Self-consistent study of the resonant-tunneling diode

N. C. Kluksdahl, A. M. Kriman, and D. K. Ferry
Center for Solid State Electronics Research, Arizona State University, Tempe, Arizona 85287-6206

C. Ringhofer
Department of Mathematics, Arizona State University, Tempe, Arizona 85287
(Received 8 August 1988)

Quantum transport in the resonant-tunneling diode (RTD) is modeled here with the Wigner formalism including self-consistent potentials for the first time. We examine the computational aspects of the Wigner-function approach and the boundary conditions for the model. The calculated I-V characteristics show an intrinsic bistability in the negative-differential-conductivity region of the curve. Intrinsic bistability results from charge storage and the subsequent shifting of the internal potential of the device. The cathode region of the RTD shows a strong depletion and quantization of electrons in a deep triangular potential well, which reduces the barrier height to a ballistic electron injected from the cathode, enhancing the valley current and reducing the peak-to-valley ratio. Undoped spacer layers prevent the formation of a deep quantum well at the cathode barrier, and the distribution does not deplete as sharply as without the spacer layer. The I-V curve with the spacer layers shows a much lower negative resistance, and a sharper bistable region. A finite relaxation time for the electrons increases the negative resistance, reduces the peak-to-valley ratio of the current, and causes a "soft" hysteresis in the bistable region. A zero-bias anomaly is found to result from high-momentum tails in the distribution at the barrier interface. These high-momentum tails contribute a small high-conductance current. The transient current during switching from the peak to the valley of the I-V curve shows inductive behavior and negative resistance for frequencies below 2 THz.

I. INTRODUCTION

Microfabrication technology has advanced rapidly since the dawn of the semiconductor era, with each advance giving a sizable reduction in the size of individual features. Where tens of micrometers were once the common size, devices fabricated with metalorganic chemical-vapor deposition (MOCVD) and molecular-beam epitaxy (MBE) now have features as small as a few nanometers. In addition, electron-beam lithography can now be used to make working field-effect transistors (FET's) with gates as short as 25 nm.[1] On these spatial scales, quantization effects are quite evident; electrons in a high-electron-mobility transistor (HEMT) travel in a two-dimensional sheet, a result of perpendicular quantization.[2] Fabrication of a gridlike gate extends the quantization to all three dimensions.[3] In this structure, Bloch oscillations of the electrons are expected due to the quantization effects.

One quantum structure that has been extensively studied recently is the resonant-tunneling diode (RTD).[4-6] In this structure, a thin GaAs quantum well is grown between two $Al_xGa_{1-x}As$ barriers. Each barrier, and the well, is a few nanometers thick. Thick GaAs bulk layers are grown on either side of this barrier structure. Contacts are made to the top and bottom bulklike layers. The two $Al_xGa_{1-x}As$ barriers and the GaAs well constitute a resonant-tunneling system, with a resonant energy marked by preferential tunneling (Fig. 1). In this situation the existence of a resonant level in the GaAs well provides a textbook example.[7] The I-V characteristic of this two-terminal device has a strong region of negative differential conductivity (NDC), which results from quantum tunneling in the device.

Despite the great strides made in fabrication technology, the detailed theory behind these quantum devices has not kept pace, although realization of simple quantum device models is quite easy. The first attempts at modeling the tunneling structures incorporated simple quantum tunneling.[8-11] For a given potential $V(x)$, a tunneling probability can be calculated as a function of energy $T(E)$ using the WKB approximation[7] or transfer matrices.[8,11] The current through the device is then calculated by integrating the charge density times the tunneling probability and the carrier velocity $v(E)$. For a one-dimensional system, the current (from left to right) is

$$J_{1r}=q\int dE\ T_{1r}(E)n(E)f(E)v(E)\ , \tag{1}$$

where $f(E)$ is the Fermi-Dirac distribution function and $n(E)$ is the density of states. An analogous second equation is obtained for J_{rl}, and the total current is then difference between the current from left to right and the current from right to left. The current is a function of the differences in the distribution functions and Fermi energies in the two sides.[12]

A major criticism of this approach is that it requires knowledge of the distribution of the electrons at each side of the tunneling interface, rather than the bulklike distribution far from the tunneling interface, although the

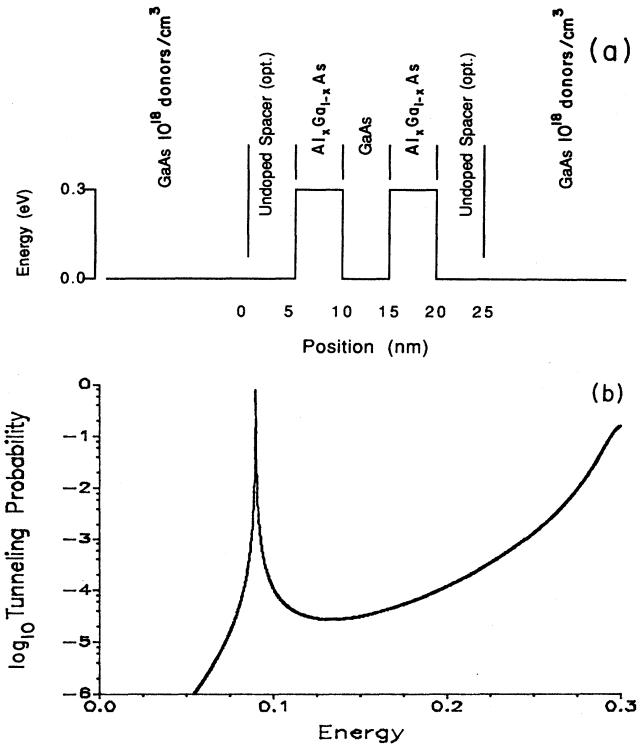

FIG. 1. Resonant-tunneling system. (a) The Γ-valley conduction band for the GaAs/$Al_xGa_{1-x}As$ resonant tunneling system. E_r is a resonant energy where the tunneling probability approaches 1. (b) The tunneling probability for the resonant-tunneling system.

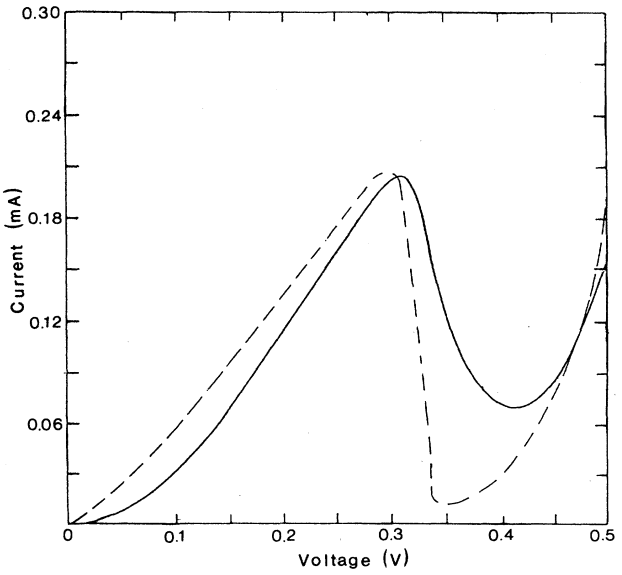

FIG. 2. Experimental (solid) and theoretical (dashed) *I-V* curves for a resonant-tunneling diode. The theoretical curve is calculated by the simple tunneling model.

latter is usually used. The distribution at the interface is completely different from the bulk distribution due to quantum repulsion from the barrier.[13,14] Failure to account for this difference leads to erroneous results.

None of the above treatments have incorporated self-consistent–potential effects. It is known that self-consistent bending of the potentials within the device may lead to the formation of bound states, or other types of quantization in the interface region.[15] Accumulation or depletion layers, and their effects upon the potential and, consequently, the tunneling probability, can drastically affect the local distributions and hence the tunneling current.

The primary failure of the simple tunneling models, though, lies in the extremely poor fit of the predicted *I-V* curve to that observed experimentally. Figure 2 shows a comparison of an *I-V* curve calculated by the tunneling approach and the experimentally observed *I-V* curve for a device with the same structural parameters. The current predicted for the valley of the curve is far too small. The model cannot account for barrier lowering or ballistic injection over the barriers, both of which can be important experimentally.

Despite the quantitative failures of the tunneling model, this simple theory provides tremendous insight into the appropriate role for a contact.[16–21] An "ideal" contact has been defined as an infinite reservoir of thermally distributed carriers, which are injected as needed into the device. For carriers leaving the device, the "ideal" contact acts as a perfect sink, absorbing all carriers incident upon the contact without reflection. This contact serves as a boundary for quantum correlation; all incoming carriers are randomly distributed, and outgoing carriers are immediately randomized within the contact. Büttiker[21] has used this role of the contact reservoir to introduce dissipation into ballistic structures, and subsequently addressed the role of contact resistance, a point to which we return below.

In this paper, we will present a fully-self-consistent model of the RTD based upon the quantum-mechanical Wigner function. Though the Wigner function has been around for quite some time, and its properties have been well investigated,[22–25] it has only recently been applied to electronic transport.

The Wigner formalism offers many advantages for quantum modeling. First, it is a phase-space description, similar to classical Boltzmann distributions. In the Wigner formalism, scattering is a local phenomenon.[25] Because of the phase-space nature of the distribution, it is conceptually possible to use the correspondence principle to determine where quantum corrections enter a problem. At the boundaries the phase-space description permits separation of incoming and outgoing components of the distribution, which thus permits modeling an ideal contact, and hence an open system. Still another advantage is that the Wigner function is purely real, which simplifies calculation and interpretation of results. By coupling the Wigner-function equation of motion to the Poisson equation, we obtain a fully-self-consistent model of the RTD. This then allows us to examine charge-redistribution effects and many properties of the devices.

In the field of quantum optics, Wigner functions have been widely used. The quantum phase-space distribution has been used to describe coherence of optical fields and

to describe polarization and transient superposition effects.[26] Quite naturally, this description has been applied to finding quantum-mechanical solutions for a laser master equation (the Fokker-Planck equation[27]) and for describing quantum noise in lasers.[28] More recently, Wigner functions have been applied to optical systems and signals, where they provide a link between Fourier optics and geometric optics.[29,30] Two-dimensional Wigner optical distributions have been generated,[31] as well as slices of four-dimensional Wigner optical distributions.[32] The Wigner-function description of optical signals has been investigated for use in elementary pattern recognition.[33] From this, we may easily recognize the ability of the Wigner function to exhibit phase interference and quantum resonances.

