PHYSICAL REVIEW B VOLUME 54, NUMBER 3 15 JULY 1996-I

Envelope-function formalism for electrons in abrupt heterostructures with material-dependent basis functions

Bradley A. Foreman
Department of Physics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom

(Received 12 October 1995; revised manuscript received 25 March 1996)

An envelope-function model is derived for electrons in abrupt semiconductor heterostructures. It uses material-dependent basis functions that diagonalize the bulk zone-center Hamiltonian in each unit cell of the crystal. The initial formalism is exactly equivalent to the one-electron Schrödinger equation; approximations suitable for abrupt junctions are then developed. The abrupt change in microscopic potential at an ideal interface is shown to introduce no new interband coupling; all such coupling arises from the kinetic energy, specifically from a momentumlike matrix element containing the gradient of the basis functions with respect to changes in material composition. This generates interface effects not included in conventional envelope-function theories, such as zone-center coupling between heavy and light holes. An effective-mass equation is derived for the conduction band of a layered zinc-blende structure; it exhibits both envelope discontinuities and δ-function potentials, in agreement with the transfer matrices derived from other microscopic theories. [S0163-1829(96)10027-8]

I. INTRODUCTION

The envelope-function method, by virtue of its simplicity and ease of interpretation, is by far the most popular method for calculating the properties of electrons in semiconductor heterostructures. The standard model used for such problems[1–5] is based on bulk $\mathbf{k}\cdot\mathbf{p}$ theory[6,7] and neglects the change in band-edge Bloch functions at an interface, characterizing the heterostructure in terms of its bulk properties alone. However, it is now recognized that such a treatment is not always adequate; for example, in the valence bands of zincblende structures, microscopic calculations[8–12] show a significant heavy-hole–light-hole (HH-LH) coupling at $\mathbf{k}_{\parallel}=\mathbf{0}$, which is not included in the standard envelope-function theory. Possible experimental evidence for this effect has also been found.[13,14] To describe such coupling, one must include properties of the interface itself, which are determined by the difference in Bloch functions between the two media. A number of envelope-function models incorporating this difference have been proposed in recent years,[9–11,15–26] but most of these require extensive microscopic calculations. What has been missing is a straightforward analytical derivation of an envelope-function model that offers a clear, simple, yet complete description of how the material dependence of the basis functions alters the interface connection rules. The object of this paper is to present such a derivation, starting from the one-electron Schrödinger equation for a lattice-matched system without spin.

The two key steps in the derivation are the choice of basis functions and the choice of approximation technique. An intuitive description of position-dependent material properties is most easily obtained from a localized basis such as the Wannier functions.[27] However, there are also distinct advantages to using a Luttinger-Kohn basis,[28] such as the simplicity of the equations for degenerate bands and the natural appearance of energy-dependent effective masses for the treatment of nonparabolic effects. In this paper a hybrid basis is chosen in which localized Wannier functions are defined not in terms of $\mathbf{k}$-dependent Bloch waves, but in terms of zone-center Luttinger-Kohn functions. This material-dependent basis is defined so that the bulk zone-center Hamiltonian is diagonal in each unit cell of the crystal.

An effective technique for approximating the heterostructure matrix elements may be developed by assuming that the envelopes of the wave function and material properties are small in the outer half of the Brillouin zone, i.e., that they are ‘‘slowly varying’’ in $\mathbf{k}$ space.[18] Such a $\mathbf{k}$-space approximation remains valid even when these functions change abruptly in $\mathbf{r}$ space, provided the abrupt discontinuities are not too close together. This technique therefore permits the treatment of problems that cannot be handled using the conventional assumption that the envelopes vary slowly in $\mathbf{r}$ space.

Under these conditions, it is shown that the envelope-function equations take the form of local, spatially varying $\mathbf{k}\cdot\mathbf{p}$ equations, but with additional terms describing the explicit effects of the interface. The fundamental parameter characterizing the interface properties is a momentumlike matrix element that depends on the gradient of the band-edge Bloch functions with respect to changes in *material composition* (rather than the ordinary spatial gradient in the bulk momentum matrix). This term has a δ-like character at an abrupt interface and, in tandem with the bulk momentum matrix, gives rise to a δ-function potential coupling the heavy- and light-hole bands. Such a matrix element was first derived by Karavaev and Tikhodeev[20] for heterostructures with slow spatial variations. The present work extends their theory to structures containing abrupt discontinuities.

