

## Single and multiband modeling of quantum electron transport through layered semiconductor devices

Roger Lake, Gerhard Klimeck, R. Chris Bowen, and Dejan Jovanovic

Citation: *J. Appl. Phys.* **81**, 7845 (1997); doi: 10.1063/1.365394

View online: <http://dx.doi.org/10.1063/1.365394>

View Table of Contents: <http://jap.aip.org/resource/1/JAPIAU/v81/i12>

Published by the [American Institute of Physics](#).

---

### Related Articles

Light trapping in solar cells: Analytical modeling

*Appl. Phys. Lett.* **101**, 151105 (2012)

Analysis of the external and internal quantum efficiency of multi-emitter, white organic light emitting diodes

*Appl. Phys. Lett.* **101**, 143304 (2012)

Analysis of the external and internal quantum efficiency of multi-emitter, white organic light emitting diodes

*APL: Org. Electron. Photonics* **5**, 224 (2012)

Impact of oxygen annealing on high-k gate stack defects characterized by random telegraph noise

*Appl. Phys. Lett.* **101**, 122105 (2012)

Ozone-exposure and annealing effects on graphene-on-SiO<sub>2</sub> transistors

*Appl. Phys. Lett.* **101**, 121601 (2012)

---

### Additional information on J. Appl. Phys.

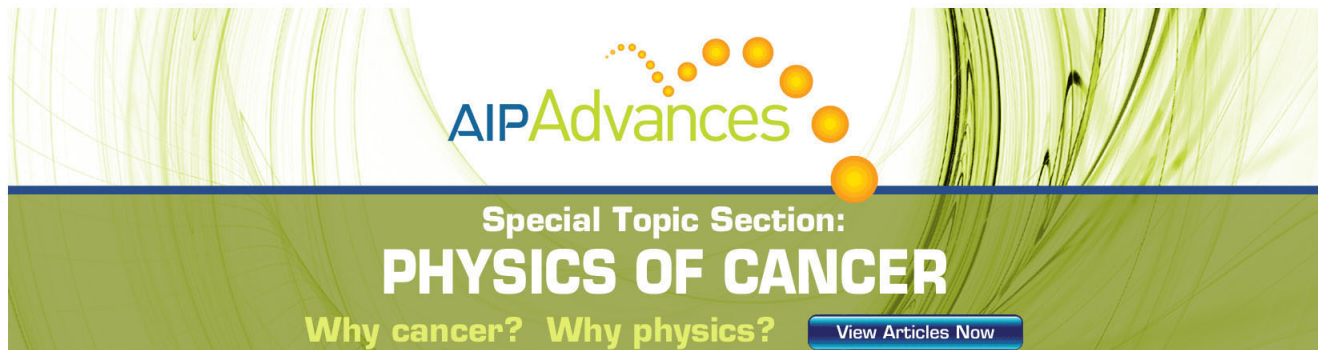
Journal Homepage: <http://jap.aip.org/>

Journal Information: [http://jap.aip.org/about/about\\_the\\_journal](http://jap.aip.org/about/about_the_journal)

Top downloads: [http://jap.aip.org/features/most\\_downloaded](http://jap.aip.org/features/most_downloaded)

Information for Authors: <http://jap.aip.org/authors>

## ADVERTISEMENT



The advertisement banner features a green background with abstract, flowing lines. At the top, the text "AIPAdvances" is displayed in a stylized font, with "AIP" in blue and "Advances" in green. Below this, the text "Special Topic Section:" is in white, followed by "PHYSICS OF CANCER" in large, bold, white capital letters. At the bottom, the text "Why cancer? Why physics?" is in yellow, and a blue button with white text says "View Articles Now".

# Single and multiband modeling of quantum electron transport through layered semiconductor devices

Roger Lake,<sup>a)</sup> Gerhard Klimeck, R. Chris Bowen, and Dejan Jovanovic  
Corporate Research Laboratories, Texas Instruments Incorporated, Dallas, Texas 75243

(Received 13 November 1996; accepted for publication 18 February 1997)

Non-equilibrium Green function theory is formulated to meet the three main challenges of high bias quantum device modeling: self-consistent charging, incoherent and inelastic scattering, and band structure. The theory is written in a general localized orbital basis using the example of the zinc blende lattice. A Dyson equation treatment of the open system boundaries results in a tunneling formula with a generalized Fisher-Lee form for the transmission coefficient that treats injection from emitter continuum states and emitter quasi-bound states on an equal footing. Scattering is then included. Self-energies which include the effects of polar optical phonons, acoustic phonons, alloy fluctuations, interface roughness, and ionized dopants are derived. Interface roughness is modeled as a layer of alloy in which the cations of a given type cluster into islands. Two different treatments of scattering; self-consistent Born and multiple sequential scattering are formulated, described, and analyzed for numerical tractability. The relationship between the self-consistent Born and multiple sequential scattering algorithms is described, and the convergence properties of the multiple sequential scattering algorithm are numerically demonstrated by comparing with self-consistent Born calculations. © 1997 American Institute of Physics. [S0021-8979(97)00411-8]

## I. INTRODUCTION

The design of resonant tunneling based quantum devices requires accurate modeling of the quantum charge, resonant levels, and scattering effects in complicated and varied potential profiles made possible by hetero-epitaxial based band engineering. The modeling of such structures poses combined theoretical and numerical challenges. The two great theoretical challenges are to include band-structure and scattering effects. What makes the theory particularly challenging is that it must be formulated to be numerically tractable for the simulation of realistic quantum devices. To make clear what we mean by a realistic device, we refer the reader to Figs. (1) and (2) of Ref. 1. Because of the length of the devices and the extreme thermalization combined with the quantum structure in the emitters, such devices cannot be modeled well by the standard tunneling approaches<sup>2,3</sup> or the approaches for including scattering; see, for example, Refs. 3–7 and citations therein.

The ability to model extended devices required a novel application of the theory developed by Caroli *et al.*<sup>8</sup> for including the effects of the contacts. We refer to our use of the theory as a generalized treatment of the open system boundaries which is described briefly in Refs. 1 and 9–11 and in full detail in Sec. (IV). Our approach allows one to treat large regions of the structure as emitter and collector reservoirs *even when there are spatially varying potentials in these regions*. Sec. (IV B) incorporates the theory into a standard coherent tunneling simulator<sup>2</sup> making such a simulator much more versatile. It provides a simple, tractable solution to the problems considered by Frensley<sup>3</sup> and Fiig and Jauho.<sup>12</sup> Numerical examples of the boundary conditions used in single band calculations are given in Refs. 1, 11, and

13–16. In Refs. 14 and 17, the boundary conditions are used in  $sp^3s^*$  calculations.

While our treatment of the contacts is useful when implemented as a coherent tunneling device simulator, it becomes crucial when we include scattering. The theory allows us to partition a long structure into two large reservoirs and a short device. The computational and memory intensive calculations required to include scattering are then performed only within the short device region. This application was the reason for developing the novel treatment of the boundaries. Numerical examples of this application are given in Refs. 13–16 and in Sec. (VII).

Screened polar optical phonon scattering, acoustic phonon scattering, alloy scattering, interface roughness scattering, and ionized dopant scattering are treated in Born-type approximations. The self-consistent Born (SCB) approximation and the single-electron approximation within the self-consistent Born approximation are described in Sec. (V). We show in Secs. (V B) and (V C) that for incoherent elastic scattering or for inelastic scattering in the one electron approximation, the self-consistent Born treatment leads to an infinite continued fraction expansion of the retarded Green function and a power series expansion of  $G^<$ . For reasons of numerical tractability, we then consider finite order treatments of scattering.

We describe an approach based on a truncation of the self-consistent Born expansions which conserves current. The truncation leads to a Green function version of a multiple sequential scattering (MSS) algorithm<sup>6</sup> in Sec. (VI). The first order truncation leads to expressions for the self-energy and current identical to ones written down by Roblin and Liou.<sup>6</sup> The MSS approximation is not without its own drawbacks; it violates detailed balance in equilibrium, but it consistently appears to give reasonable results for high-bias simulations. We then describe both MSS and self-consistent

<sup>a)</sup>Electronic mail: r-lake@ti.com

Born treatments of the elastic scattering mechanisms combined with first order treatments of the polar optical phonon scattering in Sec. (VI).

Self-energy expressions for screened polar optical phonon scattering, acoustic phonon scattering, alloy scattering, interface roughness scattering, and ionized dopant scattering are derived in Appendix A. We take a fresh look at interface roughness and model it as a layer of alloy in which the cation species of a given type cluster into islands. The model smoothly approaches the usual alloy model as the clustering becomes a homogeneous mixture of the two cation species. This does not appear to be the case for most models found in the literature, e.g., Ref. 7, except for the work of Ting *et al.*<sup>18</sup> Interface roughness self energies are calculated for both Gaussian and exponential autocorrelation models. Numerical examples comparing the different interface roughness models are given in Ref. 15. Numerical examples showing the effect of polar optical phonon scattering and interface roughness scattering on the valley current of a resonant tunneling diode (RTD) are given in Refs. 14–16.

## II. HAMILTONIAN AND BASIS

The general form of the Hamiltonian is shown in Eq. (1) where  $H_o$  contains the kinetic energy and the effects of the band structure, the applied potential, and the Hartree potential. The five terms to the right represent the potential felt by the electrons due to polar optical phonons, acoustic phonons, interface roughness, alloy disorder, and ionized dopants, respectively. The underbrace and  $\Sigma$  indicate that these terms will be included through self-energies shown in Appendix A.

$$H = H_o + \underbrace{H_{pop} + H_{ac} + H_{ir} + H_{al} + H_{id}}_{\Sigma} \quad (1)$$

$H_o$  is broken down into five terms,

$$H_o = \underbrace{H_o^D + H_o^L + H_o^R + H_o^{LD} + H_o^{RD}}_{\Sigma^B}, \quad (2)$$

which represent the Hamiltonian of the device, the left contact, the right contact, the coupling of the left contact to the device, and the coupling of the right contact to the device, respectively. The under brace indicates that the effects of the contacts on the device will also be taken into account through a self-energy,  $\Sigma^B$ . The self-energy treatment of the scattering Hamiltonians is perturbative. The self-energy treatment of the contacts is exact.

For zinc-blende compounds, a layer contains both a layer of cations and a layer of anions. An example is GaAs where the cations and anions are, respectively, Ga and As. A vector  $\mathbf{R}^L$ , pointing to a cation in layer L has a  $z$  component  $L\Delta$  and a transverse component  $\mathbf{R}_t^L$  where  $\Delta$  is the layer spacing (half the width of the conventional cubic cell). The anions in the anion plane are shifted in position from the cations in the neighboring cation plane by the lattice vector  $\mathbf{v} = \frac{A}{2}(111)$ . A localized cation orbital at position  $\mathbf{R}^L$  is

$|c, L, \mathbf{R}_t^L\rangle$ . An orbital of the corresponding anion at position  $\mathbf{R}^L + \mathbf{v}$  is  $|a, L, \mathbf{R}_t^L\rangle$ . The indices  $a$  and  $c$  run over the set of orbitals (and, in the presence of spin orbit coupling, over the spins). For example, in the  $sp^3$  model there are 1  $s$  and 3  $p$  orbitals neglecting spin. From the localized orbital basis, construct transverse  $\mathbf{k}$  states corresponding to Bloch sums over the transverse plane.

$$|c, L, \mathbf{k}\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}_t^L} e^{i\mathbf{k} \cdot \mathbf{R}_t^L} |c, L, \mathbf{R}_t^L\rangle, \quad (3)$$

$$|a, L, \mathbf{k}\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}_t^L} e^{i\mathbf{k} \cdot (\mathbf{R}_t^L + \mathbf{v})} |a, L, \mathbf{R}_t^L\rangle. \quad (4)$$

Note that  $\mathbf{k}$  is, and will remain throughout the rest of the article, a purely *transverse* wave vector. In our nomenclature, *longitudinal* represents the direction of current flow and *transverse* represents the direction perpendicular to current flow. The field operator in this basis is

$$\psi(\mathbf{r}) = \sum_{\mathbf{k}, L} \left[ \sum_c \langle \mathbf{r} | c, L, \mathbf{k} \rangle c_{c, L, \mathbf{k}} + \sum_a \langle \mathbf{r} | a, L, \mathbf{k} \rangle c_{a, L, \mathbf{k}} \right], \quad (5)$$

where  $c_{a, L, \mathbf{k}}$  is the destruction operator for an electron in state  $|a, L, \mathbf{k}\rangle$  and  $c_{c, L, \mathbf{k}}$  is the destruction operator for an electron in state  $|c, L, \mathbf{k}\rangle$ .

The matrix elements of  $H_o$  are, in general,

$$\langle \alpha, L, \mathbf{k} | H_o | \alpha', L', \mathbf{k}' \rangle = D_{\alpha, \alpha'; L}(\mathbf{k}) \delta_{L, L'} - t_{\alpha, L; \alpha', L'}(\mathbf{k}) \delta_{L', L \pm j \neq 0}. \quad (6)$$

The diagonal block contains the orbital energies, the electrostatic potential, and the anion-cation matrix elements.  $H_o$  of Eq. (2) is partitioned such that layers  $1, \dots, N$  span  $H_o^D$  and correspond to the device, layers  $-\infty, \dots, 0$  span  $H_o^L$ , and layers  $N+1, \dots, \infty$  span  $H_o^R$ . We will refer to *block matrix* notation in which the orbital indices are suppressed and the elements such as  $t_{L, L'}$  are  $m \times m$  matrices where  $m$  is the number of orbitals (anion plus cation). We will also refer to *full matrix* notation in which all indices are suppressed and matrices are the size of  $H_o^D$ .

There are two special cases that have been heavily studied, the tight-binding single-band<sup>4,6</sup> and  $sp^3s^*$  models.<sup>19,20</sup> In the  $sp^3s^*$  model with transport in the (100) direction, a cation orbital only has matrix elements between itself and other anion orbitals (and vice versa for an anion orbital).  $H_o$  is block tri-diagonal with the blocks being half the size of the layer basis, i.e.,  $5 \times 5$  instead of  $10 \times 10$ . The diagonal blocks are themselves diagonal in the absence of spin orbit coupling.<sup>21</sup>

In the single-band tight-binding model, there is only one orbital per layer (the anion and cation are lumped into a single orbital). The layer uniquely specifies the orbital so that all orbital indices can be dropped from Eqs. (3–6) and any term containing anion indices in Eqs. (3–5) is discarded. The single-band tight-binding matrix elements of  $H_o$  are related to the discretized effective mass Hamiltonian in the usual way.<sup>22</sup> The effective mass Hamiltonian is

$$H_o = \frac{-\hbar^2}{2} \frac{d}{dz} \frac{1}{m^*(z)} \frac{d}{dz} + V_k(z) + \frac{\hbar^2 k^2}{2m_L^*}, \quad (7)$$

where  $m_L^*$  is the effective mass in the left lead and

$$V_k(z) = V(z) + \frac{\hbar^2 k^2}{2m_L^*} \left( \frac{m_L^*}{m^*(z)} - 1 \right). \quad (8)$$

Note that the spatial dependence of the transverse energy has been transferred into a transverse momentum dependent potential, and that we have assumed a parabolic (instead of cosine) transverse dispersion.

The tight-binding parameters in Eq. (6) corresponding to the finite difference form of Eq. (7) are

$$t_{i,j} = \frac{\hbar^2}{(m_i + m_j)\Delta^2}, \quad (9)$$

$$D_i(k) = \frac{\hbar^2}{2\Delta^2} \left( \frac{1}{m^-} + \frac{1}{m^+} \right) + V_i(k), \quad (10)$$

$$m^- = \frac{m_{i-1} + m_i}{2} \quad \text{and} \quad m^+ = \frac{m_i + m_{i+1}}{2}, \quad (11)$$

where in Eqs. (9–11)  $m_i$  is the effective mass at mesh site “i”,  $V_i(k)$  is the potential at site “i” due to conduction band offsets, the applied potential, the Hartree potential and the transverse momentum dependent part of  $V_k(z)$ , and  $\Delta$  is the mesh spacing.

### III. NON-EQUILIBRIUM GREEN FUNCTIONS

#### A. Definitions

The non-equilibrium Green function formalism provides a method for calculating the non-equilibrium statistical ensemble average of the single particle correlation operators,

$$\begin{aligned} G_{\alpha,L;\alpha',L'}^<(\mathbf{k};t,t') &= \frac{i}{\hbar} \langle c_{\alpha',L'}^\dagger(t') c_{\alpha,L}(t) \rangle \\ G_{\alpha,L;\alpha',L'}^>(\mathbf{k};t,t') &= \frac{-i}{\hbar} \langle c_{\alpha,L}(t) c_{\alpha',L'}^\dagger(t') \rangle \\ G_{\alpha,L;\alpha',L'}^R(\mathbf{k};t,t') &= \Theta(t-t') [G_{\alpha,L;\alpha',L'}^>(\mathbf{k};t,t') \\ &\quad - G_{\alpha,L;\alpha',L'}^<(\mathbf{k};t,t')] \\ &= -i\Theta(t-t') A_{\alpha,L;\alpha',L'}(\mathbf{k};t,t') \\ G_{\alpha,L;\alpha',L'}^A(\mathbf{k};t,t') &= \Theta(t'-t) [G_{\alpha,L;\alpha',L'}^<(\mathbf{k};t,t') \\ &\quad - G_{\alpha,L;\alpha',L'}^>(\mathbf{k};t,t')] \\ A_{\alpha,L;\alpha',L'}(\mathbf{k};t,t') &= i[G_{\alpha,L;\alpha',L'}^>(\mathbf{k};t,t') \\ &\quad - G_{\alpha,L;\alpha',L'}^<(\mathbf{k};t,t')] \\ &= i[G_{\alpha,L;\alpha',L'}^R(\mathbf{k};t,t') \\ &\quad - G_{\alpha,L;\alpha',L'}^A(\mathbf{k};t,t')], \end{aligned} \quad (12)$$

where the indices  $\alpha$  and  $\alpha'$  include both the anion and cation orbitals.

We are concerned with steady state, so we Fourier transform the time difference coordinate,  $(t-t')$ , to energy in Eqs. (12) and work with, for example, the quantity  $G_{\alpha,L;\alpha',L'}^<(\mathbf{k};E) = \int d(t-t') e^{iE(t-t')/\hbar} G_{\alpha,L;\alpha',L'}^<(\mathbf{k};t-t')$ .

