Semiconductor-heterostructure-interface connection rules

Claudio Aversa and J. E. Sipe
Department of Physics and Ontario Laser and Lightwave Research Center, University of Toronto, Toronto, Ontario, Canada M5S 1A7
(Received 7 August 1992)

We introduce a technique for the study of semiconductor heterostructures which is more general than the usual envelope-function approximation. This scheme accounts for material dependences in the Bloch functions, and also allows a full zone treatment of the constituent semiconductors. However, it involves only small-dimensional matrices, and thus remains computationally efficient. The presented formalism offers, at each energy, what can be considered a "best choice" flux-conserving interface connection rule between two semiconductors. As an illustration of the technique we consider the Γ-X scattering that occurs in GaAs/AlAs heterostructures.

I. INTRODUCTION

Advances in growth technology have permitted the fabrication of heterostructures composed of thin layers of different semiconductors with nearly abrupt interfaces. The flexibility in the choice of materials and layer widths allows one to virtually engineer the electronic and optical response of these heterostructures. A reliable theoretical prediction of such responses requires a physically reasonable quantum mechanical model to describe the nature of the relevant electronic states. In this article we present a scheme for the description of these states.

In their review of the theory of the electronic structure of semiconductor heterostructures, Smith and Mailhiot[1] point out that the simplest approximate models can be classified as empirical boundary condition techniques, in which we are interested here. Since in these models the heterostructure problem is divided into two parts, a treatment of the bulk constituent semiconductors followed by the matching of wave functions across the interfaces, small-dimensional matrices are typically involved. In this paper we are primarily concerned with the appropriate semiconductor heterostructure interface connection rules that define the wave-function matching. From the point of view of boundary condition techniques, the basic building block of any semiconductor heterostructure is the heterojunction formed by two distinct materials. Hence, we organize our presentation in this paper around this elemental unit; a successful model of the heterojunction allows more complicated geometries to be easily investigated.[2]

Consider the abrupt heterojunction formed by growing material B, in the z direction, on a substrate of material A. We wish to solve for the wave function at a given energy, E, and a given value of crystal momentum parallel to the interface (xy plane), $\mathbf{k}_{||}$. The envelope function approximation (EFA) is a popular approach used to treat such heterostructures.[3] In the EFA, the wave function in material α is written in a fashion similar to the bulk $\mathbf{k}\cdot\mathbf{p}$ formalism,

$$\psi_\alpha(\mathbf{r}) = e^{i\mathbf{k}_{||}\cdot\mathbf{r}} \sum_{n=1}^{N} f_\alpha^n(z) u_{n0}^\alpha(\mathbf{r}), \tag{1}$$

where the functions $u_{n0}^\alpha(\mathbf{r})$ are a set of bulk Bloch functions at $\mathbf{k}=0$. The unknown expansion coefficients, f_α^n, now z dependent, are the so-called "envelope functions." The main objective of envelope function schemes is to derive effective Schrödinger equations for the envelope functions in each material, along with a corresponding set of effective boundary conditions that these functions and their derivatives satisfy across the heterojunction.

There are two main assumptions adopted in the EFA which greatly simplify this procedure. First, the periodic basis functions $u_{n0}^\alpha(\mathbf{r})$ in Eq. (1) are assumed to be the same in each material. Secondly, the envelope functions are assumed to vary slowly over a crystal lattice period. These assumptions allow one to obtain effective boundary conditions across the interface which are consistent with the physical requirement of flux conservation.[3] While these approximations appear to be well founded for many heterostructures in energy ranges of interest, there are situations which are not easily tractable with such a model. In particular, the case of Γ-X coherent scattering in certain GaAs/AlAs heterostructures has not been rigourously formulated within a simple EFA type scheme.[1,4,5] Further, it would be convenient to have a simple model to investigate the implications of material differences in the $u_{n0}^\alpha(\mathbf{r})$, particularly as novel material systems arise.[6]

In this paper, we present a scheme which does not involve the assumptions inherent to the EFA, while still providing flux-conserving interface boundary conditions. For a fixed number of bulk basis functions in each material, this scheme supplies, at each energy, what we argue are the best interface connection rules for the abrupt interface. The formalism is independent of the particular method one adopts to describe the bulk band structure. Further, the scheme remains computationally efficient, since it can involve only small-dimensional matrices, typically of the same size as those in the EFA technique.

In Sec. II we summarize the general formalism used to treat an abrupt interface,[1] highlighting the typical problem which arises regarding the lack of flux conservation across the interface. In Sec. III we recast the problem in the language of scattering theory. This more easily exposes the origin of the flux-conservation problem, and also suggests an approximate flux-conserving scheme which can be followed. Finally, in Sec. IV we present some sample calculations, mainly to illustrate that this simple flux-conserving scheme can easily reproduce phenomena expected from more computationally intensive calculations.