Numerous works have expounded on the virtues of the Wigner-function formalism for quantum electronic transport, with an eye toward the advantages offered by a phase-space representation.[23–25,34,35] Some early attempts at modeling transport in the RTD used the Wigner formalism,[14,36] while other models advocated use of the density matrix.[37] The density-matrix approach, though quantum-mechanically correct, possesses some serious drawbacks. The density matrix is complex, and except for the diagonal terms which yield the particle density $n(x)$ the density matrix does not correspond to any well-known classical distribution function. Additionally, scattering is nonlocal in the density-matrix formalism, which complicates the inclusion of scattering in the transport model.

In the next section we will review the development of the Wigner function and its equation of motion. The roles of nonlocality and correlation in the Wigner function will be discussed. The critical nature of a quantum-mechanically correct initial distribution, and its subsequent temporal development, will be examined.

Major concerns in modeling are the stability and the convergence of the numerical techniques used in the model, which will then be discussed. Because the model is one of an open system, the nature of the boundaries will be treated at length. Poisson's equation will be coupled with the Wigner-function equation of motion to include fully-self-consistent potentials, and simple scattering will be added to the model.

Finally, the self-consistent Wigner-function model will be applied to the RTD, and the steady-state and transient behavior will be examined. Bistability, a source of controversy in the RTD, is shown to occur, and the conditions for this will be explored with the model. A zero-bias anomaly, observed previously in tunneling structures, is found to occur primarily as a result of the distribution functions near the barriers. Finally, the effect of variations in doping of the RTD structure will be explored.

II. THE WIGNER DISTRIBUTION FUNCTION

A. Nonlocality of the Wigner function

Suppose a system is in a general state described by the density matrix. If we use a set of position eigenstates, the density matrix is

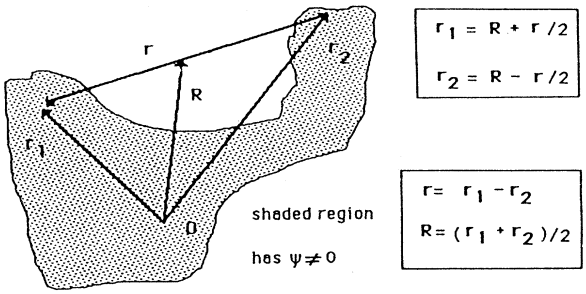

FIG. 3. Illustration of the nonlocal nature of the Wigner function (Ref. 47). Outside the shaded region, the wave function is zero. In center-of-mass coordinates, a distribution may be nonzero in regions that are forbidden.

$$\rho(x_1,x_2)=\langle x_2|\hat{\rho}|x_1\rangle \ , \tag{2}$$

where $\hat{\rho}$ is the density operator. The density matrix is clearly a function of the two positions x_1 and x_2. If we use a relative vector x between the points x_1 and x_2,

$$x=x_1-x_2 \ , \tag{3}$$

and a "center-of-mass" vector X to the mean of the positions x_1 and x_2,

$$X=\tfrac{1}{2}(x_1+x_2) \ , \tag{4}$$

then the density matrix may be described in terms of these two new coordinates as $\rho(X+\frac{1}{2}x,X-\frac{1}{2}x)$. The issue of the "locality" of the chosen coordinates must be examined.[25,34] In Fig. 3 it is evident that a distribution defined in terms of the "center-of-mass" coordinates may be nonzero at positions X where the wave functions are zero.

The Wigner function may be defined in terms of momentum eigenstates and the density matrix. In a device such a definition is useful, since the device is usually described by a mixed state. The density matrix $\rho(x_1,x_2)$ may be written in terms of the new coordinates as

$$\rho(X,x)=\rho(X+\tfrac{1}{2}x,\, X-\tfrac{1}{2}x) \ . \tag{5}$$

The relative vector x is a natural choice for transformation to the momentum representation (the transformation used is the Wigner-Weyl transformation), which leads to the definition of the Wigner function,[22–24]

$$f_W(X,p)=\frac{1}{2\pi\hbar}\int dx\, e^{ipx/\hbar}\rho(X+\tfrac{1}{2}x,\, X-\tfrac{1}{2}x) \ . \tag{6}$$

It is easily seen that there is no requirement in the definition of the Wigner function, (6), which requires it to be a positive quantity. For this reason, the Wigner function interpretation as a probability distribution function must be handled carefully;[38] nevertheless, the Wigner function is quite useful.[22]

In order to use the Wigner function to study devices, an equation describing the response of the Wigner function to changing conditions is required. The equation of motion of the Wigner function is derived from the quantum Liouville equation by applying the Wigner-Weyl transformation. For a general noninteracting Hamiltonian,

$$H=\frac{\hat{p}^2}{2m}+V(x)\ , \tag{7}$$

application of the Wigner-Weyl transformation gives the equation of motion for the Wigner function:[22–25]

$$\frac{\partial}{\partial t}f_W(x,p,t)-\frac{p}{m}\frac{\partial}{\partial x}f_W(x,p,t)$$
$$=\frac{1}{2\pi\hbar}\int dP\, M(x,y,P)f_W(x,p+P,t)\ , \tag{8}$$

where

$$M(x,y,P)=\int dy\, e^{iPy/\hbar}[V(x+\tfrac{1}{2}y)-V(x-\tfrac{1}{2}y)]\ . \tag{9}$$

The kinetic term is identical to the kinetic term of the Boltzmann equation. The potential term in (9), however, is nonlocal in the position of the potential and in the momentum of the distribution function. These nonlocalities give rise to the quantum corrections in the equation of motion.[34]

For a density matrix represented by one set of the conjugate coordinates of position and momentum, evaluating the Wigner-Weyl transform is at the very least tedious when dealing with Hamiltonian terms which involve the other set of coordinates, i.e., evaluating the kinetic term when the density matrix is in the position representation. For terms which are mixed, as when working with a magnetic field in the Landau gauge, evaluation of the Wigner-Weyl transform is quite involved. The contribution of a term from the Hamiltonian to the equation of motion may be found by means of another formalism.[23] For a general Hamiltonian H, the equation of motion is

$$\frac{\partial}{\partial t}f(x,p,t)=\frac{i}{\hbar}\int d\eta\int d\xi\int d\Theta\int d\tau\, e^{i[\tau(\eta-p)+\Theta(\xi-x)]}f(\xi,\eta,t)[H(\xi-\tfrac{1}{2}\hbar\tau,\eta+\tfrac{1}{2}\hbar\Theta)-H(\xi+\tfrac{1}{2}\hbar\tau,\eta-\tfrac{1}{2}\hbar\Theta)]\ . \tag{10}$$

Using this powerful equation, otherwise difficult terms of a Hamiltonian are relatively easily transformed.

The Wigner function is closely related to the correlation Green's function $G^<$ of the quantum Boltzmann equation.[39–41] The difference between the two lies in the fact that the Wigner function is defined as a one-time, two-position function, whereas the Green's function $G^<$ is a two-time, two-position function, i.e., $G^<(x_1,t_1,x_2,t_2)$, or $G^<(X,T,x,\tau)$ in terms of the center-of-mass coordinates. The relative coordinates x and τ may be Fourier transformed into momentum and frequency variables, giving $G^<(X,T,p,\omega)$. The single-time limit of $G^<$ is found by integrating over the frequency ω, giving the Wigner function. Since $G^<$ is explicitly a correlation function, the single-time limit of $G^<$ is also a correlation function; the Wigner function thus automatically incorporates spatial correlations. Earlier, it was shown that the Wigner function may be nonzero in nonphysical regions; this component of the Wigner function exhibits the correlation between the allowed wave functions. (See Fig. 3.)

B. Operators

In a quantum-mechanical system desired information is often found by evaluating expectation values of operators. Since the Wigner formalism is a fully quantum formalism, expectation values of operators contain much useful information. The use of Wigner functions permits easy evaluation of expectation values. For an operator $\hat{A}$, the expectation value is

$$\langle\hat{A}\rangle=\int dx\ \hat{A}\rho(X+\tfrac{1}{2}x,X-\tfrac{1}{2}x)\ . \tag{11}$$

By introducing

$$1=\frac{1}{2\pi\hbar}\int dp\, e^{ipx}\ , \tag{12}$$

Eq. (11) becomes

$$\langle\hat{A}\rangle=\int dx\int dp\ \hat{A}f_W(x,p)\ . \tag{13}$$

It has been demonstrated that (13) holds for powers of operators which are linear functions of the operators $\hat{x}$ and $\hat{p}$.[22]

An operator of extreme importance in device modeling is the current operator. The current may be defined through the operator

$$\hat{j}=e\hat{p}/m\ . \tag{14}$$

Within a quantum device, however, the properties of the Wigner distribution function require care in interpreting current. In a classical system a local current density $J(x)$ may be exactly defined, since the distribution is a true probability distribution, and both position and momentum are exactly known. As dictated by the Heisenberg uncertainty principle, however, exact knowledge of both position and momentum in a quantum system is impossible. This manifests itself in the Wigner function as negative-valued distribution, which may be thought of as negative probability.[38] While we can use (14) in many cases, we must be cautious that its use is limited.

C. Sensitivity to initial condition

A serious consideration of the Wigner-function equation of motion (8) is the entry of quantum-mechanical effects through the nonlocal potential term. It is a trivial exercise to show that, for a potential which is quadratic, the potential term of (8) reduces to

$$[V,\rho]_{WW}=e\mathbf{E}\cdot\nabla_p f_W(x,p,t)\ , \tag{15}$$

which is exactly the classical term from the Boltzmann equation.[42,43] The source of quantum corrections has vanished, and the equation of motion is classical. The quadratic potential, however, could describe a purely quantum harmonic oscillator. Using the Wigner-function equation of motion alone to determine the state of this

system will not reproduce the required quantum-mechanical eigenstates of the system.

A correct quantum-mechanical steady-state solution to the problem may be found by specifying the correct boundary conditions and solving (8) with the time derivative set to zero. The fallacy in this procedure is subtle. The correct quantum-mechanical boundary conditions presuppose knowledge of the state of the system at the boundaries, which are a function of the internal potential. Thus, knowledge of the boundary conditions implies knowledge of the solution without the Wigner-function equation of motion.

In order to include all orders of quantum corrections, one of two things can be done. The first is to extend the computational domain sufficiently far from the source of quantum effects that the system is classical. A classical distribution may then be used as the boundary. It has been demonstrated that quantum corrections "heal" over several thermal wavelengths.[13] In a reasonable GaAs device at 300 K, this length is nearly 100 nm.

The second approach to incorporate quantum effects is to use an adjoint equation to determine the initial Wigner function. The adjoint equation must be quantum mechanically correct to give a proper initial Wigner function. Equation (8) may then be numerically integrated to find steady-state solutions which are certain to include all orders of quantum corrections. This is the approach we will take.