Several useful relationships that apply to the Fourier transformed quantities are (in full matrix notation)

$$G^A = [G^R]^\dagger, \quad (13)$$

$$A = i[G^R - G^A] = A^\dagger. \quad (14)$$

Fourier transforming the definition of  $G^R$  in Eq. (12) results in

$$G^R(\mathbf{k},E) = P \int \frac{dE'}{2\pi} \frac{A(\mathbf{k};E')}{E-E'} - \frac{i}{2} A(\mathbf{k};E). \quad (15)$$

The relationships between the various self-energies that we will discuss are identical to the relationships between the various Green functions. A few that we will use are

$$\begin{aligned} \Gamma(\mathbf{k};E) &= i[\Sigma^>(\mathbf{k};E) - \Sigma^<(\mathbf{k};E)] \\ &= i[\Sigma^R(\mathbf{k};E) - \Sigma^A(\mathbf{k};E)] \end{aligned} \quad (16)$$

and

$$\Sigma^R(\mathbf{k},E) = P \int \frac{dE'}{2\pi} \frac{\Gamma(\mathbf{k};E')}{E-E'} - \frac{i}{2} \Gamma(\mathbf{k};E). \quad (17)$$

#### B. Electron density and current

Once the correlation function  $G^<$  is known, it immediately provides the electron density,

$$\begin{aligned} n_L &= \frac{-2i}{A\Delta} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \sum_{\alpha} G_{\alpha,L;\alpha,L}^<(\mathbf{k},E) \\ &= \frac{-2i}{A\Delta} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \text{tr}\{G_{L,L}^<(\mathbf{k},E)\} \end{aligned} \quad (18)$$

and the current density,

$$\begin{aligned} J_L &= \frac{2e}{\hbar A} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \sum_{\alpha,\alpha'} \sum_{L_1 \leq L} \sum_{L_2 > L} [t_{\alpha,L_1;\alpha',L_2} \\ &\quad \times G_{\alpha',L_2;\alpha,L_1}^<(\mathbf{k},E) - t_{\alpha,L_2;\alpha',L_1} G_{\alpha',L_1;\alpha,L_2}^<(\mathbf{k},E)] \\ &= \frac{2e}{\hbar A} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \sum_{L_1 \leq L} \sum_{L_2 > L} 2\text{Re}\{\text{tr}[t_{L_1;L_2} G_{L_2;L_1}^<(\mathbf{k},E)]\} \end{aligned} \quad (19)$$

throughout the device.  $J_L$  is the current crossing the plane between layer  $L$  and  $L+1$ . In Eqs. (18) and (19),  $A$  is the cross sectional area, and  $e$  is the electron charge. The factors of 2 are for spin degeneracy. In the absence of spin degeneracy, the trace is also taken over spins. In the second line of Eqs. (18) and (19), we use block matrix notation for  $t$  and  $G^<$  and  $\text{tr}\{\dots\}$  indicates a trace over the orbital indices. In the nearest neighbor tight-binding models,

$$J_L = \frac{2e}{\hbar A} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} 2\text{Re}\{\text{tr}[t_{L;L+1} G_{L+1;L}^<(\mathbf{k},E)]\}. \quad (20)$$

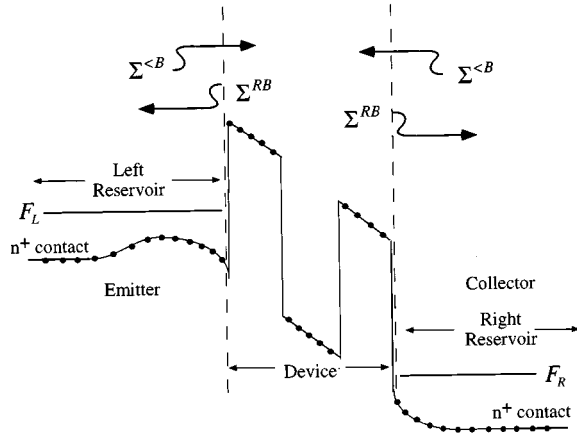


FIG. 1. The structure is partitioned into two large reservoirs and a short device. The self-energy  $\Sigma^{RB}$  accounts for loss from the device to the contacts. The self-energy  $\Sigma^{<}$  accounts for injection from the contacts to the device.

One should keep in mind that the goal of all of the different approaches that we discuss in the following sections is the calculation of Eqs. (18) and (19).

#### IV. SINGLE ELECTRON TUNNELING THEORY

##### A. Generalized open system boundary conditions

In this section, we describe how to partition the device such that large regions, even where there are spatially varying potentials, can be treated as reservoirs. We have found this development to be the single most important and useful feature of the theory presented in this article. The approach is based on the Dyson equation treatment of the contacts developed by Caroli *et al.*<sup>8</sup> and subsequently used by a number of authors.<sup>19,23–28</sup>

The novelty of our approach is twofold. The first is the manner in which we apply the theory. All previous applications of the theory injected carriers only from the flatband regions of the contacts. See, for example, appendix B and Fig. (4) of Ref. 19. In contrast, we use the theory to treat large regions of the structure, even where there are spatially varying potentials, as reservoirs. Thus, we are capable of injecting electrons into the device from mixed, non-asymptotic states in the emitter (see Figs. (1) and (2) of Ref. 1). This is in contrast to standard scattering theory, and it is the reason that the problem of injection from emitter quasi-bound states was never properly formulated without this theory.<sup>12</sup> Our approach is also the key which allows us to include scattering in long devices.

The second novelty is our small but essential modification of the original theory to include an imaginary potential in the contacts. This potential plays the role of the imaginary part of the retarded self-energy in the contacts resulting from scattering. For injection into the device from plane-wave states, the effect of the imaginary potential is small. For injection from emitter states that lie below the continuum on the left (see Fig. (1)), the imaginary potential is crucial. This is discussed and illustrated numerically in Ref. 11.

The theory is illustrated by Fig. (1). The effect of the left and right regions are collapsed into the self energies  $\Sigma^{RB}$  and  $\Sigma^{<B}$ .  $\Sigma^{RB}$  accounts for the spectrum shift and loss in the device resulting from the coupling of the device to the contacts.  $\Sigma^{<B}$  accounts for the inscattering from the contacts to the device.

For completeness, we derive the expressions for the boundary self-energies below. We first include exactly the effects of the contacts on the device using Dyson's equation, then, in the following sections, we include the effects of scattering in the device. We consider Dyson's equation for  $G^{<}$  since it will give us the self-energies which account for both the dynamics and the kinetics of the leads. To avoid excessive subscripts and summations, we derive the equations for the nearest neighbor tight-binding model, and, at the end, write down the answers for the general multi-neighbor coupling models. Unless stated otherwise, block matrix notation will be used throughout. All subscripts refer to layer indices. We calculate  $G^{<}$  in the device by including the coupling of the contacts to the device,  $t_{0,1}$ ,  $t_{1,0}$ ,  $t_{N,N+1}$ , and  $t_{N+1,N}$ , exactly using Dyson's equation. The derivation for  $G^{>}$  is identical and obtained by replacing the superscript  $<$  with  $>$  and the Fermi factors  $f_{e(c)}$  with  $1 - f_{e(c)}$  in the following equations. For any two layers  $i, j \in \{1, \dots, N\}$  in the device, the Dyson equation for  $G_{i,j}^{<}(\mathbf{k}, E)$  is (cf. Eq. (2.25) of Langreth<sup>29</sup>)

$$G_{i,j}^{<} = g_{i,j}^{<} + g_{i,1}^R(-t_{1,0})G_{0,j}^{<} + g_{i,1}^{<}(-t_{1,0})G_{0,j}^A + g_{i,N}^R(-t_{N,N+1})G_{N+1,j}^{<} + g_{i,N}^{<}(-t_{N,N+1})G_{N+1,j}^A, \quad (21)$$

where the arguments  $(\mathbf{k}, E)$  have been suppressed. In Eq. (21), the lower case  $g$ 's are calculated with the device-contact couplings,  $t_{0,1}$ ,  $t_{1,0}$ ,  $t_{N,N+1}$ , and  $t_{N+1,N}$  set to zero, and the upper case  $G$ 's are exact. The middle term of Langreth's Eq. (2.25) is zero since the  $t$ 's are local in time (see Appendix of Caroli *et al.*<sup>8</sup>). The full Green functions which cross the device-contact boundaries,  $G_{0,j}^{<}$ ,  $G_{N+1,j}^{<}$ ,  $G_{0,j}^A$ , and  $G_{N+1,j}^A$  are obtained by writing a second Dyson equation,

$$G_{0,j}^{<} = g_{0,0}^R(-t_{0,1})G_{1,j}^{<} + g_{0,0}^{<}(-t_{0,1})G_{1,j}^A$$

$$G_{N+1,j}^{<} = g_{N+1,N+1}^R(-t_{N+1,N})G_{N,j}^{<} + g_{N+1,N+1}^{<}(-t_{N+1,N})G_{N,j}^A \quad (22)$$

and

$$G_{0,j}^A = g_{0,0}^A(-t_{0,1})G_{1,j}^A$$

$$G_{N+1,j}^A = g_{N+1,N+1}^A(-t_{N+1,N})G_{N,j}^A. \quad (23)$$

Substituting Eqs. (22) and (23) into Eq. (21) results in

$$\begin{aligned}
G_{i,j}^< &= g_{i,j}^< + \underbrace{g_{i,1}^R t_{1,0} g_{0,0}^R t_{0,1}}_{\Sigma_{1,1}^{RB}} G_{1,j}^< + \underbrace{g_{i,1}^R t_{1,0} g_{0,0}^< t_{0,1}}_{\Sigma_{1,1}^{<B}} G_{1,j}^A \\
&+ \underbrace{g_{i,1}^< t_{1,0} g_{0,0}^A t_{0,1}}_{\Sigma_{1,1}^{AB}} G_{1,j}^A \\
&+ \underbrace{g_{i,N} t_{N,N+1} g_{N+1,N+1} t_{N+1,N}}_{\Sigma_{N,N}^B} G_{N,j} \quad \text{terms}
\end{aligned} \quad (24)$$

From Eq. (24), we read off the boundary self-energies which take into account the effect of the semi-infinite left and right contacts on the device.

$$\begin{aligned}
\Sigma_{1,1}^{RB} &= t_{1,0} g_{0,0}^R t_{0,1} \\
\Sigma_{N,N}^{RB} &= t_{N,N+1} g_{N+1,N+1}^R t_{N+1,N} \\
\Sigma_{1,1}^{AB} &= t_{1,0} g_{0,0}^A t_{0,1} \\
\Sigma_{N,N}^{AB} &= t_{N,N+1} g_{N+1,N+1}^A t_{N+1,N}
\end{aligned} \quad (25)$$

and

$$\begin{aligned}
\Sigma_{1,1}^{<B} &= t_{1,0} g_{0,0}^< t_{0,1} \\
\Sigma_{N,N}^{<B} &= t_{N,N+1} g_{N+1,N+1}^< t_{N+1,N}.
\end{aligned} \quad (26)$$

Since the contacts are by definition in equilibrium,

$$\begin{aligned}
g_{0,0}^< &= i f_e a_{0,0} \\
g_{N+1,N+1}^< &= i f_c a_{N+1,N+1},
\end{aligned} \quad (27)$$

where  $a = i(g^R - g^A)$  is the spectral function and  $f_{e(c)}$  is the Fermi factor of the emitter (collector) contact. Defining  $\Gamma^B$  as

$$\begin{aligned}
\Gamma_{1,1}^B &= i(\Sigma_{1,1}^{RB} - \Sigma_{1,1}^{AB}) = t_{1,0} a_{0,0} t_{0,1} \\
\Gamma_{N,N}^B &= i(\Sigma_{N,N}^{RB} - \Sigma_{N,N}^{AB}) = t_{N,N+1} a_{N+1,N+1} t_{N+1,N}
\end{aligned} \quad (28)$$

we obtain the final form for  $\Sigma^{<B}$ .

$$\begin{aligned}
\Sigma_{1,1}^{<B} &= i f_e \Gamma_{1,1}^B \\
\Sigma_{N,N}^{<B} &= i f_c \Gamma_{N+1,N+1}^B.
\end{aligned} \quad (29)$$

In the nearest neighbor tight-binding model, all of the boundary self-energies are zero for layers  $\{i,j\} \neq \{1,1\}$  or  $\{N,N\}$ . While the self-energies  $\Sigma^{RB}$  and  $\Gamma^B$  are valid in general, the self-energy  $\Sigma^{<B}$ , Eq. (29), is valid only if the reservoir regions are well equilibrated with the  $n^+$  contacts (see Fig. (1)).

To obtain the equation of motion for  $G_{i,j}^<$  for  $i,j \in \{1, \dots, N\}$ , we operate on Eq. (24) from the left with  $(E - H_o^D)$  using  $(E - H_o^D)g^R = 1$  and  $(E - H_o^D)g^< = 0$  to obtain (in full matrix notation)

$$(E - H_o^D - \Sigma^{RB})G^< = \Sigma^{<B}G^A, \quad (30)$$

where  $(E - H_o^D)$  is an  $N \times N$  block tridiagonal matrix and  $\Sigma^{RB}$  and  $\Sigma^{<B}$  are  $N \times N$  block matrices with two non-zero blocks, one in the upper left corner and one in the lower right corner.

The matrix on the left of  $G^<$  is  $G^{R-1}$ , so from Eq. (30) we also get the equation of motion for the retarded Green function in the device (in full matrix notation),

$$(E - H_o^D - \Sigma^{RB})G^R = 1 \quad (31)$$

Derivations of  $\Sigma^{RB}$  using the more commonly known Dyson equation for  $G^R$  can be found in Refs. 10 and 30.

An explicit representation of the retarded Green function for any layer  $\{i,j\} \in \{1, \dots, N\}$  is thus (for  $N=3$ ),

$$[G^R] = \begin{bmatrix} E - D_1 - \Sigma_{1,1}^{RB} & t_{1,2} & 0 \\ t_{1,2} & E - D_2 & t_{2,3} \\ 0 & t_{3,2} & E - D_3 - \Sigma_{3,3}^{RB} \end{bmatrix}^{-1}. \quad (32)$$

In general, the contact regions will be regions with spatially varying potentials where it is appropriate to approximate the occupation factor with an equilibrium Fermi-Dirac factor. An example is the lightly doped emitter lead and resulting emitter quasi-bound state of a resonant tunneling diode.<sup>31</sup> This region can be 200 nm long. If a lattice spacing of 2.83 Å is used, 700 nodes would be required for this region. By including its effect through a self-energy, the size,  $N$ , of the matrix,  $H_o^D$ , is reduced by 700. Once we include scattering, computation time scales as  $N^2$  for local interactions and  $N^3$  for non-local interactions, this reduction therefore becomes essential.

To calculate the self-energies  $\Sigma_{1,1}^{RB}$  and  $\Sigma_{3,3}^{RB}$  in Eq. (32), we need  $g_{0,0}^R$  and  $g_{4,4}^R$ . Taking, for example, a left contact consisting of three layers with indices  $j = \{-2, -1, 0\}$  and assuming that the potential at layers  $\{-3, \dots, -\infty\}$  is identical to that at layer  $-2$ ,  $g_{0,0}^R$  is found from

$$g_{0,0}^R = \begin{bmatrix} E - D_{-2} - \Sigma_{-2,-2}^{RB} & t_{-2,-1} & 0 \\ t_{-1,-2} & E - D_{-1} & t_{-1,0} \\ 0 & t_{0,-1} & E - D_0 \end{bmatrix}_{0,0}^{-1}, \quad (33)$$

where

$$\Sigma_{-2,-2}^{RB} = t_{-2,-3} g_{-3,-3}^R t_{-3,-2} \quad (34)$$

and  $g_{-3,-3}^R$  is the surface Green function of the semi-infinite bulk terminated at layer  $-3$ , i.e., the semi-infinite bulk consisting of layers  $\{-\infty \dots -3\}$ . For nearest neighbor tight-binding models the surface Green function is<sup>30</sup>

$$g_{-3,-3}^R = [E - D_{-3} + t_{-4,-3} \chi Z \chi^{-1}]^{-1} \quad (35)$$

and the boundary self energy is

$$\Sigma_{-2,-2}^{RB} = -t_{-3,-2} \chi Z \chi^{-1}, \quad (36)$$

where  $Z$  is the diagonal matrix of propagation factors and  $\chi$  is the matrix of Bloch states propagating toward the device.

For the single-band model, the  $t$ 's are real scalars which we denote as  $t_e$  in the left lead so that

$$g_{-3,-3}^R = -e^{i\gamma_e \Delta} / t_e \quad (37)$$

and

$$\Sigma_{-2,-2}^{RB} = -t_e e^{i\gamma_e \Delta}, \quad (38)$$

where  $\gamma_e$  is the longitudinal wave vector in the emitter contact. Using the single-band dispersion relation

$$E = D_{-2} - 2t_e \cos(\gamma_e \Delta) \quad (39)$$

the upper left corner element of the matrix in Eq. (33) becomes  $-t_e e^{-i\gamma_e \Delta}$ .

True bound states may be present in the spectrum of  $g_{0,0}^R$  and  $g_{N+1,N+1}^R$  occurring in Eqs. (25–29) since the matrix elements  $t_{0,1}$  and  $t_{N,N+1}$  are set to zero (see Eq. (33)). To give a realistic energy width to these states, an energy dependent optical potential,  $i\eta$ , is added to the diagonal elements of  $D_i$  in the leads in matrix (33). The optical potential is given an exponential decay for energies below the conduction band edge to avoid unrealistic band tails. Physically, the optical potential represents the scattering induced broadening; if an emitter quasi-bound state is to act as a reservoir, it must be coupled to the continuum of states to the left through inelastic channels, and the coupling to the continuum must be greater than the coupling to the states to the right through the emitter barrier. The optical potential is an approximation for the imaginary part of the retarded self-energy resulting from scattering in the reservoirs. The optical potential is non-zero only in the reservoirs and zero in the device. Therefore, in the device, where current is calculated, current is conserved.

In general, for arbitrary neighbor coupling,

$$\Sigma_{e,L,L'}^{RB} = \sum_{L_1 \leq 0} \sum_{L_2 \leq 0} t_{L,L_1} g_{L_1,L_2}^R t_{L_2,L'}, \quad (40)$$

$$\Sigma_{c,L,L'}^{RB} = \sum_{L_1 \geq N+1} \sum_{L_2 \geq N+1} t_{L,L_1} g_{L_1,L_2}^R t_{L_2,L'}, \quad (41)$$

where we have used block matrix notation. The subscripts  $e$  and  $c$  label the self-energies resulting from coupling to the emitter and collector contacts, respectively. The number of terms in the sums is determined by the number of non-zero off-diagonal block matrix elements in  $H_o$ . The term  $g_{L_1,L_2}^R$  is calculated with all matrix elements coupling the lead to the device set to zero.  $\Gamma^B$  resulting from coupling to the (left/right) contact is defined as

$$\Gamma^{B(e/c)} = i[\Sigma_{(e/c)}^{RB} - [\Sigma_{(e/c)}^{RB}]^\dagger]. \quad (42)$$

$\Sigma^{<B}$  resulting from coupling to the (emitter/collector) contact is given by

$$\Sigma_{(e/c)}^{<B} = i f_{(e/c)} \Gamma^{B(e/c)}. \quad (43)$$

The central equations of this section are the expressions for the self-energies which account for both the dynamics, Eqs. (25), (28), (40), (41), and (42), and the kinetics, Eqs. (29) and (43), of the contacts and the equations of motion for  $G^{<}$ , Eq. (30), and  $G^R$ , Eq. (31).

