the terms and display the equations that will be invoked.

A statistically mixed quantum state is described by a density matrix[12] which, in a real-space basis, can be written as

$$\rho(x,x') = \sum_{\{|i\rangle\}} w_i \langle x | i \rangle \langle i | x' \rangle \ , \tag{2}$$

where $\{|i\rangle\}$ represents a complete set of states and w_i is a probability. The time evolution of the density matrix is given by the quantum Liouville equation:

$$\frac{\partial \rho}{\partial t} = (1/i\hbar)[H,\rho] \equiv (L/i\hbar)\rho$$

$$= \frac{1}{i\hbar} \left[-\frac{\hbar^2}{2m} \left[\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial x'^2} \right] \rho \right.$$

$$\left. + [v(x) - v(x')]\rho \right] \ , \tag{3}$$

where H is the Hamiltonian and L is the Liouville (super)operator. The potential v will include contributions from the device structure in the form of heterojunction band discontinuities and from the electrostatic (Hartree) potential due to mobile electrons, ionized impurities and externally applied fields. In the present work the independent-electron model will be assumed, so that only the single-electron reduced density matrix or distribution function is required.

The Wigner distribution function[13] is obtained from the density matrix by changing the independent variables to $\mathcal{X} = \frac{1}{2}(x + x')$ and $\xi = x - x'$. The classical position is then identified with $\mathcal{X}$ and the Fourier transform of ξ is taken to obtain the classical momentum variable:

$$f(\mathcal{X},k) = \int_{-\infty}^{\infty} d\xi \, e^{-ik\xi} \rho(\mathcal{X} + \tfrac{1}{2}\xi, \mathcal{X} - \tfrac{1}{2}\xi) \ . \tag{4}$$

This transformation of variables and its effect on the boundary conditions for finite systems is illustrated in Fig. 2. The Liouville equation for the Wigner function can then be written

$$\frac{\partial f}{\partial t} = -\frac{\hbar k}{m} \frac{\partial f}{\partial \mathcal{X}} - \frac{1}{\hbar} \int_{-\infty}^{\infty} \frac{dk'}{2\pi} V(\mathcal{X}, k - k') f(\mathcal{X},k') \ , \tag{5}$$

where the kernel of the potential operator is given by

$$V(X,k)=2\int_0^\infty d\xi \sin(k\xi)[v(X+\tfrac{1}{2}\xi)-v(X-\tfrac{1}{2}\xi)] . \qquad (6)$$

The nonlocal potential is the means by which interference between alternative paths enters the Wigner-function formalism.[15]

The density matrix, and thus the Wigner function, may be normalized so as to represent the particle density. Thus the electron density $n(X)$ (in units of particles per cm^3, for example) can be found from

$$n(X)=\rho(X,X)=\int_{-\infty}^\infty \frac{dk}{2\pi}f(X,k) . \qquad (7)$$

The continuity equation can be derived from the Liouville equation (5) by integrating with respect to k. The contribution from the potential operator vanishes by antisymmetry and one obtains an expression for the current density:

$$j(X)=\int_{-\infty}^\infty \frac{dk}{2\pi}\frac{\hbar k}{m}f(X,k) . \qquad (8)$$

Now let us consider the effect of boundary conditions on the eigenvalue spectrum of the Liouville superoperator. The Hermiticity of L follows directly from that of the Hamiltonian H for a closed, conservative system. Such a system can only display oscillatory behavior. If we change the boundary conditions so as to allow particles to pass into or out of the system, we violate the Hermiticity of Liouville operator, because its Hermiticity depends upon a relation derived from (1). This introduces imaginary parts into at least some of the eigenvalues of L, or
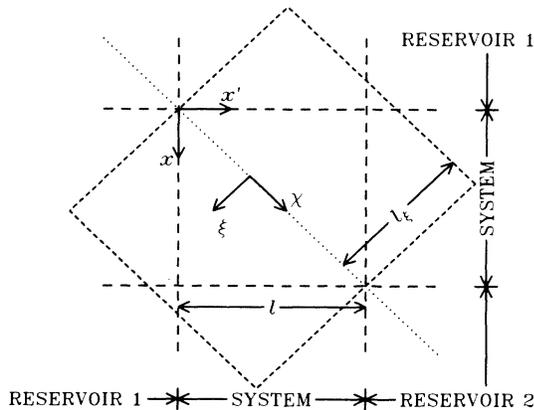
RESERVOIR 1 →✳← SYSTEM →✳← RESERVOIR 2

FIG. 2. Domain of the density-matrix and Wigner-distribution function calculations. The arguments of the density matrix are x and x'. The Wigner function is obtained by transforming to the coordinates X and ξ, followed by a Fourier transform with respect to ξ. The long-dashed lines indicate the system-reservoir boundaries, and they partition the domain into regions corresponding to various system-system, system-reservoir, and reservoir-reservoir correlations. The short-dashed lines represent the boundaries of the domain of the Wigner-distribution-function calculation. Note that the Wigner-distribution-function domain includes regions which represent correlation with the reservoirs.

real parts into some of the eigenvalues of $(L/i\hbar)$, giving a real exponential character to the time dependence of f. If the open-system boundary conditions are time reversible, then the real parts of the eigenvalues of $(L/i\hbar)$ occur symmetrically. That is, there is a positive real part, corresponding to an unstable solution, for every negative real part.[16] An example of such a boundary condition may be applied to the density matrix.[17] It is $\partial\rho/\partial X=0$, along boundaries parallel to the x and x' axes (shown as the long-dashed lines in Fig. 2). This is a plausible boundary condition for an open system, because it leads to a constant density at the boundary, approximating the effect of a fixed chemical potential at the boundary. Fixing the chemical potential is the usual way to represent an ohmic contact in classical semiconductor-device analyses.[18,19] In the present case, however, its time-reversal symmetry leads to an unphysical exponentially growing solution of the Liouville equation.