II. THE ABRUPT INTERFACE: GENERAL FORMALISM

Following Smith and Mailhiot,[1] we write the wave function for the abrupt heterojunction formed by materials A and B, with interface at $z = z_0$, as

$$\psi(\mathbf{r}) = \begin{cases} \psi_a(\mathbf{r}), & z < z_0 \\ \psi_b(\mathbf{r}), & z > z_0, \end{cases} \tag{2}$$

with the most general wave function in each material written as a linear superposition of bulk Bloch waves. For example, in material A we have

$$\psi_a(\mathbf{r}) = \sum_{j=1}^{2m} F_a^j \psi_a^j(\mathbf{r}), \tag{3}$$

where the bulk Bloch waves are treated as our known basis and take the familiar form

$$\psi_a^j(\mathbf{r}) = e^{ik_j^a z} e^{i\mathbf{k}_{||}\cdot\mathbf{r}} u_j^a(\mathbf{r}). \tag{4}$$

The $u_j^a(\mathbf{r})$ are periodic functions, and the k_j^a are in general complex. For each bulk Bloch wave which can be considered traveling towards the interface, there always exists a wave traveling away from the interface. Expanding over $2m$ such waves in each material leaves the expansion coefficients F_a^j, and the corresponding F_b^j in material B, as the only unknowns. These unknowns are determined by satisfying the appropriate boundary conditions.

We note at this point that the variables F_α^j of the general formalism are always formally related to the unknowns in the EFA.[7] Specifically, carefully comparing Eq. (1) with Eqs. (3) and (4) reveals a bulk matrix W_α connecting the envelope functions and their derivatives, in material α, to the F_α^j,

$$\begin{pmatrix} f_\alpha \\ \frac{df_\alpha}{dz} \end{pmatrix} = W_\alpha F_\alpha, \tag{5}$$

in obvious vector notation. Thus, one set of unknowns determines the other. In particular, if we adopt a rule connecting F_a and F_b across the interface, we have also determined the relevant connection rule for the envelope functions; the decision to work with Eq. (1) or Eq. (3) is simply a choice of basis. We choose to work with Eq. (3), since we feel that those bulk Bloch waves are the more natural basis. For example, far from the interface only a few of the F_α^j are relevant, as the great majority of the $\psi_\alpha^j(\mathbf{r})$ decay rapidly. In contrast, all of the envelope functions are important, even infinitely far from the interface, since they must account for the bulk band mixing of the $\mathbf{k} = 0$ levels away from $\mathbf{k} = 0$, a property already incorporated into the $\psi_\alpha^j(\mathbf{r})$ basis. In addition, the current density operator, from which we construct a scattering theory in Sec. III, assumes a very simple form in the $\psi_\alpha^j(\mathbf{r})$ basis.

The boundary conditions across the interface require the continuity of the wave function and its normal derivative at $z = z_0$, which can be written compactly as[8]

$$J_z(x, y, z_0)|\psi_a\rangle = J_z(x, y, z_0)|\psi_b\rangle \tag{6}$$

using the z component of the current density operator,

$$J_z(\mathbf{r}) \equiv \frac{1}{2m} \left(|\mathbf{r}\rangle\langle\mathbf{r}|p_z + p_z|\mathbf{r}\rangle\langle\mathbf{r}| \right). \tag{7}$$

Equation (6) must be satisfied at each point (x, y) on the interface, implying an infinite number of equations. However, to address the problem in practice we have limited ourselves to only $4m$ unknowns, namely the F_a^j and F_b^j. Further, there are always $2m$ boundary conditions that must be satisfied at $z = \pm\infty$. Hence, the "best" we can do is to reduce Eq. (6) to $2m$ equations. This can be achieved by performing some sort of projection onto a set of $2m$ states. A natural projection is to multiply by *either* the $\langle\psi_a^j|$ *or* the $\langle\psi_b^j|$ and to integrate over the interface.

In either case, this natural projection leads to a transfer matrix connecting the unknowns across the interface. In particular, projecting onto the set of $\langle\psi_a^j|$ results in a transfer matrix T_1,

$$F_a = T_1 F_b, \tag{8a}$$

whereas projecting onto the $\langle\psi_b^j|$ provides a different transfer matrix T_2,

$$F_a = T_2 F_b. \tag{8b}$$

These matrices have been identified previously[1] and can be explicitly written as

$$T_1 = (J_{aa})^{-1} J_{ab}, \tag{9a}$$

$$T_2 = (J_{ba})^{-1} J_{bb}, \tag{9b}$$

with

$$(J_{\alpha\beta})_{ij} \equiv \langle\psi_\alpha^i|J|\psi_\beta^j\rangle, \tag{10}$$

where

$$J \equiv \iint J_z(x, y, z_0)\, dx\, dy \tag{11}$$

is a Hermitian operator. These matrices are easily calculated once the bulk Bloch waves are obtained from a bulk band-structure calculation.

The main difficulty with the approach summarized above is that the matrices T_1 and T_2 are not equal, which

seems to imply an ambiguity or inconsistency. This is, in fact, a manifestation of the lack of formal flux conservation across the interface in this general formalism.[1] To see this, note first that the conservation of flux across the interface, namely the condition

$$\langle\psi_a|J|\psi_a\rangle = \langle\psi_b|J|\psi_b\rangle, \tag{12}$$

requires that

$$F_a^+ J_{aa} F_a = F_b^+ J_{bb} F_b. \tag{13}$$

Flux conservation thus implies a condition that any proposed matrix T must satisfy if it is to be a physically meaningful transfer matrix. Namely, if $F_a = TF_b$, we should have

$$T^+ J_{aa} T = J_{bb}. \tag{14}$$

Neither T_1 nor T_2 as constructed above satisfies this equation in general. In fact, substituting T_1 or T_2 into Eq. (14) reveals that flux conservation requires $T_1 = T_2$, which is not guaranteed in the general formalism presented above. Therefore, even if the bulk materials could be perfectly described, so that the $2m\times 2m$ matrices T_1 and T_2 defined in Eqs. (9a) and (9b) were contructed exactly, we could not guarantee flux conservation in a model using T_1 or T_2 as the interface connection rule. In the next section, we recast this problem in a framework that elucidates why this somewhat curious result arises.