## B. Current

### 1. Tunneling formula

Starting with the expression for  $J_0$ , Eq. (19), using Dyson's equation for  $G_{e,d}^{<}$ , which crosses the device emitter-contact boundary where the subscript  $e$  indicates a layer in the emitter contact (to the left) and the subscript  $d$  indicates a layer in the device (in block matrix notation),

$$G_{e,d}^{<} = \sum_{e',d'} [g_{e,e'}^R (-t_{e',d'}) G_{d',d}^{<} + g_{e,e'}^{<} (-t_{e',d'}) G_{d',d}^A] \quad (44)$$

and the fact that  $G_{d,e}^{<} = -[G_{e,d}^{<}]^\dagger$ , we can rewrite the current Eq. (19) to obtain Eq. (5) of Meir and Wingreen,<sup>27</sup>

$$J_0 = \frac{2e}{\hbar A} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \sum_{L=1}^N \sum_{L'=1}^N \text{tr} \{ \Gamma_{L,L'}^{Be} [f_e A_{L',L} + i G_{L',L}^{<}] \}. \quad (45)$$

As they noted, Eq. (45) is valid even when scattering processes and many-body interactions are present in the device. With no scattering in the device, Eq. (45) can be re-written as a generalized Fisher-Lee tunneling formula.<sup>32</sup> Using the relations (in full matrix notation),

$$G^{<} = G^R \Sigma^{<B} G^A \quad (46)$$

and

$$A = G^R \Gamma^B G^A, \quad (47)$$

Eq. (45) becomes

$$J = \frac{2e}{\hbar A} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \sum_{L_1=1}^N \sum_{L_2=1}^N \sum_{L_3=1}^N \sum_{L_4=1}^N \times \text{tr} \{ \Gamma_{L_1,L_2}^{Be} G_{L_2,L_3}^R \Gamma_{L_3,L_4}^{BR} G_{L_4,L_1}^A \} (f_e - f_c). \quad (48)$$

For the single-band model, Eq. (48) becomes

$$J = \frac{2e}{\hbar A} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \Gamma_{1,1}^{Be} \Gamma_{N,N}^{Bc} |G_{1,N}^R|^2 (f_e - f_c). \quad (49)$$

Defining  $\text{Tr}\{\dots\}$  as being the trace over all device states, Eq. (48) can be written as

$$J = \frac{2e}{\hbar A} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \text{Tr} \{ \Gamma^{Be} G^R \Gamma^{Bc} G^A \} (f_e - f_c), \quad (50)$$

where we are using full matrix notation exactly as written by Meir and Wingreen.<sup>27</sup> The difference is that our  $\Gamma$ 's contain the effects of band bending and quantized states in the leads. Eq. (50) is the tunneling formula with a generalized Fisher-Lee form of the transmission coefficient.<sup>32</sup> For the first time, current flowing from continuum states and current flowing from emitter quasi-bound states is treated on an equal footing in a Tsu-Esaki-type tunneling<sup>33</sup> formula.<sup>12</sup>

### 2. A numerically more efficient tunneling formula

Although Eq. (50) is formally pleasing, it is numerically inefficient since it requires the calculation of the far off-diagonal elements of  $G^R$ . The most efficient formulation would require only the corner diagonal elements of  $G^R$ . We can rewrite Eq. (50) to satisfy this requirement. Using the expression for  $A$  (in full matrix notation)

$$A = G^R \Gamma^{Be} G^A + G^R \Gamma^{Bc} G^A \quad (51)$$

the transmission probability in Eq. (50) can be written as

$$T(\mathbf{k}, E) = \text{Tr}[\Gamma^{Be} [A - G^R \Gamma^{Be} G^A]]. \quad (52)$$

Since, in Eq. (52) the Green functions only couple to  $\Gamma^{BL}$ , only the left corner elements are needed. As an example, for nearest neighbor tight binding (in block matrix notation),

$$J_L = \frac{2e}{\hbar A} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \text{tr}[\Gamma_{1,1}^{Be} [A_{1,1} - G_{1,1}^R \Gamma_{1,1}^{Be} G_{1,1}^A]] (f_e - f_c) \quad (53)$$

only the first diagonal block is needed ( $A_{1,1}$  is obtained from Eq. (14)). For a typical recursive Green function algorithm, Eq. (53) can be twice as fast as Eq. (48).

### C. Electron density

The approach just described to speed up the current calculation can also be used to speed up the calculation of the electron density when no scattering is present in the device. To calculate the electron density from Eq. (18), we need

$$-iG_{L,L}^< = \sum_{L_1, L_2} [f_e G_{L,L_1}^R \Gamma_{L_1, L_2}^{Be} G_{L_2, L}^A + f_c G_{L,L_1}^R \Gamma_{L_1, L_2}^{Bc} G_{L_2, L}^A]. \quad (54)$$

For the simplest case of nearest neighbor tight binding, Eq. (54) requires both the left and right block columns of  $G^R$ . We can rewrite Eq. (54) so that only the block diagonal elements and left or right block columns are needed. In the recursive Green function algorithm, this change can reduce the calculations by a factor of 2.

To acquire the left block columns, we define

$$A_{L,L}^{\mathcal{L}} = \sum_{L_1, L_2} G_{L,L_1}^R \Gamma_{L_1, L_2}^{Be} G_{L_2, L}^A. \quad (55)$$

Eq. (54) then becomes

$$-iG_{L,L}^< = f_e A_{L,L}^{\mathcal{L}} + f_c [A_{L,L} - A_{L,L}^{\mathcal{L}}]. \quad (56)$$

Eq. (56) requires only the block diagonal elements of  $G^R$  and the elements which couple to the left contact through  $\Gamma^{Be}$ . As an example, for nearest neighbor tight binding,

$$A_{L,L}^{\mathcal{L}} = G_{L,1}^R \Gamma_{1,1}^{Be} G_{1,L}^A. \quad (57)$$

In this example, only the diagonal blocks and the left block column of  $G^R$  are needed.

Once the device is under sufficiently high bias such that at higher energies  $f_R$  can be neglected, one only needs the left column of  $G^R$ . This can be obtained from one sweep from right to left across the device with the recursive Green function algorithm, Eq. (59), followed by a walk down the left column.

Since the contacts are in equilibrium, we calculate  $G_{L,L}^<$  in the contacts using the equilibrium relation

$$-iG_{L,L}^< = f A_{L,L}, \quad (58)$$

where  $f$  is the Fermi factor of the contact and  $G^R$  is the exact Green function of the *connected* lead. Therefore, in the contacts, we need to calculate the diagonal elements of  $G^R$ . The

recursive Green function algorithm described in the next subsection provides an extremely efficient means for obtaining all of the information necessary for calculating the electron density and the current.

### D. Recursive Green function algorithm

The recursive Green function algorithm is most powerful for the nearest neighbor tight-binding model. We will limit our discussion to that example in this section. For next nearest-neighbor models, one can use all of the equations as stated by simply doubling the block size.

In the following derivation, all upper case  $G$ 's and lower case  $g$ 's are retarded Green functions and hence we will drop the superscript  $R$ . Upper case  $G$ 's are reserved for the exact Green function elements. The term  $g_{L,L}^{\mathcal{L}}$  indicates that the Green function takes into account everything exactly on the left with the coupling elements immediately on the right,  $t_{L,L+1}$  and  $t_{L+1,L}$ , set to zero. The term  $g_{L,L}^r$  indicates that the Green function takes into account everything on the right exactly with the coupling elements immediately to the left,  $t_{L,L-1}$  and  $t_{L-1,L}$ , set to zero.

To calculate the electron density, we need the diagonal and left block column of  $G^R$  for Eqs. (56) and (57). We start with the expression for  $g^r$  in the flatband lead, see Eq. (35), and, using the expression for the open system boundary self-energy, Eq. (25), walk from right to left across the device creating  $g_{L,L}^r$ :

$$g_{L,L}^r = [E - D_L - t_{L,L+1} g_{L+1,L+1}^r t_{L+1,L}]^{-1}. \quad (59)$$

At the left end of the left lead (say layer 0), we have  $g_{0,0}^{\mathcal{L}}$  from Eq. (35) and we create the exact diagonal block  $G_{1,1}^R$  from

$$G_{1,1} = [E - D_1 - t_{1,0} g_{0,0}^{\mathcal{L}} t_{0,1} - t_{1,2} g_{2,2}^r t_{2,1}]^{-1}. \quad (60)$$

If we only needed the current, Eq. (53), we would stop here. With  $G_{1,1}$ , we march back across the device creating  $G_{L,L}$  using

$$G_{L,L} = g_{L,L}^r + g_{L,L}^r t_{L,L-1} G_{L-1,L-1} t_{L-1,L} g_{L,L}^r, \quad (61)$$

which is derived by combining the Dyson equation for  $G_{L,L}$ ,

$$G_{L,L} = g_{L,L}^r + g_{L,L}^r (-t_{L,L-1}) G_{L-1,L}, \quad (62)$$

with the alternate form of the Dyson equation for  $G_{L-1,L}$ ,

$$G_{L-1,L} = G_{L-1,L-1} (-t_{L-1,L}) g_{L,L}^r. \quad (63)$$

Finally, we walk down the left column creating  $G_{L,1}$  using

$$G_{L,1} = g_{L,L}^r (-t_{L,L-1}) G_{L-1,1}. \quad (64)$$

### V. SELF-CONSISTENT BORN EQUATIONS OF MOTION AND SOLUTION

In the previous section, we used Dyson's equation to include exactly the effect of the contact Hamiltonians in Eq. (2). We now use perturbation theory to include approximately the effects of the scattering potentials in Eq. (1). The derivation of the self-energies in a self-consistent Born approximation is given in Sec. (A).



The equations of motion for  $G^R$  and  $G^<$  which include the effects of incoherent scattering are (in full matrix notation)

$$(E - H_o - \Sigma^R - \Sigma^{RB})G^R = 1 \quad (65)$$

and

$$(E - H_o - \Sigma^R - \Sigma^{RB})G^< = (\Sigma^< + \Sigma^{<B})G^A, \quad (66)$$

where  $\Sigma^R = \Sigma_{\text{pop}}^R + \Sigma_{\text{ap}}^R + \Sigma_{\text{ir}}^R + \Sigma_{\text{al}}^R + \Sigma_{\text{id}}^R$  and  $\Sigma^< = \Sigma_{\text{pop}}^< + \Sigma_{\text{ir}}^< + \Sigma_{\text{al}}^< + \Sigma_{\text{id}}^<$  are the sum of the individual scattering self-energies for each type of scattering process: polar optical phonon, acoustic phonon, alloy disorder, interface roughness, and ionized dopant.  $\Sigma^{RB}$  and  $\Sigma^{<B}$  are the open system boundary self-energies. If we multiply Eq. (66) by  $G^R$  on the left, we obtain Eq. (17) of Caroli *et al.*<sup>34</sup>

We note that in the self-consistent Born approximation and, also, for the single-electron approximation within the self-consistent Born approximation that we discuss in Sec. (V C), current is conserved. For  $G^<$  calculated from Eq. (66), the divergence of the current is

$$J_L - J_{L-1} = \frac{2e}{\hbar A} \sum_k \int \frac{dE}{2\pi} \text{tr}[\Sigma^R G^< + \Sigma^< G^A - G^R \Sigma^< - G^< \Sigma^A] = 0, \quad (67)$$

where the notation  $\Sigma G \Rightarrow \Sigma_{L'} \Sigma_{L,L'} G_{L',L}$ .

### A. Retarded self-energy, $\Sigma^R$

The retarded self-energy,  $\Sigma^R$ , is related to  $\Sigma^<$  and  $\Sigma^>$  by Eq. (17). For the elastic self-energies due to acoustic phonons in the high temperature approximation, alloy scattering, interface roughness, and ionized dopants, the retarded self-energies are obtained from the expressions for  $\Sigma^<$  in Eqs. (A13), (A23), (A29), and (A40) in Appendix A, by simply replacing  $G^<$  with  $G^R$ . We illustrate this with the example of acoustic phonon scattering (in full matrix notation).

$$\begin{aligned} \Gamma(E) &= i(\Sigma^>(E) - \Sigma^<(E)) \\ &= D_{\text{ap}} \int \frac{d^2 k}{4\pi^2} i(G^>(k, E) - G^<(k, E)) \\ &= D_{\text{ap}} \int \frac{d^2 k}{4\pi^2} A(k, E). \end{aligned} \quad (68)$$

In Eq. (68),  $D_{\text{ap}}$  contains all of the prefactor terms in Eq. (A13). Substituting the expression for  $\Gamma(E)$  into Eq. (17) and using Eq. (15), we obtain

$$\Sigma^R(E) = D_{\text{ap}} \int \frac{d^2 k}{4\pi^2} G^R(k, E). \quad (69)$$

We note that since the total scattering rate  $\Gamma$  depends solely on the total number of states,  $A$ , and not their occupation, Pauli exclusion plays no role in incoherent elastic scattering.

Therefore, for incoherent elastic processes, single electron theory<sup>35</sup> and non-equilibrium Green function theory are identical.

For the inelastic self-energy resulting from polar optical phonons, there is no simple relationship such as Eq. (69) by which one can calculate  $\Sigma^R$  from  $G^R$ .  $\Sigma^R$  and  $\Gamma$  do depend on the occupation statistics which are contained in  $G^<$  and  $G^>$ . To calculate  $\Sigma^R$  for polar optical phonons, one is forced to perform the principal value integral in Eq. (17) numerically. That is a daunting task that we wish to avoid. In Sec. (V C), we discuss how to approximate the principal value integral analytically.

### B. Elastic scattering

First, we discuss the solution of Eqs. (65) and (66) if only elastic scattering mechanisms are present so that  $\Sigma^R = \Sigma_{\text{ap}}^R + \Sigma_{\text{ir}}^R + \Sigma_{\text{al}}^R + \Sigma_{\text{id}}^R$  and  $\Sigma^< = \Sigma_{\text{ap}}^< + \Sigma_{\text{ir}}^< + \Sigma_{\text{al}}^< + \Sigma_{\text{id}}^<$ . For only elastic self-energies, the two Eqs. (65) and (66) decouple. First, we perform a self-consistent solution for  $G^R$  and  $\Sigma^R$ . Then we self-consistently solve for  $G^<$  and  $\Sigma^<$ . Furthermore, and perhaps more importantly, each total energy decouples from the others, so that we perform the complete self-consistent calculations one energy at a time and then move on to the next energy. The energy decoupling reduces memory requirements.

The self-consistent solution  $G^R$  and  $\Sigma^R$  generates a continued fraction expansion which converges very fast. The self-consistent solution of  $G^<$  and  $\Sigma^<$  generates a power series expansion which converges very slowly. All of the elastic self-energies, Eqs. (A13), (A23), (A29), and (A40) have the form

$$\Sigma^R = D \otimes G^R \quad (70)$$

and

$$\Sigma^< = D \otimes G^<, \quad (71)$$

where  $D$  contains the matrix element squared of the scattering potential and  $\otimes$  indicates convolution over transverse momentum and matrix manipulation of the orbital indices (see, for example, Eq. (A23)). Substituting Eq. (70) into Eq. (65) and expanding, we get the continued fraction expansion.

$$\begin{aligned} G^R &= (E - H_o - \Sigma^{RB} - D \otimes (E - H_o - \Sigma^{RB} - D \\ &\quad \otimes (E - H_o - \Sigma^{RB} - D \otimes \dots)^{-1})^{-1})^{-1}. \end{aligned} \quad (72)$$

To demonstrate the power series expansion of Eq. (66) we first multiply through on the left by  $G^R$  to rewrite it as

$$G^< = G^R \Sigma^< G^A + G^R \Sigma^{<B} G^A. \quad (73)$$

The second term on the right is a source term due to injection from the contacts. We denote it as  $S = G^R \Sigma^{<B} G^A$ . To make the result most transparent, we write the following equations for  $G^<$  using the single-band model for diagonal self-energies. Then, the block matrices are scalars, the order is not important, and the factors of  $G^R$  and  $G^A$  can be com-

binned into  $|G^R|^2$ . We denote the diagonal elements of  $G^<$  and  $S$  with a single subscript as  $G_L^<$  and  $S_L$ . With this notation, Eq. (73) becomes

$$G_L^<(k) = \sum_{L'} |G_{L,L'}^R(k)|^2 \sum_{k'} D_{L',L}(k,k') G_{L'}^<(k') + S_L(k), \quad (74)$$

where we have explicitly written out the convolution over transverse momentum in Eq. (71). Suppressing the layer indices and iterating, Eq. (74) becomes

$$G^< = S + |G^R|^2 D \otimes S + |G^R|^2 D \otimes [|G^R|^2 D \otimes S] + \dots \quad (75)$$

Note that the form of Eq. (75) has a physical interpretation. The electron density at layer  $L$  is composed of multiple contributions. The first term,  $S$ , represents flux injected directly from the contacts. The next term,  $|G^R|^2 D \otimes S$ , represents the contribution from flux injected from the contacts that scattered at various layers  $L'$  and then propagated from layer  $L'$  to layer  $L$  according to  $|G_{L,L'}^R|^2$ . The third term represents the contribution from flux that scattered twice in getting from the contact to layer  $L$  and so on. We can formally write down the solution to Eqs. (74) and (75) as

$$G^< = [1 - |G^R|^2 D]^{-1} S. \quad (76)$$

For  $N_L$  layers and  $N_k$  transverse momenta,  $G^<$  and  $S$  are vectors of length  $N_L N_k$  and  $[1 - |G^R|^2 D]^{-1}$  is a  $(N_L N_k) \times (N_L N_k)$  full matrix. Typical numbers that we use in single-band simulations are  $N_L = 50$  and  $N_k = 200$  resulting in a full matrix with  $10^8$  elements which is too large to allow a direct solution. The method that we have used to solve Eq. (74) is Jacobi iteration combined with successive over relaxation (SOR).

### C. Inelastic (polar optical phonon) scattering

With polar optical phonon scattering, the self-consistent solutions of Eqs. (65) and (66) combined with the self-energy expressions (A9), (A10), (16), and (17) are fully coupled. Worse yet, all energies are coupled through the principal value integral in the calculation of  $\Sigma^R$ , Eq. (17). Furthermore, the self-energies are non-local, full  $N_L \times N_L$  matrices in a single band model. Thus, a self-consistent solution requires the storage of four double-complex, four-dimensional functions,  $G_{L,L'}^R(k,E)$ ,  $G_{L,L'}^<(k,E)$ ,  $\Sigma_{L,L'}^R(k,E)$ , and  $\Sigma_{L,L'}^<(k,E)$ . For our typical energy grid size of  $N_E = 150$  in a single band calculation with  $N_L$  and  $N_k$  as above, 4.2 GB of memory are required. For this reason, we have developed a number of different approximations to treat the polar optical phonon self energy. The remainder of the article discusses various approximate methods for calculating the self-energies.

First, we use the single-electron or low-density approximation to calculate the polar optical phonon retarded self-energy,  $\Sigma^R$ . In the calculation of  $\Gamma$  in Eq. (16),  $\Sigma^<$  is set to zero and  $\Sigma^>$  is calculated from Eq. (A7) replacing  $G^>$  with  $-iA$ . This approximation is valid for low electron densities, when the electron density per unit energy is much less than the density of states. This condition is valid for low doping, high temperatures, or for devices under bias when charging is negligible. The last two conditions apply to symmetric, room temperature RTDs.  $\Gamma$  is now of the form

$$\Gamma = D_{\text{pop}}^> \otimes A, \quad (77)$$

where  $D_{\text{pop}}^> \propto n_B \delta(E + \omega) + (n_B + 1) \delta(E - \omega)$  contains the matrix element squared of  $H_{\text{ep}}$ ,  $n_B$  is the Bose-Einstein factor,  $\omega$  is the polar optical phonon energy, and  $\otimes$  now represents convolution over both transverse momentum and total energy. Using this  $\Gamma$  in Eq. (17) to calculate  $\Sigma^R$  results in

$$\Sigma^R = D_{\text{pop}}^> \otimes G^R. \quad (78)$$

In other words, the retarded self-energy is now obtained from Eq. (A7) replacing  $G^>$  with  $G^R$ .

In this approximation, the non-equilibrium Green function theory no longer accounts for Pauli blocking and it becomes equivalent to a single electron theory.<sup>36,35</sup> There are two large benefits. The first is that the principal value integral in Eq. (17) is performed analytically. The second is that the coupling between the equations for  $G^R$  and  $G^<$  is removed allowing them to be solved independently as discussed for the case of elastic scattering in Sec. (V B). Since only  $G^R$  is needed to obtain  $\Sigma^R$ , the self-consistent equations for  $G^R$  and  $\Sigma^R$  form a closed loop independent of the equation for  $G^<$  and  $\Sigma^<$ .