One might expect that the unstable eigenvalues would be removed by the inclusion of the damping which results from coupling the system to a heat bath. (The damping in semiconductors is due to random scattering of electrons by phonons.[20]) Within the simple models which I have studied,[16] such is not the case. If one uses a simple Fokker-Planck operator[21] to approximate the effects of phonon scattering, then damping coefficients which are appropriate to a high-mobility material such as GaAs are not sufficient to render the time-dependent behavior stable. Damping coefficients about an order of magnitude too large are required to remove the positive eigenvalues. Thus, the stability of a model of an open system should be guaranteed by the open-system boundary conditions themselves, and this means that these boundary conditions must be time irreversible.

The required irreversibility can be obtained in a physically appealing way by assuming that the reservoirs to which the device is connected have properties analogous to those of a black body: the distribution of electrons emitted into the device from the reservoir is characterized by the thermal equilibrium distribution function of the reservoir, and all electrons impinging upon a reservoir from the device are absorbed by the reservoir without reflection. To implement this picture, we must be able to distinguish the sense of the velocity of an electron at the position of the boundary. Thus the Wigner function is the natural representation for an open system. Let the interface between the device and the left-hand reservoir occur at $X=0$, and the interface between the device and the right-hand reservoir occur at $X=l$. Then we may write the open-system boundary conditions as

$$f(0,k)\,|_{k>0}=f_l(k) , \qquad (9a)$$

$$f(l,k)\,|_{k<0}=f_r(k) . \qquad (9b)$$

Note that these boundary conditions are appropriate for the Liouville equation in the form of (5). The derivative is first order in X and so one value of f must be specified for each k. The dependence on k is expressed as an integral, so no boundary conditions need be specified in the k direction. The reservoirs are characterized by the chemical potential μ and reciprocal temperature β. The equilib-

rium distribution functions of the reservoirs can be written (after integrating over the transverse momenta) as

$$f_{l,r}(k) = (m/\pi\hbar^2\beta_{l,r})$$

$$\times \ln\{1 + \exp[-\beta_{l,r}(\hbar^2k^2/2m - \mu_{l,r})]\} . \quad (10)$$

The physical picture which underlies these boundary conditions is of course well known. It is invoked in most forms of transport calculation, including scattering analyses of conductance[8] and tunneling,[22-24] and classical calculations such as those performed by Monte Carlo methods.[25,26] The characteristic feature of such calculations, however, is that the interaction between the system and the reservoir is treated *implicitly* within an algorithmic procedure. The usefulness of this approach is inherently limited. We will see that the *explicit* statement of the boundary conditions (9) leads to new insights and calculational capabilities.

The most important insight is that the boundary conditions (9) assure the stability of the solutions of the Liouville equation. This may be demonstrated by proving that the real parts of the eigenvalues of $(L/i\hbar)$ are all nonpositive. We may do so by evaluating the expectation value of the homogeneous part of $(L/i\hbar)$ for an arbitrary distribution function f. Note that because L is a superoperator, the expectation value is taken between operators (the distribution functions). If we define the inner product of distribution functions in the obvious way, the expectation value is readily evaluated. Again the contribution from the potential operator vanishes by antisymmetry and the gradient operator can be integrated to obtain

$$\langle f,(L/i\hbar)f\rangle = \int dx \int dk\, f(L/i\hbar)f$$

$$= (\hbar/2m) \int dk\, k[f^2(0,k) - f^2(l,k)]$$

$$= (\hbar/2m) \left[\int_{-\infty}^{0} kf_s^2(0,k)dk + \int_{0}^{\infty} kf_l^2(0,k)dk - \int_{-\infty}^{0} kf_r^2(l,k)dk - \int_{0}^{\infty} kf_s^2(l,k)dk \right] . \quad (11)$$

Here I have used the notation f_s to denote the part of the distribution function that is a property of the system, that is, not specified by the boundary conditions. For the homogeneous case, $f_l = f_r = 0$. Then the terms containing these quantities vanish, and the two terms containing f_s are clearly nonpositive. Therefore the real parts of the eigenvalues of $(L/i\hbar)$ must all be nonpositive. The physical interpretation of this argument is that the electrons in an open system will eventually escape and the internal density will approach zero if there is no inward current flow from the reservoirs.

IV. DISCRETE MODEL

The problems which are of interest in studying the behavior of quantum devices do not fall into the domain of analytically soluble problems. The results of the present model must therefore be evaluated numerically. This requires that the infinite number of mathematical operations implied by the continuum formulation of the problem must be reduced to a finite number. A natural way to do this is to replace the continuous domain of the problem by a mesh or lattice of discrete points in phase space.

The position coordinates x will be taken to be elements of a uniformly spaced set with mesh spacing Δ_x: $x \in \{0, \Delta_x, 2\Delta_x, \ldots, l\}$. The number of mesh points in the x dimension is thus $N_x = l/\Delta_x + 1$. The underlying wave functions and the operators which act upon them are assumed to be defined only on a mesh point [although for the purpose of evaluating the potential operator (6) the mesh may be extended into the reservoirs]. The definition of the Wigner distribution function (4) may be discretized as follows: The position argument X takes discrete values from the set defined above. (Henceforth I will neglect the distinction between the position argument of the Wigner function X and the position argument of the wave func-

tion x, and refer to both of these quantities as x.) The relative coordinate ξ, from which the momentum argument is obtained, must be treated more carefully. Because we have assumed that the potential is only defined an integral multiples of Δ_x and because ξ appears in (6) with a coefficient of $\frac{1}{2}$, we must constrain ξ to take only even multiples of Δ_x: $\xi \in \{0, 2\Delta_x, 4\Delta_x, \ldots, l_\xi\}$. Then when we take the Fourier transform with respect to ξ as in (4) or (6), the resulting function is periodic in k with a period of π/Δ_x. The maximum value of ξ in this scheme, l_ξ, is not necessarily related to any of the previously defined quantities. It determines the maximum distance over which quantum correlations are taken into account, thus determining in some sense the number of alternative paths which are allowed to interfere.[15] In the present calculations, l_ξ has been taken to be equal to l.

