Coherent electron transport across semiconductor heterojunctions with rough interfaces

W. T. Dietze and R. B. Darling
Solid-State Laboratory, Department of Electrical Engineering, Box 352500, University of Washington, Seattle, Washington 98195
(Received 14 September 1995)

A treatment of coherent electron transport in planar media with rough interfaces is developed to first order within the effective-mass approximation. This approach starts with the exact boundary conditions of a single rough interface and derives a perturbational expansion dependent upon the magnitude and slope of the roughness. The first-order results are cast into a convenient transfer-matrix formalism for ease of use and interpretation. The model is derived for the purpose of modeling the effects of interface roughness in epitaxial systems and is demonstrated using double-barrier resonant tunneling devices.

I. INTRODUCTION

Interface roughness in epitaxial structures and its effects on electron transport have been studied by many groups and are particularly important in structures with carrier confinement on the scale of an electron wavelength. Rough interfaces are an inevitable result of epitaxial growth kinetics, exacerbated by the fact that the ideal growth conditions for the requisite two materials that define the interface are rarely the same.[1,2] Treatments of the effects of rough interfaces typically reduce to three cases: the analysis of transport normal to[3–10] or in the plane of the interface[11] and of excitons in roughened quantum wells.[7,12,13] This paper concentrates exclusively upon coherent transport across heteroepitaxial structures with rough interfaces. As such, it covers the basic process of tunneling through single- and multiple-barrier devices upon which inelastic processes are parasitic.

The effects of roughness upon tunneling are expected to be important when the deviation in thickness is a sizable fraction of a tunneling barrier, electron wavelength, or quantum-well width. Quantum-well width variations are particularly important because the energy of quasibound states (resonances) is defined by the round-trip phase change in the well. The effects of roughness on tunneling through heteroepitaxial barriers have been measured in photoluminescence and photoluminescence excitation spectra in quantum wells with thin barriers[12,13] and in electron transport through resonant tunneling devices (RTD's).[9] This latter effort is mainly concerned with the effects of substrate orientation on the *I-V* curve of the RTD.

In typical high-quality epitaxy, particularly when the materials are lattice matched, interface roughness is usually characterized by monolayer-high island formation and the typical descriptive parameters are the average island size and percent of coverage. Another typical characterization is to use a statistical description of the height-height autocorrelation function of the rough interface.[14] Such idealizations are used because the precise nature of interface roughness is unknown and sensitive to many epitaxial growth parameters such as temperature, growth rate, stop growth times, substrate orientation, growth constituent flux ratios, alloy composition, etc.[15] In addition, the roughness at a solid/solid interface has proven particularly hard to characterize; extremely sensitive techniques are required, such as high-resolution transmission electron spectroscopy[16] and precision x-ray diffraction measurements.[17] Roughness on the growing epitaxial surface is typically monitored by grazing incidence electron beam techniques;[18] however, these techniques are really only qualitative. Finally, *surface* roughness can be precisely characterized by scanning tunneling microscopy but the exposed surface may be critically different from the heterointerface. In studies of quenched-growth epitaxial surfaces, scanning tunneling microscopy reveals several types of rough surfaces that are difficult to parametrically characterize.[2]

Methods used to model the effects of interface roughness on coherent tunneling generally approach the problem in one of three ways: (1) transfer-matrix techniques that treat each rough interface separately,[5,19] (2) diagrammatic techniques within the Born approximation,[3,20] and (3) diagrammatic techniques within diffusive perturbation theory that include the effects of all ladder and maximally crossed Feynman diagrams often from within a weak localization approximation.[10,21]

The transfer-matrix method of Liu and Coon is a partial application of the techniques developed in this paper to the problem of electron scattering at solid-state interfaces.[5] Unfortunately, that treatment only modifies one of the interfacial boundary conditions in the presence of roughness. In addition, that work uses δ functions to model the roughness. Since no effective mass can be associated with the δ function, the scattering associated with the effective mass difference is omitted.

Diagrammatic techniques made within the Born approximation include multiple scattering (from the perturbing potential) with each term in the perturbative expansion. This has been used to model single and double scattering through RTD's due to point impurities,[20] alloy scattering,[22] and interface roughness.[3]

Diffusive perturbation theory allows a multiple-scattering approach by selecting certain types of interaction diagrams and summing over all of them so as to produce an average Green's function for the system. This approach has been used by Berkovits for scattering due to point impurities in the well region of a RTD and, by analogy, the resonant cavity of a Fabry-Pérot interferometer. In that application the purely

specular term for resonant tunneling (no interactions with the roughness of the interface) is destroyed. It yields a self-energy shift in the resonant energy and a slight broadening of the resonance.[21] Johansson has adopted this technique to model surface roughness from wide monolayer islands as point impurities using three-dimensional δ-function potentials, weighted to reflect island size.[10] Such a reduction seems inappropriate for large-size islands.

The perturbative techniques utilize an eigenstate expansion defined by planar interfaces located at the average interface positions. This is problematic in all roughened quantum-well situations in which bound and quasibound states exist. The energies for such states are determined by a round-trip phase matching condition and thus directly affected by the variable interface location. These two techniques are most suitable when they are used for scattering due to fixed impurities or alloy clustering.[20–22] In these situations the perturbing potential is assumed weak and the planar-interface eigenfunctions are a good approximation to the actual ones.

This paper adopts the transfer-matrix approach by converting the rough interface/homogeneous boundary condition problem to that of a planar interface with inhomogeneous boundary conditions. Rather than only modeling the rough interface as a δ-function sheet potential at the average interface position,[5] we adapt optical techniques used to explain anomalous surface plasmon scattering from metallic surfaces.[19] By using a transfer-matrix formalism, the effects of scattering in multilayer devices, particularly in resonant ones, are easily and naturally calculated. Our transfer-matrix technique is an extension of the envelope-function transfer-matrix type of calculation made within the effective-mass approximation used with planar interfaces.[23] As such, it is a technique that does not suffer the extreme numerical instabilities of transfer-matrix calculations based on tight-binding calculations.[24,25]

We approach the problem of writing approximate interface matching conditions at a plane by expanding the wave function on either side of the interface in a perturbative series and writing each term in a Taylor expansion around a reference plane at $z=z_1$. To do this, the wave functions are continued to or truncated at $z=z_1$ as if the interface were actually at that plane. The interface is characterized by the band edge potential $V(x,y,z)=V_b\Theta(z-z_1-f(x,y))$, where $\Theta(z)$ is the unit step function. At each point (x,y,z_1), the boundary conditions account for both the deviation of the interface from the reference plane $f(x,y)$ and the slope of the roughness.

It is important to note that the continued wave function, evaluated at the reference plane, does not represent the actual wave function at that point, so the new boundary conditions are not expected to represent (even in a limiting sense) any realizable potential or effective-mass feature at z_1. At this level, the boundary conditions that we derive are a mathematical expedient that produce the correct wave functions on either side of the actual interface. In this way the electron wave problem differs from the optical one in which the new boundary conditions can be interpreted as sheet currents at the reference plane.[19]

This approach treats only the coherent transport of conduction band electrons between two semiconductors with conduction band edge and/or effective-mass discontinuities. Only the transport of electrons at the same symmetry point is considered and spherically symmetric, parabolic bands are assumed for simplicity of exposition. This approach will be solely concerned with the envelope portion of the full electron wave function. The lattice periodic portions of the full wave function in each material are assumed to be largely the same for a given wave vector and the scattering due to the roughness will be assumed to be weak in the sense that only small changes in the wave vector relative to the Brillouin zone edge are supported with each interaction.

The effects of differing band structures are felt only in the variation of the conduction band edge minima and effective mass across the interface and in the gradient boundary condition, for which the wave function gradient is weighted by the inverse effective mass.[26,27] The interfaces are assumed to be abrupt in the sense that the conduction band edge potential and effective masses are treated as step variations. We do not include the scattering between band minima due to the detailed interface matching conditions derived from tight-binding and pseudopotential calculations of the full wave function.[24,28]

This paper proceeds in Sec. II by reviewing the transfer-matrix formalism in the context of rough interfaces. In Sec. III we derive a set of boundary conditions at the average interface plane that mimic the usual conditions valid at the actual interface. A perturbative expansion is used for the wave functions on either side of the interface and the boundary conditions are written for like terms in these expansions. The first-order derivation is outlined in Sec. IV, followed in Sec. V with an analysis of the first-order results. In Sec. VI the effects of island size on the transmission characteristic are computed for a RTD with one rough well/barrier interface.

II. THE TRANSFER-MATRIX FORMALISM

Before proceeding to the detailed solution of the problem we outline a few points about the transfer-matrix formalism used here. As shown in Fig. 1, an epitaxial structure such as a resonant tunneling device is made up of nominally planar layers with interface quality determined by the epitaxial growth dynamics. Figure 2 shows the conduction band edge profile of a double-barrier resonant tunneling device utilizing heavily doped n-type electrode regions. The roughened interfaces are indicated schematically.