D. Computation of the initial state

One method of finding a quantum-mechanically correct initial distribution is to Fourier transform the density matrix $\rho(x,x')$. The density matrix, in turn, may be calculated using a scattering-state basis.[13] If the potential approaches a constant as $x \rightarrow \pm\infty$, i.e., $V(x)=V^-$ for $x<x^-$ and $V(x)=V^+$ for $x>x^+$, the basis states are plane waves, i.e., $\psi \propto e^{\pm ikx}$. For equilibrium, where $V^-=V^+$, this gives, for states incident from the left with $k>0$,

$$\psi_k(x)=\left[\frac{1}{2\pi}\right]^{1/2}[e^{ikx}+r(k)e^{-ikx}] \tag{16}$$

for $x<x^-$, and

$$\psi_k(x)=\left[\frac{1}{2\pi}\right]^{1/2}t(k)e^{ikx} \tag{17}$$

for $x>x^+$, where $t(k)$ is the transmission coefficient and $r(k)$ is the reflection coefficient for wave vector k. In a similar manner, states incident from the right are defined by

$$\psi_k(x)=\left[\frac{1}{2\pi}\right]^{1/2}[e^{-ikx}+r(k)e^{ikx}] \tag{18}$$

for $x>x^+$, and

$$\psi_k(x)=\left[\frac{1}{2\pi}\right]^{1/2}t(k)e^{-ikx} \tag{19}$$

for $x<x^-$. The density matrix is then defined by

$$\rho(x,x')=\frac{1}{Z}\left[\sum_n \psi_n(x)\psi_n^*(x')f(E_n) + \int dk\ \psi_k(x)\psi_k^*(x')f(E(k))\right], \tag{20}$$

where Z is the partition function, the sum over n is over bound states with energies E_n, $E(k)$ is the energy of a scattering state, and $f(E)$ is the distribution function (Boltzmann or Fermi-Dirac).

From an unnormalized basis function, and using translation matrices, an unnormalized state may be computed on the entire domain. These states are normalized by applying scattering theory, in which wave functions in the presence of a scatterer are compared to those in a reference space. These are related through the Lippmann-Schwinger equation. A useful consequence of this equation is that the scattering states satisfy precisely the same orthonormality relations as unperturbed states.[13] Each state contributes to the density matrix according to the thermal distribution function $f(E)$. The partition function is found by considering the limit of $x \rightarrow x^-$, also a consequence of the normalization conditions. It is defined by

$$Z^{-1}=e^{\beta V}2\sqrt{\pi\beta}\lim_{x\rightarrow-\infty}\rho(x)\ . \tag{21}$$

An algorithm for computing the density matrix is thus available.[13] A set of points $\{x\}$ is chosen at which the density matrix is desired. The potential is taken to be piecewise constant between the points. Energies are randomly sampled according to the distribution function. Each energy gives a wave vector k for a left-incident state and a right-incident state. The states are translated and normalized and their contributions added to the density matrix. The density matrix is normalized with the partition function, and the resultant density matrix is Fourier transformed to the Wigner function. The computational region of the density matrix satisfies the conditions outlined above.

The density matrix and the resultant Wigner distribution are calculated for the resonant-tunneling diode.[43] Figure 3 shows the Wigner distribution function. The Wigner function calculated by scattering states is characterized by a thermal distribution far from the barriers. Oscillations in the distribution near the barrier are evident. These oscillations are a result of quantum repulsion from the barrier.[13,14] The quantum repulsion causes the density of carriers to deplete, although this depletion is not a depletion layer in the classical sense. The depletion is caused by the quantum repulsion rather than by band bending. The quantum repulsion is, in a sense, complementary to barrier penetration: just as a nonzero density penetrates a finite distance into a classically forbidden region, a density deficit extends a finite distance into a classically allowed region. This charge then accumulates a short distance from the barrier, and overall charge neutrality in the device is maintained (see Fig. 4).

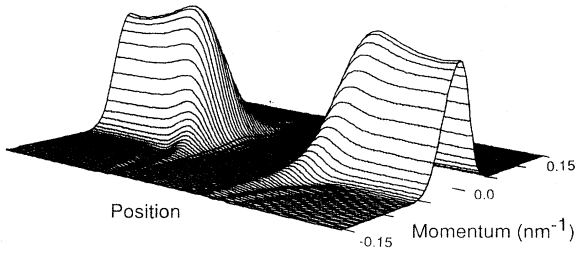

FIG. 4. Self-consistent equilibrium Wigner function for a resonant-tunneling diode, calculated from a scattering-state basis. Depletion at the barrier interface stems from quantum repulsion; this is compensated by an accumulation further from the barrier.

III. NUMERICAL TECHNIQUES

The validity of a numerical solution to an equation depends to a large extent upon the choice of numerical approximations, discretization, and the boundary conditions. We discuss these effects in this section.

A. Discretization

The Wigner-function equation of motion is evaluated on a two-dimensional discretized grid for the position and momentum variables x and p. The modeled region is a domain from $x=0$ to $x=L$, and is divided into a spatial mesh, with a mesh size Δx chosen so that features of interest, such as potential barriers, are adequately represented. Since the potential in a RTD varies over a distance of a few atomic monolayers, an appropriate spatial mesh Δx is of the order of a unit cell of the GaAs lattice, and we approximate that with $\Delta x=0.25$ nm. The mesh size for the momentum variable is found by considering the Fourier transform in (3) which defines the Wigner function. The discrete Wigner function is periodic in momentum, with a period of $\hbar\pi/\Delta x$. (For convenience, the momentum is expressed in terms of the wave vector $k=p/\hbar$.) This period in momentum is discretized into a convenient number of meshes, and ranges from $-\pi/2\Delta x$ to $\pi/2\Delta x$. The discretized grid may conveniently be split into two regions, one half with positive momentum and the other half with negative momentum.[36,44] This will become important in the following discussion.

B. Stability and convergence

A common numerical technique for solving discretized temporal equations is explicit differencing. Stability of an explicit scheme requires that error in the discretized equation remains bounded. Fourier analysis of the growth of error in these schemes has led to the Courant-Friedrichs-Lewy stability criterion[45] (CFL condition) for explicit finite-difference approximation schemes: for a mesh size Δx and a time step Δt a necessary and sufficient condition for stability is that

$$\frac{\Delta x}{\Delta t} \leq v \ , \tag{22}$$

where v is the velocity of the fastest component of the solution. For a time step Δt which satisfies (22) for the maximum velocity, convergence and stability of the solution is assured for linear problems.

Equation (8) is discretized using Lax-Wendroff explicit time differencing.[46] The Lax-Wendroff method retains the second-order terms in the Taylor expansion of $f(t+\Delta t)$. This introduces a second-order spatial difference term into the equation of motion. This second-order term represents an artificial diffusion, which acts to counteract spurious numerical diffusion that always arises from the first-order terms. The Lax-Wendroff scheme has proven to be the technique of choice of numerical solutions of the Schrödinger equation.[14,47]

Each point in a discretized equation has a characteristic direction, which is to say that each point has a direction of propagation. For a phase-space representation, the velocity which defines propagation is directly proportional to the momentum. For positive momentum, propagation is in the positive x direction, while negative momentum causes propagation in the negative x direction. For any given point, the characteristic direction is defined by the local momentum. In the discretized Wigner function, there is a "slice" of the mesh over which the momentum is constant. These slices of the mesh can be viewed as systems of equations which are coupled through the potential term of (8). Each "slice" of the mesh is characterized by the same characteristic velocity; for positive momentum, information flows into the domain from the boundary at $x=0$, and moves toward the boundary at $x=L$, where it leaves the domain. For negative momentum the characteristic direction and the roles of the incoming and outgoing boundaries are reversed. Recognition of these characteristics directions solves an inherent problem.

A difficulty with the Lax-Wendroff discretization involves the boundaries. A second-order finite-difference term at a point x_i involves the points x_{i-1} and x_{i+1}. In the interior of the device this creates no problem, nor does it cause difficulty with the incoming boundary, where the distribution is specified. The problem occurs on the outgoing boundaries. The function is not known beyond the boundary, making the definition of a second-order difference at the boundary impossible. The solution to this dilemma is to use first-order upwind differencing[46] to propagate the function to the outgoing boundary along the characteristic direction, which is

$$\frac{\partial f}{\partial x} = \begin{cases} \dfrac{f(x_i)-f(x_{i-1})}{\Delta x} & \text{if } p(x_i)>0 \ , \quad (23a) \\ \dfrac{f(x_{i+1})-f(x_i)}{\Delta x} & \text{if } p(x_i)<0 \ . \quad (23b) \end{cases}$$

Stability of the first-order terms is also dependent upon satisfying the CFL condition. The discretization is illus-

trated in Fig. 5. The characteristic directions on the mesh are indicated, as are the incoming and outgoing boundaries on the regions of the mesh and the type of discretization used for each mesh point.

Since virtually every elementary quantum-mechanics textbook studies the evolution of a Gaussian wave packet (GWP), convergence of the equation of motion is tested using this reference.[43,48] The Wigner-function description of a Gaussian wave packet is

$$f_W(x,k)=\frac{1}{2\pi\hbar}\frac{1}{a\sqrt{2\pi}}e^{-x^2/4a^2}e^{-4(k-k_0)^2a^2} . \quad (24)$$

This distribution is Gaussian in both position and momentum.[49] A further property of the resulting distribution is that it satisfies the Heisenberg uncertainty principle. For a Gaussian distribution in position, the uncertainty is the spread in position. In (24), Δx is $2a$. The uncertainty in momentum is similarly the momentum spread, which is $\Delta k = 1/2a$. This gives

$$\Delta x\,\Delta k = 2a\frac{1}{2a}=1 . \quad (25)$$

With this analytic expression for the Wigner function of a Gaussian wave packet, the properties of the equation of motion may be investigated.