The domain is similarly discretized in the k dimension. The domain $-\pi/2\Delta_x < k \leq \pi/2\Delta_x$ may be discretized into an arbitrary number of mesh points N_k. However, the numerical solution of the Liouville equation is simplified if we choose the mesh so that $k = 0$ is not one of the mesh values. [At $k = 0$ the gradient term of (5) degenerates, leading to zeros on the diagonal of the superoperator. Such zeros require that row interchanges be performed in the Gaussian elimination procedure. If the mesh is chosen so that these zeros do not occur, it has been observed that no interchanges are required.] Thus we require that N_k be even, so as to have equal numbers of points for positive and negative k, and let the mesh straddle $k = 0$: $k \in \{(\pi/\Delta_x)[(j - \frac{1}{2})/N_k - \frac{1}{2}], \quad j = 1, 2, \ldots, N_k\}$. The k mesh spacing is therefore $\Delta_k = \pi/N_k\Delta_x$. Equation (6) for the potential operator then takes the discretized form:

$$V(x,k) = \frac{2}{N_k} \sum_{\{\xi\}} \sin(k\xi)[v(x + \tfrac{1}{2}\xi) - v(x - \tfrac{1}{2}\xi)] . \quad (12)$$

Discretization of the Liouville equation (5) requires some care, as the gradient (drift) term must be replaced by a finite-difference approximation. There is a potential ambiguity in this procedure, because the gradient can be expressed as either a left-hand difference: $(\nabla f)(x) = [f(x) - f(x - \Delta_x)]/\Delta_x$, or as a right-hand difference: $(\nabla f)(x) = [f(x + \Delta_x) - f(x)]/\Delta_x$. This ambiguity is naturally resolved by considering the boundary conditions. Let us imagine that the boundary conditions (9) will be represented by extra rows of mesh points placed just outside of the domain of f, as illustrated in Fig. 3. The values of f along these boundary rows will be regarded as fixed. There will be a row along $x = -\Delta_x$ for $k > 0$ and a row along $x = l + \Delta_x$ for $k < 0$. Now consider $k > 0$. If the boundary conditions are to be coupled into the domain at all, we must use the left-hand difference for the

gradient, at least next to the left-hand boundary. If we want the discrete version of the fundamental theorem of calculus to hold for integration with respect to x, we must then use the left-hand difference for all x. A similar argument leads to the use of the right-hand difference for all x for $k < 0$. This sort of scheme for discretization of the gradient is the phase-space equivalent of a scheme which is known as an "upstream" or "upwind" difference in the context of classical fluid dynamic calculations.[27] Similar schemes have been employed in neutron-transport calculations.[28] The use of this upwind difference is what permits the argument of Eq. (11) to hold in the discrete case, and is thus the key element in obtaining stable solutions for the Wigner distribution function. The discretized Liouville equation can thus be written as

$$\frac{\partial f(x,k)}{\partial t} = -\frac{1}{\hbar} \sum_{\{k'\}} V(x, k - k') - \frac{\hbar k}{m \Delta_x} \times \begin{cases} [f(x + \Delta_x) - f(x)], & k < 0 \\ [f(x) - f(x - \Delta_x)], & k > 0 . \end{cases} \tag{13}$$

The discretized continuity equation is derived from the discretized Liouville equation (13). In a simple discretization scheme, the vector quantities such as electric field or current density are most naturally associated with the *intervals* between mesh points rather than with the mesh points themselves.[29] Thus we may expect the discretized continuity equation to have the form:

$$\frac{\partial n(x)}{\partial t} = [j(x + \tfrac{1}{2}\Delta_x) - j(x - \tfrac{1}{2}\Delta_x)]/\Delta_x . \tag{14}$$

If the density is defined as

RESERVOIR→|←———DEVICE———→|←RESERVOIR

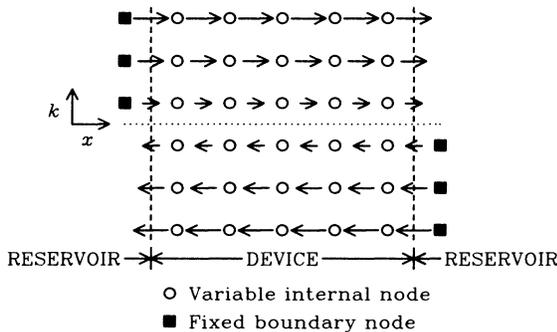O Variable internal node
■ Fixed boundary node

FIG. 3. Discretization scheme for the gradient (or drift) term of the Liouville equation. The flow of probability between mesh points is indicated by the arrows, which also indicate the sense of the finite-difference approximation for the gradient. The upwind difference means that a flow toward the right implies a left-hand difference approximation, and *vice versa*. Note that the sense of the finite difference is uniquely determined by the form of the boundary conditions.

$$n(x) = \sum_{\{k\}} \frac{\Delta_k}{2\pi} f(x,k) , \tag{15}$$

then the form of the discretized Liouville equation (13) requires that the current density be

$$j(x + \tfrac{1}{2}\Delta_x) = \frac{\Delta_k}{2\pi} \left[\sum_{k<0} \frac{\hbar k}{m} f(x + \Delta_x, k) + \sum_{k>0} \frac{\hbar k}{m} f(x,k) \right] . \tag{16}$$

Using this definition of j, the current density calculated for a steady-state solution $(\partial f / \partial t = 0)$ is independent of x.

The derivation of the discrete continuity equation (14) from the discrete Liouville equation (13) requires that

$$\sum_{\{k\}} V(x,k) = 0 , \tag{17}$$

for all x. This condition follows from the antisymmetry of V and from the Fourier completeness relation:

$$\sum_{\{k\}} e^{ik\xi} = N_k \delta_{\xi,0} . \tag{18}$$

This is significant because one is very tempted to employ a smaller set of k values to reduce the computation time. If the k's included in the computation do not span the "Brillouin zone" defined by the ξ discretization [and thus satisfy (18)], however, the continuity equation will not be satisfied. This can lead to steady-state solutions in which the current density is not constant, and to transient solutions in which there is a significant gain or loss of particle density inside the device.