The method of solution is similar to that described for the case of elastic scattering in Sec. (V B) except that  $N_\omega = (E_{\text{max}} - E_{\text{min}})/\omega$  total energies are coupled where  $E_{\text{max}}$  and  $E_{\text{min}}$  are the maximum and minimum energies of the total energy grid (we are assuming dispersionless polar optical phonons). The solution of  $G^R$  and  $\Sigma^R$  can still be cast as a continued fraction expansion, Eq. (72), except now  $\otimes$  represents convolution over both transverse momentum and total energy and  $D \propto n_B \delta(E + \omega) + (n_B + 1) \delta(E - \omega)$ . In the self-consistent solution of  $G^R$  and  $\Sigma^R$ , we need to store two double-complex four-dimensional functions  $G_{L,L'}^R(k,E)$  and  $\Sigma_{L,L'}^R(k,E)$ . For the single band model, the functions are of size  $N_L^2 \times N_k \times N_\omega$  which, for our typical numbers of  $N_L = 50$ ,  $N_k = 200$ , and  $N_\omega = 13$ , results in a memory requirement of 104 MB.

Once a solution of  $G^R$  at the  $N_\omega$  energies is obtained,  $G_{L,L'}^R(k,E)$  is stored and used in the self-consistent solution of  $G^<$  and  $\Sigma^<$  at that same set of  $N_\omega$  energies. The equation for  $G^<$  has the same form as before (cf. Eq. (74)),

$$G_{L,L'}^<(k,E) = S_{L,L'}(k,E) + \sum_{L_1,L_2} G_{L,L_1}^R(k) \sum_{k'} D_{L_1,L_2}^<(k,k') [n_B G_{L_1,L_2}^<(k',E - \omega) + (n_B + 1) G_{L_1,L_2}^<(k',E + \omega)] G_{L_2,L'}^A(k,E), \quad (79)$$

except that now the self-energy is non-local and couples  $G_{L,L'}^<(k,E)$  to  $G_{L,L'}^<(k,E \pm \omega)$ . After a self-consistent solution for  $G^<$  is obtained for one set of  $N_\omega$  total energies, a new set is chosen and the calculations are performed again. This is repeated until the value for the current, Eq. (19), converges.

The self-consistent solution of  $G^R$  and  $G^<$  for multiple coupled energies is still computationally intensive. Also, a self-consistent Born approximation requires a converged solution of the  $\Sigma$ 's and the  $G$ 's to conserve current. This can be numerically problematic. For these reasons, we have sought out finite order, non-self-consistent treatments of scattering.

We tried a first Born treatment.<sup>9</sup> However, we discovered that the approach is only valid for calculating the valley current of an RTD. The problem arises because the first Born treatment approximates  $(1 - g^R \sigma^R)^{-1}$  as  $1 + g^R \sigma^R$  which is only valid for  $g^R \sigma^R \ll 1$ . We found, numerically, that for any typical RTD, a first Born treatment of scattering resulted in the spectral function becoming negative near resonance (an explanation is given in Ref. 10) with a corresponding negative current density in that narrow energy range. References 37 and 38 are examples of the use of first Born approximations. For a general purpose device calculation, it is essential to use a finite order treatment of scattering which keeps  $\sigma^R$  in the denominator of  $G^R$ .

## VI. MULTIPLE SEQUENTIAL SCATTERING

The only finite order approach that we have found which can be used for both the calculation of the resonant and off resonant current is based on an algorithm which allows one to truncate the continued fraction expansion of  $G^R$ , Eq. (72), at any order and still conserve current. The corresponding equation for  $G^<$ , Eq. (75) or (79), becomes a truncated power series expansion. The equations resulting from a first order truncation can be shown to be identical to equations written down by Roblin and Liou.<sup>6</sup> The approach is a single electron theory which ignores the Pauli exclusion principle in the calculation of the polar optical phonon self-energy. From an applied point of view the approach has strong positive aspects: it allows fast first order calculations, it conserves current, and it appears to give reasonable physical results for devices under high bias. From a theoretical point of view the approach is unsatisfying and incorrect; it violates detailed balance in equilibrium giving rise to non-zero equilibrium current flow for devices that are not symmetric. Nevertheless, it is the only finite order approach that we have found that consistently gives reasonable results for a variety of high bias devices.

The greatest difficulty in deriving a non-self-consistent approximation for a self-energy is ensuring current conservation. Diagrammatic perturbation theory offers rules that can be mechanically followed which ensure current conservation. However, diagrammatic perturbation theory is based on a power series expansion of the Green function. Since a partial power series expansion is not useful to us for modeling RTDs, we are forced to work with a partial continued fraction expansion, and there are no rules, of which we are aware, that can be followed which ensure current conserva-

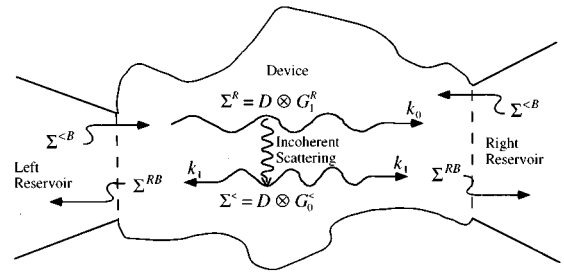


FIG. 2. Physical interpretation of the self-energies for single sequential scattering.  $\Sigma^R$  accounts for the out scattering and energy shifts and  $\Sigma^<$  accounts for the in scattering.

tion. Thus, we are forced to derive approximations for the self-energies using physical, intuitive arguments, and then check afterwards that the approximations do conserve current.

The general physical picture which guides us is the multiple sequential scattering picture which informs the work of Roblin and Liou.<sup>6</sup> The particular picture which leads to our final equations is the following. A plane wave,  $\psi_0$ , injected from the contact at energy  $E$  propagates into the device and scatters either elastically due to the random potential of interface roughness, alloys, ionized dopants, or acoustic phonons, or inelastically due to polar optical phonons. Flux is removed from the incident wave and fed into the scattered waves,  $\psi_1$ , at energies  $E$  and  $E \pm \omega$ , which have no phase coherence with  $\psi_0$ . The term  $\omega$  is the polar optical phonon energy. The wave  $\psi_1$  now propagates and scatters elastically giving rise to  $\psi_2$ , etc. An infinite number of elastic scattering events and only one inelastic scattering event are allowed to occur. When one extends the multiple sequential scattering algorithm to infinite sequential scattering, it becomes the self-consistent Born approximation. Our derivations use the above picture combined with a physical interpretation of the self energies.<sup>39,40</sup> In the non-equilibrium Green function formalism, the effect of outscattering is contained in  $\Sigma^R$  and the in scattering is contained in  $\Sigma^<$ .

### A. Single sequential elastic scattering

#### 1. Equations of motion

We first derive equations for treating incoherent *elastic* scattering in the single sequential scattering approximation.<sup>10</sup> In Appendix B, we use these simple equations to demonstrate current conservation, the violation of detailed balance, and the non-zero equilibrium current. Eqs. (80–83) can be shown to lead to identical expressions for the current and self-energy written down by Roblin and Liou in Secs. V and VI of Ref. 6.

For single sequential elastic scattering, we have an incident wave  $\psi_0(\mathbf{k})$ , whose propagation is governed by  $G_0^R$ , scattering and creating a wave  $\psi_1(\mathbf{k}')$  which has no phase coherence with  $\psi_0(\mathbf{k})$ . Since  $\psi_1(\mathbf{k}')$  suffers no further scattering, its propagation is governed by the bare Green function,  $G_1^R$ . The equations that we write down for the Green functions are

$$(E - H_0^D - \Sigma^{RB} - D \otimes G_1^R) G_0^R = 1, \quad (80)$$

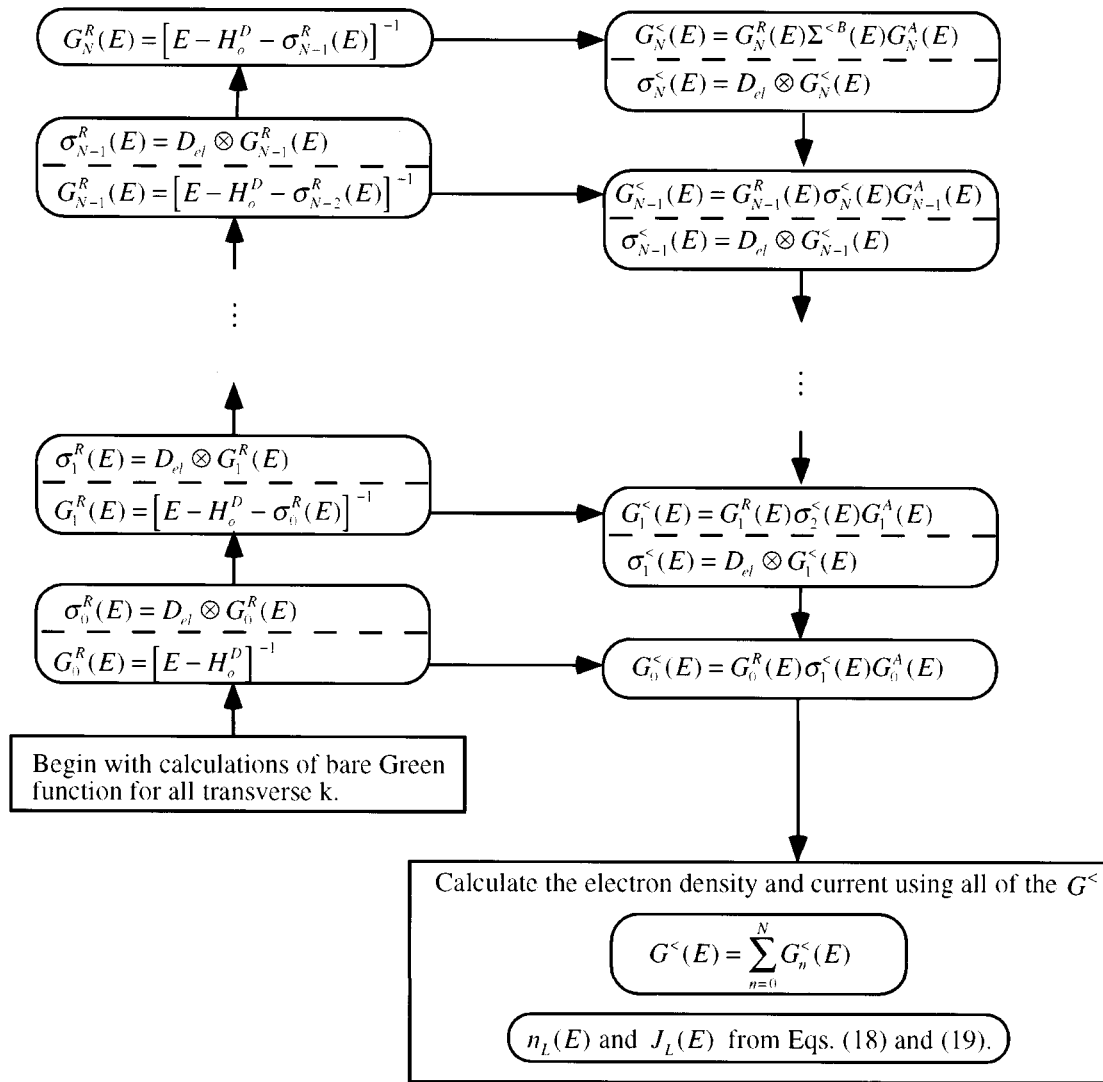


FIG. 3. Flow chart showing the multiple sequential scattering algorithm for incoherent *elastic* scattering. The equations are numbered so that the bare Green function is  $G_0^R$ .

$$(E - H_o^D - \Sigma^{RB})G_1^R = 1. \quad (81)$$

The self-energy,  $D \otimes G_1^R$ , in the equation of motion for  $G_0^R$  accounts for the loss and spectrum shift resulting from the outscattering of state  $\psi_0$  to  $\psi_1$ . The equations that we write down for  $G_0^<$  and  $G_1^<$  are

$$G_0^< = G_0^R \Sigma^{<B} G_0^A, \quad (82)$$

$$G_1^< = G_1^R [D \otimes G_0^<] G_1^A. \quad (83)$$

The source term for  $G_0^<$ ,  $\Sigma^{<B}$  is due to injection from the contacts. The source term for  $G_1^<$ ,  $D \otimes G_0^<$ , is due to in-scattering from  $G_0^<$ . Figure (2) illustrates the physical interpretation of the self-energies. The left reservoir injects a state  $k_0$  into the device ( $\Sigma^{<B}$ ). Flux can leave state  $k_0$  by either leaving the device ( $\Sigma^{RB}$ ) or incoherently out scattering into

another state ( $\Sigma^R$ ). The out scattered flux from  $k_0$  is scattered into state  $k_1$  ( $\Sigma^<$ ). State  $k_1$  leaks out into the contacts ( $\Sigma^{RB}$ ).

This set of equations is solved by first calculating the bare Green function  $G_1^R$ , then the self-energy,  $D \otimes G_1^R$ , and then the Green function  $G_0^R$ . With  $G_0^R$ , one then calculates  $G_0^<$ , then  $D \otimes G_0^<$ , and finally  $G_1^<$ . The electron density and current is found from summing the contributions from both  $G_0^<$  and  $G_1^<$  using Eqs. (18) and (19), respectively.

## 2. Truncated expansions

To observe the relationship between Eqs. (80) and (81) and the truncated continued fraction expansion of Eq. (72), write  $\sigma_1^R$  as  $D \otimes (E - H_o^D - \Sigma^{RB})^{-1}$  and substitute it into Eq. (80) to obtain

$$G_0^R = (E - H_o^D - \Sigma^{RB} - D \otimes (E - H_o^D - \Sigma^{RB})^{-1})^{-1}, \quad (84)$$

which is the first order truncation of Eq. (72).

To observe the relationship between Eqs. (82) and (83) and the truncated power series expansion of Eq. (75), note that the solution of the diagonal elements of  $G_0^<$  is (in the single band model using the notation of Eq. (75))  $G_0^< = |G_0^R|^2 \Sigma^{<B} = S_0$ . The solution for the diagonal elements of  $G_1^<$  is (again, using the notation of Eq. (75))  $G_1^< = |G_1^R|^2 D \otimes S_0$ . Defining  $G^< = G_0^< + G_1^<$ , we get

$$G^< = S_0 + |G_1^R|^2 D \otimes S_0 \quad (85)$$

similar to the first two terms of Eq. (75).

The difference between simple truncation of Eqs. (75) and (85) is that there are two different Green functions,  $G_0^R$  and  $G_1^R$ , contained on the right hand side of Eq. (85).  $G_1^R$  is shown explicitly, and  $G_0^R$  is contained in  $S_0$ . This is essential to conserve current. However, it makes impossible the equilibrium relationship between  $G^<$  and the spectral functions or sum of spectral functions. In other words, there is no equilibrium relationship  $G^< = ifA$  in the multiple sequential scattering theory.<sup>6</sup> This is the source of the violation of detailed balance and of the non-zero equilibrium current.

## B. Multiple sequential elastic scattering

The extension of Eqs. (80–83) to include multiple scattering events is straightforward. The equations for the retarded Green functions governing the propagation of the successively scattered waves are

$$\begin{aligned} [E - H_o^D - \Sigma^{RB} - D \otimes G_1^R(E)] G_0^R(E) &= 1 \\ [E - H_o^D - \Sigma^{RB} - D \otimes G_2^R(E)] G_1^R(E) &= 1 \\ &\vdots \\ [E - H_o^D - \Sigma^{RB} - D \otimes G_N^R(E)] G_{N-1}^R(E) &= 1 \\ [E - H_o^D - \Sigma^{RB}] G_N^R(E) &= 1. \end{aligned} \quad (86)$$

The physical interpretation of Eqs. (86) is as follows.  $G_0^R$  governs the propagation of the initial injected wave,  $\psi_0$ . The self-energy  $D \otimes G_1^R(E)$  accounts for the spectrum shift and loss due to the incoherent elastic out scattering into state  $\psi_1$ .  $G_1^R$  governs the propagation of state  $\psi_1$ , and  $G_1^R$  is dressed by the self-energy  $D \otimes G_2^R(E)$  resulting from the out scattering of  $\psi_1$  into state  $\psi_2$ , etc.

The set of equations for the  $G^<$  are

$$G_0^<(E) = G_0^R \Sigma^{<B} G_0^A(E)$$

$$G_1^<(E) = G_1^R(E) [D \otimes G_0^<(E)] G_1^A(E)$$

$\vdots$

$$G_{N-1}^<(E) = G_{N-1}^R(E) [D \otimes G_{N-2}^<(E)] G_{N-1}^A(E)$$

$$G_N^<(E) = G_N^R(E) [D \otimes G_{N-1}^<(E)] G_N^A(E). \quad (87)$$

Eqs. (87) are the Green function analogue to the wave function equations for the  $\psi_i$ . Their physical interpretation is as follows.  $G_0^<$  corresponds to  $\psi_0$ . The source term,  $\Sigma^{<B}$ , is due to injection from the contacts and propagation is governed by  $G_0^R$ .  $G_1^<(E)$  corresponds to  $\psi_1(E)$ . The source term,  $D \otimes G_0^<(E)$ , results from the incoherent elastic inscattering of  $G_0^<(E)$  and propagation is governed by  $G_1^R(E)$ .  $G_1^<$  elastically out scatters providing the source term for  $G_2^<(E)$ , etc. The electron density and current are found from summing the contributions from all of the  $G_i^<$  using Eqs. (18) and (19), respectively.

In Fig. (3), we show a flow chart displaying the order in which the equations are solved. In the figure, we reverse the numbering system of the equations to correspond to the numbering system used in the numerical solution procedure. The bare Green function becomes  $G_0^R$ . The number of events  $N$  can be set in advance, or it can be determined by a convergence criterion for  $G_N^R$ .

We have used the simple examples of single sequential and multiple sequential elastic scattering to investigate the properties of the multiple sequential scattering algorithm. We now derive the equations which we find most useful for our device simulator in which elastic scattering is treated in the multiple sequential scattering or self-consistent Born approximation and the polar optical phonon scattering is treated in the single sequential scattering approximation.

## C. Multiple sequential elastic scattering and single sequential polar optical phonon scattering

We write down the series of equations which are the starting point for both the MSS and self-consistent Born treatment of the elastic scattering processes. If we truncate the series of equations at some finite order, we get an MSS algorithm. If we let the series of equations go to infinite order, we can analytically sum the series to obtain a self-consistent Born treatment of the elastic scattering.

### 1. Equations of motion for $G^R$

The physical picture described in the paragraph preceding Sec. (VI A) leads one to write down the following series of equations for Green functions which govern the propagation of the various scattered waves,  $\psi_i$ .

$$\left[ E - H_o^D - \Sigma^{RB} - \underbrace{D_{el} \otimes G_1^R(E)}_{\Sigma_{el}^R} - \underbrace{D_{pop}^{>em} \otimes G_1^R(E - \omega) - D_{pop}^{>ab} \otimes G_1^R(E + \omega)}_{-\sigma_{pop}^R} \right] G_0^R(E) = 1. \quad (88)$$

$$\begin{aligned}
[E - H_o^D - \Sigma^{RB} - D_{el} \otimes G_2^R(E)] G_1^R(E) &= 1 \\
[E - H_o^D - \Sigma^{RB} - D_{el} \otimes G_3^R(E)] G_2^R(E) &= 1. \\
&\vdots
\end{aligned} \tag{89}$$

$$\begin{aligned}
[E + \omega - H_o^D - \Sigma^{RB} - D_{el} \otimes G_2^R(E + \omega)] G_1^R(E + \omega) &= 1 \\
[E + \omega - H_o^D - \Sigma^{RB} - D_{el} \otimes G_3^R(E + \omega)] G_2^R(E + \omega) &= 1. \\
&\vdots
\end{aligned} \tag{90}$$

$$\begin{aligned}
[E - \omega - H_o^D - \Sigma^{RB} - D_{el} \otimes G_2^R(E - \omega)] G_1^R(E - \omega) &= 1 \\
[E - \omega - H_o^D - \Sigma^{RB} - D_{el} \otimes G_3^R(E - \omega)] G_2^R(E - \omega) &= 1 \\
&\vdots
\end{aligned} \tag{91}$$

In Eqs. (88–91),  $\Sigma^{RB}$  accounts for the open system boundary conditions,  $\otimes$  indicates convolution over transverse momentum,  $D_{el}$  contains the matrix element squared due to the elastic scattering processes,  $D_{pop}^{>em} \propto n_B + 1$  contains the matrix element squared governing the out scattering due to the emission of polar optical phonons, and  $D_{pop}^{>ab} \propto n_B$  contains the matrix element squared governing the out scattering due to the absorption of polar optical phonons (see Eq. (A9)). The equations are written in full matrix notation and transverse momentum  $k$  has been suppressed for compactness of notation.