III. A SCATTERING THEORY

A. Revised formalism

We first choose a convenient normalization for the bulk Bloch waves. Consider the bulk matrices J_{aa} and J_{bb}, which have a particularly simple form,[8]

$$(J_{\alpha\alpha})_{ij} \equiv \langle\psi_\alpha^i|J|\psi_\alpha^j\rangle = (J_\alpha^i)\delta_{k_i,k_j^*}. \tag{15}$$

We choose to normalize the $|\psi_\alpha^j\rangle$ in each material so that $J_\alpha^i = \pm 1$. The bulk Bloch basis waves $\psi_\alpha^j(\mathbf{r})$ corresponding to purely real k_j values (propagating waves) will then carry "unit flux current," with the sign of J_α^i indicating the direction of that current flow. One advantage of this particular normalization is that the simple product $(F_\alpha^j)^* F_\alpha^j$ gives the magnitude of flux current carried by the propagating wave $\psi_\alpha^j(\mathbf{r})$, a quantity often of interest in open quantum structures.

We now proceed to transform to a new basis which will ultimately simplify the flux conservation expression. First note that, since the matrices J_{aa} and J_{bb} are Hermitian, we can diagonalize them with a unitary transformation. That is,

$$U_\alpha J_{\alpha\alpha} U_\alpha^+ = D, \tag{16}$$

where D is a diagonal matrix of the eigenvalues of $J_{\alpha\alpha}$. From the form of $J_{\alpha\alpha}$ and the normalization condition, it follows that half of the eigenvalues of the $2m \times 2m$ matrix $J_{\alpha\alpha}$ are -1 while the other half are $+1$; thus, we can always organize the basis of which U_α consists such that D has the following block form:

$$D = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \tag{17}$$

where I is the $m\times m$ unit matrix. The unitary matrix U_α, which is easily found analytically, identifies a transformed set of variables in each material, labeled as $\tilde{F}_\alpha$ and given by

$$\tilde{F}_\alpha = U_\alpha F_\alpha. \tag{18}$$

In terms of these new variables, the flux conservation condition, Eq. (13), takes the form

$$\tilde{F}_a^+ D \tilde{F}_a = \tilde{F}_b^+ D \tilde{F}_b. \tag{19}$$

The simple form of D now makes it convenient to decompose the new vectors $\tilde{F}_a$ and $\tilde{F}_b$ into two vectors of equal length by writing

$$\tilde{F}_a = \begin{pmatrix} \tilde{F}_a^i \\ \tilde{F}_a^o \end{pmatrix}, \qquad \tilde{F}_b = \begin{pmatrix} \tilde{F}_b^o \\ \tilde{F}_b^i \end{pmatrix}. \tag{20}$$

This allows Eq. (19) to be written in a form involving only vector products,

$$(\tilde{F}_a^i)^+ \tilde{F}_a^i - (\tilde{F}_a^o)^+ \tilde{F}_a^o = (\tilde{F}_b^o)^+ \tilde{F}_b^o - (\tilde{F}_b^i)^+ \tilde{F}_b^i. \tag{21}$$

Here the vectors $\tilde{F}_a^i$ and $\tilde{F}_b^i$ can be considered the amplitudes of waves carrying flux towards (into) the interface, while $\tilde{F}_a^o$ and $\tilde{F}_b^o$ determine the flux scattered away (out). It is then natural to define the composite vectors v_i and v_o,

$$v_i = \begin{pmatrix} \tilde{F}_a^i \\ \tilde{F}_b^i \end{pmatrix}, \qquad v_o = \begin{pmatrix} \tilde{F}_a^o \\ \tilde{F}_b^o \end{pmatrix}. \tag{22}$$

It is in terms of these new vectors that the flux conservation condition assumes its simplest form,

$$v_o^+ v_o = v_i^+ v_i, \tag{23}$$

which associates with the scattering process a vector the length of which is conserved.

We next note that we can identify a new transfer matrix $\tilde{T}$ for any proposed transfer matrix T connecting F_a and F_b,

$$\tilde{T} = U_a T U_b^+ = \begin{pmatrix} \tilde{t}_{11} & \tilde{t}_{12} \\ \tilde{t}_{21} & \tilde{t}_{22} \end{pmatrix}, \tag{24}$$

which connects the transformed variables, so that $\tilde{F}_a = \tilde{T}\tilde{F}_b$. In Eq. (24) we have also written $\tilde{T}$ in terms of $m\times m$ subblocks so that we can easily define a new matrix S connecting the composite vectors v_i and v_o,

$$v_o = S v_i, \tag{25}$$

with

$$S = \begin{pmatrix} \tilde{t}_{21}(\tilde{t}_{11})^{-1} & \tilde{t}_{22} - \tilde{t}_{21}(\tilde{t}_{11})^{-1}\tilde{t}_{12} \\ \tilde{t}_{11}^{-1} & -(\tilde{t}_{11})^{-1}\tilde{t}_{12} \end{pmatrix}. \tag{26}$$

Inspecting Eq. (23) shows that the physical requirement of flux conservation translates into the condition that S be unitary,

$$S^+ S = I. \tag{27}$$

Since S connects the incoming waves v_i to the outgoing waves v_o, reminiscent of scattering theory, we call S the scattering matrix. Note that while S is similar to the scattering matrices used by other authors,[2,9] there are crucial differences. Namely, the "incoming" and "outgoing" waves are now not defined simply by whether they decay towards or away from the interface. Rather, this classification is based on the direction in which relevant waves carry flux.