V. STEADY-STATE BEHAVIOR

The steady-state behavior of the resonant-tunneling diode was numerically calculated by solving the discretized Liouville equation (13) for the condition $\partial f/\partial t = 0$. The boundary conditions (9) supply the inhomogeneous source terms. The present calculations used $N_x = 80$, $N_k = 60$. The mesh spacing was taken to be $\Delta_x = 0.565$ nm to make the assumed layer thicknesses commensurate with the atomic layer spacing of the (Al,Ga)As system. The discrete Liouville equation was solved by Gaussian elimination, which requires storage proportional to $N_x N_k^2$ and time proportional to $N_x N_k^3$.

The device structure assumed in the present calculations consists of a 4.5-nm-wide quantum well of GaAs bounded by identical 2.8-nm-wide barrier layers of $Al_{0.3}Ga_{0.7}As$. The conduction-band discontinuity was taken to be 0.60 (Ref. 30) of the total band-gap discontinuity (Ref. 31). 17.5 nm of the GaAs electrode layer was included in the simulation domain on each side of the device. Because Hartree self-consistency was not incorporated into the present calculations, the applied bias voltages were assumed to be dropped uniformly across the well and barriers, as illustrated in Fig. 1. The electron density assumed in the boundary reservoirs was 2×10^{18} cm^{-3}. (This is the only place in the computation where the doping density appears if self-consistency is not implemented.) All calculations were performed at a temperature of 300 K.

The assumed barriers are rather thin compared to those of typical experimental devices[3,6] because thicker layers result in narrower resonances in the transmission coefficient. This requires that one use more mesh points in k in order to adequately resolve the narrower resonances. The presently assumed structure is one whose Wigner function can be adequately represented by the discrete model described above.

The steady-state calculation was performed for a range of bias voltages, and the current density was evaluated from the resulting Wigner function. The current-voltage characteristic from this calculation is shown in Fig. 4,
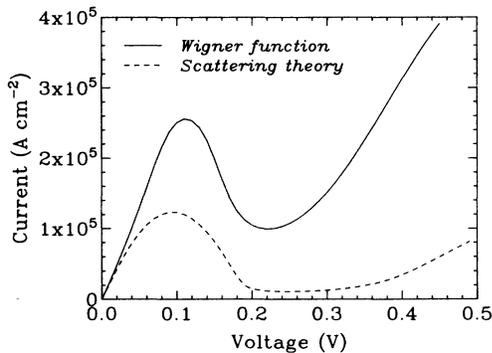
along with the curve derived from a scattering calculation which was done in a finite-difference approximation[17] with the same mesh spacing Δ_x. (In the scattering calculation the wave functions were evaluated recursively[32] using an energy-dependent initial condition which matched the wave function to a traveling wave.) The Wigner-function calculation predicts a higher current density and lower peak-to-valley current ratio than the scattering calculation. The voltages at which the peak and valley occur agree very well. The Wigner-function calculation more nearly resembles the experimental results at 300 K, but at lower temperatures it seriously underestimates the peak-to-valley ratio. Moreover, the presently neglected phonon scattering processes, which account for most of the temperature dependence, will tend to reduce the peak-to-valley ratio. Thus, at present, the scattering theory is more likely to fit the experimental data[6] when such processes are taken into account. The difference between the Wigner and scattering calculations is probably due to the difference in the boundary conditions. The boundary conditions (9) impose a constraint upon the standing-wave patterns of the underlying wave functions, whereas the energy-dependent boundary conditions of the scattering analysis do not do so. A result of this effect is that the current-voltage curves calculated from the Wigner function depend weakly on the assumed position of the boundaries.

The voltage at which the peak current occurs in both theoretical calculations of Fig. 4 is significantly lower than that at which the peak occurs in the experimental data of Fig. 1. The reason for this is the neglect of self-consistency in the evaluation of the electrostatic potential.[33] In particular, the upstream accumulation and downstream depletion layers in the experimental devices drop much of the applied voltage, and there is also the ohmic drop in the contact layers. The theoretical calculations predict a current density about an order of magni-

FIG. 4. Current vs voltage for a resonant-tunneling diode consisting of 2.8-nm layers of $Al_{0.3}Ga_{0.7}As$ bounding a 4.5-nm GaAs well, at a temperature of 300 K. The current derived from a calculation of the Wigner-distribution-function (solid line) is compared to that derived from a more conventional scattering calculation (dashed line).
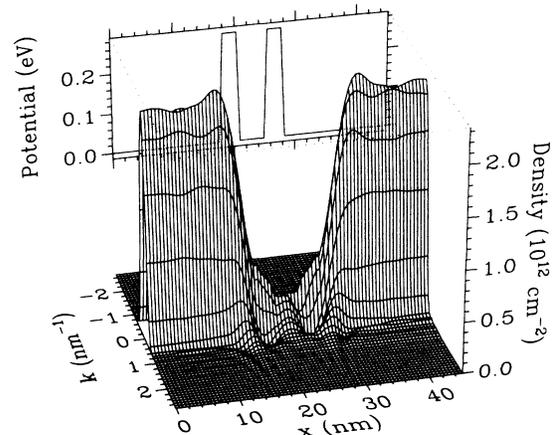

FIG. 5. Wigner-distribution-function for the steady-state calculation of Fig. 4, at zero-bias voltage (thermal equilibrium). In the electrode regions, the distribution is approximately Maxwellian (as a function of k). The density is significantly reduced in the vicinity of the quantum well due to size-quantization effects.
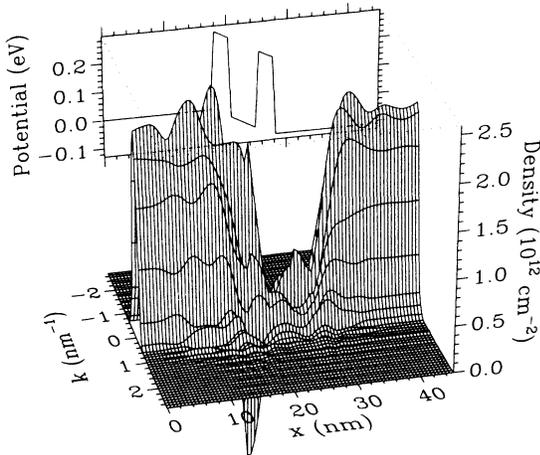
FIG. 6. Wigner-distribution-function for 0.11-V bias, corresponding to the peak current in Fig. 4. The standing-wave patterns and strong negative peak indicate that significant quantum-interference effects are occurring.