The physical interpretation of Eqs. (88–91) is as follows.  $G_0^R$  governs the propagation of the initial injected wave,  $\psi_0$ .  $\Sigma_{el}^R$  and  $\sigma_{pop}^R$  account for the spectrum shift and loss due to the incoherent elastic and inelastic scattering, respectively. The wave  $\psi_0(E)$  can be scattered into states  $\psi_1$  at three different energies,  $E$ , and  $E \pm \omega$ . The states  $\psi_1$  can then be scattered elastically an infinite number of times into states  $\psi_2$ ,  $\psi_3$ , etc. The propagation of states  $\psi_1 - \psi_\infty$  is governed by  $G_1^R - G_\infty^R$ .

## 2. Equations of motion for $G^<$

The physical picture described in the paragraph preceding Sec. (VI A) leads one to write down the following series of equations for  $G^<$ .

$$G_0^<(E) = G_0^R(E) \Sigma^{<B} G_0^A(E), \tag{92}$$

$$G_1^<(E) = G_1^R(E) [D_{el} \otimes G_0^<(E)] G_1^A(E),$$

$$G_2^<(E) = G_2^R(E) [D_{el} \otimes G_1^<(E)] G_2^A(E),$$

$$\vdots \tag{93}$$

$$G_1^<(E + \omega) = G_1^R(E + \omega) [D_{pop}^{<ab} \otimes G_0^<(E)] G_1^A(E + \omega)$$

$$G_2^<(E + \omega) = G_2^R(E + \omega) [D_{el} \otimes G_1^<(E + \omega)] G_2^A(E + \omega)$$

$$G_3^<(E + \omega) = G_3^R(E + \omega) [D_{el} \otimes G_2^<(E + \omega)] G_3^A(E + \omega),$$

$$\vdots \tag{94}$$

$$G_1^<(E - \omega) = G_1^R(E - \omega) [D_{pop}^{<em} \otimes G_0^<(E)] G_1^A(E - \omega)$$

$$G_2^<(E - \omega) = G_2^R(E - \omega) [D_{el} \otimes G_1^<(E - \omega)] G_2^A(E - \omega)$$

$$G_3^<(E - \omega) = G_3^R(E - \omega) [D_{el} \otimes G_2^<(E - \omega)] G_3^A(E - \omega).$$

$$\vdots \tag{95}$$

Eqs. (92–95) for  $G_i^<$  are the Green function analogue to the wave function equations for the  $\psi_i$ . Their physical interpretation is as follows.  $G_0^<$  corresponds to  $\psi_0$ . The source term,  $\Sigma^{<B}$ , is due to injection from the contacts and propagation is governed by  $G_0^R$  (note that  $G^A$  is the hermitian conjugate of  $G^R$ ).  $G_1^<(E)$  corresponds to  $\psi_1(E)$ . The source term,  $D_{el} \otimes G_0^<(E)$ , results from the elastic in scattering of  $G_0^<(E)$  and propagation is governed by  $G_1^R(E)$ .  $G_1^<$  elastically out scatters providing the source term for  $G_2^<(E)$ , etc.  $G_1^<(E \pm \omega)$  corresponds to  $\psi_1(E \pm \omega)$ . The source term,  $D_{pop}^{<ab,em} \otimes G_0^<(E)$ , results from the inelastic in scattering of  $G_0^<(E)$  due to the polar optical phonons (see Eq. (A9)) and propagation is governed by  $G_1^R(E \pm \omega)$ .  $G_1^<(E \pm \omega)$  elastically out scatters providing the source term for  $G_2^<(E \pm \omega)$ , etc.

If we truncate the chains of Eqs. (89–91) and (93–95) to any finite order, we have a multiple sequential scattering treatment of the incoherent elastic scattering and a single sequential scattering treatment of the polar optical phonon scattering. To solve the truncated series of equations, one begins by solving for the bare retarded Green functions at energies  $E$  and  $E \pm \omega$  and at all transverse  $k$ . Then one moves up the three chains of Green function equations solving for the  $G_i^R$ . At the top, one calculates  $G_0^R$  and then moves over to the equations for  $G^<$  getting  $G_0^<$  and then down the three chains of equations for  $G_i^<$  at energies  $E$  and  $E \pm \omega$ . The electron density and current is found from summing the contributions from all of the  $G_i^<$  at energies  $E$  and  $E \pm \omega$  using Eqs. (18) and (19), respectively.

In Fig. (4), we show a flow chart displaying the order in which the equations are solved. In the figure, we again reverse the numbering system of the equations to correspond to the numbering system used in the numerical solution procedure. The bare Green function becomes  $G_0^R$ . The number of events  $N$  can be set in advance, or it can be determined by a convergence criterion for  $G_N^R$  independently at each energy,  $E$  and  $E \pm \omega$ .

## D. Self-consistent born elastic scattering and single sequential polar optical phonon scattering

### 1. Equations of motion for $G^R$ and $G^<$

If we let the series of Eqs. (89–91) become infinite, all of the  $G_i^R$ 's,  $i \in \{1, \dots, \infty\}$ , at a given energy and transverse momentum, are identical. Therefore, the infinite set of Eqs. (88–91) becomes

$$[E - H_o^D - \Sigma^{RB} - \Sigma_{el}^R(E) - \sigma_{pop}^R(E)] G_0^R(E) = 1, \tag{96}$$

$$[E - H_o^D - \Sigma^{RB} - \Sigma_{el}^R(E)] G_{el}^R(E) = 1, \tag{97}$$

$$[E + \omega - H_o^D - \Sigma^{RB} - \Sigma_{el}^R(E + \omega)] G_{el}^R(E + \omega) = 1, \tag{98}$$

$$[E - \omega - H_o^D - \Sigma^{RB} - \Sigma_{el}^R(E - \omega)] G_{el}^R(E - \omega) = 1, \tag{99}$$

where in Eqs. (96–99),  $\Sigma_{el}^R(E) = D_{el} \otimes G_{el}^R(E)$  is the self-energy due to the elastic scattering mechanisms and it is calculated self-consistently with  $G_{el}^R$ , and  $\sigma_{pop}^R(E) = D_{pop}^{>em} \otimes G_{el}^R(E - \omega) + D_{pop}^{>ab} \otimes G_{el}^R(E + \omega)$ .



The physical observables of electron density and current are calculated by summing the contributions from all of the  $G_i^<$  in Eqs. (92–95). We define

$$G_{el}^<(E) = \sum_{i=1}^{\infty} G_i^<(E). \quad (100)$$

Since all of the  $G_i^R = G_{el}^R$  are identical for a given energy and transverse momentum, we can perform the sum. Summing the infinite series of Eqs. (93), we have

$$\begin{aligned} G_{el}^<(E) &= G_{el}^R(E) \left[ D_{el} \otimes G_0^<(E) + D_{el} \otimes \sum_{i=1}^{\infty} G_i^<(E) \right] G_{el}^A(E) \\ &= G_{el}^R(E) \left[ \underbrace{D_{el} \otimes G_0^<(E)}_{\sigma_{el}^{<0}} + \underbrace{D_{el} \otimes G_{el}^<(E)}_{\Sigma_{el}^<(E)} \right] G_{el}^A(E). \end{aligned} \quad (101)$$

Summing the infinite series of Eqs. (94), we have

$$\begin{aligned} G_{el}^<(E+\omega) &= G_{el}^R(E+\omega) \left[ D_{pop}^{<ab} \otimes G_0^<(E) + D_{el} \otimes \sum_{i=1}^{\infty} G_i^<(E+\omega) \right] G_{el}^A(E+\omega) \\ &= G_{el}^R(E+\omega) \left[ \underbrace{D_{pop}^{<ab} \otimes G_0^<(E)}_{\sigma_{pop}^{<ab}} + \underbrace{D_{el} \otimes G_{el}^<(E+\omega)}_{\Sigma_{el}^<(E+\omega)} \right] G_{el}^A(E+\omega) \end{aligned} \quad (102)$$

and finally, summing the infinite series of Eqs. (95), we have

$$\begin{aligned} G_{el}^<(E-\omega) &= G_{el}^R(E-\omega) \left[ D_{pop}^{<ab} \otimes G_0^<(E) + D_{el} \otimes \sum_{i=1}^{\infty} G_i^<(E-\omega) \right] G_{el}^A(E-\omega) \\ &= G_{el}^R(E-\omega) \left[ \underbrace{D_{pop}^{<em} \otimes G_0^<(E)}_{\sigma_{pop}^{<em}} + \underbrace{D_{el} \otimes G_{el}^<(E-\omega)}_{\Sigma_{el}^<(E-\omega)} \right] G_{el}^A(E-\omega). \end{aligned} \quad (103)$$

Equations (96–99), (92), and (101–103) are the final set of equations that we need to solve.

## 2. Solving the equations

We now write down the equations in the order in which they are solved. First, the self-energies,  $\Sigma_{el}^R$  in Eqs. (97–99) are solved self-consistently with their corresponding  $G_{el}^R$  at energies  $E$  and  $E \pm \omega$  and for all transverse momentum,  $k$ . The equations are

$$[E - H_o^D - \Sigma^{RB} - \Sigma_{el}^R(E)] G_{el}^R(E) = 1, \quad (104)$$

$$\Sigma_{el}^R(E) = D_{el} \otimes G_{el}^R(E), \quad (105)$$

$$[E + \omega - H_o^D - \Sigma^{RB} - \Sigma_{el}^R(E + \omega)] G_{el}^R(E + \omega) = 1, \quad (106)$$

$$\Sigma_{el}^R(E + \omega) = D_{el} \otimes G_{el}^R(E + \omega), \quad (107)$$

$$[E - \omega - H_o^D - \Sigma^{RB} - \Sigma_{el}^R(E - \omega)] G_{el}^R(E - \omega) = 1, \quad (108)$$

$$\Sigma_{el}^R(E - \omega) = D_{el} \otimes G_{el}^R(E - \omega). \quad (109)$$

With the solution of Eqs. (104–109), we now have the elastic self-energy,  $\Sigma_{el}^R(E)$ , from Eq. (105) that goes into the matrix on the left hand side of Eq. (96), and we calculate the inelastic self-energy,  $\sigma_{pop}^R(E)$ , from  $G_{el}^R(E + \omega)$  and  $G_{el}^R(E - \omega)$ :

$$\sigma_{pop}^R(E) = D_{pop}^{>em} \otimes G_{el}^R(E - \omega) + D_{pop}^{>ab} \otimes G_{el}^R(E + \omega). \quad (110)$$

Knowing the self-energies,  $\sigma_{pop}^R(E)$  and  $\Sigma_{el}^R(E)$ , we then calculate  $G_0^R(E)$  from (96):

$$[E - H_o^D - \Sigma^{RB} - \Sigma_{el}^R(E) - \sigma_{pop}^R(E)] G_0^R(E) = 1. \quad (111)$$

With  $G_0^R(E)$ , we immediately have  $G_0^<(E)$  from (92):

$$G_0^<(E) = G_0^R(E) \Sigma^{<B}(E) G_0^A(E). \quad (112)$$

We can now calculate the source terms for Eqs. (101–103),

$$\sigma_{el}^{<0} = D_{el} \otimes G_0^<(E), \quad (113)$$

$$\sigma_{pop}^{<ab} = D_{pop}^{<ab} \otimes G_0^<(E), \quad (114)$$

$$\sigma_{pop}^{<em} = D_{pop}^{<em} \otimes G_0^<(E) \quad (115)$$

and then we self-consistently solve Eqs. (101–103) for  $G_{el}^<$  and  $\Sigma_{el}^<$  using Jacobi iteration combined with SOR as described in Sec. (V B).

$$G_{el}^<(E) = G_{el}^R(E) [\sigma_{el}^{<0} + \Sigma_{el}^<(E)] G_{el}^A(E), \quad (116)$$

$$\Sigma_{el}^<(E) = D_{el} \otimes G_{el}^<(E), \quad (117)$$

$$G_{el}^<(E + \omega) = G_{el}^R(E + \omega) [\sigma_{pop}^{<ab} + \Sigma_{el}^<(E + \omega)] G_{el}^A(E + \omega), \quad (118)$$



$$\Sigma_{el}^<(E+\omega) = D_{el} \otimes G_{el}^<(E+\omega), \quad (119)$$

$$G_{el}^<(E-\omega) = G_{el}^R(E-\omega) [\sigma_{pop}^{<em} + \Sigma_{el}^<(E-\omega)] G_{el}^A(E-\omega), \quad (120)$$

$$\Sigma_{el}^<(E-\omega) = D_{el} \otimes G_{el}^<(E-\omega). \quad (121)$$

The current and electron density per unit total energy,  $E$ , in the device is then calculated by summing the contributions from  $G_0^<(E)$ ,  $G_{el}^<(E)$ , and  $G_{el}^<(E \pm \omega)$  using Eqs. (19) and (18) (ignoring the integral over energy). After the calculations are performed for one total energy, another total energy is chosen, the calculations are repeated and the new current per unit total energy is added to the running total. This is repeated until the integrated total current and electron density is obtained. Current conservation is shown in the same way as in Sec. (VI A) summing the contributions to the divergence of the current from  $G^<$  calculated in Eqs. (112), (116), (118), and (120).

A flow chart of the solution procedure is shown in Fig. (5). Self-consistent calculations are indicated by boxes drawn with heavier linestyles.

For the self-consistent potential calculation, we also need the charge density in the leads. This is calculated using  $-iG^< = fA_0$  in Eq. (18) where  $A_0 = i(G_0^R - G_0^A)$  with  $G_0^R$  given by Eq. (88).  $A_0$  is used because it is the most “dressed” spectral function best representing the true spectral function of the leads. The charge density is then used for generating the Jacobian in the Newton-Raphson scheme as described in Appendix C.

## E. Current

Once one is confident that the numerical solutions do conserve current, considerable numerical efficiency can be obtained by calculating the current using Eq. (45) modified in a similar manner as Eq. (B1) in Appendix B. These equations require only the block corner elements of  $G^R$  and  $G^<$ . This information already exists from the MSS or self-consistent Born calculations. Eq. (19) uses the off-diagonal blocks of  $G^<$  which require an extra calculation. For the MSS approaches described in Secs. (VI B) and (VI C), the current per unit incident energy is found from (for a nearest neighbor model)

$$J_0(E) = \frac{2e}{\hbar A} \sum_k \text{tr} \left\{ \Gamma_{1,1}^B(k, E) \left[ f_e A_{0,1,1}(k, E) + i \left( G_{0,1,1}^<(k, E) + \sum_{n=1}^N G_{n,1,1}^<(k, E) \right) \right] \right\} \\ + \frac{2e}{\hbar A} \sum_k \text{tr} \left\{ \Gamma_{1,1}^B(k, E+\omega) i \sum_{n=1}^N G_{n,1,1}^<(k, E+\omega) \right\} + \frac{2e}{\hbar A} \sum_k \text{tr} \left\{ \Gamma_{1,1}^B(k, E-\omega) i \sum_{n=1}^N G_{n,1,1}^<(k, E-\omega) \right\}. \quad (122)$$

For the combined self-consistent Born calculation described in Sec. (VI D), the current per unit incident energy is found from (for a nearest neighbor model)

$$J_0(E) = \frac{2e}{\hbar A} \sum_k \text{tr} \{ \Gamma_{1,1}^B(k, E) [f_L A_{0,1,1}(k, E) + i(G_{0,1,1}^<(k, E) + G_{el,1,1}^<(k, E))] \} \\ + \frac{2e}{\hbar A} \sum_k \text{tr} \{ \Gamma_{1,1}^B(k, E+\omega) i G_{el,1,1}^<(k, E+\omega) \} + \frac{2e}{\hbar A} \sum_k \{ \Gamma_{1,1}^B(k, E-\omega) i G_{el,1,1}^<(k, E-\omega) \}, \quad (123)$$

where  $A_0 = i(G_0^R - G_0^A)$ . The term  $f_L A_0$  results from the contact source term in the equation for  $G_0^<$  and represents the inflow of current from the contact to the device. The  $G^<$  terms represent the backflow from the device to the contact.

## VII. NUMERICAL COMPARISON OF MSS AND SCB CALCULATIONS

In this section, we numerically illustrate the convergence of the multiple sequential scattering algorithm with the self-consistent Born algorithm. We consider two different RTDs at  $T=4.2$  K which differ only in the barrier material. One RTD has AlAs barriers and the other has  $\text{Al}_{0.35}\text{Ga}_{0.65}\text{As}$  barriers. The device structure consists of 19.2 nm intrinsic GaAs spacer layers, 3.4 nm barriers, a 5.66 nm GaAs well, and Si doped  $10^{18} \text{ cm}^{-3}$  GaAs contacts. The AlAs barrier RTD is the same device modeled in Refs. 14–16 using the algorithm illustrated in Fig. (4). Here, we show the convergence of the elastic multiple sequential scattering algorithm illustrated in

Fig. (3) with the self-consistent Born calculation which consists of Eqs. (65), (66), (70), and (71). The interface roughness self-energy, Eq. (A35), is used with a correlation length of  $\Lambda = 10$  nm.

Figure (6) shows how the MSS algorithm converges to the SCB result at the peak and valley current regions of the GaAs / AlAs RTD. Figure 6(a) shows the overall  $I$ - $V$  curve calculated using the SCB algorithm and the MSS algorithm with one scattering event (Fig. (3) with  $N=1$ ). As discussed in Ref. 15, the electrostatic potential is calculated self-consistently with the quantum charge in the absence of incoherent scattering. Once the potential is obtained, it is then used for the calculation which includes incoherent scattering. This approach is used to reduce CPU time since we calculate an extreme number of scattering events ( $N=4000$  in Fig. (3)) to demonstrate convergence of the MSS and SCB algorithms.