The appearance of a unitary scattering matrix highlights the origin of the lack of flux conservation encountered in the general formalism. An exact solution to the heterojunction problem would involve an infinite dimensional scattering matrix, S_∞, which of course would be unitary. However, if we use Eqs. (24) and (26) to form a $2m \times 2m$ scattering matrix S from a transfer matrix T, which itself is obtained from a projection of Eq. (6), then S will not be unitary. This should not be surprising since even an exact $2m \times 2m$ subblock of a unitary S_∞ is itself not necessarily unitary. Flux conservation cannot be guaranteed in the general formalism once we have chosen a finite basis of bulk Bloch waves in each material. This is a consequence of the fact that the separate bases of $2m$ Bloch waves, the $\psi_a^j(\mathbf{r})$ and $\psi_b^j(\mathbf{r})$, are "distinct," because each basis will in general span a different subspace of the functions of x and y. In essence, a projection of Eq. (6) can never lead to a flux-conserving interface connection rule if the state on the left-hand side of Eq. (6) is built from basis states spanning a subspace which is different than the subspace spanned by the basis states used on the right-hand side. In particular, it is clear that the transfer matrices T_1 and T_2 presented in Sec. II cannot in general be expected to guarantee the conservation of flux.

B. An approximate flux-conserving model

A physically reasonable model of semiconductor heterostructures demands a flux-conserving interface connection rule. The above argument shows that even if the $2m$ bulk basis functions in each material are known exactly, we will not obtain a flux-conserving interface connection rule from the general formalism. Therefore, the relevant issue to be addressed does not concern the "correct" interface connection rule, as often implied in envelope-function discussions. Rather, since any practical model will involve a finite basis, we must be satisfied to find a "best choice" interface connection rule.

A "best choice" interface connection rule can be obtained as follows. First, we describe each bulk material as well as possible, and choose a basis of $2m$ bulk Bloch waves. Then, a $2m \times 2m$ scattering matrix S is formed through Eqs. (24) and (26), using either transfer matrix T_1 or T_2 of Eqs. (9a) and (9b). Finally, we construct an effective scattering matrix S_e from this S. We demand that it be unitary $S_e^+ S_e = I$, and choose it to be "close" to S by minimizing the Hilbert-Schmidt norm of the difference $S_e - S$,

$$\{\mathrm{Tr}[(S_e - S)^+ (S_e - S)]\}^{1/2}, \tag{28}$$

An analytic solution to this problem can be obtained, yielding the result

$$S_e = SGP^{-1}G^+, \tag{29}$$

where G is a unitary matrix which diagonalizes the positive definite Hermitian matrix S^+S, and P^2 is the corresponding diagonal matrix of eigenvalues,

$$G^+S^+SG = P^2. \tag{30}$$

This choice for S_e preserves the time reversal symmetries present in the original S, which should be required on physical grounds. Having obtained S_e, an effective transfer matrix T_e can be uniquely identified by inverting Eqs. (24) and (26).

The unitary effective scattering matrix S_e can be considered a new version of S, slightly adjusted to ensure flux conservation. As Eq. (29) shows, S_e is obtained from S by post-multiplying by the Hermitian transformation $GP^{-1}G^+$. This transformation reduces to the identity if P itself is the identity; $P - I$ serves as an indicator of the lack of unitarity of the original scattering matrix S. As the value of m increases, more of the allowed scattering channels (k_j) in each material are accounted for, and the eigenvalues of S^+S, and hence the diagonal elements of P, are expected to get closer to unity. The value of m should be chosen large enough so that post-multiplying by the matrix $GP^{-1}G^+$ can be considered only a small adjustment to S. If S is not "almost" unitary we cannot expect the results predicted by S and S_e to be close.

It is obvious that a different norm than Eq. (28) may lead to a completely different choice for S_e. This is not our concern here since we only wish to stress the spirit of the technique and hence view our constructed S_e as only one possible choice. Finally, we point out that we have formed effective interface connection rules by considering the scattering matrices rather than the transfer matrices directly. This results because flux conservation implies a condition (i.e., unitarity) on scattering matrices which is more transparent and easy to take advantage of than the condition (14) on transfer matrices.

The scheme outlined above may appear overly complicated, considering that the EFA itself already provides a flux-conserving interface connection rule. The EFA transfer matrices, however, involve only bulk parameters,[3,7] whereas transfer matrices T_1 and T_2 of the general formalism suggest that envelope-function connection rules should involve nonbulk matrix elements. Therefore, we argue that the assumptions inherent to the EFA, while leading to flux-conserving interface connection rules, result in simplistic transfer matrices.[8] We are not suggesting that the EFA interface model is wholly inadequate, but rather, from the point of view of our approximate scheme, suggest that perhaps the EFA connection rules are in general not the "best choice." The transfer matrices suggested in this paper inherently account for material differences in the Bloch functions, since each material is described individually and as accurately as possible. Another scheme has been suggested to account for such material differences;[8] however, it is a perturbative technique and does not ensure flux conservation. A

further advantage of our scheme is that there is no use of a slowly varying envelope approximation, in contrast to the EFA, so that all k_j values throughout the Brillouin zone can be naturally included. Hence, Γ-X scattering problems can be easily treated.