A single-band effective-mass equation is derived from the $\mathbf{k}\cdot\mathbf{p}$ formalism by using Löwdin perturbation theory[29] to eliminate the interband coupling. The resulting equation has all the features of more complex microscopic theories (i.e., envelope-function discontinuities and δ-function potentials), and it offers a clearly understandable explanation of how these effects arise. It therefore justifies the phenomenological technique of using effective-mass equations at an abrupt interface by providing an unambiguous prescription for the form of such a Hamiltonian. As a consequence, the present theory affords a conceptual unification of the microscopic and phenomenological approaches to heterostructure envelope-function theory, approaches that were viewed in the past as fundamentally different in principle,[19] but are now seen to be equivalent (if formulated correctly).

The derivation begins in Sec. II, where an envelope-function formalism with material-dependent basis functions is developed in a form exactly equivalent to the Schrödinger equation. To simplify the exact theory, an approximation technique is proposed in Sec. III and applied to the Hamiltonian in Sec. IV. The approximate model is used to derive a single-band effective-mass equation and interface connection rules in Sec. V. A comparison of the model with other envelope-function theories is presented in Sec. VI, and the main results of the paper are reviewed and discussed in Sec. VII.

II. MATERIAL-DEPENDENT BASIS FUNCTIONS

In this section, an exact envelope-function formalism is developed by defining a material-dependent basis set that diagonalizes the bulk zone-center Hamiltonian in each unit cell of the crystal. The starting point is the one-electron Schrödinger equation:

$$\left[\frac{\mathbf{p}^2}{2m}+V(\mathbf{r},t)\right]\psi(\mathbf{r},t)=i\hbar\frac{\partial}{\partial t}\psi(\mathbf{r},t), \tag{2.1}$$

where $\mathbf{p}=-i\hbar\nabla$, m is the free-electron mass, and V is a nonperiodic microscopic potential. This potential describes the intrinsic properties of the heterostructure, along with any space charges or external fields.

To construct appropriate basis functions for this Hamiltonian, it will be assumed that we have complete information about the bulk media making up the structure. That is, for every unit cell $\mathbf{R}$ (where $\mathbf{R}$ is the position of a Bravais lattice site), we know the bulk potential (or pseudopotential) $V_0(\mathbf{r},\mathbf{R})$ for the medium occupying that cell. This potential is by definition periodic in $\mathbf{r}$ [i.e., $V_0(\mathbf{r},\mathbf{R})=V_0(\mathbf{r}+\mathbf{R}',\mathbf{R})$], so we can solve the corresponding Schrödinger equation by standard methods:[30,31]

$$\left[\frac{\mathbf{p}^2}{2m}+V_0(\mathbf{r},\mathbf{R})\right]\psi_{n\mathbf{k}}(\mathbf{r},\mathbf{R})=E_{n\mathbf{k}}(\mathbf{R})\psi_{n\mathbf{k}}(\mathbf{r},\mathbf{R}), \tag{2.2}$$

where $\psi_{n\mathbf{k}}(\mathbf{r},\mathbf{R})$ is a Bloch function satisfying $\psi_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}',\mathbf{R})=e^{i\mathbf{k}\cdot\mathbf{R}'}\psi_{n\mathbf{k}}(\mathbf{r},\mathbf{R})$, and $E_{n\mathbf{k}}(\mathbf{R})$ is the associated eigenenergy. In this equation, $\mathbf{R}$ should be viewed as merely an index labeling the different possible bulk solutions that occur at different locations in the heterostructure.