In typical applications of a *planar* effective-mass transfer-matrix analysis, the structure is further divided into many regions of constant (or linear) potential. A set of boundary conditions is then used to link the electron wave function across each interface (either artificial or actual) and the energy eigensolutions are found. The important portions of the full eigensolutions, as far as the global transmission and reflection coefficients are concerned, are the coefficients of the two plane-wave components in each of the bounding regions. A unity amplitude (or flux) plane wave incident from one side is assumed and the resulting plane-wave coefficients corresponding to the outbound waves yield, with a little manipulation, the transmission and reflection coefficients. These

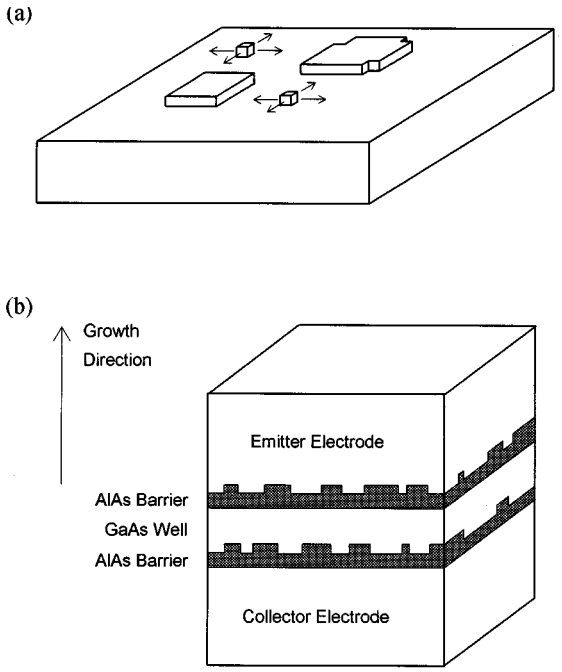

FIG. 1. (a) The dynamics of epitaxial growth and the resulting rough interfaces. Mobility of the atoms on the surface, residence time, surface free energy, and other parameters determine the average island size. (b) In the $Al_xGa_{1-x}As$ system the Al atoms have the least mobility on the growing surface resulting in two interfaces that are rougher than the others in a typical double-barrier structure.

coefficients are then applied to suitably constructed (narrow in energy) wave packets[29] or to the spectral decomposition of an arbitrary wave packet.[30]

We modify the above procedure by using the following construction for the full wave function in each constant potential region:

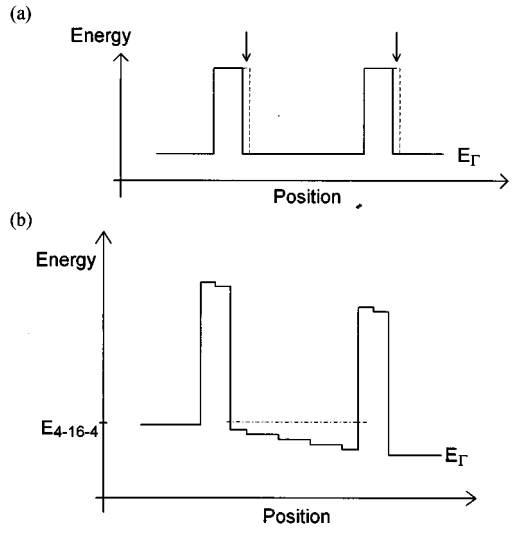

FIG. 2. (a) The conduction band edge energy for a double-barrier structure shown schematically for a structure with two roughened interfaces. (b) The constant potential approximation to the actual conduction band for a device biased near resonance.

$$\Psi_j=\sum_\nu (a_{j\nu}e^{ik_{j\nu}z}+b_{j\nu}e^{-ik_{j\nu}z})e^{i\mathbf{q}_\nu\cdot\mathbf{r}_\parallel},$$

where j is the region index and runs from 1 to N, ν indexes a set of transverse wave vectors, $\mathbf{q}_\nu=(q_{x\nu},q_{y\nu})$ is the transverse wave vector, $k_{j\nu}$ is the wave number for the component of the total wave vector normal to the average interface plane for the indicated $\mathbf{q}_\nu$, and $\mathbf{r}_\parallel$ is the in-plane coordinate vector. Here, the average interface plane is taken as normal to the z axis. In the above, each set of coefficients can be written in vector form as $\mathbf{a}_j$ and $\mathbf{b}_j$ where the number of elements in the vector represents the subset $\{\mathbf{q}_\nu\}$ retained in the calculation. The vector $\mathbf{a}_j$ corresponds to the forward propagating states in region j, and $\mathbf{b}_j$ to the backward propagating ones. Because we seek energy eigenfunctions, $k_{j\nu}$ and $\mathbf{q}_\nu$ are related via

$$\frac{\hbar^2(k_{j\nu}^2+q_\nu^2)}{2m_j}=E-V_j,$$

where E is the total electron energy, $q_\nu=|\mathbf{q}_\nu|$, and m_j, $k_{j\nu}$, and V_j are the effective mass, z wave number, and potential, respectively, in region j.

The above form of solution forces a strict plane-wave basis for the transverse portion of the wave function. This is reasonable when the average planes for each heterointerface in the structure are parallel. This plane-wave basis for the transverse portion of the wave function can be cumbersome; however, with enough terms the solution can be represented to high accuracy. It should be noted that the above form for the transverse wave function implies a finite crystal size in the transverse direction defined as L_x and L_y. We assume periodic boundary conditions at the edges of the crystal and the set $\{\mathbf{q}_\nu\}$ is then limited to the first Brillouin zone.

The transfer matrix we shall use, M, relates the coefficients in the bounding regions according to

$$\begin{bmatrix}\mathbf{a}_1\\ \mathbf{b}_1\end{bmatrix}=\mathsf{M}\begin{bmatrix}\mathbf{a}_N\\ \mathbf{b}_N\end{bmatrix}. \tag{1}$$

In a minimum analysis, N equals the number of distinct layers in the epitaxial structure and M is constructed from the product of $N-1$ interface transfer matrices, to be derived in Sec. IV. N will be larger if any of the layers is divided into multiple regions of constant potential; however, the ensuing regions should not be thinner than the magnitude of the roughness.

To model plane-wave incidence, set $\mathbf{a}_1=\mathbf{e}_\mu$ and $\mathbf{b}_N=\mathbf{0}$, where $\mathbf{e}_\mu$ is a vector with all zero elements except one unity value corresponding to a state incident with energy E and transverse wave vector $\mathbf{q}_{\text{in}}=\mathbf{q}_\mu$. Typically the set $\{\mathbf{q}_\nu\}$ is symmetric, spanning the range $\mathbf{q}_{\text{in}}-\mathbf{q}_{\max}$ to $\mathbf{q}_{\text{in}}+\mathbf{q}_{\max}$ with Δq_x and Δq_y, the granularity of $\{\mathbf{q}_\nu\}$, equal to $2\pi/L_x$ and $2\pi/L_y$, respectively.

Writing

$$\mathsf{M}=\begin{bmatrix}\mathsf{M}_A & \mathsf{M}_B\\ \mathsf{M}_C & \mathsf{M}_D\end{bmatrix},$$

we get the usual results

$$\mathbf{a}_1=\mathbf{e}_\mu=\mathsf{M}_A\mathbf{a}_N,$$

$$\mathbf{b}_1 = \mathsf{M}_C \mathbf{a}_N ,$$

which can be solved first for $\mathbf{a}_N$ and then $\mathbf{b}_1$. The first operation can be done by matrix inversion or, more efficiently, by lower triangular/upper triangular matrix (LU) decomposition of M_A and appropriate forward and back substitution.

The transmission and reflection coefficients are given by

$$T(E,\mathbf{q}_\mu \rightarrow \mathbf{q}_\nu) = T_{\mu\nu} = |\mathbf{a}_{N\nu}|^2 \frac{m_1}{m_N} \frac{k_{N\nu} + k_{N\nu}^*}{2k_{1\mu}}$$

and

$$R(E,\mathbf{q}_\mu \rightarrow \mathbf{q}_\nu) = R_{\mu\nu} = |\mathbf{b}_{1\nu}|^2 \frac{k_{1\nu} + k_{1\nu}^*}{2k_{1\mu}},$$

where it is assumed that $k_{1\mu}$ is real valued. Total transmission is given by $T_{\text{tot}} = \Sigma_\nu T_{\mu\nu}$ and "specular" transmission by $T_{\text{spec}} = T_{mm}$. Total transmission into scattered states is therefore $T_{\text{scatt}} = T_{\text{tot}} - T_{\text{spec}}$.