Finally, we wish to stress that our suggested scheme is an attempt to construct a formalism which can still involve small-dimensional matrices. Of course, the accuracy of the results should improve as the basis set gets larger (larger m) and such computationally intensive schemes have been discussed.[9,10] However, even with large m one will find a lack of formal flux conservation, and a unitary effective scattering matrix should be adopted on physical grounds. Our main interest is in a computationally efficient scheme, and hence we are willing to exchange some accuracy in return for a small m. We assume that m is large "enough" if S^+S is "almost" unitary. The main point is that only a few of the Bloch waves, those with k_j having a small imaginary part, are expected to be strongly excited in the problems of interest. Hence, the relevant physics should still be captured if only these waves are included. However, the neglected bulk Bloch waves, those with k_j having large imaginary parts, may be important for the description of localized interface states. Further, as in any boundary condition technique, as the distance between interfaces decreases the neglect of these waves should be reconsidered.

IV. EXAMPLES: Γ-X SCATTERING

In this section we present some calculations that will illustrate our technique. For this purpose we wish to investigate the Γ-X coherent interfacial scattering problem in the GaAs/AlAs material system[2,4,5,11] and so we consider heterostructures grown in the [001] direction. While the scheme is applicable to arbitrary $\mathbf{k}_{||}$ we only consider the case $\mathbf{k}_{||}=0$ for simplicity. We also neglect any spin-orbit coupling and strain effects, but this is not a fundamental limitation. As discussed elsewhere,[12] the calculated results will be sensitive to the choice of location for the abrupt interfacial plane. In this paper we have taken the abrupt interface at an As plane. For all heterostructures we take $\Delta E_v = 0.3\Delta E_g$ to be the valence-band offset, where ΔE_g is the band-gap difference.

In order to do any calculations we must adopt a band-structure model for the bulk materials. The choice here is not limited by the boundary matching technique in any way; for illustration we adopt a simple numerical scheme, performing an empirical pseudopotential calculation[13] for the energy levels and states at some point $\mathbf{k}_0$ in the Brillouin zone. We then choose a subset of these states as a basis set and expand the wave function for arbitrary k as

$$\psi_k(\mathbf{r}) = e^{ikz}e^{i\mathbf{k}_{||}\cdot\mathbf{r}}\sum_{n=1}^{N} C_{kn}u_{nk_0}(\mathbf{r}). \tag{31}$$

As discussed in detail elsewhere[1] this choice of basis set is advantageous for complex band-structure calculations since it leaves the energy eigenvalue problem in a simple form,

$$(H_2k^2 + H_1k + H_0)C_k = EC_k, \tag{32}$$

where the H_i are Hermitian matrices. In a complex band-structure calculation it is the energy and $\mathbf{k}_{||}$ which act as input while all allowed complex k are the sought after results. The quadratic form of Eq. (32) allows this problem to be solved in a numerically efficient manner.

We have used pseudopotentials and lattice constants given elsewhere[14] and performed two such calculations, one at the Γ point and another at the X point, for reasons explained below. We used 181 and 230 plane waves at the Γ point and X point, respectively. The 27 lowest energy states at each of these points were then used as the bases required in Eq. (31). Part of the calculated complex band structure for GaAs is shown in Fig. 1. The 27 states used are a much larger number than is necessary to obtain reasonable bands. We have worked with such a large basis here in order to ensure that the bulk wave functions, and hence T_1 and T_2, are well described so that any flux discrepancy can be assumed to result primarily from the use of different bases in each material.

An interface is completely characterized by the matrix J_{ab}, and a reliable estimate of this matrix requires a good approximation of the bulk band structures and states involved. While these bulk states will never be completely accurate, it is crucial that they are constructed in a manner that ensures they possess all expected symmetries, in order for J_{ab} to reflect these symmetries. The use of the finite basis in Eq. (31) does lead to some problems concerning this point.[1] In particular, the calculated wave functions will not have the $\mathbf{k}$-space periodicity expected of Bloch states,

$$\psi_{\mathbf{k}+\mathbf{g}}(\mathbf{r}) = \psi_{\mathbf{k}}(\mathbf{r}), \tag{33}$$

where $\mathbf{g}$ is a reciprocal lattice vector. In our specific case this implies that a calculation at the Γ point ($\mathbf{k}_0 = 0$) will not provide bands which have the correct symme-

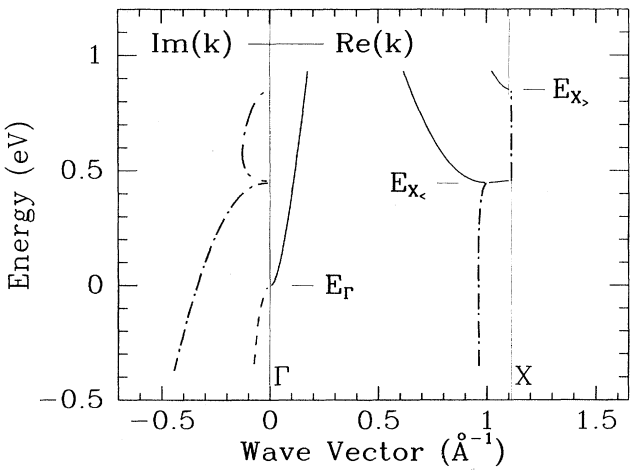

FIG. 1. The calculated complex bulk band structure of GaAs, $\mathbf{k}_{||}=0$, for energies around the conduction-band minima. Purely real (imaginary) k values are shown as solid (dashed) lines, while complex k values are represented as dash-dotted lines.

try around the X point, and vice versa. We circumvent this problem by performing separate calculations around the Γ and X points to obtain those portions of the band structure close to these respective points. Note, however, that with this model one must be careful to use a consistent treatment of the waves representing $\psi_k(\mathbf{r})$ and $\psi_{-k^*}(\mathbf{r})$ in order to ensure that time reversal symmetry is incorporated into J_{ab}.