Following Luttinger and Kohn,[28,30] we will now focus attention on the $\mathbf{k}=\mathbf{0}$ solutions $u_n(\mathbf{r},\mathbf{R})\equiv\psi_{n\mathbf{0}}(\mathbf{r},\mathbf{R})$ and $E_n(\mathbf{R})\equiv E_{n\mathbf{0}}(\mathbf{R})$. The functions $u_n(\mathbf{r},\mathbf{R})$ are periodic in $\mathbf{r}$, and at any $\mathbf{R}$ they provide a complete orthonormal basis for the expansion of any cell-periodic function (see Appendix A). Although these functions are material dependent, they are not localized, so they do not offer the most natural description of spatially varying material properties. To introduce such localization,[27] consider the following unit-cell δ function, which is simply a superposition of all plane waves in the first Brillouin zone (BZ):

$$\delta_B(\mathbf{r})=(2\pi)^{-3}\int B(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}d^3k, \tag{2.3}$$

where

$$B(\mathbf{k})=\begin{cases}1, & \mathbf{k}\in\mathrm{BZ}\\ 0, & \mathbf{k}\notin\mathrm{BZ}\ .\end{cases} \tag{2.4}$$

The function $\delta_B(\mathbf{r})$ has unit area and a spatial extent roughly equal to the volume v of a primitive unit cell; its properties are discussed further in Appendix B. At the lattice sites, this function satisfies

$$\delta_B(\mathbf{R}-\mathbf{R}')=v^{-1}\delta_{\mathbf{R},\mathbf{R}'}\ . \tag{2.5}$$

This equation is valid regardless of the shape of the BZ,[31] so we can choose a shape different from the conventional Wigner-Seitz cell if desired. (Other choices are often convenient for treating surface and interface problems.[32])

We can define a localized basis similar to the familiar Wannier functions[27,30] by taking the product of this unit-cell δ function with the periodic Luttinger-Kohn functions:

$$w_n(\mathbf{r},\mathbf{R})\equiv v\,\delta_B(\mathbf{r}-\mathbf{R})u_n(\mathbf{r},\mathbf{R}). \tag{2.6}$$

These modified Wannier functions may be completely different in adjacent unit cells, but they still form a complete orthonormal basis:

$$\begin{gathered}\sum_{n,\mathbf{R}} w_n(\mathbf{r},\mathbf{R})w_n^*(\mathbf{r}',\mathbf{R})=\delta(\mathbf{r}-\mathbf{r}'),\\ \int w_n^*(\mathbf{r},\mathbf{R})w_{n'}(\mathbf{r},\mathbf{R}')d^3r=\delta_{nn'}\delta_{\mathbf{R},\mathbf{R}'}\ ,\end{gathered} \tag{2.7}$$

as shown in Appendix A. Since the basis (2.6) is complete, we can use it to represent any wave function $\psi(\mathbf{r})$:

$$\begin{gathered}\psi(\mathbf{r})=v^{1/2}\sum_{n,\mathbf{R}} f_n(\mathbf{R})w_n(\mathbf{r},\mathbf{R}),\\ f_n(\mathbf{R})=v^{-1/2}\int w_n^*(\mathbf{r},\mathbf{R})\psi(\mathbf{r})d^3r,\end{gathered} \tag{2.8}$$

where $f_n(\mathbf{R})$ is a discrete envelope function for energy band n and unit cell $\mathbf{R}$. In terms of these envelope functions, the Schrödinger equation (2.1) takes the form

$$\sum_{n',\mathbf{R}'} H_{nn'}(\mathbf{R},\mathbf{R}')f_{n'}(\mathbf{R}')=i\hbar\frac{\partial}{\partial t}f_n(\mathbf{R}), \tag{2.9}$$

where the nonlocal matrix elements of the Hamiltonian are

$$H_{nn'}(\mathbf{R},\mathbf{R}')=\int w_n^*(\mathbf{r},\mathbf{R})\left[\frac{\mathbf{p}^2}{2m}+V(\mathbf{r})\right]w_{n'}(\mathbf{r},\mathbf{R}')d^3r. \tag{2.10}$$

In the above equations, the time dependence of the various functions (V,ψ,f_n) has been suppressed for brevity.