A GWP in free space spreads as it evolves in time. Further, when a wave packet is defined with some initial nonzero wave vector k_0, the centroid of the wave packet moves with a velocity $\hbar k_0/m$, the group velocity of the packet.[7] In the absence of a potential, (8) becomes

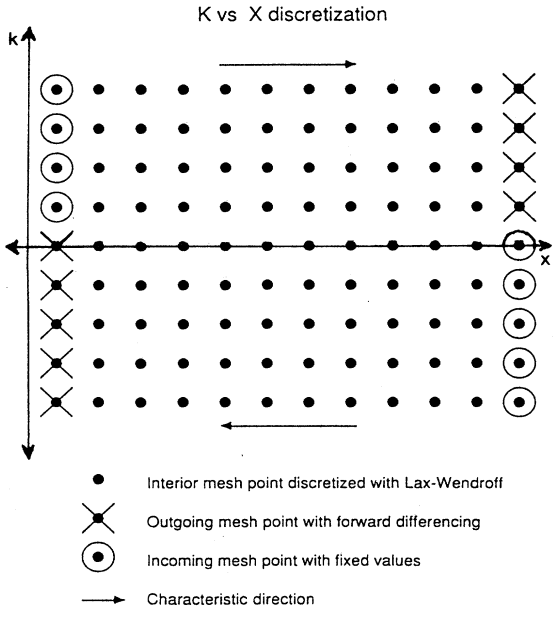

FIG. 5. Phase-space discretization for the Wigner function. Each row of the grid has a uniform velocity, and can be thought of as a subsystem, or slice, of the distribution.

$$\frac{\partial}{\partial t}f_W(x,p,t)-\frac{p}{m}\nabla_x f_W(x,p,t)=0 . \quad (26)$$

A Wigner function of a GWP may be used as the initial condition for (8), which is numerically stepped forward in time. The resulting Wigner functions are integrated over momentum, giving a distribution in position only. The wave function spreads in position, as expected, and the centroid of the wave packet moves with the group velocity, $\hbar k_0/m$. The Wigner function is integrated over both position and momentum at all time steps. Any variance in the value of this integral is solely due to numerical errors in the algorithm, since the analytic solution conserves distribution. The GWP can be used to evaluate this stability. After 200 time steps, the Lax-Wendroff scheme shows less than 1% numerical error. As a comparison, simple first-order finite differencing of (8) shows more than 10% error after the same number of time steps.[50] The hybrid discretization demonstrates good stability and convergence.

Next, the GWP is allowed to interact with a pair of quantum barriers separated by a quantum well.[43] The barriers are each 3 nm thick and 0.3 eV high, while the well is 5 nm wide. Figure 6(a) shows the initial wave packet as it begins to interact with the barriers, with the positions of the barriers indicated by the dark lines. In Fig. 6(b) the majority of the wave packet has reflected from the barrier, while a small portion has tunneled through. Finally, in Fig. 6(c) most of the wave packet has reflected from the barrier. Some of the wave packet has tunneled through the barriers, and is leaving the region of simulation.

Of particular note in Fig. 6(c) is the rapidly oscillating structure along the $k=0$ axis. This structure contains the important correlation between the transmitted and reflected wave packets. As long as this information is retained, the system has time-reversal symmetry. When the correlation is removed, the original wave packet can no longer be recovered.[43,47] Any dissipative processes will act to remove this correlation, and thus destroy reversibility.

Equation (8) may readily be used to check the correctness of an initial distribution. The initial potential and distribution are inserted into (8), and the system is stepped forward in time. The correct equilibrium distribution will not vary under the equation of motion. Because the numerical scheme is stable and convergent, no errors will be introduced through the equation. The only source of error, therefore, is from an incorrect initial distribution. We check this by using the initial distribution calculated in the preceding section. The system is stepped 500 fsec in time, or 5000 time steps, with zero applied bias, after which the state of the system is compared to the initial state. For an initial Wigner function with a maximum value of 1.0, the maximum error in the time-evolved Wigner function was 2.18×10^{-6}, and the L^2 norm of the error was less than 10^{-10}. As a further check on the temporal stability of the initial Wigner function, the current at the contacts was calculated. With zero applied bias, the current calculated at the left contact was 0.038 pA/cm^2, while the current at the right

contact was -0.013 pA/cm^2, in comparison to the normal operating currents of milliamperes/cm^2. Thus, the integrated error in current is about 10^{-5} of the actual current.

C. Contacts

Simulation of a real device includes some model of the interface between the device and an external circuit. At the very minimum, the external circuit consists of a "battery" which fixes a potential across the device, and "wires" which carry current from the "battery" to the device. These circuit parameters are usually included in device simulations as boundary conditions. In many models, the contact serves as an infinite reservoir of thermally distributed carriers.[16–21] This reservoir maintains a fixed distribution at the contact where the particles enter the simulation domain.

Conditions at the contact must be consistent with physical reality. If a current is flowing through the device, current continuity requires that an identical current be flowing through the external circuit. This implies that the external circuit, and thus the contact, must be characterized by some distribution which reflects current flow, such as a shifted thermal distribution.[51]

In (8), we have left and right boundaries, at $x=0$ and $x=L$, respectively. Mathematical constraints permit us to specify only one boundary, however. As discussed earlier, the model may be thought of as a coupled set of systems, each of which is a slice of constant momentum (see Fig. 5). Each of these momentum slices has a boundary, or contact, from which electrons enter the slice, and they leave the slice from the opposite boundary.

The "entering" contact is constrained by the current. Just inside the device a current exists. Because an identical current must be present in the contact, and the distribution within the contact is assumed to be a shifted thermal distribution, the amount of shift and therefore the distribution within the contact are known. The distribution within the contact is then matched to the corresponding momentum slice of the model. This procedure must be self-consistently carried out, since the current within the device is a function of the boundary conditions, and the boundary conditions are functions of the current within the device.[51] Because each momentum slice has only one boundary prescribed, the boundary conditions are mathematically consistent with (8).

A second consideration of the contact is that it remove all carriers which are leaving the device.[16–21] This defines the "leaving" contact. In the so-called "ideal" contact the carriers are perfectly extracted from the device without reflection. Studies of (8) with a Gaussian wave packet have shown that the momentum nonlocality in the equation couples the incoming and outgoing boundaries at the contacts, and serves as a source of artificial numerical reflections.[36] Such phenomena have been observed in other well-posed numerical systems,[52] and arise strictly from the coupling at the boundaries. Such artificial reflections mar the concept of an ideal contact.

A mathematical decoupling algorithm[52] has been adapted for (8). Artificial reflections are effectively removed, and the ideality of the contacts is restored. In the model of an ideal contact, electrons which leave the device are instantly scattered within the contact into a thermal distribution.[12] The scattering removes correlation and introduces irreversibility. In an actual device ideal contacts are not physically realizable, and this decoupling procedure may not be the best choice from a physical standpoint; nevertheless, modeling ideal contacts is useful in that it removes size dependencies by breaking up correlation at the boundaries. This is important, since Fig. 5(c) illustrates that correlation can extend over a long range. In the contact scheme used here, the correlation length is limited to the device length.

D. Self-consistent potentials

Fully-self-consistent potentials are introduced by adding Poisson's equation to the system. A self-consistent initial Wigner distribution is calculated which includes the background charge from the ionized donors. Knowledge of the temperature and doping level establishes the Fermi level within the bulk GaAs,[53] and thus specifies the ionized-donor concentration and the electron concentration far from the barriers. Since the electron density is a function of the potential, and the ionized-donor charge and potentials are interdependent, an iterative procedure is required to find the correct initial distribution. A guess is made for the potential, which is used to calculate the density $n(x)$, which is the trace of the density matrix found from the scattering states. The potential is also used to calculate the ionized charge density $N_d^+(x)$. The total charge is then used to solve Poisson's equation to get a new potential. Because the ionized charge density is exponentially dependent upon the potential, the two quantities must be iteratively made consistent for the electron charge $n(x)$. The new potential is then used to calculate the density. The entire process is iterated until full consistency between $n(x)$, $V(x)$, and $N_d^+(x)$ is achieved. The density matrix is then transformed into the Wigner function, which together with the background charge $N_d^+(x)$ and the self-consistent potential $V(x)$ is input into (8) and Poisson's equation. Solution proceeds by stepping (8) forward, then adjusting the potential via Poisson's equation. In addition to an unchanging Wigner distribution, steady-state conditions may be tested by examining the time-variant fluctuations in the internal potentials. When both the Wigner function and the potentials become invariant, steady-state conditions have been reached.

E. Scattering

In a quantum-mechanical system proper inclusion of dissipation is nontrivial. The simplest possible approach is the relaxation-time approximation. A second approach is the use of the classical scattering rates,[54] such as are used in Monte Carlo calculations. The most difficult approach is the use of Levinson's formalism,[24] which is the Wigner-Weyl transformation of the interaction terms of the Hamiltonian. Although this method properly introduces the intracollisional field effect,[42] the transformation

is entirely nontrivial, even for simple interactions. The quantum Boltzmann equation[39,40] includes scattering terms in its formulation, which could be transformed into the Wigner formalism with some difficulty. A major drawback of this approach is that the scattering terms involve a second function, which is not clearly related to the Wigner function, so that another equation would be necessary with this approach.

We choose to model scattering with the relaxation-time approximation, which lumps all dissipation processes into one macroscopic parameter.[55] The relaxation-time approximation is expressed as

$$\left[\frac{\partial f}{\partial t}\right]_{\text{collision}} = \frac{f - f_0}{\tau} , \tag{27}$$

where f is the distribution function, f_0 is the equilibrium distribution, and τ is the relaxation time. Although crude in comparison to the other formalisms, the resulting expression is trivially evaluated. This approach has been shown to effectively remove correlation and to shift the I-V characteristics of earlier models of the RTD.[43]

IV. APPLICATION TO THE RESONANT-TUNNELING DIODE

A. Structure to be simulated

The self-consistent Wigner-function model has been applied to a resonant-tunneling diode (RTD). The structure under consideration has two $Al_xGa_{1-x}As$ quantum barriers which are 0.3 eV high and 5 nm thick. These barriers are separated by a 5-nm GaAs quantum well. Outside the barriers, the device is GaAs, doped with 10^{18} donors cm^{-3}. The barriers and the quantum well are undoped. The structure is illustrated in Fig. 6. The relaxation time used to model scattering is chosen to correspond to a mobility of 3000 cm^2/V sec.

B. I-V characteristics

The steady-state I-V curve of the device is calculated by applying an incremental bias potential to the cathode contact, then stepping (8) and Poisson's equation until steady-state conditions are reached. Earlier work with the Wigner-function model of the RTD showed that large-signal transients decay exponentially, and that steady-state conditions are achieved in a few hundred femtoseconds.[43,56] We step the system to a time of 1.5 psec to ensure that steady state has been achieved, while checking for invariance of the distribution function and the potential. A transient analysis (to be discussed in detail later) with the self-consistent model shows that transients have essentially vanished within 700 fsec.[57] From the steady-state conditions, the current in the device is calculated, and the bias potential is again incremented. After the bias has been increased to its maximum, the potential is decremented toward zero, with the current being calculated along the way. The resultant I-V curve, shown in Fig. 7, shows a peak-to-valley ratio of approximately 2:1 at 300 K.