III. THE ROUGH INTERFACE BOUNDARY CONDITIONS

In this section boundary conditions correct to first order are derived. We do this for a single interface between materials 1 and 2, located at $z = z_1 + f(x,y)$, at which both the effective mass and band edge potential can change. It is not required that the average value of roughness $\langle f(x,y) \rangle$ but we do make the stipulation that $f(x,y)$ is single valued and continuous. This latter requirement is needed below in the definition of the normal vector in the boundary conditions. In practical usage, discontinuities in $f(x,y)$ are removed because of the truncation of the Fourier expansion of $f(x,y)$ used below to calculate the scattering coefficients. In this context z_1 serves as the (somewhat arbitrary) reference plane at which new boundary conditions will be derived.

The actual boundary conditions, within the effective-mass approximation, are

$$\Delta[\Psi(x,y,z_1+f)] = 0, \tag{2a}$$

$$\Delta\left(\frac{1}{m}\boldsymbol{\nabla}\Psi(x,y,z_1+f)\right)\cdot\mathbf{n} = 0, \tag{2b}$$

where the operator Δ extracts the difference between the function evaluated on each side of the interface, viz., $\Delta(h) = h_1(x,y,(z_1+f)^-) - h_2(x,y,(z_1+f)^+)$. The vector $\mathbf{n}$ is normal to the rough surface and given by

$$\mathbf{n} = (\partial_x f, \partial_y f, -1).$$

Here ∂_x and ∂_y indicate partial derivatives with respect to the coordinates x and y, respectively.

The above gradient boundary condition is peculiar to the effective-mass approximation within the envelope-function formalism. It is an extension of the condition on the time-independent Schrödinger equation which requires $\nabla^2\Psi$ to be finite valued everywhere (for finite valued potentials). This results in the continuity of $\boldsymbol{\nabla}\Psi$. Since the effective mass makes a step change across the interface $(1/m)\boldsymbol{\nabla}\Psi$ is assumed continuous in the direction normal to the interface. This choice is not unique;[31] however, it has been borne out by various theoretical calculations for such material systems as $\text{Al}_x\text{Ga}_{1-x}\text{As}$.[27,28]

Unfortunately, in materials or at wave vectors for which the effective-mass approximation is invalid, the form of boundary conditions is much less clear and tight-binding or pseudopotential methods need to be used to derive the proper connection relations for the wave function at the interface.[24,28] Since the effective-mass equation is only valid near a band extremum this approach is similarly restricted. Of course $(1/m)$ may be a tensor quantity; however, it is assumed scalar for clarity of exposition.

The rough interface is introduced perturbationally using the parameter A with $z = z_1 + Af(x,y)$, where $0 \leqslant A \leqslant 1$. The parameter A is used to keep track of the order of the perturbation and is later set to unity. Assume that the wave function in each region can be expressed as a Rayleigh-Schrödinger perturbational series

$$\Psi_1 = \Psi_1^{(0)} + A\Psi_1^{(1)} + A^2\Psi_1^{(2)} + \cdots \tag{3a}$$

and

$$\Psi_2 = \Psi_2^{(0)} + A\Psi_2^{(1)} + A^2\Psi_2^{(2)} + \cdots, \tag{3b}$$

where each function $\Psi_j^{(i)}$ is a solution to Schrödinger's equation. The index j indicates the material in which the solution is valid and i indicates the order of the term. The normal vector is also split into zeroth- and first-order parts,

$$\mathbf{n} = \mathbf{n}^{(0)} + A\mathbf{n}^{(1)} = (0,0,-1) + A(\partial_x f, \partial_y f, 0), \tag{4}$$

and is exact at first order. Note that splitting $\mathbf{n}$ in this way is somewhat arbitrary as the slope of a general function $f(x,y)$ is unrelated to its magnitude and expanding $\mathbf{n}$ into these two parts assumes that the roughness is slowly varying. This seems appropriate on two counts: first, the typical roughness in high-quality epitaxy is dominated by island growth so the steps are typically one monolayer in height and, second, change in the conduction band edge is distributed over a distance of one or two monolayers so the transverse steplike potential at an island edge is expected to be gradual rather than abrupt. This has implications for the "bandwidth" of the Fourier transform of $f(x,y)$, keeping that transform largely within the first Brillouin zone.

Since boundary conditions are sought at the plane $z = z_1$ that mimic the actual ones at $z = z_1 + f(x,y)$, the appropriate material in region 1 or 2 is extended up to that plane and each of the $\Psi_j^{(i)}$ is expanded around z_1 in a Taylor series,

$$\Psi_j^{(i)}(x,y,z_1+f) = \Psi_j^{(i)}(x,y,z_1) + (Af)\partial_z\Psi_j^{(i)}(x,y,z_1) + \frac{1}{2}(Af)^2\partial_z^2\Psi_j^{(i)}(x,y,z_1) + \cdots,$$

where $j = 1,2$ and $i = 0,1,2,\ldots$. It is this step that reduces the rough interface problem with homogeneous boundary conditions to a planar interface with the inhomogeneous boundary conditions derived below and its use mimics the Rayleigh hypothesis first used in rough-surface acoustic scattering.[35] These expansions are substituted into the expressions for the full wave functions, (3), and used with (4) in the actual boundary conditions at $z = z_1 + f(x,y)$, (2). One

then extracts relations for the continuity and gradient boundary conditions for the functions evaluated at z_1. The actual continuity boundary condition to second order is (all terms are evaluated at $z=z_1$)

$$\Delta(\Psi^{(0)})+Af\Delta(\partial_z\Psi^{(0)})+\frac{1}{2}(Af)^2\Delta(\partial_z^2\Psi^{(0)})+A\Delta(\Psi^{(1)})$$
$$+A^2f\Delta(\partial_z\Psi^{(1)})+A^2\Delta(\Psi^{(2)})=0.$$

Comparing terms of like order, the continuity boundary conditions on the zeroth-, first-, and second-order terms are

$$\Delta(\Psi^{(0)})=0,$$
$$\Delta(\Psi^{(1)})=-f\Delta(\partial_z\Psi^{(0)}),$$
$$\Delta(\Psi^{(2)})=-f\Delta(\partial_z\Psi^{(1)})-\frac{1}{2}f^2\Delta(\partial_z^2\Psi^{(0)}).$$

The continuity boundary condition for the ith-order term in the wave functon expansion is

$$\Delta(\Psi^{(i)})=-\sum_{\alpha=1}^{i}\frac{f^\alpha}{\alpha!}\Delta(\partial_z^\alpha\Psi^{(i-\alpha)}).$$

Note that the order of the Δ and ∂_z operations is not interchangeable.

Setting $A=1$, the above relations can be combined to generate the continuity boundary condition for the *total* wave function correct to zeroth, first, and second order:

$$\Delta(\Psi)^{(0)}=\Delta(\Psi^{(0)})=0, \tag{5a}$$
$$\Delta(\Psi)^{(1)}=\Delta(\Psi^{(0)}+\Psi^{(1)})=-f\Delta(\partial_z\Psi^{(0)}), \tag{5b}$$
$$\Delta(\Psi)^{(2)}=\Delta(\Psi^{(0)}+\Psi^{(1)}+\Psi^{(2)})$$
$$=-f\Delta(\partial_z[\Psi^{(0)}+\Psi^{(1)}])-\frac{1}{2}f^2\Delta(\partial_z^2\Psi^{(0)}). \tag{5c}$$

Here the superscript $^{(i)}$ is moved outside the parentheses, $\Delta(\cdots)^{(i)}$, to denote a boundary condition applicable to the total wave function, correct to ith order. These expressions are suitable for iterative calculations, but awkward for direct calculations of $\Delta(\Psi)^{(i)}$. This can be fixed by adding selected higher-order terms to the right hand sides of (5) without losing accuracy with respect to the prescribed order of each expression. The result is the following boundary conditions at $z=z_1$ accurate to zeroth, first, and second order:

$$\Delta(\Psi)^{(0)}=0, \tag{6a}$$
$$\Delta(\Psi)^{(1)}=-f\Delta(\partial_z\Psi), \tag{6b}$$
$$\Delta(\Psi)^{(2)}=-f\Delta(\partial_z\Psi)-\frac{1}{2}f^2\Delta(\partial_z^2\Psi). \tag{6c}$$

To arbitrary order, the continuity boundary condition is ($i\geqslant 1$)

$$\Delta(\Psi)^{(i)}=-\sum_{\alpha=1}^{i}\frac{f^\alpha}{\alpha!}\Delta(\partial_z^\alpha\Psi).$$