As yet the envelopes f_n are defined only at the lattice sites $\mathbf{R}$. To interpolate these functions between the lattice sites, we can apply the sampling theorem,[33] which gives a unique interpolation $f_n(\mathbf{r})$ as follows:[34–37]

$$f_n(\mathbf{r}) = v\sum_{\mathbf{R}} f_n(\mathbf{R})\,\delta_B(\mathbf{r}-\mathbf{R}). \tag{2.11}$$

This definition is unique because its Fourier transform $f_n(\mathbf{k})$ is identically zero outside the BZ.[28,38,18] First-zone functions such as (2.11) are called ''quasicontinuum'' functions[35] because the discrete and continuum representations are completely interchangeable. One could also define quasicontinuum Wannier functions, but the discrete basis (2.6) is simpler for most calculations, so this topic will not be pursued here.

Equations (2.9)–(2.11) provide a concise, exact envelope-function representation of the one-electron Schrödinger equation. However, in order to extract useful results from these equations, we need to invoke some form of approximation. The most frequently used approximation in this context is the slowly varying amplitude approximation,[39–46] i.e., the assumption that the material properties and envelope functions do not vary rapidly on the scale of the unit cell. Such an approximation is totally inapplicable at an abrupt junction, where the amplitudes of these functions change by a large amount in a single monolayer. Nonetheless, by defining a slowly varying amplitude in **k** space[18,40,45] rather than **r** space, we can develop an extremely useful approximation that is valid even in cases of abrupt discontinuity.

III. GLOBALLY SLOWLY VARYING AMPLITUDE APPROXIMATION

A detailed treatment of the Hamiltonian (2.10) is rather complex algebraically (see Appendixes C and D). However, the following simple example suffices to illustrate all of the basic features of the **k**-space approximation technique and its advantages over **r**-space methods. Consider the convolution integral

$$f_B(\mathbf{r}) = \int \delta_B(\mathbf{r}-\mathbf{r}')f(\mathbf{r}')d^3r', \tag{3.1}$$

in which $f(\mathbf{r})$ is an arbitrary function, limited only by the assumption that it can be represented in the Fourier integral form

$$\begin{aligned} f(\mathbf{r}) &= (2\pi)^{-3/2}\int f(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}\,d^3k,\\ f(\mathbf{k}) &= (2\pi)^{-3/2}\int f(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}\,d^3r. \end{aligned} \tag{3.2}$$

Integrals similar to (3.1) are ubiquitous in effective-mass theories (see, e.g., Wannier and Slater[27] or Luttinger and Kohn[28]), and the question that arises is as follows: Under what conditions is it permissible to replace the integral $f_B(\mathbf{r})$ with the original function $f(\mathbf{r})$?

The standard approach to this problem[27,28] is based on the observation that $\delta_B(\mathbf{r})$ is a δ-like function localized about a unit cell. Therefore, if $f(\mathbf{r})$ varies slowly over distances comparable to the lattice spacing, we can treat $\delta_B(\mathbf{r})$ as a Dirac δ function, so to the lowest order of approximation

$$f_B(\mathbf{r}) \approx f(\mathbf{r}), \tag{3.3}$$

and we need not bother with the details of the integral (3.1). This approximation is at the very heart of bulk effective-mass theory,[27,28] and is also widely used in similar theories for slowly graded inhomogeneous media.[39–44,46]

This **r**-space technique is very useful, but it also leads immediately to several further questions. How rapidly can $f(\mathbf{r})$ vary before the approximation (3.3) breaks down? What corrections to (3.3) should be included in the next order of approximation? These issues are of only minor interest in the bulk theory, but they become critically important at an abrupt junction.

A precise answer to these questions may be found by transforming Eq. (3.1) to **k** space, which yields the well-known identity[47] that a convolution in **r** space is simply a product in **k** space:

$$f_B(\mathbf{k}) = B(\mathbf{k})f(\mathbf{k}). \tag{3.4}$$

This means that $f_B(\mathbf{k})$ is the same as $f(\mathbf{k})$ inside the BZ, but is identically zero outside. Hence for quasicontinuum functions such as (2.11), which are already zero outside the BZ, Eq. (3.3) is not just a good approximation, it is exact. This conclusion, although rudimentary, has profound implications. A discrete function $f(\mathbf{R})$ can change *arbitrarily* rapidly from one unit cell to the next, and yet the quasicontinuum interpolation $f(\mathbf{r})$ satisfies (3.3) *exactly*.