Further problems related to the failure to satisfy Eq. (33) should be mentioned as well. Consider a matrix element of operator J between states $|\psi_k^\alpha\rangle$ and $|\psi_q^\alpha\rangle$ which includes the factor[1]

$$\sum_{g_\perp} e^{ig_\perp z_0}(C_k^\alpha)^+\{H_2(k^*+q+g_\perp)+H_1\}C_{q+g_\perp}^\alpha, \tag{34}$$

where $g_\perp\hat{\mathbf{z}}$ is a reciprocal lattice vector in the growth direction. It is easy to show[1] from Eq. (32) that the following bulk orthogonality relation exists:

$$(C_k^\alpha)^+\{H_2(k^*+q)+H_1\}C_q^\alpha=\delta_{k^*,q}. \tag{35}$$

This orthogonality condition implies that all $g_\perp \neq 0$ terms in J_{aa} and J_{bb} should vanish exactly. However, since Eq. (33) does not hold, these terms are not exactly zero in the calculation. We simply ignore these small $g_\perp \neq 0$ terms in J_{aa} and J_{bb}. Even with this simplification, the matrices J_{aa} and J_{bb} still have small components which should vanish according to the orthogonality condition (35). These artifacts appear due to our use of two different Hamiltonians in the separate Γ point and X point calculations. They are also ignored. We assume any errors in the calculated J_{ab} associated with the above effects will be small.

Regardless of the above problems related to the choice of basis in Eq. (31), we choose to work with this scheme due to the numerically efficient algorithm it provides for the complex band structure. Further, this type of basis is known to provide good estimates of effective masses, which suggests that the calculated J_{ab} may be satisfactory. Lastly, in this scheme the only empirical parameters needed are the pseudopotentials and the adopted band offset. Any other bulk band-structure calculation is acceptable, and it may be that a localized orbital technique[15,16] circumvents many of the encountered problems.

It has been suggested[17] that within a specific energy range the GaAs/AlAs heterostructures may be modeled by considering only six bulk Bloch waves ($m=3$) in each layer. As Fig. 1 suggests, this set consists of two waves whose k values have real parts close to the Γ point and another four waves whose real k parts are near the X point. Over the energy range of interest any of these six waves may possess purely real k values (propagating waves). Propagating waves can always carry flux away from an interface, and hence should always be included in an analysis. In contrast, bulk Bloch waves having a k value with a large imaginary part are not expected to be strongly excited at a single interface by an incoming propagating wave. Even in the multiple interface case, if the interfaces are sufficiently far apart then very little flux is expected to be carried away from any one interface by such nonpropagating waves. Thus, it seems plausible that scattered fluxes could be approximately predicted by neglecting these strongly evanescent waves.

A. GaAs/AlAs heterojunction

The heterojunction formed by GaAs and AlAs is the simplest example to consider. A schematic drawing of the energy minima lineups is shown in Fig. 2. We choose our conditions at $z=\pm\infty$ so that one Γ-like Bloch wave, $\mathrm{Re}(k)\approx 0$, carrying unit flux is incident on the interface from the GaAs side. For energies below $E_{X_<}^{\mathrm{AlAs}}$ we expect all the incident flux to be reflected back into the left propagating Γ-like wave. For higher energies, it becomes possible for some of the incident flux to be scattered into any of the other existent propagating channels. The results of the heterojunction calculation are presented in Figs. (3a) and (3b), which show the reflected and transmitted fluxes, respectively. Total reflection is verified for energies below $E_{X_<}^{\mathrm{AlAs}}$ as expected, while near E_Γ^{AlAs} the shape of the curves is similar to that expected for a free particle[18] and similar to an EFA calculation.[3] A detailed analysis of the scattered fluxes would entail a deeper investigation into the influence of the location of the abrupt interfacial plane within the unit cell, which is not the main concern here.

In the above calculation flux conservation is exactly satisfied at all energies due to the use of the unitary effective scattering matrix S_e. Also, the above results are insensitive to the choice of the initial transfer matrix. That is, obtaining S_e from either T_1 or T_2 of Eqs. (9a) and (9b) leads to the same curves in Fig. 3. A calculation that simply adopts T_1 or T_2 as the transfer matrix without "unitization" readily illustrates a lack of flux conservation. The results of such calculations, presented in Fig. 4, clearly show that the sum of all scattered fluxes is not always equal to the flux incident on the interface. These calculations suggest that the effective unitary scattering matrix provides a flux-conserving compromise between the T_1 and T_2 results.