A similar set of operations leads to the following gradient boundary conditions for the total wave function, correct to zeroth, first, and second order:

$$\Delta\left(\frac{1}{m}\partial_z\Psi\right)^{(0)}=0, \tag{7a}$$
$$\Delta\left(\frac{1}{m}\partial_z\Psi\right)^{(1)}=(\partial_x f)\Delta\left(\frac{1}{m}\partial_x\Psi\right)+(\partial_y f)\Delta\left(\frac{1}{m}\partial_y\Psi\right)$$
$$-f\Delta\left(\frac{1}{m}\partial_z^2\Psi\right), \tag{7b}$$
$$\Delta\left(\frac{1}{m}\partial_z\Psi\right)^{(2)}=(\partial_x f)\left[\Delta\left(\frac{1}{m}\partial_x\Psi\right)+\Delta\left(\frac{1}{m}\partial_x(f\partial_z\Psi)\right)\right]$$
$$+(\partial_y f)\left[\Delta\left(\frac{1}{m}\partial_y\Psi\right)+\Delta\left(\frac{1}{m}\partial_y(f\partial_z\Psi)\right)\right]$$
$$-f\Delta\left(\frac{1}{m}\partial_z^2\Psi\right)-\frac{1}{2}f^2\Delta\left(\frac{1}{m}\partial_z^3\Psi\right). \tag{7c}$$

To arbitrary order the gradient boundary condition is ($i\geqslant 1$)

$$\Delta\left(\frac{1}{m}\partial_z\Psi\right)^{(i)}=(\partial_x f)\sum_{\alpha=0}^{i-1}\frac{1}{\alpha!}\Delta\left(\frac{1}{m}\partial_x(f^\alpha\partial_z^\alpha\Psi)\right)$$
$$+(\partial_y f)\sum_{\alpha=0}^{i-1}\frac{1}{\alpha!}\Delta\left(\frac{1}{m}\partial_y(f^\alpha\partial_z^\alpha\Psi)\right)$$
$$-\sum_{\alpha=1}^{i}\frac{f^\alpha}{\alpha!}\Delta\left(\frac{1}{m}\partial_z^{\alpha+1}\Psi\right),$$

and is equal to zero when $i=0$.

IV. THE FIRST-ORDER TRANSFER MATRIX

Now that the problem has been reduced to a planar interface we make the following notation change while discussing a single interfacial transfer matrix. The wave functions in regions 1 and 2 are written as

$$\Psi_1=\sum_\nu\ (a_\nu e^{ik_{1\nu}z}+b_\nu e^{-ik_{1\nu}z})e^{i\mathbf{q}_\nu\cdot\mathbf{r}_\parallel}, \tag{8a}$$
$$\Psi_2=\sum_\nu\ (c_\nu e^{ik_{2\nu}z}+d_\nu e^{-ik_{2\nu}z})e^{i\mathbf{q}_\nu\cdot\mathbf{r}_\parallel}. \tag{8b}$$

Here the index of each summation specifies the same set of transverse wave vectors $\{\mathbf{q}_\nu\}$.

We note in passing the effect of an effective-mass change upon specular tunneling, that is, tunneling across a planar interface with no change in $\mathbf{q}$. When $\mathbf{q}=\mathbf{0}$,

$$\frac{\hbar^2k_2^2}{2m_2}=\frac{\hbar^2k_1^2}{2m_1}-V_b$$

or

$$E_{z2}=E_{z1}-V_b,$$

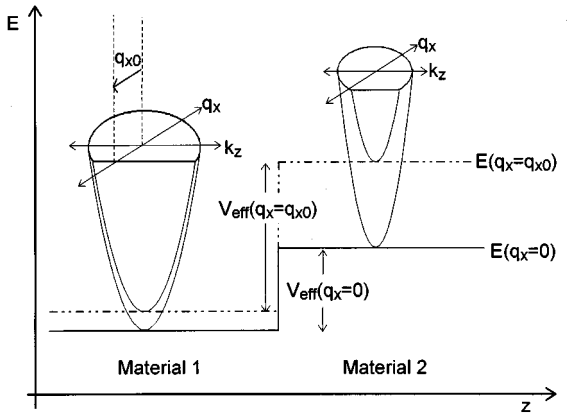

FIG. 3. The transverse-wave-vector-dependent effective potential step. Shown is the conduction band edge depicting a barrier formed by a heterostructure with a conduction band discontinuity. Material 2 has an effective mass smaller than material 1. The parabolic dispersion relations are indicated for each material. For a given transverse wave vector q_{x0}, a new barrier potential is specified by the cut of the dispersion relations defined by $q_x=q_{x0}$. In this case the potential step is effectively increased by nonzero transverse wave vector. When the effective masses are the reverse of those shown above, the effective potential step decreases with increasing transverse wave vector.

where $V_b=V_2-V_1$ is the step in conduction band edge potential energy between materials and E_{z1} and E_{z2} are the kinetic energies associated with motion in the z direction in each material. When $\mathbf{q}\neq\mathbf{0}$, this relation is written as

$$E_{z2}=E_{z1}-V_{\text{eff}}(q),$$

where

$$V_{\text{eff}}(q)=V_b+\frac{\hbar^2q^2}{2}\left(\frac{1}{m_2}-\frac{1}{m_1}\right).$$

$V_{\text{eff}}(q)$ is the effective potential step at the interface acting upon an electron incident with the transverse wave number $\mathbf{q}$ (see Fig. 3). This expression reflects one of the major differences that effective mass and band edge potential play in tunneling. An effective-mass change acts much like the index of refraction in optics and produces effects akin to refraction and total internal reflection. A step in the band edge potential by itself leaves the transverse and normal motion completely uncoupled and has no simple optical analog.

Note that the transverse wave numbers q_x and q_y are always real valued. While there may be localization due to quasibound quantum-well states in the plane of a device, the periodic boundary conditions at the edges of the crystal do not allow exponential transverse functions. Localization must be represented by a superposition of transverse plane waves. On the other hand the z wave numbers k_1 and k_2 can be either real or imaginary. For a given value of $\mathbf{q}$, three classes of solution are immediately apparent: when k_1 and k_2 are both real, only one of k_1 and k_2 is real, or both are imaginary. In the first case the wave is both transmitted and reflected, in the second case the wave is reflected with evanescent penetration into the second material, and in the third case a bound interface state is created. All of these cases exist at a rough interface in order to support the actual boundary conditions.

In single-interface problems the bound interface states are relatively unimportant as they are not included in the asymptotic wave function far from the interface. In multilayer structures these states can be important due to their nonvanishing amplitudes at the next interface.

To generate a set of algebraic relations between the sets of coefficients **a** and **b** with **c** and **d**, substitute (8) into the first-order boundary conditions (6b) and (7b), and then take the two-dimensional Fourier transform in the plane of the interface. The transformed continuity condition is then used in the transformed gradient condition to remove one set of coefficients. This process is detailed in the Appendix. The results are expressions for each of a_μ and b_μ in terms of **c**, **d**, and the Fourier coefficients of $f(x,y)$. The resulting transfer matrix of (1) for a single interface is

$$\mathsf{M}=\begin{bmatrix}\Phi^- & 0\\ 0 & \Phi^-\end{bmatrix}\mathsf{m}\begin{bmatrix}\Phi^+ & 0\\ 0 & \Phi^+\end{bmatrix}. \tag{9}$$

where m is the reduced transfer matrix, and Φ_1^- and Φ_1^+ are diagonal submatrices that contain the reference plane phase information, with

$$\Phi^-{}_{\mu\nu}=e^{ik_{1\mu}z_1}\delta_{\mu\nu},$$

$$\Phi^+{}_{\mu\nu}=e^{ik_{2\mu}z_1}\delta_{\mu\nu}.$$

If there are p elements in each set of coefficients then these submatrices are each $p\times p$. Noting the form of the total wave function in (3), it is expected that the reduced transfer matrix can also be expanded as a perturbational sum with

$$\mathsf{m}=\mathsf{m}^{(0)}+\mathsf{m}^{(1)}+\mathsf{m}^{(2)}+\cdots.$$

The first-order results of the Appendix represent the first two terms in this series, the first of which is

$$\mathsf{m}^{(0)}=\frac{1}{2}\begin{bmatrix}\mathsf{I}+\mathsf{R} & \mathsf{I}-\mathsf{R}\\ \mathsf{I}-\mathsf{R} & \mathsf{I}+\mathsf{R}\end{bmatrix},$$

where I is the identity matrix and R is a diagonal matrix with elements specified by

$$R_{\mu\nu}=\frac{m_1k_{2\mu}}{m_2k_{1\mu}}\delta_{\mu\nu}.$$

This term does not generate any scattering and reproduces the specular result for a planar interface located at the reference plane.