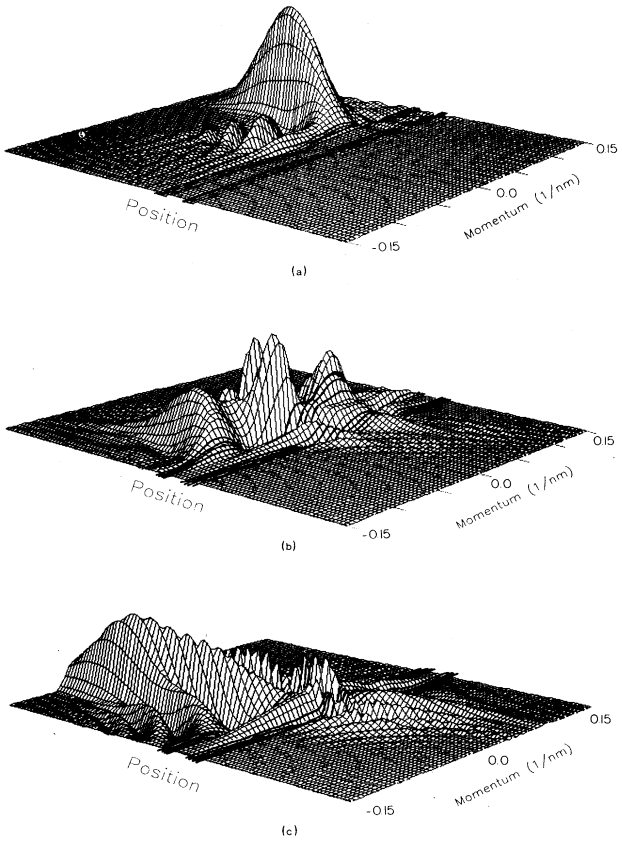

FIG. 6. Gaussian wave packet interacting with resonant quantum potential barriers. The barriers are indicated by the dark bands. (a) The incident wave packet, moving from left to right, is just beginning to interact with the barriers. (b) Gaussian wave packet during reflection. The incident and reflected components are visible, as is the correlation centered around $k=0$. Part of the packet is tunneling through the barrier. (c) Gaussian wave packet after reflection. Most of the wave packet has been reflected. The tunneling packet is visible to the right of the barriers.

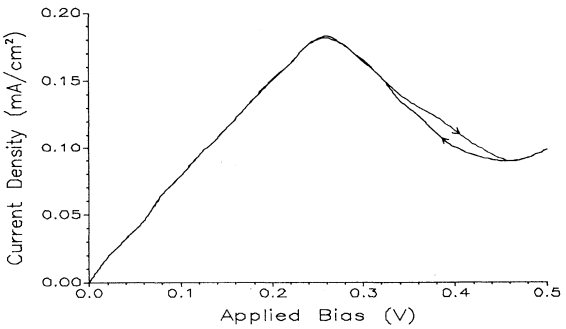

FIG. 7. Self-consistent I-V curve for the RTD. The potential is increased to a maximum, then decreased. The I-V curve shows a hysteresis resulting from redistribution of the potential and charges.

Early modes of the RTD assumed that all potential was dropped across the barrier-well structure.[43,44,48,58] The self-consistent internal potential of the device, for an applied bias near the peak of the *I-V* curve, illustrated in Fig. 8, shows that only about one-third of the potential is dropped across the barriers. Of the remaining potential, a majority is dropped across the cathode end of the device. At the cathode end of the device the Wigner distribution shows heavy depletion, as seen in Fig. 9. As bias is applied, this region is pulled down, and would be expected to show accumulation of electrons. In fact, at very low biases, this region does accumulate slightly. As the bias is increased, however, the cathode-barrier–interface region is lowered in energy even more, until it forms a shallow triangular quantum well with quantization of electrons, similar to the quantization in an enhancement metal-oxide semiconductor field-effect transistor[15,59] (MOSFET) or a HEMT.[60] An overall depletion of electrons in the cathode occurs. As the bias increases further, the quantum state becomes evident in the Wigner distribution illustrated in Fig. 10. Quantization causes the formation of the ring structure to the left of the barriers. In a phase-space representation, the structure of the quantized state and the oscillatory nature of electrons in the state are quite evident.

The deep triangular well at the cathode is a physical result of cathode fields and an inherent contact resistance. It has implications for the $I-V$ curve as well. At the valley of the *I-V* curve, where the quantized state is plainly evident, Fig. 8 shows the potential within the device. Because the well is very deep, the top of the barrier is at nearly the same energy as the cathode contact. Electrons which are injected from the cathode and travel to the barrier without scattering are able to travel over the quantum barriers. This process leads to a greatly enhanced valley current, and reduces the peak-to-valley ratio of the device.

C. Bistability

A controversial aspect of the RTD is the source of observed bistability in the *I-V* curve in the negative-differential-conductivity (NDC) regime. Intrinsic bistability is thought to result from charge storage within the quantum well, which changes the potential and field near the barriers.[5] This modifies the positions of the conduction-band edges on either side of the barriers and of the resonant level in the well. Since the current is strongly affected by the positions of these bands and the resonant level, subtle shifts in stored charge can greatly affect the current tunneling through the resonant level between the barriers. Extrinsic bistability, on the other hand, results from the external circuit and may either cause, or be a result of, the device oscillating at very high frequency about the bias point.[6] Such bistability is then a function of the external circuit rather than the quantum-well structure.

While experiments have observed bistability in the operation of RTD's, there is general disagreement as to its source.[5,6,61–64] If all oscillations of the device could be suppressed, any observed bistability would have to be intrinsic. Because the negative resistance R_n is low, however, debate centers around whether all oscillation has, in fact, been suppressed, and intrinsic bistability observed.

In earlier studies the potential was dropped uniformly across the barrier structure.[43,44,48,58] Treatments which specify a fixed electrostatic potential cannot account for the effects of charge storage or self-consistent–potential shifts outside the barriers. Since intrinsic bistability is thought to result from changes in internal potentials, it is not surprising that rigid-potential modes do not show bistability.

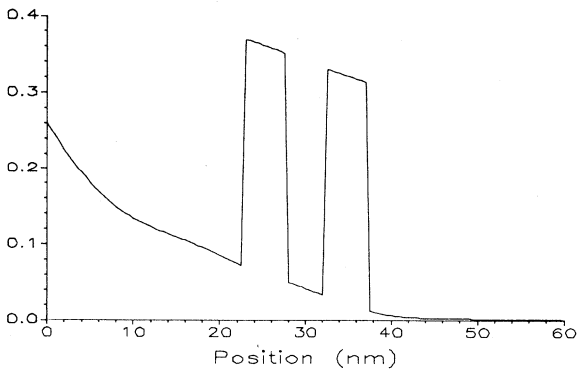

FIG. 8. Potential for the RTD for an applied bias of 0.22 V, the peak of the *I-V* curve. Much of the potential is dropped across the cathode end of the device, forming a triangular potential well. Depletion due to contact resistance causes the upward bending of the potential.

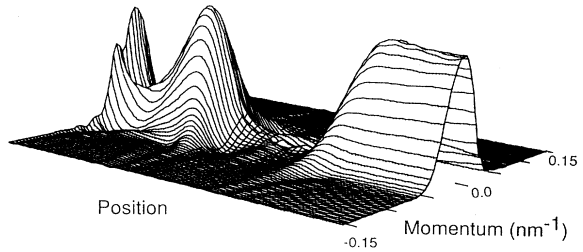

FIG. 9. Steady-state Wigner distribution at the peak of the *I-V* curve. Depletion in the cathode region is evident. At the left contact the incoming distribution appears as a shifted Fermi-Dirac distribution.

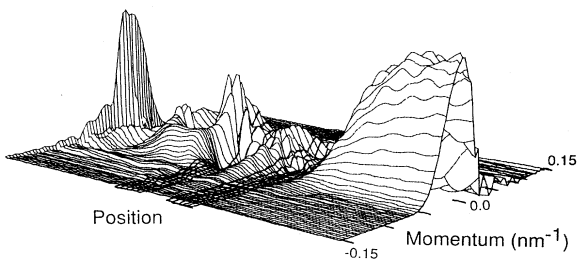

FIG. 10. Steady-state Wigner distribution at the valley of the *I-V* curve. Depletion is strongly evident in the cathode region. The distribution in the cathode-barrier–interface region forms a quantized state, the ring structure to the left of the barriers.

The calculation of the I-V curve does not include external circuit effects. Any bistability which is found, therefore, must be intrinsic in nature. In fact, one theoretical approach has shown that an intrinsic bistability does exist.[64] Scattering is modeled by a relaxation-time approximation, and its primary effect is in determining the potential drop in the resistive regions outside the barriers. This resistive load also influences the slope of the NDC region, which is proportional to $1/R_n$. Since the mobility at 300 K is density dependent, the carrier concentration in the device is critical to R_n.

Earlier work, without self-consistent potentials, showed a shoulder in the NDC region of the curve.[43] With the addition of fully-self-consistent potentials, the shoulder is not seen, but a soft bistability, similar to ferromagnetic hysteresis, is found in the same region of the I-V curve. This model shows one region of bistability with a finite R_n, resulting strictly from differences in the self-consistent potentials and the distributions on either side of the bistable region. Both effects are related to the charge stored in the quantum well. The maximum difference in current in the bistable region is found at an applied potential of 0.36 V. In Fig. 11 the difference between the steady-state Wigner distribution for increasing potential and the steady-state distribution for decreasing potential in the bistable region is plotted. For decreasing potential, less current flows through the resonant structure, and the injected carriers accumulate on the cathode side of the structure, shown as the peak in Fig. 11. Since less current is flowing in the entire device, the anode end shows a slight depletion of current-carrying distribution.

In an experimental circuit the RTD is loaded, which changes the slope of the NDC region. Experiments have shown two bistable regions of the I-V curve, each of which has a very large R_n. What may be observed experimentally is the intrinsic bistability being heavily modified by the external circuit parameters.

The self-consistent potential within the RTD is shown in Fig. 12. Two potentials are shown, corresponding to the two stable values of current for an applied bias of 0.36 V. The difference in the potentials is indicative of a difference in the charge distributions and currents.

D. Zero-bias anomaly

Some experimental work with tunneling devices has observed an increased conductivity near zero bias.[65] This

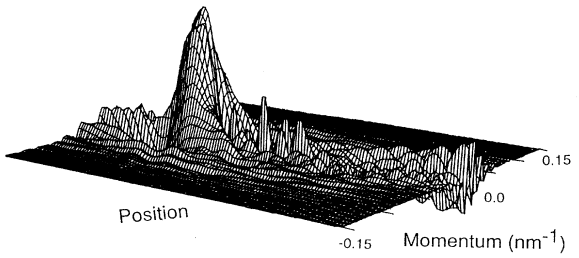

FIG. 11. Difference between the bistable Wigner distributions at a bias of 0.36 V. The quantized state in the cathode well has more carriers. More current is flowing, indicated by the "ridge" in the distribution.