This heterojunction Γ-X scattering problem has been treated previously by including many bulk Bloch waves in each material,[19] which ensured that any flux discrepancy would be small. A main concern in our work is to be able to include only a small number of Bloch waves within a flux-conserving scheme. We assume that the error induced by neglecting the strongly evanescent waves

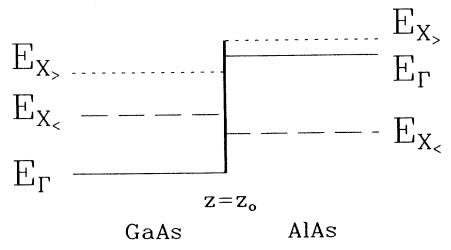

FIG. 2. A schematic drawing of the relevant energy minima for the GaAs/AlAs heterojunction. The energies $E_{X_<}$ and $E_{X_>}$ are defined by the minima labeled in Fig. 1.

will be small. We can test this hypothesis by including a greater number of Bloch waves in our calculation. For this purpose the results of a simulation involving 32 bulk Bloch waves ($m = 16$) in each material, 22 near the Γ point, and 10 near the X point, are shown as points in Fig. 3. These latter results clearly suggest that the simpler calculation with only six waves offers reasonable predictions of the scattered fluxes.

For several reasons, including differences in the calculated band structures, a detailed comparison of our work with earlier results[19] cannot be made. In particular, the transfer matrices used in Ref. 19 are not obtained from matrices similar to J_{ab} but rather from a significantly different projection of the interface boundary conditions. Nonetheless, both results show many of the same qualitative features.

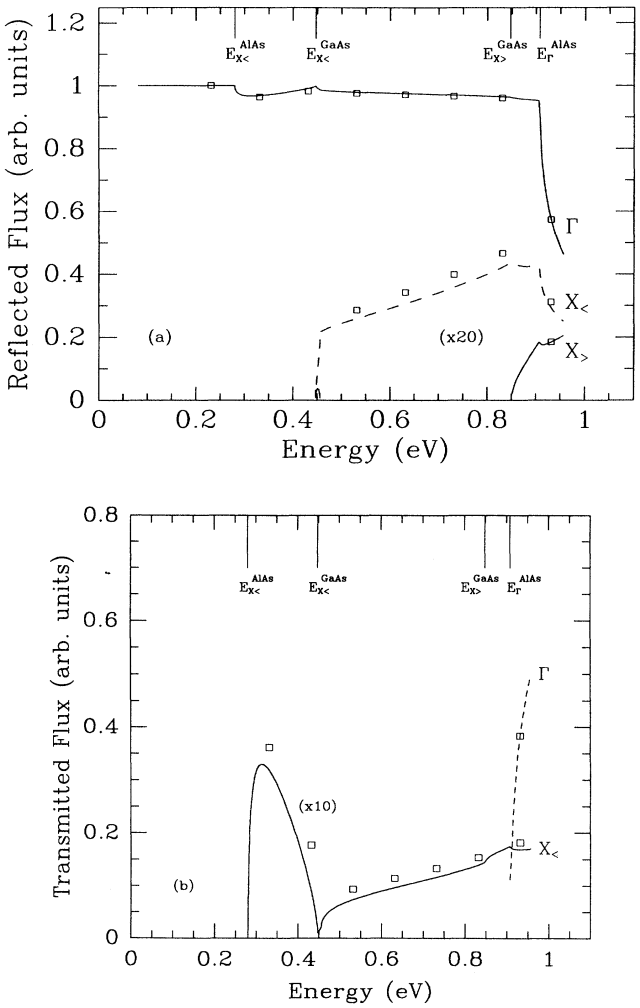

FIG. 3. Calculated scattered fluxes for the GaAs/AlAs heterojunction with energies measured from the Γ minima of GaAs; (a) reflected fluxes in GaAs and (b) transmitted fluxes in AlAs. Solid lines give the results of the 6×6 model while the squares represent the 32×32 model calculations. Note that all X-like fluxes have been multiplied 10 or 20 times for clarity.

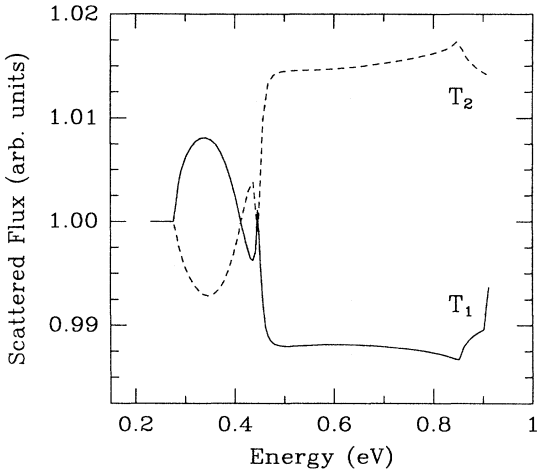

FIG. 4. Total scattered flux for the GaAs/AlAs heterojunction for different adopted transfer matrices, T_1 and T_2. For these calculations the abrupt interface was taken at a cation plane.

B. GaAs/AlAs double barrier

As a second example we calculate the transmitted flux, for the same incident conditions, through a symmetric double barrier structure composed of AlAs barriers and a central GaAs quantum well. The barrier and well thicknesses are five and ten lattice constants, respectively. The use of a unitary effective scattering matrix for each interface ensures flux conservation throughout the structure. However, as discussed elsewhere,[2] the thickness of the heterostructure requires that a numerically stable routine be implemented.