The first-order matrix is written in terms of a set of three scattering operators ($p\times p$ submatrices) according to

$$\mathsf{m}^{(1)}=\frac{i}{2}\begin{bmatrix}(\Omega_\nabla+\Omega_S+\Omega_C) & (\Omega_\nabla+\Omega_S-\Omega_C)\\ -(\Omega_\nabla+\Omega_S-\Omega_C) & -(\Omega_\nabla+\Omega_S+\Omega_C)\end{bmatrix},$$

where the elements of the scattering operators within this unity amplitude formalism are given by

$$\Omega_{C\mu\nu}=\frac{m_2-m_1}{m_2}k_{2\nu}\mathscr{F}_{\mu\nu},$$

$$\Omega_{\nabla\mu\nu}=\frac{1}{m_2k_{1\mu}}(m_1k_{2\nu}^2-m_2k_{1\nu}^2)\mathscr{F}_{\mu\nu},$$

$$\Omega_{S\mu\nu}=\frac{m_2-m_1}{m_2k_{1\mu}}S_{\mu\nu}\mathscr{F}_{\mu\nu}.$$

Here $\mathscr{F}_{\mu\nu}$ are two-dimensional Fourier series coefficients specified in (A1) with $\mathbf{q}=\mathbf{q}_\mu-\mathbf{q}_\nu$ and $S_{\mu\nu}$, given in (A4), contains the dependence of the interface scattering on the slope of the roughness.

V. ANALYSIS OF THE DERIVED RESULTS

The above interfacial transfer matrix is a particularly pleasing result. At first order, the rough interface couples inbound states (propagating or evanescent) to multiple outbound states according to the specified scattering operators above. This coupling takes the form of weighted convolutions with the Fourier transform of the surface roughness. The weighting factors collectively account for the interface deviation, the effective potential step, and the slope of $f(x,y)$.

It is interesting to note some of the properties of the scattering operators.

Ω_C: This operator includes all the effects of the continuity boundary condition and depends only upon the deviation of the rough interface from the reference plane $z=z_1$.

Ω_∇: This operator includes the effects of the gradient boundary condition dependent upon the deviation of the rough interface from the reference plane.

Ω_S: This operator includes the effects of the gradient boundary condition due to the slope of the interface roughness.

Some properties of the first-order interfacial transfer matrix *in general* are also worthy of note.

(1) When $f(x,y)=d$, a constant, the Fourier coefficients reduce to δ-functions $\delta_{\mu\nu}$, and there is no scattering to other transverse wave numbers as the electron interacts with the interface. Each submatrix in the above relations is diagonal and the problem reduces to that of the usual two-channel specular case.

(2) Also, when $f(x,y)=d$, the reflection coefficient and phase change upon reflection are, to first order in kd, identical to those of the exact result. This is deemed particularly important when modeling large islands in RTD epitaxial structures; particularly as regards the different resonant energies due to large scale well-width variations.

(3) When $m_1=m_2$ and $V_1=V_2$ the scattering operators are all zero and there is no reflected wave. This is the appropriate result as $f(x,y)$ no longer represents an interface; scattering results when $m_1\neq m_2$ or $V_1\neq V_2$.

(4) When $m_1=m_2$ both $\Omega_S=0$ and $\Omega_C=0$. This is an interesting result: to first order, only the effective-mass variation across the interface causes scattering due to the slope of $f(x,y)$ and the continuity boundary condition. When the effective-mass change is zero, only the potential step and the interface deviation from $z=z_1$ contribute to the first-order scattering via Ω_∇.

(5) The operator Ω_∇ has the alternate form

$$\Omega_{\nabla\mu\nu}=\frac{2im_1}{\hbar^2k_{1\mu}}V_{\text{eff}}(\mathbf{q}_\nu)\mathscr{F}_{\mu\nu}. \tag{10}$$

It is interesting that the effective potential for specular tunneling of the final wave vector acts as the weighting factor for each term in this scattering operator. When $m_1=m_2$ then $V_{\text{eff}}=V_b$, producing the result obtained when the surface roughness is modeled as a planar interface with a variable δ-function potential at the reference plane.[5]

Aside from the above general considerations, it is also worthwhile to comment on the accuracy and applicability of the method. When $f(x,y)=d$, these boundary conditions are accurate as long as $|k|d\ll 1$ in each layer. This condition is easily met in most high-quality epitaxial structures. In structures with large-band-gap semiconductors it becomes somewhat questionable. For GaAs/AlAs structures $|k|_{\max}\approx 2\times 10^9\,\text{m}^{-1}$ (for electrons in the barrier). The maximum allowable reference plane shift from the average interface is therefore expected to be about one-half a monolayer ($d\approx 1.4$ Å) if this method is to remain accurate.

When $f(x,y)$ is *not* planar, the overall validity of the method becomes less clear. It is required for all coefficients that are "appreciable" that $|k_{j\nu}\mathscr{F}_{\mu\nu}|\ll 1$. [Another version of this requirement is $|k_{j\nu}|\max(|f(x,y)|)\ll 1$, which is not as precise but easier to interpret.] While vague, this is a disturbing requirement because the scattering operators can run to the Brillouin zone edge where, under the assumption of parabolic bands, $|k_{j\nu}|$ is large. Fortunately the method is most inaccurate precisely where the coefficients are expected to be smallest.

Another consideration with this method is the number of transverse wave vectors to retain in the calculation. Because of the resonant nature of the problem this set can be larger than expected. Essentially, each island size requires a certain transverse wave vector to produce the localized state. If the interfaces in question have both very large- and very small-sized islands and valleys then a large number of terms need to be retained, making simulations of these interfaces computationally difficult.

From the above, it is seen that this technique is limited to systems with Fourier coefficients in the central portion of the Brillouin zone. This is a redundant condition as the effective-mass approximation requires a similar restriction. Such a restriction is not as limiting as it might seem as it sets the minimum island size to around 20 Å. The limitations of this method are therefore found in large-band-gap barrier materials (in which $|k|_{\max}$ can be quite large), very rough surfaces [for which $\max(|f|)$ is large], the assumption of slowly varying $V(z)$ within any layer of homogeneous material, the parabolic band approximation, the use of simple boundary conditions, and the simulation of interfaces with a large variation in feature size.

It should also be noted that the first-order scattering matrix produces a set of scattering coefficients that obey time reversal symmetry. However, because the method does not include all terms, it does not completely conserve particle flux. In the simulations that follow, overall particle conservation deviates from unity by less than 0.6% in all but the last simulation for which it is 1.5%. This deviation occurs

only in the peaked portion of the transmission characteristic and reduces to zero far from resonance.

Finally, we discuss the problematic use of δ functions to represent the surface roughness. Some treatments use δ functions that condense a three-dimensional island to a point with the same integrated perturbation.[3,8,10] This allows impurity-type formalisms to be used; however, it seems questionable whether such a formalism can accurately model transverse localization and transmission via these localized states, especially in the limit of large island size.

In addition, the transfer-matrix formalism, as implemented by Liu and Coon, approximates the interface roughness as a sheet δ function at the reference plane in an application of the Taylor series[5]

$$V(x,y,z) = V_b \Theta(z - z_1 - f(x,y)),$$
$$\approx V_b \Theta(z - z_1) - V_b f(x,y)\,\delta(z - z_1). \quad (11)$$

This expansion does not converge to $V(x,y,z)$ between z_1 and $z_1 + f(x,y)$ and using the first two terms of this expansion merely preserves the integrated value of the potential across the interface. Using (11) results in an equation of the form of (10) but with V_{eff} replaced by V_b. This leaves out the effects of the effective mass upon scattering: when $m_1 \neq m_2$ and $V_1 = V_2$ no scattering is predicted. This deficiency is also present for the point δ-function condensation. Any use of δ functions to approximate interface roughness should therefore be considered perilous at best.

VI. MODEL CALCULATIONS: RANDOM MONOLAYER ROUGHNESS

In this section, numerical results for normally incident electrons on an AlAs/GaAs double-barrier structure are briefly reported. The target device has a double-barrier structure with four monolayer barriers and a 16-monolayer well region which is referred to as the 4-16-4 structure. A 1.0 eV conduction band edge offset is used and the total, specular and total scattered transmission coefficients are calculated based on a biased device with the potential shown in Fig. 2(b). This structure has two resonances below the barrier energy. The lowest of these is responsible for the peaked *I-V* curve and the following simulations concentrate exclusively upon it.

The rough interfaces are all assumed to have 50% coverage by one-monolayer-high islands. No effort is made to model any realistic epitaxial growth model, but rather equal radius islands are distributed at random across a surface five times the size of the islands. That is, the simulation with 50 Å islands covers a square virtual crystal that is 250 Å on a side. Where islands overlap, no additional height is added to the rough interface. The resulting interface has a variety of (simulated) island sizes and shapes. A typical rough interface generated by this process is shown in Fig. 4. This is the roughness function used in all cases below.