tude larger than the experimental data. This is due to the thinner barriers assumed in the theoretical calculations. The current density decreases exponentially with increasing barrier thickness.[34,35]

The steady-state Wigner functions for different bias voltages are plotted in Figs. 5–7. The zero-bias (equilibrium) case is shown in Fig. 5. The large electron density in the electrode regions, and much smaller density in the vicinity of the barriers, is evident. Small oscillations appear in various regions, reflecting the underlying quantum interference. Figure 6 shows the steady-state Wigner function for a bias of 0.11 V, which corresponds to the peak of the $I(V)$ curve. The oscillations are in this case much more pronounced, including a strong negative peak. This is evidence that strong quantum-interference effects are present, as expected. In contrast, the Wigner function (shown in Fig. 7) for a *larger* bias voltage, 0.22 V, which
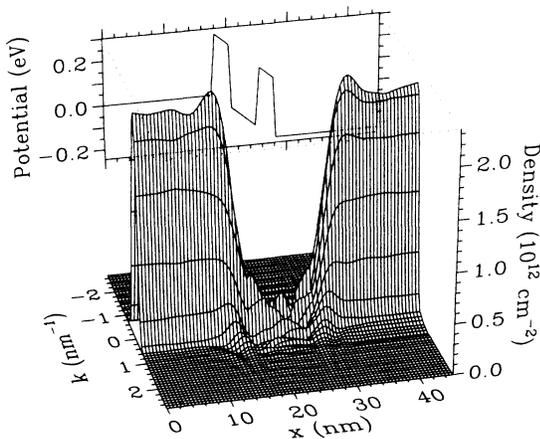

FIG. 7. Wigner-distribution-function for 0.22-V bias, corresponding to the valley current in Fig. 4. Note the similarity of this plot to that of the equilibrium case in Fig. 5.
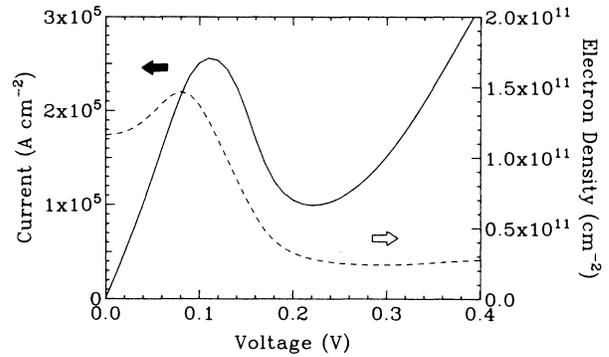

FIG. 8. Current density (solid line) and electron density in the quantum well (dashed line) as functions of bias voltage. The electron density in the well is a measure of the occupation of the resonant state.

corresponds to the current "valley," is quite similar to that of the equilibrium case.

Much more information can be obtained from the steady-state calculations than just the $I(V)$ curve. An example is the electron density in the quantum well as shown in Fig. 8. The maximum electron density of 1.5×10^{11} cm^{-2} occurs at a somewhat lower bias voltage than the current peak. The shift in the electrostatic potential due to the occupation of the well is therefore of the order of 0.02 V if one assumes that the effective separation of the charge dipole is 10 nm. (The charge in the well will be neutralized by a change in the mobile charge density at that location which minimizes the voltage shift. In the case of the RTD this will be in the accumulation layer that forms against the upstream barrier. The distance from the center of the well to the centroid of the accumulation layer is of the order of 10 nm.) The 0.02-V shift due to the charge in the quantum well is properly compared to the voltage axis of Figs. 1, 4, or 8, and is readily seen to be negligible.

VI. TRANSIENT BEHAVIOR

To obtain the transient response of a quantum device, the discretized Liouville equation (13) may be integrated. This requires a further discretization of the equation with respect to time. The way in which this discretization is performed also has a profound effect on the numerical stability of the resulting solutions, but the considerations here are well known within the context of "stiff" differential equations.[36] To illustrate the connection between the numerical approach and the more familiar techniques of quantum-mechanical calculations, let us discuss this issue using a more quantum-mechanical notation.

The general solution of the Liouville equation can be written as

$$f(t) = \left[\exp \left[\frac{1}{i\hbar} \int_{t_0}^{t} L(t')dt' \right] \right] f(t_0) . \qquad (19)$$

The conventional approach to evaluating this expression is

to find a part of L whose effect can be integrated analytically, and then to expand the remaining part of L in a perturbation series. The simple nonperturbative numerical approaches approximate the integral in the exponential by a finite sum, thereby obtaining a product expansion of the operator exponential

$$\exp\left[\frac{1}{i\hbar}\int_{t_0}^{t}L\,dt'\right] \approx \exp\left[\frac{1}{i\hbar}\sum_{t'=t_0}^{t}L(t')\Delta_t\right]$$

$$\approx \prod_{t'=t_0}^{t}\left[1+\frac{L(t')\Delta_t}{i\hbar}\right], \quad (20)$$

where $t'\in\{t_0,t_0+\Delta_t,t_0+2\Delta_t,\ldots,t\}$, to obtain a simple iterative scheme (in this case Euler's formula) which I will call the explicit scheme:

$$f(t+\Delta_t)=f(t)+\Delta_t(L/i\hbar)f(t). \quad (21)$$

(Note that this approach to evaluate the time-development operator trivially resolves the issue of time ordering the operator product.[37]) The usefulness of this scheme is limited by the existence of large negative eigenvalues of $(L/i\hbar)$. Consider the eigenfunction f_λ corresponding to such an eigenvalue λ. Each iteration of Eq. (21) multiplies the f_λ component of the solution by a factor $(1+\lambda\Delta_t)$. If the modulus of this factor is greater than unity, the solution will diverge, even though the criterion for stability of the continuum problem [$\mathrm{Re}(\lambda)\le0$] is satisfied. The modulus of the factor $(1+\lambda\Delta_t)$ can be made less than unity by choosing Δ_t to be sufficiently small, but the resulting values often lead to impractically large requirements for computer time.