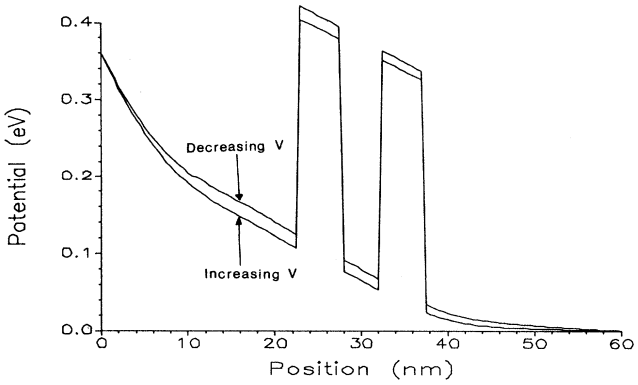

FIG. 12. Potentials corresponding to the two bistable states for an applied bias of 0.36 V. The upper curve corresponds to the upper branch of the bistable region.

"zero-bias anomaly" was believed to arise from impurity-assisted tunneling, in which the spin interaction of the tunneling electron with the impurity leads to several conducting channels.[66,67] In this model the s-d interaction in third- and higher-order perturbation theory shows a peak in the conductivity at zero bias, which results from the combined effects of spin-flip scattering in the barrier and the exclusion principle in the contacts. Enhanced conductivity is also seen near zero bias in MBE-grown tunneling structures.[68] Here, we find that an increased conductivity at zero bias results from high-momentum tails in the distribution at the barriers which have a nonthermal character.

Near a quantum barrier the electron density is reduced as a result of quantum interaction with the barrier.[13] This is not a depletion layer in the classical sense. It is not caused by band bending; this phenomenon occurs even when there are no electromagnetic fields. The wave function at a particular momentum k is repelled by the tunnel barrier. The first peak in the standing wave occurs approximately at π/k from the barrier, and is much closer for high momentum. This leads to a broadened distribution for the density, with a length scale of about λ_{th}. The quantum repulsion is, in a sense, complementary to barrier penetration: just as a nonzero density penetrates a finite distance into a classically forbidden region, a density deficit penetrates a finite distance into a classically allowed region. Figure 13 illustrates the momentum spread which results from this quantum repulsion. This momentum spread of the distribution highlights two components of current at extremely low biases. The high-momentum tails of the distribution in Fig. 13 extend above the barrier in energy. A slight perturbation of the potential is sufficient to radically alter the nature of this distribution and force a current flow. Because the distribution tails are characterized by a high momentum, the current which results may be characterized by extremely high conductance. After the momentum tails have depleted, the only remaining source of current is tunneling from the main thermal distribution, which is characterized by a much lower conductance. Thus, near zero bias we see a dramatic change in $f(E)$ that appears in (1).

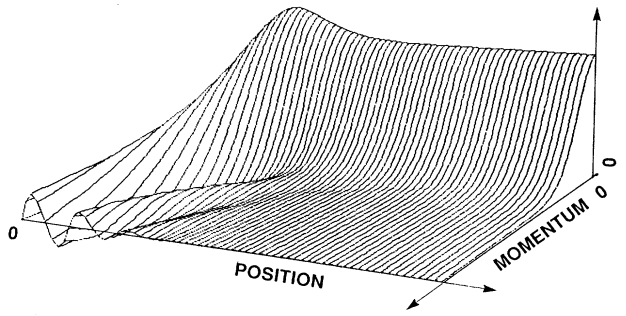

FIG. 13. Wigner distribution near an infinite barrier. This plot demonstrates the quantum repulsion by the barrier and the high-momentum tails in the distribution.

Early models of tunneling were unable to show this effect, since it was assumed that the distribution at the interface between the barriers and the bulk region is essentially thermal. In fact, the distribution at the interface possesses a momentum spread which is best characterized by an effective "temperature" much larger than that of the bath.[69] Failure to account for the proper distribution at the interface neglects the contributions these momentum tails make to the current at low biases.

The *I-V* curve calculated for Wigner functions shows the zero-bias anomaly below 0.1 mV, as shown in Fig. 14. Above 0.1 mV the current begins to show an exponential dependence, but does not extrapolate to zero. Below 20 μV the current is characterized by a conductivity much higher than that above 0.1 mV. Between these regions the conductivity changes with the applied bias.

In Fig. 15(a) the difference between the steady-state distribution with 20 μV applied bias and the equilibrium distribution is plotted (the amplitude is magnified by 15). Point *A* on this figure indicates a loss of carrier density from the high-momentum tails. Quantum repulsion near the barriers is still present, but has been reduced by the small applied bias, leading to the positive peak *D*. Point *B*, the location of the accumulation in equilibrium, shows some slight reduction as carriers begin to tunnel through

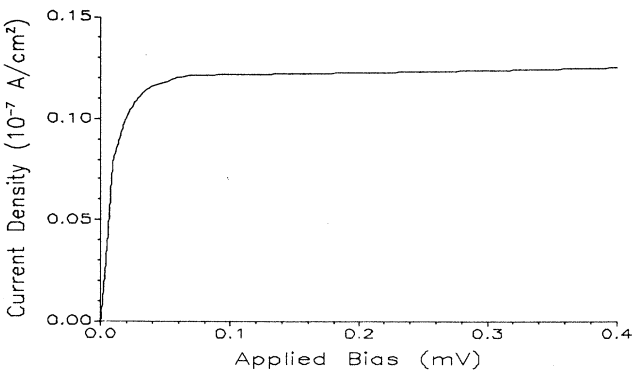

FIG. 14. *I-V* curve of the RTD for low biases. The current displays two different conductances. The exponential dependence above 0.1 mV does not extrapolate to zero.

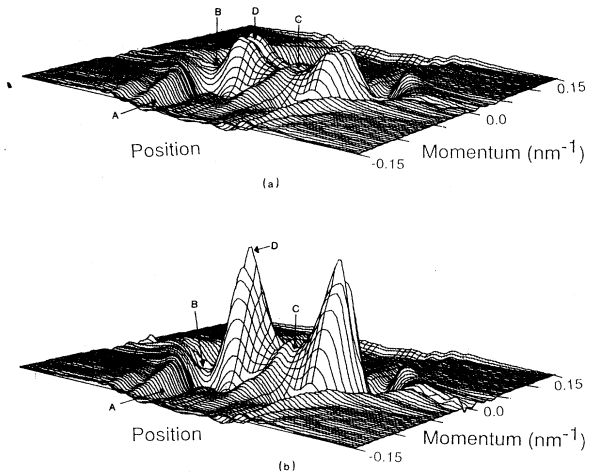

FIG. 15. Difference between the steady-state distribution and the low-bias distributions (magnified 15×). (a) Applied bias is 20 μV. Point *A* shows the depletion of the high-momentum tails. *B* and *D* show a shift of the accumulation layer toward the barriers, while *C* shows a slight enhancement of the distribution within the quantum well. (b) Applied bias is 100 μV. Point *A* shows the same loss of tails as the 20-μV distribution, while *B* and *D* show the accumulation layer shifted clearly into the barriers. The distribution within the well, point *C*, is greatly enhanced.

the barriers. Within the quantum well, at point *C*, the distribution is enhanced by the tunneling carriers. At this small bias nearly all of the current in the barrier region comes from the high-momentum carriers. These carriers have sufficient energy to tunnel easily through (or over) the quantum barriers. Tunneling from the thermal distribution does not yet contribute significantly to the current.

In Fig. 15(b) the difference between the distribution at a higher applied bias of 100 μV and the equilibrium one is shown. Comparison between Figs. 15(a) and 15(b) reveals that there is little difference at point *A*, as the high-momentum tails have been completely depleted. The differences between Figs. 15(a) and 15(b) show clearly at point *B*. The accumulation layer is moved considerably nearer the barrier, from point *B* to point *D*. The distribution at point *D* clearly reflects the presence of a thermal distribution right up to the barriers. In fact, it extends well into the barriers, and a large distribution within the well, at point *C*, is evident. Calculation of the current in the barrier region reveals that most of the current now results from carriers at lower energies which tunnel through the barriers.

E. Transient behavior

The transient behavior of the RTD may be studied using Wigner functions.[43,56,57] In this particular study, the system for $t<0$ is biased at the peak of the *I-V* curve of Fig. 7. At $t=0^+$ the bias is raised instantaneously to a

value corresponding to the valley of the curve, so as to determine the large-signal transient behavior of the device. The current is recorded as a function of time as the system evolves toward its new steady state. The current and potential transients are then Fourier transformed, and the conductance of the device as a function of frequency is found.

The current transient is plotted in Fig. 16. The inset shows the I-V curve and the voltage points between which the device is switched. The oscillatory behavior of the transient is clearly evident, as is the tremendous overshoot in current. The overshoot probably arises from a rapid discharge of the trapped charge in the potential well, and the oscillations are caused when the internal potential adjusts to the changing distribution, and sets up plasma oscillations within the device. Examination of the time-varying distributions clearly shows the plasma oscillations. The large overshoot of current is related to an inherent inductive effect of the carriers.

The current transient has been calculated for three values of the relaxation time, corresponding to mobilities of 1500, 3000 sec, and 4200 cm^2/V sec, to ensure that the relaxation time is not the factor limiting the frequency of operation of the device. The magnitude of the conductance for all three cases is plotted in Fig. 17. The conductance shows a peak in all cases at approximately 1.5 THz, with a smaller second-harmonic peak at 3 THz. A reduction in mobility reduces the magnitude of the peak.

The imaginary part of the conductance is plotted in Fig. 18(a). For low frequencies the system is inductive in nature. As the relaxation time and the mobility decrease, the magnitude of the inductive contribution decreases, but the peak remains unchanged. The electrons essentially move ballistically over a mean free path, and their inertia resists sudden changes in potentials. As the mobility decreases, the electrons are ballistic over a shorter mean free path and less inductive. For frequencies above 2 THz the system is capacitive in nature, which is a result of the charging and discharging of the resonant energy level within the quantum well. Experimental evidence supports the buildup of a space-charge layer in the tunneling structure,[6] and earlier studies[48] have shown a charge storage within the quantum well, characterized by a time of the order of 250 fsec.