Simulations are performed using a single rough well interface with three island sizes: 50, 500, and 5000 Å. The results of these simulations are shown in Figs. 5 through 7. Plotted in these figures are the total probability of transmission, the probability of specular transmission, and transmission into a scattered state. The roughness is located at the emitter-barrier/well interface and corresponds to extra growth of the barrier layer. The sum of the emitter barrier and well widths is kept constant so the ensuing structure has lateral regions where the barrier is 4 or 5 monolayers wide and the well is 16 and 15 monolayers wide, respectively. The resulting transmission line shapes are therefore expected to be loosely bounded by the transmission characteristics of the two structures taken separately: the planar 4-16-4 and the 5-15-4 monolayer structures.

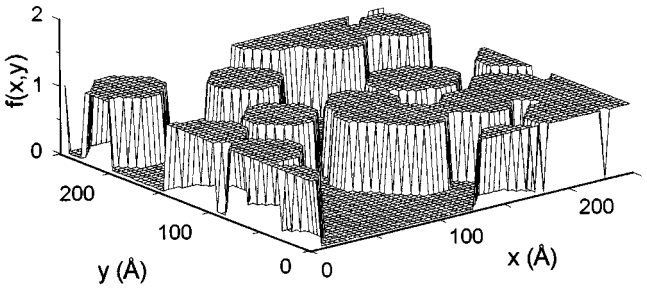

FIG. 4. A mesh plot of the roughness function used for the simulations of single-monolayer two-dimensional roughness. The *x* and *y* dimensions are arbitrary and scale with the simulations. Here they indicate the dimensions used for the 50 Å island case. The vertical scale is in monolayers.

Figure 5 shows the transmission characteristic in the presence of 50 Å islands. Included in this figure are the transmission coefficients for the planar 4-16-4 and 5-15-4 structures. The 50 Å roughness shifts the transmission peak and slightly broadens it. Clearly the roughness is on such a small scale that only the average interface is important. Note that the peak of T_{tot} is slightly wider than the two bounding planar characteristics. In this regime, the broadening is akin to a correction due to the imaginary part of the self-energy.[32]

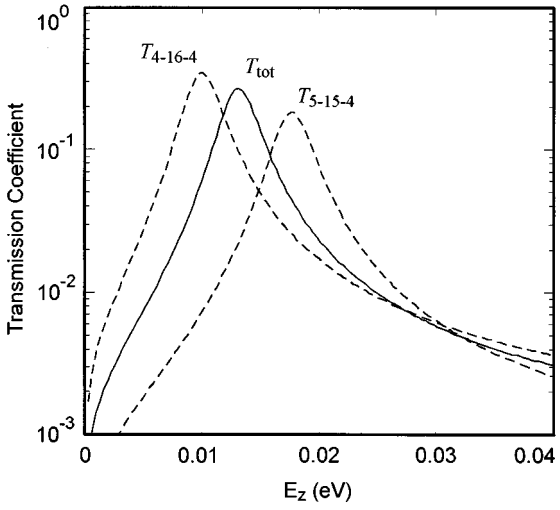

FIG. 5. A plot of the total, specular, and total scattered transmission coefficients for a biased RTD with one rough well interface characterized by 50% coverage by 50 Å islands. This simulation uses the central 9×9 Fourier coefficients of a 250 Å sample. In this case, $T_{\text{tot}} \approx T_{\text{spec}}$ and T_{scatt} is too small to appear in this plot. Also shown (dashed lines) are the planar-interface 4-16-4 and 5-15-4 resonances.

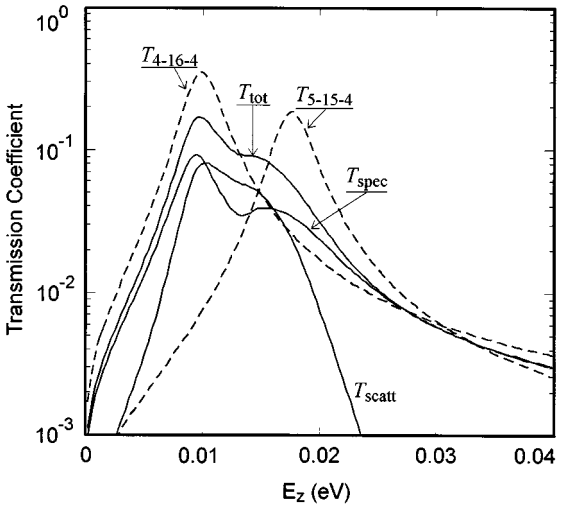

FIG. 6. A plot of the total, specular, and total scattered transmission coefficients for a biased RTD with one rough well interface characterized by 50% coverage by 500 Å islands. This simulation uses the central 13×13 Fourier coefficients of a 2500 Å sample. Also shown (dashed lines) are the planar-interface 4-16-4 and 5-15-4 resonances.

The overwhelming part of the transmission across the resonance is specular. The lack of a large scattered component is due to the large transverse wave vector picked up by those electrons that do scatter. The characteristic island wave vector magnitude is $\pi/50$ Å^{-1} around which $\mathscr{T}_{\mathbf{q}}$ is large. For this size wave vector the scattered states are nonpropagating in all layers.

Figure 6 shows the transmission characteristic for scattering from the same roughness function used above, but scaled for a 500 Å island size. At this larger size, the transmission

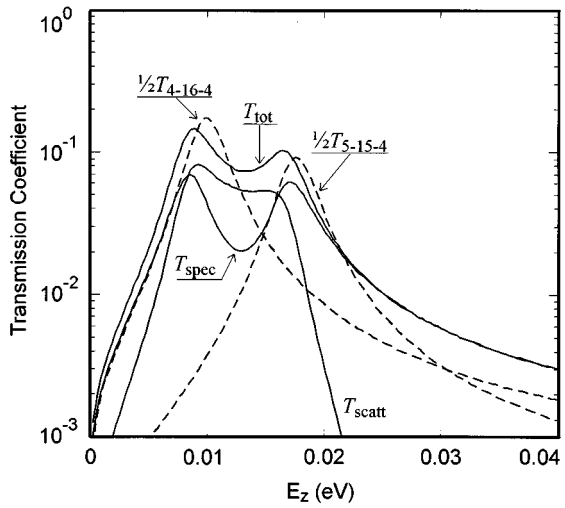

FIG. 7. A plot of the total, specular, and total scattered transmission coefficients for a biased RTD with one rough well interface characterized by 50% coverage by 5000 Å islands. This simulation uses the central 13×13 Fourier coefficients of a 25 000 Å sample. Also shown (dashed lines) are the *half-height* planar-interface 4-16-4 and 5-15-4 resonances.

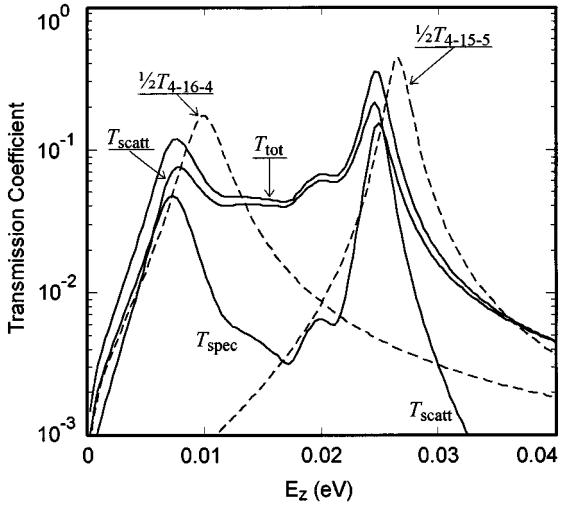

FIG. 8. Monolayer roughness at the well/collector-barrier interface for a biased RTD and 5000 Å islands. Also shown (dashed lines) are the *half-height* planar-interface 4-16-4 and 4-15-5 resonances.

characteristic has clearly split into two peaks, but it has done so unevenly, even though the coverage is 50%. Superimposed upon the figure are the two planar transmission line shapes for the 4-16-4 and 5-15-4 monolayer structures. Note in particular the dominance of transmission to scattered states between the peaks and the weak specular component of the high-energy peak (due to the narrow-well portions of the device). The wide-well regions form laterally localized states while the narrow-well regions form loosely bound continuum states. In the lateral direction, confinement to these regions is poor and the ensuing resonance is not as strong, causing a smaller high-energy peak in the transmission characteristic.

Figure 7 shows the transmission characteristic for the same roughness function scaled for a 5000 Å island size. Here it is seen that transmission via the islands and valleys is becoming much more independent: the double-peak characteristic more nearly approximates the sum of two half-height 4-16-4 and 5-15-4 peaks. Between the peaks, the transmission to scattered states still dominates.