An absolutely stable integration scheme is the implicit, or backward, Euler method. It is obtained by expanding the operator exponential as

$$\exp\left[\frac{1}{i\hbar}\int_{t_0}^{t}L\,dt'\right] \approx \prod_{t'-t_0}^{t}[1-L(t')\Delta_t/i\hbar]^{-1}. \quad (22)$$

This leads to the following iteration equation:

$$(1-\Delta_t L/i\hbar)f(t+\Delta_t)=f(t). \quad (23)$$

In this scheme each eigenfunction of the Liouville operator is multiplied by a factor of $(1-\lambda\Delta_t)^{-1}$ which has a modulus less than unity for any value of Δ_t if $\mathrm{Re}(\lambda)<0$. A consequence of this stability is that one must solve a linear system of equations at each time step (computation proportional to $N_x N_k^3$), as opposed to simply multiplying by a matrix as in the explicit scheme (computation proportional to $N_x N_k^2$). The use of larger time steps Δ_t, however, makes the implicit method the more effective one. This scheme was employed in the present calculations.

The stability of the calculation is guaranteed in the implicit method for arbitrarily large time steps Δ_t, but there is also the issue of the accuracy of the calculation. The potential inaccuracy arises because one is trying to approximate $\exp(\lambda\Delta_t)$ by $(1-\lambda\Delta_t)^{-1}$, which is clearly inadequate when $|\lambda\Delta_t|>1$. [Recall that $\mathrm{Re}(\lambda)\le0$.] If all of the eigenfunctions of the Liouville operator were present in the Wigner-function solution, this condition for accuracy of the implicit scheme is equivalent to the cri-

terion for stability of the explicit scheme. The benefits of the implicit scheme are due to the fact that not all of the eigenfunctions are present in the desired solution. In particular, the eigenfunctions of the large eigenvalues (the "stiff" components) are not significantly present. These eigenfunctions correspond to the occupation of higher-energy states. We may estimate the magnitude of the important eigenvalues by estimating the maximum occupied energy levels in a given situation. In the case of the resonant-tunneling diode this would be the Fermi level, plus a few times kT, plus the applied bias, or the bias voltage plus 0.1 to 0.2 V for the assumed design at room temperature. The present calculations employed a time step $\Delta_t=1$ fs, which corresponds to an energy of 0.6 eV. Thus this time step is small enough to realistically represent the transient response.

The results of the calculations of the transient response of the resonant-tunneling diode model are shown in Figs. 9 and 10. Since the negative-resistance characteristic is the interesting feature of this device, the transient response calculations were performed for switching events across this region of the $I(V)$ curve. Figure 9 shows the current density in the device as a function of position and time for an event in which the initial bias of 0.11 V (corresponding to the peak in the current) was suddenly switched to 0.22 V (corresponding to the bottom of the valley) at $t=0$. More specifically, the steady-state Wigner function for a bias of 0.11 V was used as an initial value, and the time evolution under the Liouville operator for 0.22-V bias was evaluated. The response of the current is complex, as might be expected, but shows some features that are readily interpreted. The current density initially increases throughout the structure, so that the device
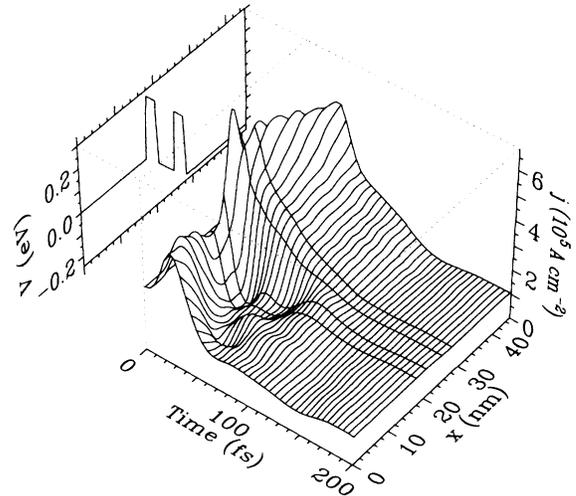

FIG. 9. Transient response of the resonant-tunneling diode of Fig. 4. Current density is plotted as a function of time and position within the device. The potential profile illustrates the device structure. At $t=0$, the voltage was suddenly switched from 0.11 V (corresponding to the peak current) to 0.22 V (corresponding to the valley current). After an initial peak, the current density approaches the lower steady-state value in 100–200 fs.
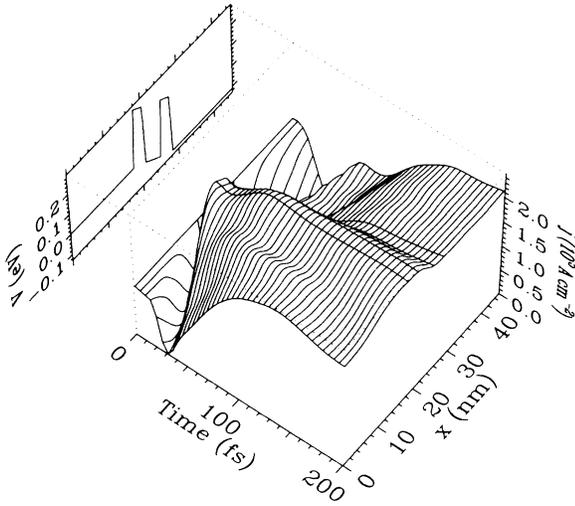
displays a positive resistance over a short time. The destructive interference which underlies the negative resistance takes some tens of femtoseconds to manifest itself. The current has settled quite near to its steady-state value after 200 fs. Of course the response of real devices will be limited by the time required to charge the device capacitance through the parasitic series resistance of the contacts. Such effects were deliberately omitted from the present model in order to observe the intrinsic response of the tunneling process itself.