Figure (6b) shows the convergence of the MSS and SCB algorithms in the valley current region of the  $I$ - $V$ . The cur-

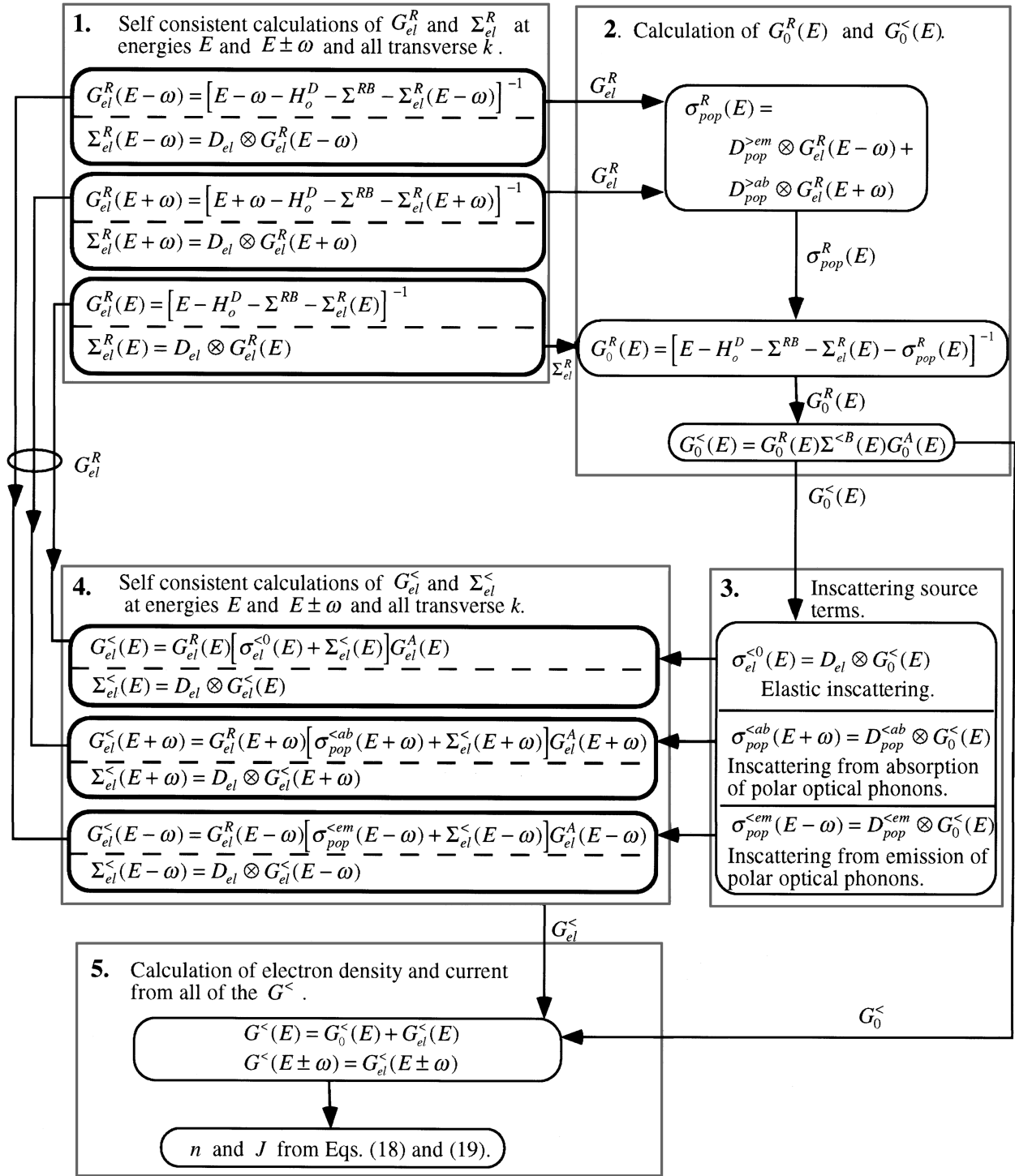


FIG. 5. Flow chart illustrating the algorithm for self-consistent Born elastic scattering and single sequential inelastic scattering. Self-consistent calculations are indicated by boxes drawn with heavier line styles.

rent calculated in a second order multiple sequential scattering approximation is essentially identical to the current calculated in a self-consistent Born approximation. The first order calculation is very good. This is consistent with the “capture” point of view of Chevoir and Vinter<sup>7</sup> and the numerical result of Johansson.<sup>41</sup>

The convergence properties of the MSS algorithm at the peak current are quite different. We believe that the conver-

gence of the MSS algorithm at peak current scales as  $\Gamma/\Gamma_i$  where  $\Gamma$  is the fully dressed width of the resonance and  $\Gamma_i$  is the intrinsic width of the resonance. This quantity is simply the scattering rate times the resonant state lifetime. The intrinsic resonance width of the AlAs barrier RTD is  $3.0 \mu\text{eV}$ . The dressed resonance width in the self-consistent Born approximation is  $2.7 \text{ meV}$ . Figure (6c) shows the peak current calculated with the MSS algorithm using  $N=1, 2, 4, 8$ , and

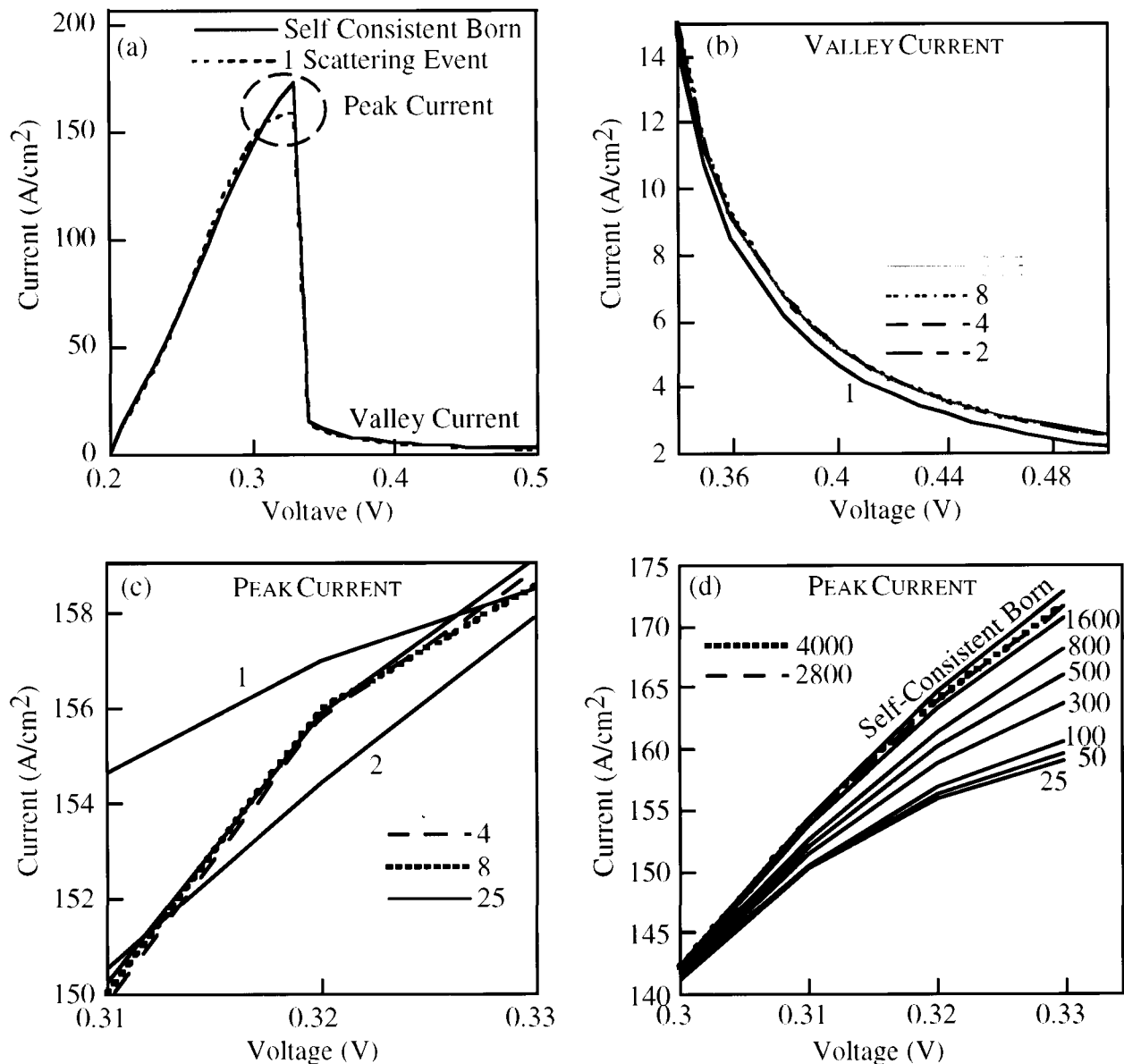


FIG. 6. Numerical demonstration of the convergence of the multiple sequential scattering algorithm and the self-consistent Born algorithm. (a) I-V of GaAs / AlAs RTD calculated using the self-consistent Born algorithm and the MSS algorithm with one scattering event. (b-d) Convergence of the MSS algorithm to the SCB algorithm: (b) in the valley current region, (c) in the peak current region for few scattering events,  $N=1-25$ , and (d) in the peak current region for  $N=25-4000$ .

25 scattering events. Initially, the value of the current tends to oscillate as the number of scattering events is increased. There is almost no change in the current as the number of scattering events is increased from four to 25. After 25 scattering events, the MSS algorithm begins monotonically converging as shown in Fig. (6d). At large  $N$  the convergence of the MSS algorithm begins to saturate with a current slightly below the SCB result (compare the  $N=2800$  and the  $N=4000$  curves in Fig. (6d)).

For the  $\text{Al}_{0.35}\text{Ga}_{0.65}\text{As}$  barrier RTD, the intrinsic resonant width is  $780 \mu\text{eV}$  which is 260 times larger than the AlAs barrier RTD. Figure 7 shows that the convergence of the MSS algorithm at peak current is very fast. After  $N=5$  scattering events, the MSS calculation and the self-consistent Born calculation are essentially identical.

## VIII. SUMMARY AND CONCLUSION

We have presented theory which can be used at a number of different levels of sophistication and complexity for modeling high bias quantum devices. The simplest level, summarized by Eq. (50), is a generalized tunneling formula which treats injection from both emitter continuum and quasi-bound states. It is just as fast as the usual tunneling approaches<sup>2</sup> and much more flexible, allowing one to model devices such as those shown in Ref. 1. This is the level heavily used in device design to quickly obtain  $I$ - $V$  characteristics. At this level we have implemented several different nearest neighbor models: single-band,<sup>1</sup> two-band,<sup>14</sup> and  $sp^3s^*$ .<sup>14,17</sup> The next level is to include scattering in the device. We have numerically implemented in a single band

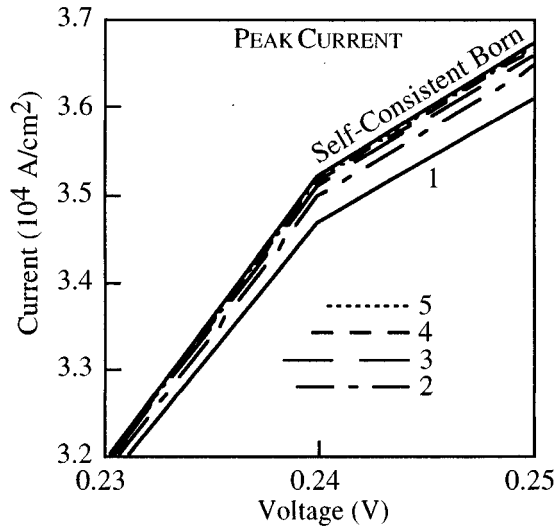


FIG. 7. Numerical demonstration of the convergence of the multiple sequential scattering algorithm and the self-consistent Born algorithm at peak current of a GaAs / Al<sub>0.35</sub>Ga<sub>0.65</sub>As RTD.

model the combined MSS treatment of elastic scattering and single sequential scattering treatment of polar optical phonon scattering described in Sec. (VIC), and the combined self-consistent Born and single sequential scattering treatment of polar optical phonon scattering described in Sec. (VID). The valley current of an RTD calculated using the multiple sequential scattering algorithm converges within a few scattering events to the valley current calculated with the self-consistent Born algorithm. The convergence of the multiple sequential scattering algorithm at a bias corresponding to the peak current of an RTD is much slower.

## ACKNOWLEDGMENTS

Selman Hershfield introduced us to the single electron approximation for the self-energy described in Sec. VC. This was the key that allowed us to bring together the self-consistent Born approximation and the multiple sequential scattering theory of Roblin and Liou. Our understanding of the multiple sequential scattering theory has benefited from many conversations and correspondences with Patrick Roblin. The work was supported in part by Texas Instruments Incorporated.

## APPENDIX A: SCATTERING SELF-ENERGIES

To calculate the scattering self-energy, we calculate the path ordered Green function in the interaction representation,

$$G_{\alpha,L;\alpha',L'}^P(\mathbf{k};t,t') = \frac{-i}{\hbar} \langle P e^{-\frac{i}{\hbar} \int_C ds H'(s)} c_{\alpha,L,\mathbf{k}}(t) c_{\alpha',L',\mathbf{k}}^\dagger(t') \rangle, \quad (\text{A1})$$

where  $P$  is the path ordering operator,  $C$  is the Keldysh contour, and  $H'(s)$  is the perturbing Hamiltonian. The brackets  $\langle \dots \rangle$  indicate the non-equilibrium ensemble average<sup>29,42</sup> and an ensemble average over the random potential distribution resulting from interface roughness, alloy disorder, and ion-

ized dopants. The scattering self-energies result from the second order term in the expansion of the exponential. For phonon scattering, the first order term is zero since  $\langle a_{\mathbf{q}} \rangle = \langle a_{\mathbf{q}}^\dagger \rangle = 0$ . For alloy disorder and interface roughness, the first order term results in the virtual crystal approximation (VCA) Hamiltonian and, for ionized dopants, the first order term is included through Poisson's equation. Using Wick's theorem to group the terms,<sup>43</sup> one obtains an equation of the form (in block matrix notation)

$$G_{L;L'}^P(\mathbf{k};t,t') = g_{L;L'}^P(\mathbf{k};t,t') + \int_C ds \int_C ds' \sum_{L_1} \sum_{L_2} [g_{L;L_1}^P(\mathbf{k};t,s) \times \Sigma_{L_1;L_2}^P(\mathbf{k};s,s') g_{L_2;L'}^P(\mathbf{k};s',t')]. \quad (\text{A2})$$

In the self-consistent Born approximation the bare  $g^P$  that occurs in  $\Sigma^P$  in Eq. (A2) is replaced with the full  $G^P$ . The real-time self-energies that we need,  $\Sigma^<$ ,  $\Sigma^>$ , and  $\Sigma^R$ , are obtained from the  $\Sigma^P$  using the relations of Langreth.<sup>29</sup>

## 1. Phonons

The potential felt by the electrons due to bulk phonons has the general form of

$$V_{\text{ep}} = \frac{1}{\sqrt{V}} \sum_{\mathbf{q}} U_{\mathbf{q}} e^{i\mathbf{q} \cdot \mathbf{r}} (a_{\mathbf{q}} + a_{-\mathbf{q}}^\dagger), \quad (\text{A3})$$

where  $a_{\mathbf{q}}$  is the destruction operator for a phonon in mode  $\mathbf{q}$ . The second quantized Hamiltonian  $H_{\text{ep}}$  is

$$H_{\text{ep}} = \int d^3r \psi^\dagger(\mathbf{r}) V_{\text{ep}}(\mathbf{r}) \psi(\mathbf{r}), \quad (\text{A4})$$

where  $\psi(\mathbf{r})$  is the field operator of Eq. (5). In the long wavelength approximation where  $e^{i\mathbf{q} \cdot \mathbf{r}}$  is assumed to be slowly varying on the order of a localized orbital,

$$H_{\text{ep}} = \sum_{L,\mathbf{k}} \frac{1}{\sqrt{V}} \sum_{\mathbf{q}} U_{\mathbf{q}} e^{iq_z \Delta L} \left[ e^{iq_z \Delta L/2} \sum_a c_{a,L,\mathbf{k}}^\dagger c_{a,L,\mathbf{k}-\mathbf{q}} + \sum_c c_{c,L,\mathbf{k}}^\dagger c_{c,L,\mathbf{k}-\mathbf{q}} \right] (a_{\mathbf{q}} + a_{-\mathbf{q}}^\dagger). \quad (\text{A5})$$

We write the expressions for  $\Sigma^>$  and  $\Sigma^<$ . The general forms of  $\Sigma^<$  and  $\Sigma^>$  are

$$\Sigma_{\alpha,L;\alpha',L'}^<(k;E) = \frac{1}{V} \sum_{\mathbf{q}} |U_{\mathbf{k}-\mathbf{q}}|^2 e^{iq_z \Delta(L-L'+v_{\alpha,\alpha'})} \times [n_{\mathbf{q}} G_{\alpha,L;\alpha',L'}^<(\mathbf{q}_t; E - \hbar \omega_{\mathbf{q}}) + (n_{\mathbf{q}} + 1) G_{\alpha,L;\alpha',L'}^<(\mathbf{q}_t; E + \hbar \omega_{\mathbf{q}})] \quad (\text{A6})$$

and

$$\Sigma_{\alpha,L;\alpha',L'}^>(k;E) = \frac{1}{V} \sum_{\mathbf{q}} |U_{\mathbf{k}-\mathbf{q}}|^2 e^{iq_z \Delta(L-L'+v_{\alpha,\alpha'})} \times [(n_{\mathbf{q}} + 1) G_{\alpha,L;\alpha',L'}^>(\mathbf{q}_t; E - \hbar \omega_{\mathbf{q}}) + n_{\mathbf{q}} G_{\alpha,L;\alpha',L'}^>(\mathbf{q}_t; E + \hbar \omega_{\mathbf{q}})], \quad (\text{A7})$$

where the indices  $\alpha$  and  $\alpha'$  run over the anion and cation states,  $n_{\mathbf{q}}$  is the Bose-Einstein factor for mode  $\mathbf{q}$ , and

$$\nu_{\alpha,\alpha'} = \begin{cases} \frac{1}{2} & \alpha=a, \alpha'=c \\ -\frac{1}{2} & \alpha=c, \alpha'=a \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A8})$$

In Eq. (A8), the indices  $a$  and  $c$  indicate an anion and cation orbital, respectively.

### a. Polar optical phonons

For dispersionless polar optical phonons,  $|U_{\mathbf{q}}|^2 = \beta q^2 / (q^2 + q_o^2)^2$  where  $\beta = e^2 \hbar \omega / 2(1/\epsilon_\infty - 1/\epsilon_o)$  and  $q_o$  is the inverse screening length.<sup>44</sup> If the Green functions are only functions of the magnitude of the transverse momentum such as in a single or two band<sup>45</sup> model, then the self energies can be written as

$$\begin{aligned} \Sigma_{\alpha,L;\alpha',L'}^>(k;E) &= \frac{\beta}{\pi} \int \frac{d^2 q_t}{4\pi^2} I(|L-L' + \nu_{\alpha,\alpha'}|, k, q_t) \\ &\quad \times [(n_B + 1) G_{\alpha,L;\alpha',L'}^>(q_t; E - \hbar\omega) \\ &\quad + n_B G_{\alpha,L;\alpha',L'}^>(q_t; E + \omega)] \end{aligned} \quad (\text{A9})$$

and

$$\begin{aligned} \Sigma_{\alpha,L;\alpha',L'}^<(k;E) &= \frac{\beta}{\pi} \int \frac{d^2 q_t}{4\pi^2} I(|L-L' + \nu_{\alpha,\alpha'}|, k, q_t) \\ &\quad \times [n_B G_{\alpha,L;\alpha',L'}^<(q_t; E - \hbar\omega) \\ &\quad + (n_B + 1) G_{\alpha,L;\alpha',L'}^<(q_t; E + \hbar\omega)], \end{aligned} \quad (\text{A10})$$

where  $n_B$  is the Bose-Einstein factor for energy  $\hbar\omega$ , and

$$\begin{aligned} I(|L-L' + \nu_{\alpha,\alpha'}|, k, q_t) &= \int_0^{\pi/a} dq_z \frac{\cos[q_z \Delta(L-L' + \nu_{\alpha,\alpha'})]}{\sqrt{(q_z^2 + q_t^2 + k^2 + q_o^2)^2 - 4k^2 q_t^2}} - \int_0^{\pi/a} dq_z \\ &\quad \times \frac{\cos[q_z \Delta(L-L' + \nu_{\alpha,\alpha'})] q_o^2 (q_z^2 + q_t^2 + k^2 + q_o^2)}{[(q_z^2 + q_t^2 + k^2 + q_o^2)^2 - 4k^2 q_t^2]^{3/2}}. \end{aligned} \quad (\text{A11})$$

For all of the self-energies to follow, we write the form for  $\Sigma^<$ . The expression for  $\Sigma^>$  and  $\Sigma^R$  are obtained by substituting everywhere the superscript ' $<$ ' with ' $>$ ' or ' $R$ ', respectively. However, the expressions for  $\Sigma^>$  will not be needed, because all of the following scattering mechanisms are, or will be approximated as, elastic.