Finally, the effect of monolayer roughness at the well/collector-barrier is shown in Fig. 8. Here the total, specular, and total scattered transmission coefficients are shown for the case of 5000 Å islands. Superimposed on this graph are two half-height 4-16-4 and 4-15-5 planar-interface transmission resonances. Several effects are apparent in the figure. First, the scattering effect is stronger. This is due to the deeper potential well at this interface. Where the well is deeper, interface roughness has a larger effect on the round-trip phase change (the resonance condition in the well). Thus the rough interface transmission coefficient is now spread out over a wider range of energies. Secondly, the peak energies are slightly less than those of the planar 4-16-4 and 4-15-5 resonances. This effect is also seen in the previous simulations and is due to the limitations of the first-order theory. Thirdly, the high-energy peak is now the higher of the two. This is due to the higher symmetry of the biased 4-15-5 structure.

VII. DISCUSSION

A transfer-matrix method for calculating the coherent tunneling through heterostructures has been developed to first order, starting from the exact boundary conditions within the effective-mass approximation. The method solves for the electron wave function on either side of a rough heterointerface and includes the evanescent, or ''near field' portion of the solution. This technique can be used in single- and multiple-barrier structures. The first-order results show the different roles of effective-mass and band edge discontinuities in roughened interface scattering as well as the roles of interface deviation and slope.

The class of problems for which this theory is expected to be most useful are tunneling problems through single and double barriers. Single-barrier tunneling is sufficiently fast that the exclusion of inelastic processes is deemed quite reasonable.[33] In resonant transport though double-barrier devices such as RTD's, this is not always true.[34] The tunneling time for various RTD's is estimated to run from hundreds of femtoseconds to nanoseconds, depending on the barrier thicknesses involved.[13] For the fastest devices in this range, inelastic processes may not be important. For the slower devices, this approach outlines the baseline coherent process upon which the inelastic processes are parasitic. In addition, this approach can provide information as to the degree of localization present in quasibound two-dimensional systems and can be expanded to cover the effects of interface roughness in optical systems such as Fabry-Pérot interferometers with thickness variations in both the reflectors and cavity.

The flexibility of this method is due to the interfacial nature of the technique. Rather than starting with a complete planar structure and calculating the first-order correction to its eigenstates due to interfacial roughness, this technique finds the first-order scattering terms at each rough interface. The transfer matrices for the rough interfaces can be combined to form the global transfer matrix for the entire structure. The scattered waves are then allowed to constructively or destructively interfere or scatter again, depending on the form of the global transfer matrix. This can be applied to both single- and multiple-barrier devices without previous knowledge of the planar eigenstates.

One drawback of this technique is that the accuracy of the numerical calculations depends not only upon the validity of the first-order approximation, but also on the number and choice of transverse plane waves used in the full expansion of the wave function. Structures with both long- and short-range lateral roughness features require a large number of transverse states and it is difficult to tell ahead of time how many states should be included. This is particularly true in resonant situations where multiple scattering can excite locally bound resonant states that require a large change in the transverse wave vector of the incident plane wave. In addition, as the barriers are made thicker, the effective electron lifetime (escape time) is longer and a broader range of resonant states can be excited.

Another feature of this method is the ability to explore scattering due to correlated rough interfaces and scattering from roughness with arbitrary statistics. This is noteworthy in that many approaches treat the rough surface statistically, losing the correlations among the interfaces, and typically assume a Gaussian height-height autocorrelation.[3,4,8] Since the epitaxial growth process favors island growth, such an assumption is dubious. The method developed here is free of critical assumptions with regard to the statistical nature of the interfaces and can be used in conjunction with an accurate model of the epitaxial growth process or to predict the effects of known or presumed interfacial features.

The utility of this method has been demonstrated in a double-barrier structure. The method can be used over a large range of rough interfaces from very short-range order to very long-range order. Most importantly this method can be used in the intermediate range where the lateral features of interface roughness are on the order of the resonant wavelength.

APPENDIX: DERIVING THE FIRST-ORDER TRANSFER MATRIX

Substituting (8) into (6b) and using the orthogonality condition

$$\frac{1}{L_x L_y}\int_A d\mathbf{r}_\parallel e^{-i(\mathbf{q}_\nu-\mathbf{q}_\mu)\cdot\mathbf{r}_\parallel}=\delta_{nm}$$

yields

$$\begin{aligned} a_\mu e^{ik_{1\nu}z_1}+b_\mu e^{-ik_{1\nu}z_1}-c_\mu e^{ik_{2\nu}z_1}-d_\mu e^{-ik_{2\nu}z_1} \\ =i\sum_\nu k_{2\mu}(c_\nu e^{ik_{2\mu}z_1}-d_\nu e^{-ik_{2\mu}z_1})\mathscr{F}_{\mu\nu} \\ -i\sum_\nu k_{1\mu}(a_\nu e^{ik_{1\mu}z_1}-b_\nu e^{-ik_{1\mu}z_1})\mathscr{F}_{\mu\nu}, \end{aligned}$$

where $\mathscr{F}_{\mu\nu}=\mathscr{F}_{\mathbf{q}_\mu-\mathbf{q}_\nu}$ is one of the two-dimensional Fourier coefficients of $f(x,y)$. Note that the finite sized crystal with periodic boundary conditions used here makes $f(x,y)=f(x+nL_x,y+mL_y)$, where n and m are integers. In this case the set $\{\mathbf{q}_\nu\}$ used in (8) is used for these coefficients as well; however, the Fourier coefficients are not limited to the first Brillouin zone. The Fourier coefficients are given by the usual definition:

$$\mathscr{F}_{\mathbf{q}}=\frac{1}{L_x L_y}\int_A d\mathbf{r}_\parallel f(\mathbf{r}_\parallel)e^{-i\mathbf{q}\cdot\mathbf{r}_\parallel}. \tag{A1}$$

Absorbing the phase factors $\exp\{\pm ik_{j\nu}z_1\}$ into the coefficients for now and collecting terms with a_μ and b_μ on one side yields

$$\begin{aligned} a_\mu+b_\mu+i\sum_\nu k_{1\nu}(a_\nu-b_\nu)\mathscr{F}_{\mu\nu} \\ =c_\mu+d_\mu+i\sum_\nu k_{2\nu}(c_\nu-d_\nu)\mathscr{F}_{\mu\nu}. \end{aligned} \tag{A2}$$

By the same set of operations, the gradient boundary condition of (7b) transforms to

$$m_2k_{1\mu}(a_\mu-b_\mu)+im_2\sum_\nu S_{\mu\nu}(a_\nu+b_\nu)\mathscr{F}_{\mu\nu}$$

$$-im_2\sum_\nu k_{1\nu}^2(a_\nu+b_\nu)\mathscr{F}_{\mu\nu}$$

$$=m_1k_{2\mu}(c_\mu-d_\mu)+im_1\sum_\nu S_{\mu\nu}(c_\nu+d_\nu)\mathscr{F}_{\mu\nu}$$

$$-im_1\sum_\nu k_{2\nu}^2(c_\nu+d_\nu)\mathscr{F}_{\mu\nu}. \tag{A3}$$

Here the slope dependence is included in

$$S_{\mu\nu}=q_{x\nu}(q_{x\nu}-q_{x\mu})+q_{yn}(q_{yn}-q_{ym}). \tag{A4}$$

To arrive at this result one uses relations of the form

$$\frac{1}{L_xL_y}\int_A d\mathbf{r}_\parallel[\partial_x f(\mathbf{r}_\parallel)]e^{-i(\mathbf{q}_\mu-\mathbf{q}_\nu)\cdot\mathbf{r}_\parallel}=i(q_{x\mu}-q_{x\nu})\mathscr{F}_{\mu\nu}.$$

The sums that include $S_{\mu\nu}$ are zero when the slope of $f(x,y)$ is zero.