A simulation of the reverse transition is shown in Fig. 10. That is, the bias was suddenly switched from 0.22 to 0.11 V. The response shows a similar short-time positive resistance, as the current density initially drops. After 200 fs the steady state has not quite been reached, as indicated by the slightly higher current density through the left-hand barrier as compared to that through the right-hand barrier. This indicates that the process of filling the quantum well is not yet complete. The difference in speed of response between peak-to-valley and the valley-to-peak transitions may be understood in terms of the shape of the energy barriers as a function of bias voltage. The higher bias voltage, and thus the higher electric field, leads to more "triangular" barriers. The more triangular barriers are effectively narrower and are therefore more transparent to tunneling currents. Thus there is more tunneling current available to complete the peak-to-valley transition, and it proceeds more quickly.

VII. DISCUSSION

The present model of an open quantum system clearly reproduces the essential features of the current-voltage curve of the resonant-tunneling diode. In addition, by incorporating the open-system irreversibility, it permits detailed calculations of the time-domain response of the de-

vice to externally applied voltages.

The open-system boundary conditions (9) are equally applicable to quantum and classical systems. In particular, they are the appropriate boundary conditions for the classical Liouville and Boltzmann equations for those cases in which a transporting medium connects two or more reservoirs (which covers the great majority of cases of interest). The demonstration (11) of the stability of the system under these boundary conditions is equally valid in the classical case.

An examination of the drift term in the discrete Liouville equation (13) shows that it has the form of a master operator. [That is, it has a (gain) — (loss) form and could therefore appear in a master equation.[38]] This suggests that an implicit Markov assumption must have been made in the derivation of (13). The Markov assumption is inherent in the boundary conditions (9). By expressing the effect of the reservoir as a boundary condition, we have suppressed the internal degrees of freedom of the reservoir. In general, such a procedure leads to a non-Markovian equation for the time evolution of the system, as the effects of the suppressed degrees of freedom are "folded" into a memory functional.[39] Some further approximation is then required to derive the conventional (Markovian) macroscopic kinetic equations.[40] In the present case, it is assumed that such a Markovian approximation exists. Physically, we would expect that the ideal reservoir behavior would be obtained when the correlation time of the reservoir is very short,[41] or equivalently, that the scattering rate in the reservoir is very high. In other words, the Markovian behavior results from the loss of information about the state of an electron as soon as it passes out of the system and into a reservoir and is quickly scattered.

The shape of the domain of the Wigner function as illustrated in Fig. 2 is significant. The density matrix $\rho(x,x')$ is essentially a spatial correlation function. In Fig. 2, the long-dashed lines indicating the system-reservoir interfaces in x and x' divide the domain into regions corresponding to the various possible system-system, system-reservoir, and reservoir-reservoir correlation functions. The domain of the Wigner function (bounded by the short-dashed lines) necessarily extends into the domain of the system-reservoir correlation function. This appears to be an essential element in the description of open quantum systems. It specifically enters the present calculations in the evaluation of Eq. (6) or (12) for the potential operator. These equations require values for the potential $v(x)$ at positions $x < 0$ and $x > l$. For this purpose the potential was assumed to extend into the reservoirs with a constant value equal to that at the system-reservoir boundary.

A consequence of the abrupt change in band structure at a heterojunction is that the effective mass also changes abruptly. This must be taken into account in the construction of the effective-mass Hamiltonian,[42,43] and this effect is readily incorporated into a density-matrix calculation.[17] When the Liouville operator is transformed into the Wigner-Weyl representation, however, the effect of the effective-mass discontinuity becomes nonlocal.[44] If the discontinuity in the effective mass is small, we may use a

semiclassical (local) approximation by placing the effective-mass factor in Eqs. (5) and (13) *inside* the gradient operator. This should adequately describe the GaAs-Al$_x$Ga$_{1-x}$As heterojunction for Al mole fractions in the direct-gap range, and this approximation was used in the present calculations.

There is no dissipation due to random scattering within the device in the present model. Energy dissipation occurs through the loss of energetic electrons to the reservoirs. The inclusion of random scattering would not seriously complicate the present calculations so long as the collision superoperator $C(x,k;x',k')$ is local in the sense that it has the form $C(k,k')\delta(x-x')$. [By "not seriously complicate the calculations," I mean that the collision operator would not add to the sparsity structure of the Liouville superoperator as defined in (13). The elements of C would of course have to be evaluated and added to the corresponding elements of L, but the algorithms and data structures required to solve for the Wigner function would remain unchanged.] The collision operator has been shown to take this form in the case of a uniform field.[45,46] An obvious first approximation would be to use a classical Boltzmann collision operator. This has not yet been done.

It would be desirable to have more-detailed microscopic models of the coupling between a device and its contacts, while still treating the contact as a reservoir. Techniques such as those used to integrate out the heat-bath variables in studies of dissipative systems[7,21,39,41] might be applied to integrate out the reservoir variables. The boundary conditions used here are perhaps a crude model, but they illustrate the essential physics of the system-reservoir interaction.[8]

VIII. RELATION TO OTHER WORK

The more traditional approach to modeling tunneling devices is to evaluate stationary scattering states.[22-24,35,47-49] This is an excellent way to obtain the steady-state behavior, particularly the $I(V)$ curve. When this approach is applied to time-dependent phenomena, however, the results are much less satisfactory. The result is expressed as a single "characteristic time."[35,47,48] (It is clear from Figs. 9 and 10 that the transient response of a tunneling device actually involves several different time constants.) There is a marked contrast between the rather qualitative nature of the derivation of this characteristic time scale and the highly detailed and precise calculations of steady-state properties in the same papers. This contrast suggests that the scattering theory is inherently unsuited to the calculation of transient phenomena.

The problem with scattering theory concerns the formulation of boundary conditions for the time-dependent Schrödinger equation. As pointed out in the discussion of Eq. (1) it is not possible to describe the transient transition from one steady state to another with finite boundary conditions that leave the Hamiltonian Hermitian. Conceptually, there is no problem if the boundary conditions are allowed to approach infinity. (The irreversibility that results in such a case comes from the infinite propagation of disturbances in the wave functions.) Boundary conditions at infinity, however, are very difficult to treat computationally. Kundrotas and Dargys[50] have dealt with this problem for the case of tunneling out of the bound state of a δ-function potential. To do so, they invoked the special properties of the bound state. The present statistical approach makes no such assumptions, and in fact does not require that a distinction be made between bound and free states.