### b. Acoustic phonons

The coupling for acoustic phonons is

$$|U_{\mathbf{q}}|^2 = \frac{\hbar D^2}{2\rho c} q, \quad (\text{A12})$$

where  $D$  is the deformation potential,  $\rho$  is the semiconductor density, and  $c$  is the velocity of sound in the material. If we make the usual assumptions of low energy (elastic)

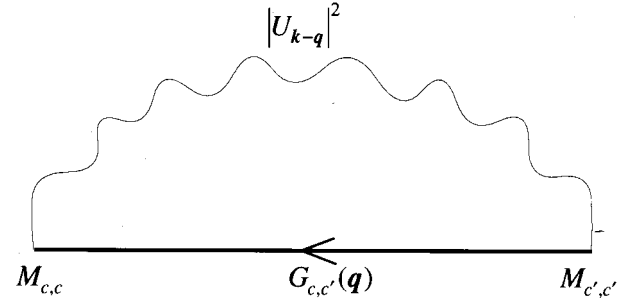


FIG. 8. Self-energy diagram  $\Sigma_{c,c'}$  for both alloy and interface roughness scattering.

scattering and high temperature<sup>44</sup> so that  $n_q \approx n_q + 1 \approx k_B T / \hbar \omega_q = k_B T / \hbar c q$  then the self-energy is

$$\Sigma_{\alpha,L;\alpha',L'}^<(E) = \delta_{L,L'} \zeta_{\alpha,\alpha'} \frac{D^2 k_B T}{\rho c^2 a} \int \frac{d^2 k}{4\pi^2} G_{\alpha,L;\alpha',L}^<(\mathbf{k}, E), \quad (\text{A13})$$

where

$$\zeta_{\alpha,\alpha'} = \begin{cases} 1 & \{\alpha, \alpha'\} = \{a, a'\} \text{ or } \{c, c'\} \\ 0 & \text{otherwise} \end{cases}. \quad (\text{A14})$$

The acoustic phonon self-energy is block diagonal in the anion and cation sub-blocks. The sub-blocks are full matrices.

## 2. Alloy disorder

We consider an alloy of the type  $A_x B_{1-x} C$  with a potential of the form

$$V(\mathbf{r}) = \sum_{\mathbf{R}_A} V_A(\mathbf{r} - \mathbf{R}_A) + \sum_{\mathbf{R}_B} V_B(\mathbf{r} - \mathbf{R}_B) + \sum_{\mathbf{R}_C} V_C(\mathbf{r} - \mathbf{R}_C). \quad (\text{A15})$$

We define an average potential

$$\begin{aligned} \bar{V}(\mathbf{r}) &= \sum_{\mathbf{R}_i \in \{\mathbf{R}_A \& \mathbf{R}_B\}} [x V_A(\mathbf{r} - \mathbf{R}_i) + (1-x) V_B(\mathbf{r} - \mathbf{R}_i)] \\ &\quad + \sum_{\mathbf{R}_C} V_C(\mathbf{r} - \mathbf{R}_C) \end{aligned} \quad (\text{A16})$$

and the perturbing potential

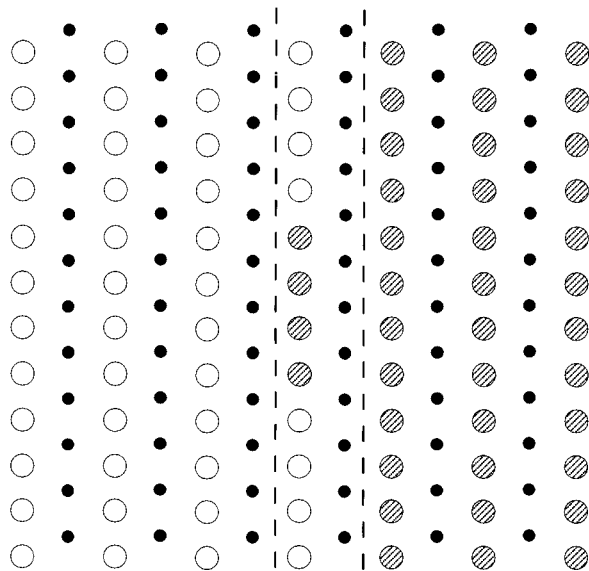
$$\begin{aligned} V_{\text{al}}(\mathbf{r}) &= V(\mathbf{r}) - \bar{V}(\mathbf{r}) \\ &= (1-x) \sum_{\mathbf{R}_A} \delta V(\mathbf{r} - \mathbf{R}_A) - x \sum_{\mathbf{R}_B} \delta V(\mathbf{r} - \mathbf{R}_B), \end{aligned} \quad (\text{A17})$$

where  $\delta V(\mathbf{r}) = V_A(\mathbf{r}) - V_B(\mathbf{r})$ . The matrix elements of the average (virtual crystal approximation) Hamiltonian are

$$\langle \alpha, L | H_{\text{VCA}} | \beta, L' \rangle = \langle \alpha, L | x H_{\text{AC}} + (1-x) H_{\text{BC}} | \beta, L' \rangle \quad (\text{A18})$$

and the matrix elements of  $\delta V(\mathbf{r})$  are

$$M_{\alpha,\alpha'} = \langle \alpha | \delta V(\mathbf{r}) | \alpha' \rangle = \langle \alpha | H_{\text{AC}} - H_{\text{BC}} | \alpha' \rangle, \quad (\text{A19})$$



Interface layer, L

Cation type B: ○  
 Cation type C: ●  
 Common anion, A: ●

FIG. 9. Model of interface roughness.

where  $H_{AC}$  and  $H_{BC}$  are the Hamiltonians of the binary compounds, AC and BC, respectively. For the following discussion, we assume that the alloy is of the cation species. For reasons of numerical tractability, we consider only the diagonal cation-cation matrix elements of  $\delta V(\mathbf{r})$  which accounts for the chemical variations and ignores the bond length variations in the calculation of the self-energy.<sup>46</sup>

The second quantized Hamiltonian representing the alloy disorder is

$$H_{\text{al}} = \int d^3r \psi^\dagger(\mathbf{r}) V_{\text{al}}(\mathbf{r}) \psi(\mathbf{r}). \quad (\text{A20})$$

Inserting  $H_{\text{al}}$  into Eq. (A1) and expanding out to second order, we are faced with evaluating  $\langle V_{\text{al}}(\mathbf{r}) V_{\text{al}}(\mathbf{r}') \rangle$  where the brackets indicate an ensemble average. In a way similar to the treatment of impurity averaging,<sup>47</sup> we define the ensemble average of a function of multiple random cation sites,  $\mathcal{F}(\mathbf{R}_1, \dots, \mathbf{R}_N)$ , as

$$\overline{\mathcal{F}} = \frac{1}{N_{\text{cat}}} \sum_{\mathbf{R}_1} \dots \frac{1}{N_{\text{cat}}} \sum_{\mathbf{R}_N} \mathcal{F}(\mathbf{R}_1, \dots, \mathbf{R}_N), \quad (\text{A21})$$

where the  $N_{\text{cat}}$  is the total number of cation sites and the sums run over all cation sites. With this definition,  $\overline{V}_{\text{al}}(\mathbf{r}) = 0$ , and

$$\langle V_{\text{al}}(\mathbf{r}) V_{\text{al}}(\mathbf{r}') \rangle = x(1-x) \sum_{\mathbf{R}_{\text{cat}}} \delta V(\mathbf{r} - \mathbf{R}_{\text{cat}}) \delta V(\mathbf{r}' - \mathbf{R}_{\text{cat}}), \quad (\text{A22})$$

where the sum runs over all cation sites. Note that unlike the treatment of impurity averaging, in which terms of order

$N^2$  are generated by the above procedure and ignored,<sup>47</sup> we have not ignored any terms in the evaluation of  $\langle V_{\text{al}}(\mathbf{r}) V_{\text{al}}(\mathbf{r}') \rangle$ , and Eq. (A22) is exact.

The self-energy for both alloy scattering and interface roughness scattering has the form shown in Fig. (8). The layer index has been suppressed. For alloy scattering,  $|U_{\mathbf{k}-\mathbf{q}}|^2 = 2\Delta^2$  which results from a factor of  $A/N_t$  where  $A$  is the cross sectional area and  $N_t = \sum_{\mathbf{R}_t}$ . The expression for the self-energy is

$$\Sigma_{c,L;c',L}^< = 2\Delta^2 x(1-x) M_{c,L;c,L} M_{c',L;c',L} \times \frac{1}{A} \sum_{\mathbf{q}} G_{c,L;c',L}^<(\mathbf{q}). \quad (\text{A23})$$

The self-energy matrix is non-zero only in the cation sub-block of the block diagonal, and, within that sub-block, it is a full matrix.

In the single band model, we follow Chevoir and Vinter to obtain the matrix element squared,  $M^2$ .<sup>7</sup> For example, with  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  we use the conduction band offset of the two different materials, e.g., the conduction band offset between GaAs and AlAs.

### 3. Interface roughness

We model the effect of substitutional disorder in an interface layer in the same way as substitutional disorder resulting from alloy composition. We confine the disorder to a single layer and, within that layer, the cations of a given type cluster into islands. This approach is valid for electron wavelengths large compared to the cluster sizes. Figure (9) illustrates our model.

The potential in the interface layer is

$$V(\mathbf{r}) = \sum_{\mathbf{R}_B} V_B(\mathbf{r} - \mathbf{R}_B) + \sum_{\mathbf{R}_C} V_C(\mathbf{r} - \mathbf{R}_C) + \sum_{\mathbf{R}_A} V_A(\mathbf{r} - \mathbf{R}_A). \quad (\text{A24})$$

$\mathbf{R}_B$  and  $\mathbf{R}_C$  are the position of the cations  $B$  and  $C$  in the interface layer,  $L$ . Let material  $B$  cover a fraction  $x$  of the interface and material  $C$  cover a fraction  $1-x$  of the interface. Then the average potential is

$$\begin{aligned} \overline{V}(\mathbf{r}) &= \sum_{\mathbf{R}=\{\mathbf{R}_B \& \mathbf{R}_C\}} [x V_B(\mathbf{r} - \mathbf{R}) + (1-x) V_C(\mathbf{r} - \mathbf{R})] \\ &+ \sum_{\mathbf{R}_A} V_A(\mathbf{r} - \mathbf{R}_A), \end{aligned} \quad (\text{A25})$$

where the sum over  $\mathbf{R}$  runs over all cation sites in layer  $L$ . The perturbing potential is

$$\begin{aligned} V_{\text{ir}}(\mathbf{r}) &= V(\mathbf{r}) - \overline{V}(\mathbf{r}) \\ &= (1-x) \sum_{\mathbf{R}_B} \delta V(\mathbf{r} - \mathbf{R}_B) - x \sum_{\mathbf{R}_C} \delta V(\mathbf{r} - \mathbf{R}_C) \\ &= \sum_{\mathbf{R}=\{\mathbf{R}_B \& \mathbf{R}_C\}} F(\mathbf{R}) \delta V(\mathbf{r} - \mathbf{R}), \end{aligned} \quad (\text{A26})$$

where

$$F(\mathbf{R}) = \begin{cases} 1-x & \mathbf{R}=\mathbf{R}_B \\ -x & \mathbf{R}=\mathbf{R}_C \end{cases} \quad (\text{A27})$$

The average of  $V_{\text{ir}}$  is zero and the autocorrelation of  $F(\mathbf{R})$  is

$$\langle F(\mathbf{R})F(\mathbf{R}') \rangle = x(1-x)A_F(|\mathbf{R}-\mathbf{R}'|). \quad (\text{A28})$$

We will consider both Gaussian and exponential forms for  $A_F(|\mathbf{R}-\mathbf{R}'|)$ . The problem is now reduced to one of a single layer of alloy in which the cations of a given species cluster into islands. The clustering gives rise to a momentum dependence in the Fourier transform of the autocorrelation function. The matrix elements of  $\delta V$  and the average (virtual crystal) Hamiltonian are the same as in Eqs. (A18) and (A19).

The interface roughness self-energy shown in Fig. (8) is

$$\begin{aligned} \Sigma_{c,L;c',L}^{\leq} &= x(1-x)M_{c,L;c,L}M_{c',L;c',L} \\ &\times \frac{1}{A} \sum_{\mathbf{q}} |U_{\mathbf{k}-\mathbf{q}}|^2 G_{c,L;c',L}^{\leq}(\mathbf{q}), \end{aligned} \quad (\text{A29})$$

where  $L$  is an interface layer.

Traditionally, a Gaussian form for  $A_F(|\mathbf{r}_t-\mathbf{r}'_t|)$  has been assumed<sup>6,7,48,49</sup>

$$A_F(|\mathbf{r}_t-\mathbf{r}'_t|) = e^{-|\mathbf{r}_t-\mathbf{r}'_t|^2/\Lambda^2} \quad (\text{A30})$$

resulting in a spectral density of

$$|U_{\mathbf{q}}|^2 = \pi\Lambda^2 e^{-\Lambda^2 q^2/4}. \quad (\text{A31})$$

If we can approximate  $G_{c_1,L;c_2,L}^{\leq}(\mathbf{k})$  as being dependent only on the magnitude of  $\mathbf{k}$ , then the self-energy in Eq. (A29) becomes

$$\begin{aligned} \Sigma_{c,L;c',L}^{\leq}(k,E) &= x(1-x)\pi\Lambda^2 e^{-\Lambda^2 k^2/4} M_{c,L;c,L} M_{c',L;c',L} \\ &\times \int \frac{d^2 q}{4\pi^2} I_0\left(\frac{\Lambda^2 k q}{2}\right) e^{-\Lambda^2 q^2/4} G_{c,L;c',L}^{\leq}(q;E), \end{aligned} \quad (\text{A32})$$

where  $I_0$  is the modified Bessel function,

$$I_0(x) = \frac{1}{\pi} \int_0^\pi d\theta e^{x \cos \theta}.$$

Recent scanning tunneling microscopy (STM) work indicates that an exponential autocorrelation may be more appropriate,<sup>50</sup>

$$A_F(|\mathbf{r}_t-\mathbf{r}'_t|) = e^{-|\mathbf{r}_t-\mathbf{r}'_t|/\Lambda} \quad (\text{A33})$$

with a spectral density of

$$|U_{\mathbf{q}}|^2 = \frac{2\pi\Lambda^2}{[1+(q\Lambda)^2]^{3/2}}. \quad (\text{A34})$$

If we can approximate  $G_{c_1,L;c_2,L}^{\leq}(\mathbf{k})$  as being dependent only on the magnitude of  $\mathbf{k}$ , then the self-energy in Eq. (A29) becomes

$$\begin{aligned} \Sigma_{c,L;c',L}^{\leq}(k,E) &= x(1-x)4\Lambda^2 M_{c,L;c,L} M_{c',L;c',L} \int \frac{d^2 q}{4\pi^2} \\ &\times \frac{E\left(\frac{\pi}{2}, \sqrt{\frac{2}{b+1}}\right) G_{c,L;c',L}^{\leq}(q;E)}{(1+\Lambda^2(q-k)^2)\sqrt{1+\Lambda^2(q+k)^2}}, \end{aligned} \quad (\text{A35})$$

where  $E$  is the complete elliptic integral of the second kind

$$E\left(\frac{\pi}{2}, x\right) = \int_0^{\pi/2} d\theta \sqrt{1-x^2 \sin^2 \theta}$$

and  $b = (1 + \Lambda^2 k^2 + \Lambda^2 q^2)/(2kq\Lambda^2)$ .

The approximations used in the single band model for alloy scattering discussed at the end of Subsec. (A 2) also apply to interface roughness scattering.

#### 4. Ionized impurities

The potential resulting from substitutional ionized dopants is

$$V(\mathbf{r}) = \sum_i v(\mathbf{r}-L_i\Delta\hat{z}-\mathbf{R}_i), \quad (\text{A36})$$

where  $v(\mathbf{r}-L_i\Delta\hat{z}-\mathbf{R}_i)$  is the potential resulting from an ionized dopant atom at lattice site  $L_i\Delta\hat{z}+\mathbf{R}_i$  and  $\mathbf{R}_i$  is the transverse component of the lattice vector. The lattice vector represents the position of a cation for n-type doping and the position of an anion for p-type doping. However, the difference in position of the anion and the cation is negligible compared to the screening length in semiconductors, so no distinction will be made between the two cases. The lattice vector is written in cylindrical form since the doping concentration is, in general, a function of the layer,  $L_i$ .

An ensemble average is performed over the transverse plane using Eq. (A21) to obtain

$$\begin{aligned} \langle V(\mathbf{r})V(\mathbf{r}') \rangle &= \sum_L \frac{N_L^I}{N_{\text{cat}}} \sum_{\mathbf{R}} v(\mathbf{r}-L\Delta\hat{z}-\mathbf{R}) \\ &\times v(\mathbf{r}'-L\Delta\hat{z}-\mathbf{R}). \end{aligned} \quad (\text{A37})$$

In Eq. (A37),  $N_L^I$  is the number of ionized dopants in layer  $L$ ,  $N_{\text{cat}}$  is the number of cations in layer  $L$ , and  $\mathbf{R}$  is the transverse component of a lattice vector.

Using a screened potential of the form

$$v(\mathbf{r}) = \frac{e^2 e^{-q_o r}}{4\pi\epsilon r}, \quad (\text{A38})$$

where  $q_o$  is the inverse screening length,  $r=|\mathbf{r}|$ , and the two-dimensional Fourier transform

$$\int d^2 R e^{-i\mathbf{k}\cdot\mathbf{R}} v(z, \mathbf{R}) = \frac{e^2}{2\epsilon} \frac{e^{-\sqrt{q_o^2+k^2}|z|}}{\sqrt{q_o^2+k^2}}, \quad (\text{A39})$$

the general form for the self-energy is

$$\Sigma_{\alpha_1, L_1; \alpha_2, L_2}^<(\mathbf{k}) = \frac{e^4}{8\epsilon^2} \sum_L n_s^L \frac{1}{A} \times \sum_{\mathbf{q}} \frac{e^{-\sqrt{q_o^2 + |\mathbf{k}-\mathbf{q}|^2} [|L_1-L| + |L_2-L|] \Delta}}{q_o^2 + |\mathbf{k}-\mathbf{q}|^2} \times G_{\alpha_1, L_1; \alpha_2, L_2}^<(\mathbf{q}), \quad (\text{A40})$$

where  $n_s^L = N_L^I/A$  is the sheet ionized dopant density in layer  $L$ .