To solve for a_μ or b_μ, substitute the continuity relation into the gradient relation. Solving (A2) for b_μ yields

$$b_\mu=c_\mu+d_\mu-a_\mu+i\sum_\nu [k_{2\nu}(c_\nu-d_\nu)-k_{1\nu}(a_\nu-b_\nu)]\mathscr{F}_{\mu\nu}$$

$$\approx c_\mu+d_\mu-a_\mu+i\sum_\nu [k_{2\nu}(c_\nu-d_\nu)+k_{1\nu}(c_\nu+d_\nu)]\mathscr{F}_{\mu\nu}$$

$$-2i\sum_\nu k_{1\nu}a_\nu\mathscr{F}_{\mu\nu}.$$

In this last step we use only the zeroth-order portion of (A2). Substituting this into (A3) results in

$$2a_\mu+2i\sum_\nu k_{1\nu}a_\nu\mathscr{F}_{\mu\nu}$$

$$=c_\mu\left(1+\frac{m_1k_{2\mu}}{m_2k_{1\mu}}\right)+d_\mu\left(1-\frac{m_1k_{2\mu}}{m_2k_{1\mu}}\right)+i\sum_\nu [k_{2\nu}(c_\nu-d_\nu)$$

$$+k_{1\nu}(c_\nu+d_\nu)]\mathscr{F}_{\mu\nu}+\frac{i}{m_2k_{1\mu}}\sum_\nu (c_\nu+d_\nu)(m_1k_{2\nu}^2$$

$$-m_2k_{1\nu}^2)\mathscr{F}_{\mu\nu}-i\frac{m_2-m_1}{m_2k_{1\mu}}\sum_\nu S_{\mu\nu}(c_\nu+d_\nu)\mathscr{F}_{\mu\nu}.$$

This is almost the desired form. To evaluate the second term on the left hand side, we need an expression, accurate to zeroth order, for $2a_\nu$ (this term is already at first order). This is readily obtained from this equation by dropping all first-order terms,

$$2a_\nu\approx c_\nu\left(1+\frac{m_1k_{2\nu}}{m_2k_{1\nu}}\right)+d_{\mathbf{q}_\mu}\left(1-\frac{m_1k_{2\nu}}{m_2k_{1\nu}}\right).$$

This results in the following first-order relation for a_μ:

$$2a_\mu=c_\mu\left(1+\frac{m_1k_{2\mu}}{m_2k_{1\mu}}\right)+d_\mu\left(1-\frac{m_1k_{2\mu}}{m_2k_{1\mu}}\right)$$

$$+i\left(1-\frac{m_1}{m_2}\right)\sum_\nu k_{2\nu}(c_\nu-d_\nu)\mathscr{F}_{\mu\nu}$$

$$+\frac{i}{m_2k_{1\mu}}\sum_\nu (c_\nu+d_\nu)(m_1k_{2\nu}^2-m_2k_{1\nu}^2)\mathscr{F}_{\mu\nu}$$

$$-i\frac{m_2-m_1}{m_2k_{1\mu}}\sum_\nu S_{\mu\nu}(c_\nu+d_\nu)\mathscr{F}_{\mu\nu},$$

A similar set of operations for b_μ results in

$$2b_\mu=c_\mu\left(1-\frac{m_1k_{2\mu}}{m_2k_{1\mu}}\right)+d_\mu\left(1+\frac{m_1k_{2\mu}}{m_2k_{1\mu}}\right)$$

$$+i\left(1-\frac{m_1}{m_2}\right)\sum_\nu k_{2\nu}(c_\nu-d_\nu)\mathscr{F}_{\mu\nu}$$

$$-\frac{i}{m_2k_{1\mu}}\sum_\nu (c_\nu+d_\nu)(m_1k_{2\nu}^2-m_2k_{1\nu}^2)\mathscr{F}_{\mu\nu}$$

$$+i\frac{m_2-m_1}{m_2k_{1\mu}}\sum_\nu S_{\mu\nu}(c_\nu+d_\nu)\mathscr{F}_{\mu\nu}.$$

These relations, along with the reference plane phase factors extracted from the coefficients, are used for M and m in (9).

[1] M. Tanaka, H. Sakaki, and J. Yoshino, Jpn. J. Appl. Phys. **25**, L155 (1986).

[2] J. Sudijono, M. D. Johnson, M. B. Elowitz, C. W. Snyder, and B. G. Orr, Surf. Sci. **280**, 247 (1993).

[3] J. Leo and A. H. MacDonald, Phys. Rev. Lett. **64**, 817 (1990).

[4] J. Leo and A. H. MacDonald, Phys. Rev. B **43**, 9763 (1991).

[5] H. C. Liu and D. D. Coon, J. Appl. Phys. **64**, 6785 (1988).

[6] H. C. Liu, J. Appl. Phys. **67**, 593 (1990).

[7] K. Ploog, in *Resonant Tunneling in Semiconductors*, edited by L. L. Chang, E. E. Mendez, and C. Tejedor (Plenum Press, New York, 1991), p. 17.

[8] B. G. R. Rudberg, Semicond. Sci. Technol. **5**, 600 (1990).

[9] P. Gueret and C. Rossel, in *Resonant Tunneling in Semiconductors* (Ref. 7), p. 71.

[10] P. Johansson, Phys. Rev. B **46**, 12 865 (1992).

[11] See, for example, A. Gold, Phys. Rev. B **41**, 8537 (1990); T. Ando

and S. Mori, Surf. Sci. **113**, 124 (1982).
[12] H. Sakaki, H. Yashimura, M. Tsuchiya, and T. Matsusue, in *Resonant Tunneling in Semiconductors* (Ref. 7), p. 307.
[13] M. K. Jackson, M. B. Johnson, D. H. Chow, T. C. McGill, and C. W. Nieh, Appl. Phys. Lett. **54**, 522 (1989).
[14] A. Steyerl, S. S. Malik, and L. R. Iyengar, Physica B **173**, 47 (1991).
[15] *Technology and Physics of Molecular Beam Epitaxy*, edited by E. H. C. Parker (Plenum, New York, 1985).
[16] J. M. Brown, N. Holonyak, Jr., M. J. Ludowise, W. T. Dietze, and C. R. Lewis, Electron. Lett. **20**, 204 (1984).
[17] L. Hart, P. F. Fewstwr, M. J. Ashwin, M. R. Fahy, and R. C. Newman, J. Phys. D **28**, A154 (1995).
[18] H. Sakaki, M. Tanaka, and J. Yoshino, Jpn. J. Appl. Phys. **24**, L417 (1985).
[19] E. Kretchmann, Z. Phys. **237**, 1 (1970); H. J. Juranek and E. Kroger, *ibid.* **233**, 324 (1970).
[20] H. A. Fertig and S. Das Sarma, Phys. Rev. B **40**, 7410 (1989); H. A. Fertig, Phys. Rev. Lett. **65**, 2321 (1990).
[21] R. Berkovits and S. Feng, Phys. Rev. B **45**, 97 (1992).
[22] E. Runge and H. Ehrenreich, Ann. Phys. (N.Y.) **219**, 55 (1992).
[23] E. O. Kane in *Tunneling Phenomena in Solids*, edited by E. Burnstein and D. Lundquist (Plenum, New York, 1969), p. 1.
[24] T. B. Boykin, Jan P. A. van der Wagt, and J. S. Harris, Jr., Phys. Rev. B **43**, 4777 (1991).
[25] K. V. Rousseau, K. L. Wang, and J. N. Schulman, Appl. Phys. Lett. **54**, 1341 (1989); J. N. Schulman and Y.-C. Chang, Phys. Rev. B **27**, 2346 (1983).
[26] D. J. BenDaniel and C. B. Duke, Phys. Rev. **152**, 683 (1966); G. Bastard and J. A. Brum, IEEE J. Quantum Electron. **22**, 1625 (1986).
[27] M. G. Burt, J. Phys. Condens. Matter **4**, 6651 (1992).
[28] H. Akera, S. Wakahara, and T. Ando, Surf. Sci. **196**, 694 (1988); T. Ando, S. Wakahara, and H. Akera, Phys. Rev. B **40**, 11 609 (1989); T. Ando and H. Akera, *ibid.* **40**, 11 619 (1989).
[29] R. Shankar, *Principles of Quantum Mechanics* (Plenum, New York, 1980), p. 178.
[30] See, for example, D. Bohm, *Quantum Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1951), p. 290; T. E. Hartman, J. Appl. Phys. **33**, 3427 (1962).
[31] W. A. Harrison, Phys. Rev. **123**, 85 (1961); H. Kroemer and Q.-G. Zhu, J. Vac. Sci. Technol. **21**, 551 (1982); O. von Roos, Phys. Rev. B **27**, 7547 (1983); R. A. Morrow and K. R. Brownstein, *ibid.* **31**, 1135 (1985).
[32] H. A. Fertig, S. He, and S. Das Sarma, Phys. Rev. B **41**, 3596 (1990).
[33] D. Esteve, J. M. Martinis, C. Urbina, E. Turlot, M. H. Devoret, H. Grabert, and S. Linkwitz, Phys. Scr. **T29**, 121 (1989); P. Gueret, in *Electronic Properties of Multilayers and Low-Dimensional Semiconductor Structures*, edited by J. M. Chamberlain, L. Eaves, and J. C. Portal (Plenum, New York, 1990), p. 317.
[34] W. T. Dietze and R. B. Darling, Phys. Rev. B **51**, 10 886 (1995).
[35] J. W. Strutt (Lord Rayleigh), *The Theory of Sound* (Dover, New York, 1945), Vol. II, p. 89.