The real part of the conductance is plotted in Fig. 18(b). For frequencies under 1.5 THz the real part of the conductance is negative, and becomes positive for frequencies over 1.5 THz. The frequency at which the tran-

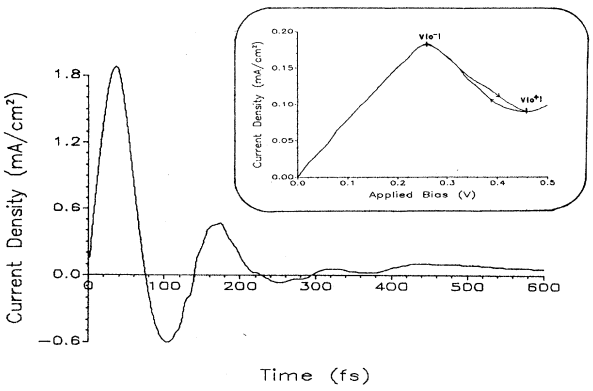

FIG. 16. Current transient of the RTD. The device is switched from the peak to the valley of the I-V curve (see inset). Large oscillations are caused by plasma oscillations and ballistic inertia of the electrons.

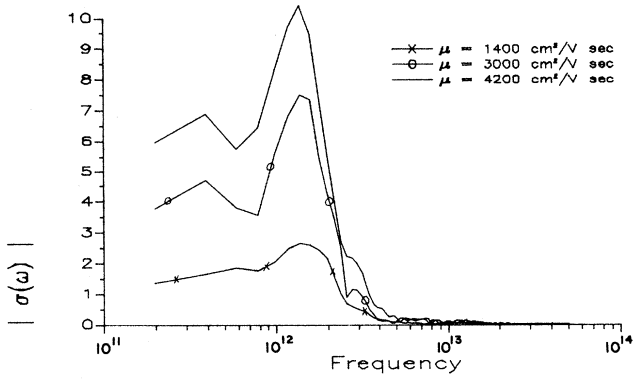

FIG. 17. Magnitude of the conductance of the RTD in the negative-differential-conductance region. Conductance is calculated for three values of mobility.

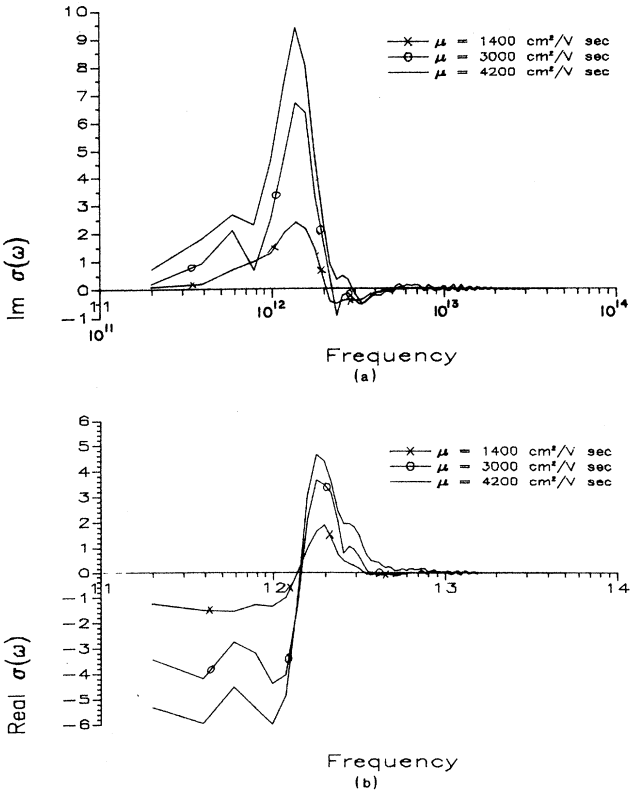

FIG. 18. Real and imaginary components of the conductance of Fig. 17. (a) Imaginary part of the conductance. Below 2 THz the system is inductive. (b) Real part of the conductance. Resistance is negative below the resonant frequency of 1.5 THz.

sition between negative and positive conductance occurs is independent of the relaxation time. We note that the peak of the reactance in Fig. 18(a) occurs at the point at which the real part is zero, as required by the Kramers-Kronig relations.

F. Modification of the RTD structure

Resonant-tunneling diodes are often fabricated with undoped spacer layers at the interface between the barriers and the bulklike contact regions. The effects which result from these spacer layers may easily be examined with the Wigner-function model. The self-consistent equilibrium Wigner-function, potential, and ionized-donor concentrations are again calculated, with the spacer layer built into the donor distribution $N_d(x)$. While the difference in the equilibrium distributions is only slight, and concentrated at the interface between the carriers and the contacts, the fundamental difference may be easily seen in the potentials $V(x)$. The self-consistent potentials for the basic structure and the RTD with a spacer layer are plotted in Fig. 19. Without a spacer layer, the potential at the barrier tends toward a triangular potential well, leading to a slight accumulation of electrons near the barrier. With a spacer layer, the potential is shifted upward, causing a much smaller tendency toward depletion of electrons. In fact, there is a small accumulation in the spacer layer. It is apparent that the spacer layer concentrates the potential near the barrier. This both reduces the contact resistance and reduces the ballistic injection so that the peak-to-valley ratio is improved.

A second difference results from the effect of the spacer layers on the scattering. Without ionized donors the electron mean free path is longer, and the relaxation time is correspondingly increased. More of the potential is dropped across the barriers, causing a higher barrier to the ballistic electrons, which reduces the valley current. The self-consistent potential for a bias at the valley is plotted in Fig. 20, showing how the potential differs from that in the RTD without the buffer layer. The effect on the valley current is clearly evident in the calculated I-V curve of Fig. 21. Another major difference caused by the buffer layer and reduced scattering is the reduced R_n,

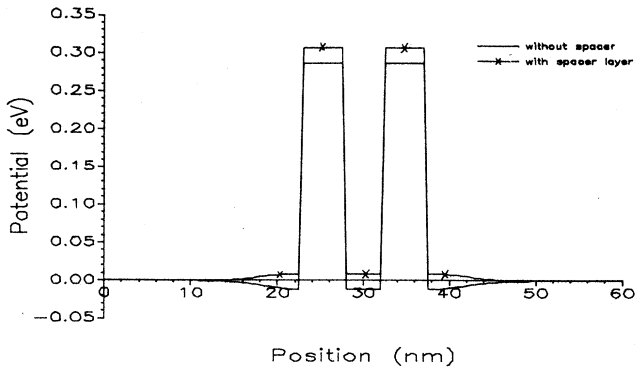

FIG. 19. Self-consistent equilibrium potentials for the RTD with and without an undoped spacer layer.

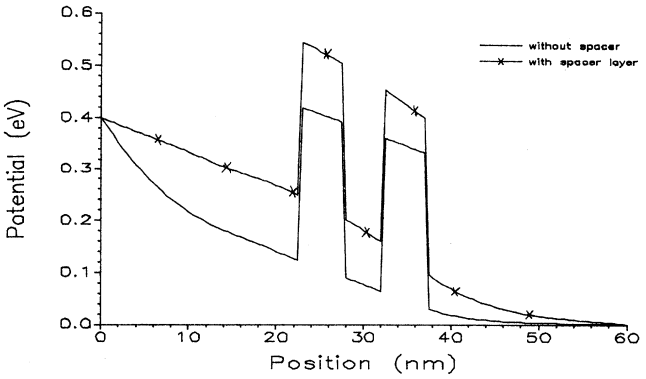

FIG. 20. Self-consistent potentials for an applied potential of 0.4 V for the RTD with and without the spacer. Without the spacer a large depletion region and triangular potential well form between the cathode contact and the barrier. The spacer layer controls the depth of the well.

which increases the slope of the NDC region of the curve.[5] Finally, the bistability of the NDC region of the I-V curve is found to be much sharper, again a consequence of the reduced R_n. We note, however, that the bistability still appears to be intrinsic. The curves are now much closer to the simpler model of Sheard and Toombs.[64]

V. SUMMARY AND CONCLUSIONS

We have presented a fully-self-consistent model of the resonant-tunneling diode based upon the quantum-mechanical Wigner function. The Wigner formalism offers the particular benefits that "ideal" boundaries are easily implemented, the transport equation is similar to the Boltzmann equation, and the nonlocal correlation effects are included. In the Wigner-function model, scattering is a local phenomenon. An adjoint equation must be used to determine the initial Wigner distribution to ensure inclusion of all orders of quantum-mechanical corrections.

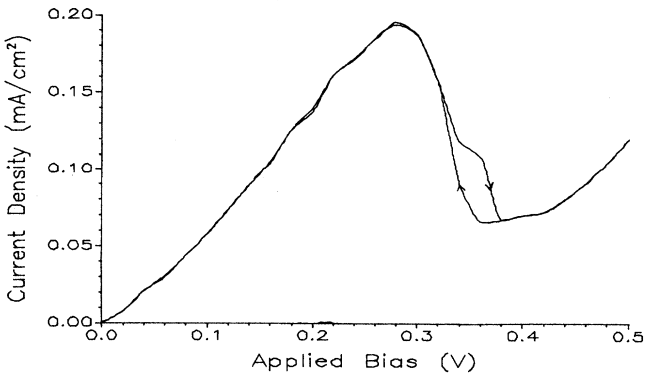

FIG. 21. I-V curve of the RTD with the spacer layer. The negative differential resistance is lower, and the bistable region is more pronounced than without the spacer layer. Peak-to-valley ratio is nearly 3:1.

The conditions on stability and convergence of the Wigner-function equation of motion have been reviewed. Lax-Wendroff explicit time differencing improves the temporal stability of the solution by including higher-order terms in the discretized equation. The time-evolution equation is unconditionally stable and convergent if the choices of space mesh Δx and the time step Δt satisfy the CFL stability criterion for the entire discretized mesh. Self-consistency is included by coupling Poisson's equation to the Wigner equation. The model includes the calculation of a self-consistent initial Wigner function, an initial potential function, and the corresponding ionized-donor distribution. Scattering is added to the model through the relaxation-time approximation, where all the scattering is combined into one relaxation-time parameter. Though admittedly crude, this effectively removes correlation and introduces irreversibility.

The model incorporates the concept of "ideal" contacts, whereby the contact serves as a reservoir of thermally distributed carriers at the boundaries where carriers enter the device, and perfectly absorbs any carriers which leave the device. The distribution within the contact must reflect the physical reality of current flow, so we enforce current continuity by using a drifted Fermi-Dirac distribution within the contacts. An algorithm has been developed to remove artificial numerical reflections from the boundaries. These reflections result from the nonlocality of the temporal equation. Removal of the reflection effectively prevents correlation from extending from the device into the contact, removing size dependencies from the model.