There are two special cases that commonly occur. The first case is when the doping is confined to the leads and zero in the device. Let the doping concentration in the left lead,  $N_e$ , be uniform in the region from sites  $-\infty$  to  $L_e$  and zero elsewhere; let the doping concentration in the right lead,  $N_c$ , be uniform in the region from sites  $L_c$  to  $\infty$  and zero elsewhere; and any layer  $L$  in the device lies between  $L_e$  and  $L_c$ . Converting the sum over  $L$  in Eq. (A40) to an integral,

and integrating over the doped regions of the left and right leads, the self-energy becomes

$$\Sigma_{\alpha_1, L_1; \alpha_2, L_2}^<(\mathbf{k}) = \frac{e^4}{8\epsilon^2} \sum_{e/c} N_{e/c}^{3D} \frac{1}{A} \sum_{\mathbf{q}} \frac{e^{-\sqrt{q_o^2 + |\mathbf{k}-\mathbf{q}|^2} [|L_1-L_{e/c}| + |L_2-L_{e/c}|] \Delta}}{[q_o^2 + |\mathbf{k}-\mathbf{q}|^2]^{3/2}} \times G_{\alpha_1, L_1; \alpha_2, L_2}^<(\mathbf{q}), \quad (\text{A41})$$

where  $N_{e/c}^{3D}$  is the doping density (1/volume) of the left/right lead.

The second case is when the doping density,  $N^{3D}$ , is uniform throughout the device region and out to some layer  $L_e$  in the left lead and  $L_c$  in the right lead. The self energy is then

$$\Sigma_{\alpha_1, L_1; \alpha_2, L_2}^<(\mathbf{k}) = \frac{e^4}{8\epsilon^2} N^{3D} \frac{1}{A} \sum_{\mathbf{q}} \frac{e^{-\beta |L_1-L_2| \Delta} [1 + 2\beta |L_1-L_2| \Delta] - e^{-\beta (L_1+L_2-2L_e) \Delta} - e^{-\beta (2L_c-L_1-L_2) \Delta}}{\beta^3} \cdot G_{\alpha_1, L_1; \alpha_2, L_2}^<(\mathbf{q}) \quad (\text{A42})$$

where  $\beta = \sqrt{q_o^2 + |\mathbf{k}-\mathbf{q}|^2}$ .

## APPENDIX B: DETAILED BALANCE AND CURRENT CONSERVATION FOR MULTIPLE SEQUENTIAL SCATTERING

### 1. Detailed balance

To show the violation of detailed balance, we use the single sequential scattering example of Sec. (VI A 1) in a single band model and start with the current expression given by Eq. (45).

$$J_0(E) = \frac{2e}{\hbar A} \sum_k \Gamma_{1,1}^B(k) \times [f_L A_{0,1,1}(k) + i(G_{0,1,1}^<(k) + G_{1,1,1}^<(k))]. \quad (\text{B1})$$

Note that a term  $f_L A_{1,1,1}$  does not appear since there is no contact source term on the right hand side of Eq. (83). One immediately sees that if the usual equilibrium relation holds,  $G^< = G_0^< + G_1^< = i f A$ , then detailed balance holds, and the equilibrium current is zero for each momentum and energy.

In the following, since we are considering equilibrium, we drop the subscript on  $f$  since  $f_L = f_R = f$  and write out the expressions for  $A_{0,1,1}$ ,  $i G_{0,1,1}^<$ , and  $i G_{1,1,1}^<$ . The expressions needed for writing down  $A_0$  are, in full matrix notation,

$$A_0 = G_0^R (\Gamma^B + \Gamma_1) G_0^A, \quad (\text{B2})$$

where  $\Gamma_1 = i(\sigma_1^R - \sigma_1^A) = D \otimes A_1$ , and  $A_1 = G_1^R \Gamma^B G_1^A$ . Writing out the indices of Eq. (B2) results in

$$A_{0,1,1}(k) = \Gamma_{1,1}^B(k) |G_{0,1,1}^R(k)|^2 + \Gamma_{N,N}^B(k) |G_{0,1,N}^R(k)|^2 + \sum_j |G_{0,1,j}^R(k)|^2 \sum_{k'} D_j(k, k') (\Gamma_{1,1}^B(k') |G_{1,1,1}^R(k')|^2 + \Gamma_{N,N}^B(k') |G_{1,j,N}^R(k')|^2). \quad (\text{B3})$$

Writing out the indices of Eq. (85), the equilibrium expression for  $i G_{1,1}^< = i(G_{0,1,1}^< + G_{1,1,1}^<)$  is

$$i G_{1,1}^<(k) = -f (\Gamma_{1,1}^B(k) |G_{0,1,1}^R(k)|^2 + \Gamma_{N,N}^B(k) |G_{0,1,N}^R(k)|^2) - f \sum_j |G_{1,1,j}^R(k)|^2 \sum_{k'} D_j(k, k') (\Gamma_{1,1}^B(k') |G_{0,j,1}^R(k')|^2 + \Gamma_{N,N}^B(k') |G_{0,j,N}^R(k')|^2). \quad (\text{B4})$$



By inspection, one can see that for the sum  $fA_0(k) + iG^<(k)$  to equal zero for each momentum  $k$ ,  $G_0^R$  must be identical to  $G_1^R$ . On the other hand, if this were the case, then current would not be conserved in non-equilibrium.

For perfectly symmetric devices, the total equilibrium current is zero. Substituting the equilibrium expressions for  $A_{0,1}$  and  $G_{1,1}^<$  back into the current expression, Eq. (B1), we obtain

$$J_0(E) = \frac{2e}{\hbar A} f \sum_j \sum_{k,k'} D_j(k,k') \Gamma_{1,1}^B(k) \Gamma_{N,N}^B(k') \\ \times [|G_{0,1,j}^R(k)|^2 |G_{1,j,N}^R(k')|^2 \\ - |G_{1,1,j}^R(k)|^2 |G_{0,j,N}^R(k')|^2] \quad (\text{B5})$$

For a perfectly symmetric device, where  $G_{1,j} = G_{N,N-j+1}$  and  $D_j = D_{N-j+1}$ , the equilibrium current summed over all transverse momenta is zero.

## 2. Current conservation

To show current conservation, we use the single sequential scattering example of Sec. (VI A 1) with a single-band model and local self-energies. We define  $J_0$  as the current from  $G_0^<$  and  $J_1$  as the current from  $G_1^<$ . Then from Eq. (67)

$$\sum_k [J_0(k) - J_{0-1}(k)] \\ = \frac{2e}{\hbar A} \sum_k [\sigma_1^R(k) G_0^<(k) - G_0^<(k) \sigma_1^A(k)] \\ = \frac{2e}{\hbar A} \sum_k G_0^<(k) (\sigma_1^R(k) - \sigma_1^A(k)) \\ = -i \frac{2e}{\hbar A} \sum_k G_0^<(k) \sum_{k'} D(k,k') A_1(k'), \quad (\text{B6})$$

where, since the elastic scattering is local, the notation  $\sigma G$  represents scalar multiplication of the diagonal elements at layer  $i$  and  $A_1 = -i(G_1^R - G_1^A)$  is the spectral function corresponding to  $G_1^R$ . Only two of the terms from Eq. (67) are present since there is no in scattering source term,  $\sigma^<$ , on the right hand side of Eq. (82). The divergence of  $J_1$  is

$$\sum_k [J_1(k) - J_{1-1}(k)] \\ = \frac{2e}{\hbar A} \sum_k [\sigma_0^<(k) G_1^A(k) - G_1^R(k) \sigma_0^<(k)] \\ = \frac{2e}{\hbar A} \sum_k [(G_1^A(k) - G_1^R(k)) \sigma_0^<(k)] \\ = i \frac{2e}{\hbar A} \sum_k \left[ A_1(k) \sum_{k'} D(k,k') G_0^<(k') \right]. \quad (\text{B7})$$

Again, only two terms are present since there is no out scattering term,  $\sigma^R$ , on the left side of Eq. (81). Since  $D(k,k') = D(k',k)$ , the sum of Eqs. (B6) and (B7) is zero

and current is conserved. The out scattered current from  $J_0$  is precisely the in scattered current into  $J_1$ , and there is no further out scattering from  $J_1$ . If we generalized this to multiple sequential scattering, we would see that the out scattered current from  $J_1$  would be the in scattered current into  $J_2$ , etc.

## APPENDIX C: SELF-CONSISTENT CALCULATION OF THE ELECTROSTATIC POTENTIAL

The quantum mechanical calculation of the charge is used in an iterative self-consistent solution of the electrostatic potential which is the Hartree potential. We use a Newton-Raphson method with a semiclassical form for the Jacobian. We begin by discretizing Poisson's equation

$$\frac{d}{dz} \epsilon(z) \frac{d}{dz} \phi(z) + q [N_D^+(z) - N_A^-(z) - n(z)] = 0 \quad (\text{C1})$$

to obtain

$$\frac{1}{a^2} (\phi_{i-1} \epsilon^- - \phi_i (\epsilon^- + \epsilon^+) + \phi_{i+1} \epsilon^+) + q [N_{D_i}^+ - N_{A_i}^- - n_i] \\ = 0, \quad (\text{C2})$$

where  $\phi_i$  is the electrostatic potential at layer  $i$ ,  $\epsilon^+ = (\epsilon_i + \epsilon_{i+1})/2$ ,  $\epsilon^- = (\epsilon_i + \epsilon_{i-1})/2$ ,  $\epsilon_i$  is the dielectric function at layer  $i$ ,  $N_{D_i}^+$  is the ionized donor concentration at layer  $i$ ,  $N_{A_i}^-$  is the ionized acceptor concentration at layer  $i$ ,  $n_i$  is electron density at layer  $i$  calculated quantum mechanically from Eq. (18), and  $q$  is the magnitude of the electron charge. We have, so far, assumed complete ionization of the dopants, although it would not be difficult to include an ionization model.

Defining  $F_i$  as the left hand side of Eq. (C2), we solve

$$\sum_j \frac{\partial F_i^m}{\partial \phi_j^m} \delta \phi_j^{m+1} = -F_i^m, \quad (\text{C3})$$

where  $i$  and  $j$  are the layer indices and  $m$  is the iteration index. The new value of  $\phi$  is then  $\phi^{m+1} = \phi^m + \delta \phi^{m+1}$ . For the calculation of  $\partial n_i / \partial \phi_j$  contained in the Jacobian,  $\partial F_i^m / \partial \phi_j^m$ , the semi-classical form is used,

$$\frac{\partial n_i}{\partial \phi_j} = \delta_{i,j} \frac{q}{k_B T} N_c \mathcal{F}_{-1/2} \left( \frac{E_{F_i} - E_{c_i} + q \phi_i}{k_B T} \right) \quad (\text{C4})$$

after first finding the quasi-Fermi level  $E_{F_i}$  by solving

$$n_i = N_c \mathcal{F}_{1/2} \left( \frac{E_{F_i} - E_{c_i} + q \phi_i}{k_B T} \right), \quad (\text{C5})$$

where  $n_i$  is the quantum charge calculated from Eq. (18). In Eqs. (C4) and (C5)  $N_c$  is the effective conduction band density of states,  $\mathcal{F}_j$  is the Fermi-Dirac integral of order  $j$ , and  $E_{c_i}$  is the energy of the conduction band edge at layer  $i$  (not including the electrostatic potential).<sup>51</sup> The quasi-Fermi levels at the boundaries are fixed by the applied potential, and the electrostatic potential at the boundaries is fixed with respect to the quasi-Fermi levels to ensure charge neutrality.

The semiclassical Jacobian has worked well leading to a converged quantum self-consistent solution,  $F_i^m = 0$ , in ap-

proximately three to ten iterations. One of the reasons that it has worked so well is that our boundary conditions fill the low notch states in the emitter lead in RTD simulations. This is similar to the model of Onishi *et al.*,<sup>52</sup> except that we calculate the quantum states in the lead exactly and then fill them according to the Fermi-Dirac factor of the lead as described by Eqs. (58) and (18).

- <sup>1</sup>G. Klimeck, R. Lake, R. C. Bowen, W. R. Frensley, and T. Moise, *Appl. Phys. Lett.* **67**, 2539 (1995).
- <sup>2</sup>D. Landheer and G. C. Aers, *Superlattices Microstruct.* **7**, 17 (1990).
- <sup>3</sup>W. R. Frensley, in *Nanostructure Physics and Fabrication*, edited by M. Reed and W. P. Kirk (Academic, New York, 1989), pp. 231–240.
- <sup>4</sup>R. Lake and S. Datta, *Phys. Rev. B* **45**, 6670 (1992).
- <sup>5</sup>N. Zou and K. A. Chou, *Phys. Rev. Lett.* **69**, 3224 (1992).
- <sup>6</sup>P. Roblin and W. Liou, *Phys. Rev. B* **46**, 2416 (1993).
- <sup>7</sup>F. Chevoir and B. Vinter, *Phys. Rev. B* **47**, 7260 (1993).
- <sup>8</sup>C. Caroli, R. Combescot, P. Nozieres, and D. Saint-James, *J. Phys. C* **4**, 196 (1971).
- <sup>9</sup>R. Lake, in *Proceedings of the Third International Workshop on Computational Electronics*, Portland, OR, May 18–20, 1994 (unpublished), pp. 239–242.
- <sup>10</sup>R. Lake, in *Quantum Transport in Ultrasmall Devices, NATO ASI Series*, edited by D. K. Ferry, H. L. Grubin, C. Jacoboni, and A. Jauho (Plenum, New York, 1995), Vol. 342, pp. 521–524.
- <sup>11</sup>R. Lake, G. Klimeck, R. C. Bowen, D. Jovanovic, P. Sotirelis, and W. R. Frensley (to appear in *VLSI Design*).
- <sup>12</sup>T. Fiig and A. Jauho, *Surf. Sci.* **267**, 392 (1992).
- <sup>13</sup>G. Klimeck, R. Lake, R. C. Bowen, W. R. Frensley, and D. Blanks, in the *1995 53rd Annual Device Research Conference Digest* (Institute of Electrical and Electronics Engineering, Inc., NJ, 1995), p. 52.
- <sup>14</sup>R. Lake, G. Klimeck, R. C. Bowen, C. Fernando, D. Jovanovic, D. Blanks, T. S. Moise, Y. C. Kao, M. Leng, and W. R. Frensley in the *1996 54th Annual Device Research Conference Digest* (Institute of Electrical and Electronics Engineering, NJ, 1996), p. 174.
- <sup>15</sup>R. Lake, G. Klimeck, R. C. Bowen, C. Fernando, M. Leng, T. Moise, and Y. C. Kao, *Superlattices Microstruct.* **20**, 279 (1996).
- <sup>16</sup>G. Klimeck, R. Lake, C. L. Fernando, R. C. Bowen, D. Blanks, M. Leng, T. Moise, Y. C. Kao, and W. R. Frensley, in *Quantum Devices and Circuits*, edited by K. Ismail, S. Bandyopadhyay, and J. P. Leburton (Imperial College Press, London, 1997), p. 154.
- <sup>17</sup>R. C. Bowen, G. Klimeck, R. K. Lake, W. R. Frensley, and T. Moise, *J. Appl. Phys.* **81**, 3207 (1997).
- <sup>18</sup>D. Z.-Y. Ting, S. K. Kirby, and T. C. McGill, *Appl. Phys. Lett.* **64**, 2004 (1994).
- <sup>19</sup>J. A. Stovneng and P. Lipavsky, *Phys. Rev. B* **49**, 164 594 (1994).
- <sup>20</sup>T. B. Boykin and J. S. Harris, *J. Appl. Phys.* **72**, 988 (1992).
- <sup>21</sup>M. Jaffe and J. Singh, *Solid State Commun.* **62**, 399 (1987).
- <sup>22</sup>W. R. Frensley, in *Heterostructures and Quantum Devices*, edited by N. Einspruch and W. R. Frensley (Academic, New York, 1994), pp. 273–303.
- <sup>23</sup>J. A. Stovneng and E. H. Hauge, *Phys. Rev. B* **44**, 13 582 (1991).
- <sup>24</sup>E. V. Anda and F. Flores, *J. Phys., Condens. Matter.* **3**, 9087 (1991).
- <sup>25</sup>M. A. Davidovich, E. V. Anda, C. Tejedor, and G. Platero, *Phys. Rev. B* **47**, 4475 (1993).
- <sup>26</sup>S. Hershfield, J. Davies, and J. Wilkins, *Phys. Rev. B* **46**, 7046 (1992).
- <sup>27</sup>Y. Meir and N. Wingreen, *Phys. Rev. Lett.* **68**, 2512 (1992).
- <sup>28</sup>G. Kim, H. Suh, and E. Lee, *Phys. Rev. B* **52**, 2632 (1995).
- <sup>29</sup>D. C. Langreth, in *Linear and Non-linear Electron Transport in Solids, NATO ASI Series*, edited by J. T. Devreese and E. van Doren (Plenum, New York, 1976), Vol. 17, pp. 3–32.
- <sup>30</sup>R. C. Bowen, W. R. Frensley, G. Klimeck, and R. K. Lake, *Phys. Rev. B* **52**, 2754 (1994).
- <sup>31</sup>J. S. Wu, C. Y. Chang, C. P. Lee, K. H. Chang, and D. G. Liu, *Solid-State Electron.* **34**, 403 (1991).
- <sup>32</sup>D. S. Fisher and P. A. Lee, *Phys. Rev. B* **23**, 6851 (1981).
- <sup>33</sup>R. Tsu and L. Esaki, *Appl. Phys. Lett.* **22**, 562 (1973).
- <sup>34</sup>C. Caroli, R. Combescot, P. Nozieres, and D. Saint-James, *J. Phys. C* **5**, 21 (1972).
- <sup>35</sup>F. Sols, *Ann. Phys.* **214**, 386 (1992).
- <sup>36</sup>N. S. Wingreen, K. W. Jakobsen, and J. W. Wilkins, *Phys. Rev. B* **40**, 11 834 (1989).
- <sup>37</sup>C. H. Grein, E. Runge, and H. Ehrenreich, *Phys. Rev. B* **47**, 12 590 (1993).
- <sup>38</sup>P. Hyldgaard, S. Hershfield, J. H. Davies, and J. W. Wilkins, *Ann. Phys.* **236**, 1 (1994).
- <sup>39</sup>L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Addison-Wesley, New York, 1989).
- <sup>40</sup>S. Datta, *J. Phys., Condens. Matter.* **2**, 8023 (1990).
- <sup>41</sup>P. Johansson, *Phys. Rev. B* **48**, 8938 (1993).
- <sup>42</sup>S. Hershfield, *Phys. Rev. Lett.* **70**, 2134 (1993).
- <sup>43</sup>R. A. Craig, *J. Math. Phys. (N.Y.)* **9**, 605 (1968).
- <sup>44</sup>B. K. Ridley, *Quantum Processes in Semiconductors*, 2nd ed. (Oxford University Press, New York, 1988).
- <sup>45</sup>G. Bastard, *Phys. Rev. B* **25**, 7584 (1982).
- <sup>46</sup>R. J. Lempert, K. C. Hass, and H. Ehrenreich, *Phys. Rev. B* **36**, 1111 (1987).
- <sup>47</sup>S. Doniach and E. H. Sondheimer, *Green's Functions for Solid State Physicists* (Benjamin, Reading, MA, 1974).
- <sup>48</sup>T. Ando, A. B. Fowler, and F. Stern, *Rev. Mod. Phys.* **54**, 437 (1982).
- <sup>49</sup>G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures* (Halstead, New York, 1988).
- <sup>50</sup>R. M. Feenstra, D. A. Collins, D. Z.-Y. Ting, M. W. Wang, and T. C. McGill, *Phys. Rev. Lett.* **72**, 2749 (1994).
- <sup>51</sup>J. S. Blakemore, *Semiconductor Statistics* (Pergamon, New York, 1962).
- <sup>52</sup>H. Ohnishi, T. Inata, S. Muto, N. Yokoyama, and A. Shibatomi, *Appl. Phys. Lett.* **49**, 1248 (1986).