The calculated transmitted flux presented in Fig. 5 shows the many resonances that are expected in such structures. Those resonances labeled A and C are associated with resonant tunneling through Γ-like states, and are also predicted by an EFA calculation. The re-

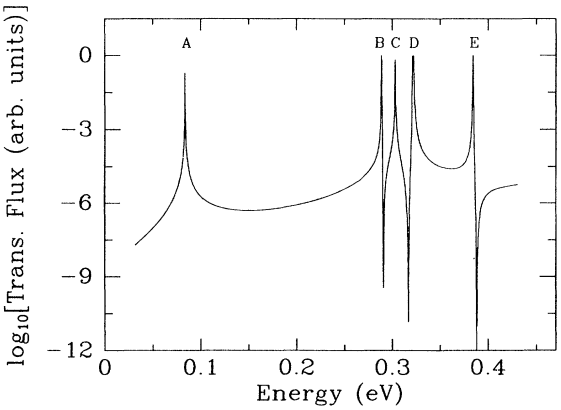

FIG. 5. Calculated flux transmitted through the symmetric double barrier structure described in the text.

maining resonances are associated with resonant tunneling through the X-like states which are not included in an EFA calculation. These resonances B, D, E are related to quasibound states localized in the AlAs layers by the effective quantum well potential defined by the X minima profile. A zero Γ-X coupling would give these states an infinite lifetime causing them to become discrete bound states. It is the nonzero coupling of these bound X-like states to the Γ-like continuum which leads to their classification as Fano resonances.[20] The shape of the curves around these resonances can be thought of as arising from a quantum mechanical interference between the different Γ-like and X-like paths that a tunneling electron can travel along, and is a signature of the nonparabolicity of the bulk band structures involved. On a higher energy resolution one can see that the B, D, E resonances are split due to the interaction of the two wells; a single barrier calculation does not show this splitting.

This geometry has also been previously investigated with a technique involving many bulk Bloch waves in each material.[2,21] While a qualitative comparison with those results shows similar resonances, the detailed structure of those resonances differs. This is not surprising since this structure will depend significantly on subtleties such as how the interface boundary conditions are projected to obtain transfer matrices.

V. SUMMARY

We have presented a flux-conserving technique which can be used in the the study of semiconductor heterostructures. The scheme was developed after identifying the relevant scattering matrix for the problem. The technique is more general than the envelope-function approximation, because material differences in the cell-periodic part of the Bloch functions are described and a full-zone treatment of the constituent semiconductors is allowed. Further, since only small-dimensional matrices need be involved the calculations can be numerically efficient. At each energy, the technique offers a "best choice" flux-conserving interface connection rule between two semiconductors.

A lack of flux conservation arises in general when the wave functions to be "matched" on opposite sides of an interface are expanded with finite bases spanning different subspaces. This is the case for semiconductor heterostructures, since the underlying bulk microscopic potentials involved are different, leading to distinct bulk band structures and Bloch waves. We would also expect an effective unitary scattering matrix approach to be of use in the study of optical waveguides or electron transport through quantum confined constrictions. In such problems, the natural finite bases for different regions of the structure will in general be distinct due to the variations in the lateral and transverse directions.

ACKNOWLEDGMENTS

We would like to thank Z. B. Wang, N. Arkani-Hamed, and A. Tameshtit for useful discussions regarding this work. This work was supported by the Natural Sciences and Engineering Research Council of Canada, and the Ontario Laser and Lightwave Research Centre.

[1] D. L. Smith and C. Mailhiot, Rev. Mod. Phys. **62**, 173 (1990).
[2] D. Y. K. Ko and J. C. Inkson, Phys. Rev. B **38**, 9945 (1988).
[3] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures* (Les Ulis, Paris, 1988).
[4] H. C. Liu, Appl. Phys. Lett. **51**, 1019 (1987).
[5] N. Pulsford *et al.*, Phys. Rev. Lett. **63**, 2284 (1989).
[6] M. Sweeny and J. Xu, Appl. Phys. Lett. **54**, 546 (1989).
[7] B. Chen, M. Lazzouni, and L. R. Ram-Mohan, Phys. Rev. B **45**, 1204 (1992).
[8] D. L. Smith and C. Mailhiot, Phys. Rev. B **33**, 8345 (1986).
[9] J. P. Cuypers and W. Van Haeringen, J. Phys. Condens. Matter **4**, 2587 (1992).
[10] A. C. Marsh and J. C. Inkson, J. Phys. C **17**, 6561 (1984).
[11] A. R. Bonnefoi *et al.*, Appl. Phys. Lett. **50**, 344 (1987).
[12] T. Ando and H. Akera, Phys. Rev. B **40**, 11 609 (1989).
[13] M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789 (1966).
[14] E. Caruthers and P. J. Lin-Chung, Phys. Rev. B **17**, 2705 (1978).
[15] D. Z. Y. Ting and Y. C. Chang, Phys. Rev. B **36**, 4359 (1987).
[16] M. K. Jackson *et al.*, Phys. Rev. B **43**, 4856 (1991).
[17] T. Ando and H. Akera, Phys. Rev. B **40**, 11 619 (1989).
[18] See, e.g., R. Eisberg and R. Resnick, *Quantum Physics* (Wiley, New York, 1974), p. 197.
[19] A. Marsh and J. C. Inkson, J. Phys. C **19**, 43 (1986).
[20] P. A. Schulz, Phys. Rev. B **44**, 8323 (1991).
[21] A. Marsh and J. C. Inkson, Semicond. Sci. Technol. **1**, 285 (1986).