The use of the Wigner distribution function in semiconductor device problems has been previously advocated by other workers. The published works on this subject are either primarily concerned with the formulation of the problem (rather than its solution),[44,46,51,52] or report calculations that suffer from numerical instabilities.[53] The missing element has been a proper formulation of the open-system boundary conditions. When these boundary conditions (and the logically consequent discretization scheme) are included, the direct solution of the Liouville equation for the Wigner distribution function becomes a practical technique for the analysis of quantum semiconductor devices.

ACKNOWLEDGMENTS

[1]R. Dingle, W. Wiegmann, and C. H. Henry, Phys. Rev. Lett **33**, 827 (1974).

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

[3]T. C. L. G. Sollner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker, and D. D. Peck, App. Phys. Lett. **43**, 588 (1983).

[4]G. H. Heilmeier, IEDM Tech. Dig., 2 (1984).

[5]W. R. Frensley, Phys. Rev. Lett. **57**, 2853 (1986).

[6]M. A. Reed, J. W. Lee, and H. L. Tsai, Appl. Phys. Lett. **49**, 158 (1986).

[7]E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976), and references therein.

[8]R. Landauer, Philos. Mag. **21**, 863 (1970); also in *Localization, Interaction and Transport Phenomena in Impure Metals*, edited by B. Kramer, G. Bergmann, and Y. Bruynseraede (Springer-Verlag, Heidelberg, 1985), p. 38.

[9]I. Langmuir and K. T. Compton, Rev. Mod. Phys. **2**, 191 (1931).

[10]A. Messiah, *Quantum Mechanics* (Wiley, New York, 1962), Vol. 1, p. 120.

[11]G. V. Chester, Rep. Prog. Phys. **26**, 411 (1963).

[12]U. Fano, Rev. Mod. Phys. **29**, 74 (1957).

[13]E. Wigner, Phys. Rev. **40**, 749 (1932).

[14]L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Benjamin/Cummings, Reading, Mass., 1962).

[15]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[16]W. R. Frensley (unpublished).

[17]W. R. Frensley, J. Vac. Sci. Technol. B **3**, 1261 (1985).

[18]W. Shockley, *Electrons and Holes in Semiconductors* (Van Nostrand, Princeton, 1950), Sec. 12.5.

[19]H. K. Gummel, IEEE Trans. Electron Devices **ED - 11**, 455 (1964).

[20]E. M. Conwell, *High Field Transport in Semiconductors* (Academic, New York, 1967).

[21]A. O. Caldeira and A. J. Leggett, Physica **A121**, 587 (1983).

[22]C. B. Duke, *Tunneling in Solids* (Academic, New York, 1969).

[23]E. L. Wolf, *Principles of Electron Tunneling Spectroscopy* (Oxford University Press, New York, 1985), Chap. 2.

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

[25]P. A. Lebwohl and P. J. Price, Appl. Phys. Lett. **19**, 530 (1971).

[26]C. Jacoboni and L. Reggiani, Rev. Mod. Phys. **55**, 645 (1983).

[27]P. J. Roache, *Computational Fluid Dynamics* (Hermosa, Albuquerque, NM, 1976), pp. 64–72.

[28]J. J. Duderstadt and W. R. Martin, *Transport Theory* (Wiley, New York, 1979), pp. 431–438.

[29]S. Selberherr, *Analysis and Simulation of Semiconductor Devices* (Springer-Verlag, Wien, 1984), Sec. 6.1.

[30]R. C. Miller, D. A. Kleinman, and A. C. Gossard, Phys. Rev. B **29**, 7085 (1984).

[31]B. Monemar, K. K. Shih, and G. D. Pettit, J. Appl. Phys. **47**, 2604 (1976).

[32]R. Haydock, in *Solid State Physics,* edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1980), Vol. 35, p. 215.

[33]H. Ohnishi, T. Inata, S. Muto, N. Yokoyama, and A. Shibatomi, Appl. Phys. Lett. **49**, 1248 (1986).

[34]W. R. Frensley, IEDM Tech. Dig., 571 (1986).

[35]D. D. Coon and H. C. Liu, Appl. Phys. Lett. **49**, 94 (1986).

[36]L. Lapidus and J. H. Seinfeld, *Numerical Solution of Ordinary Differential Equations* (Academic, New York, 1971), Chap. 6.

[37]R. P. Feynman, Phys. Rev. **84**, 108 (1951).

[38]C. W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 1985), Chap. 7.

[39]R. Zwanzig, J. Chem. Phys. **33**, 1338 (1960).

[40]H. Spohn, Rev. Mod. Phys. **53**, 569 (1980).

[41]W. H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, New York, 1973), Chap. 6.

[42]G. Bastard, Phys. Rev. B **24**, 5693 (1981).

[43]T. Ando and S. Mori, Surf. Sci. **113**, 124 (1982).

[44]J. R. Barker, D. W. Lowe, and S. Murray, in *The Physics of Submicron Structures,* edited by H. L. Grubin, K. Hess, G. J. Iafrate, and D. K. Ferry (Plenum, New York, 1984), p. 277.

[45]I. B. Levinson, Zh. Eksp. Teor. Fiz **57**, 660 (1969) [Sov. Phys.—JETP **30**, 362 (1970)].

[46]J. Lin and L. C. Chu, J. Appl. Phys. **57**, 1373 (1985).

[47]B. Ricco and M. Ya. Azbel, Phys. Rev. B **29**, 1970 (1984).

[48]P. J. Price, Superlatt. Microstruc. **2**, 593 (1986).

[49]M. O. Vassell, J. Lee, and H. F. Lockwood, J. Appl. Phys. **54**, 5206 (1983).

[50]J. Kundrotas and A. Dargys, Phys. Status Solidi B **134**, 267 (1986).

[51]G. J. Iafrate, H. L. Grubin, and D. K. Ferry, J. Phys. (Paris) Colloq. **42**, C7-307 (1981).

[52]J. R. Barker and S. Murray, Phys. Lett. **93A**, 271 (1983).

[53]U. Ravaioli, M. A. Osman, W. Pötz, N. Kluksdahl, and D. K. Ferry, Physica **134B**, 36 (1985).