The self-consistent model is applied to the RTD. The calculated I-V curve shows a peak-to-valley ratio of 2:1 and, more importantly, shows an intrinsic bistability in the NDC region of the curve. This bistability results from charge-storage effects within the quantum well, which changes the internal potentials. Bistability has been observed experimentally, and a recent theoretical model predicted that the bistability is intrinsic.[64] When we model a modified structure with undoped spacer layers adjacent to the quantum barriers, we find a peak-to-valley ratio of nearly 3:1 and a sharp decrease in the negative resistance. The bistable region is correspondingly enhanced. A deep triangular quantum well forms between the cathode contact and the quantum barrier. This well is sufficiently deep to give rise to a quantized state. Approximately half of the applied bias is found in the cathode well. At the valley of the I-V curve, the barriers are pulled down in energy, permitting ballistic electrons from the cathode to travel over the barriers, leading to an enhanced valley current and a degraded peak-to-valley ratio. A net depletion in the cathode region stems from the contact resistance of the "ideal" contact model. Inclusion of spacer layers inhibits formation of the cathode well. The spacer layer causes an upward bending of the conduction band at the barrier interface, which resists the downward bending of the deep quantum well as bias is applied. More of the potential is dropped across the barriers. The peak-to-valley ratio is enhanced by the reduction of ballistic electrons traveling over the barriers.

The transient behavior of the RTD is analyzed using the temporal equation. The device is switched from a bias at the peak of the I-V curve to the valley. Currents are calculated as the system evolves to its new steady state. The current initially increases greatly, resulting from the discharging of the quantum well. The current oscillates in response to the adjusting potentials within the device, until it decays to its new steady-state value. Fourier analysis of the current transient shows a peak in the conductance near 2 THz.

The Wigner-function model shows an anomaly in the I-V curve for very low applied bias. This zero-bias anomaly has been observed experimentally in tunneling structures, and results from quantum repulsion from the barrier. The quantum interaction leads to high-momentum tails in the distribution. The symmetry of these high-momentum tails is easily disrupted by low applied bias. Since these tails extend to energies above the barrier, they contribute to a current with high conductance. After these tails have depleted, the current comes from the thermal distribution, which is characterized by a much lower conductance.

ACKNOWLEDGMENTS

The authors would like to acknowledge helpful discussions with L. Eaves, R. Landauer, and J. Sinkonnen. Support for this work was provided by the U.S. Office of Naval Research, and one author (C.R.) was supported by the U.S. Air Force Office of Scientific Research.

[1]G. Bernstein and D. K. Ferry, IEEE Trans. Electron. Dev. **ED-35**, 887 (1988).
[2]R. Dingle *et al.*, Appl. Phys. Lett. **33**, 665 (1978).
[3]G. Bernstein and D. K. Ferry, Z. Phys. B **67**, 449 (1987).
[4]T. C. L. G. Sollner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker, and D. D. Peck, Appl. Phys. Lett. **43**, 588 (1983); T. C. L. G. Sollner, P. E. Tannenwald, D. D. Peck, and W. D. Goodhue, *ibid.* **45**, 1319 (1984).
[5]V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Phys. Rev. Lett. **58**, 1256 (1987); V. J. Goldman, D. C. Tsui, and J. E. Cunningham, *ibid.* **59**, 1623 (1987).
[6]L. L. Chang, L. Esaki, and R. Tsu, Appl. Phys. Lett. **24**, 593 (1974).
[7]See, for example, E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970).
[8]B. Ricoh and M. Azbel, Phys. Rev. B **29**, 1970 (1984).
[9]R. Tsu and L. Esaki, Appl. Phys. Lett. **22**, 562 (1973).
[10]R. Landauer, IBM J. Res. Dev. **1**, 223 (1957).
[11]W. W. Lui and M. Fukuma, J. Appl. Phys. **60**, 1555 (1986).
[12]R. Landauer, Z. Phys. B **21**, 247 (1975).
[13]A. M. Kriman, N. C. Kluksdahl, and D. K. Ferry, Phys. Rev. B **36**, 5953 (1987).
[14]U. Ravaioli, M. A. Osman, W. Pötz, N. C. Kluksdahl, and D. K. Ferry, Physica B+C **134B**, 36 (1985).

[15]S. M. Goodnick and D. K. Ferry, in *Physics and Chemistry of III-V Compound Semiconductor Interfaces,* edited by C. W. Wilunson (Plenum, New York, 1985).
[16]M. Büttiker, Phys. Rev. B **33**, 3020 (1986).
[17]R. Landauer, Philos. Mag. **21**, 863 (1970).
[18]R. Landauer and M. Büttiker, Phys. Rev. Lett. **54**, 2049 (1985).
[19]M. Büttiker, Phys. Rev. B **32**, 1846 (1985).
[20]M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, Phys. Rev. B **31**, 6207 (1985).
[21]M. Büttiker, Phys. Rev. B **35**, 4123 (1987).
[22]E. Wigner, Phys. Rev. **40**, 749 (1932).
[23]J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949).
[24]I. B. Levinson, Zh. Eksp. Teor. Fiz. **57**, 660 (1969) [Sov. Phys.—JETP **30**, 362 (1970)].
[25]G. J. Iafrate, H. L. Grubin, and D. K. Ferry, Phys. Lett. **87A**, 145 (1982).
[26]L. Mandel and E. Wolf, Rev. Mod. Phys. **37**, 231 (1965).
[27]H. Haken, H. Risken, and W. Weidlich, Z. Phys. **206**, 355 (1967).
[28]M. Lax and W. H. Louisell, IEEE J. Quantum Electron. **QE-3**, 47 (1967).
[29]M. J. Bastiaans, Opt. Commun. **25**, 26 (1978).
[30]K. H. Brenner and J. Ojeda-Castaneda, Opt. Acta **31**, 213 (1984).
[31]M. J. Bastiaans, Appl. Opt. **19**, 192 (1980).
[32]M. Conner and Y. Li, Appl. Opt. **24**, 3825 (1985).
[33]BVK Vijaya Kumar and C. W. Carroll, Opt. Eng. **23**, 732 (1984).
[34]J. R. Barker, J. Phys. C **6**, 2663 (1973).
[35]J. R. Barker and S. Murray, Phys. Lett. **93A**, 271 (1983).
[36]N. C. Kluksdahl, W. Pötz, U. Ravaioli, and D. K. Ferry, Superlatt. Microstruct. **3**, 41 (1987).
[37]W. R. Frensley, J. Vac. Sci. Technol. B **3**, 1261 (1985).
[38]R. P. Feynman, in *Quantum Implications; Essays in Honor of David Bohm,* edited by B. J. Hiley and F. David Peat (Routledge and Kegan Paul, London, 1987).
[39]L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Benjamin, New York, 1962).
[40]W. Hänsch and G. D. Mahan, Phys. Rev. B **28**, 1902 (1983); G. Mahan, J. Phys. F **13**, L257 (1983); G. Mahan and W. Hänsch, *ibid.* **13**, L47 (1983).
[41]A. P. Jauho and J. W. Wilkins, Phys. Rev. B **29**, 1919 (1984).
[42]J. R. Barker, in *Physics of Nonlinear Transport in Semiconductors,* edited by D. K. Ferry, J. R. Barker, and C. Jacoboni (Plenum, New York, 1979).
[43]N. C. Kluksdahl, A. M. Kriman, C. Ringhofer, and D. K. Ferry, Solid-State Electron. **31**, 743 (1988).
[44]W. R. Frensley, Phys. Rev. B **36**, 1570 (1987).
[45]R. Courant, K. Friedrichs, and H. Lewy, Math. Ann. **100**, 32 (1928).
[46]Y. I. Shokin, *The Method of Differential Approximation* (Springer-Verlag, Berlin, 1983).
[47]J. R. Barker, in *Proceedings of the 1987 Workshop on Quantum Transport in Semiconductors, San Miniato,* edited by C. Jacoboni and D. K. Ferry (to be published).
[48]N. C. Kluksdahl, A. M. Kriman, and D. K. Ferry, Superlatt. Microstruct. **4**, 127 (1988).
[49]P. Carruthers and F. Zachariason, Rev. Mod. Phys. **55**, 245 (1983).
[50]N. C. Kluksdahl, C. Ringhofer, and D. K. Ferry (unpublished).
[51]W. Pötz (unpublished).
[52]B. Engquist and A. Majda, Math. Comput. **31**, 629 (1977).
[53]J. S. Blakemore, *Semiconductor Statistics* (Pergamon, New York, 1962).
[54]E. M. Conwell, *High Field Transport in Semiconductors,* Suppl. 9 of *Solid State Physics,* edited by F. Seitz and D. Turnbull (Academic, New York, 1967).
[55]See, for example, K. Seeger, *Semiconductor Physics* (Springer-Verlag, Berlin, 1985).
[56]W. R. Frensley, Solid State Electron. **31**, 739 (1988).
[57]N. C. Kluksdahl, A. M. Kriman, and D. K. Ferry, IEEE Electron. Dev. Lett. **9**, 457 (1988).
[58]W. R. Frensley, Phys. Rev. Lett. **57**, 2853 (1986).
[59]A. Many, Y. Goldstein, and N. B. Grover, *Semiconductor Surfaces* (Wiley, New York, 1965).
[60]D. Delagebeaudeuf and N. T. Linh, IEEE Trans. Electron. Dev. **ED-29**, 955 (1982).
[61]V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Phys. Rev. B **35**, 9387 (1987).
[62]T. C. L. G. Sollner, Phys. Rev. Lett. **59**, 1622 (1987).
[63]L. Eaves, E. S. Alves, T. J. Foster, M. Henini, O. H. Hughes, M. L. Leadbeater, F. W. Sheard, G. A. Toombs, K. Chan, A. Celeste, J. C. Portal, G. Hill, and M. A. Pate, in *Technology of Submicron Structures,* edited by H. Heinrich, G. Bauer, and F. Kucher (Springer-Verlag, Berlin, 1989).
[64]F. W. Sheard and G. A. Toombs, Appl. Phys. Lett. **52**, 1228 (1988).
[65]L. Y. L. Shen and J. M. Rowell, Phys. Rev. **165**, 566 (1968).
[66]C. B. Duke, *Tunneling in Solids,* Suppl. 10 of *Solid State Physics,* edited by F. Seitz and D. Turnbull (Academic, New York, 1969).
[67]J. A. Appelbaum, Phys. Rev. **154**, 633 (1967).
[68]D. E. Lacklison, B. Duggan, J. J. Harris, C. T. B. Foxon, D. Hilgon, C. Roberts, and C. M. Hellon, Appl. Phys. Lett. **52**, 305 (1988).
[69]N. C. Kluksdahl and D. K. Ferry (unpublished).