In general, of course, we are interested in evaluating (3.1) for functions other than quasicontinuum functions, for which the truncation (3.4) may have a non-negligible effect. The case of primary interest is the product $f(\mathbf{r}) = f_1(\mathbf{r})f_2(\mathbf{r})$, where f_1 and f_2 are both quasicontinuum functions (usually one is an envelope function and the other is a material property). In this case $f(\mathbf{k})$ is nonzero outside the BZ, so the integral [cf. (C8) and (D8)]

$$\begin{aligned} &\int \delta_B(\mathbf{r}-\mathbf{r}')f_1(\mathbf{r}')f_2(\mathbf{r}')d^3r' \\ &\quad = (2\pi)^{-3}\int\int B(\mathbf{k}+\mathbf{k}')f_1(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}f_2(\mathbf{k}')e^{i\mathbf{k}'\cdot\mathbf{r}}d^3k\;d^3k' \end{aligned} \tag{3.5}$$

is not generally equal to $f_1(\mathbf{r})f_2(\mathbf{r})$. However, if f_1 and f_2 are zero in the outer half of the BZ (i.e., those wave vectors **k** for which $2\mathbf{k}$ is outside the BZ), then $\mathbf{k}+\mathbf{k}'$ always lies inside the BZ, and (3.5) is exactly equal to $f_1(\mathbf{r})f_2(\mathbf{r})$.

Hence we have the following simple criterion for whether (3.3) is a valid approximation: The functions f_1 and f_2 should be concentrated relatively close to the zone center, with a negligible spectrum in the outer half of the BZ. This forms the basis of what will be called the globally slowly varying amplitude approximation (GSVAA). The word *global* is used to emphasize that this **k**-space restriction makes *no reference* to whether f_1 and f_2 are slowly varying at any individual point. To require that these functions are *locally* slowly varying at each and every point in **r** space is a much more stringent condition, and goes far beyond what is actually needed. As shown below, the **k**-space definition[18] of the

GSVAA is broad enough to encompass even functions with large local discontinuities, provided these discontinuities are not too close together.

To see this, suppose we start by looking at the material (**R**) dependence of the basis functions $u_n(\mathbf{r},\mathbf{R})$ in a layered structure grown along the [001] axis of a zinc-blende crystal. For a single heterojunction at $z=0$, the material properties are proportional to the discrete step function

$$\theta_B(z_n)=\begin{cases}0, & n<0\\ 1, & n\geqslant 0,\end{cases} \tag{3.6}$$

where $z_n=(n+\frac{1}{2})a$ is the location of the nth monolayer and a is the monolayer thickness (e.g., $a=2.83$ Å in GaAs). The quasicontinuum function corresponding to (3.6) is

$$\theta_B(z)=a\sum_{n=0}^{\infty}\delta_B(z-z_n), \tag{3.7}$$

where $\delta_B(z)=\sin(\pi z/a)/\pi z$ is the one-dimensional analog of (2.3) [see also (B1)]. The Fourier transform of (3.7) is

$$\theta_B(k)=\frac{B(k)}{\sqrt{2\pi}}\left[\pi\delta(k)+\frac{a/2}{i\,\sin(ka/2)}\right], \tag{3.8}$$

where $B(k)=1$ for $|k|<\pi/a$ and zero otherwise.

The first term in (3.8) is proportional to $\delta(k)$; it describes the average value of the step function ($\frac{1}{2}$) and is not of interest at the moment. The second term is more significant, since it describes the change in amplitude across the interface. As $k\to 0$, this term diverges as k^{-1}, which means that $\theta_B(k)$ is highly concentrated near the zone center. Indeed, even without the $\delta(k)$ term, the area under $|\theta_B(k)|$ within the inner half of the BZ ($|k|<\pi/2a$) is still *infinitely* larger than that in the outer half. Thus according to the **k**-space definition, a step function is perhaps the ultimate slowly varying function aside from a constant.

The GSVAA begins to break down only when two interfaces are brought into close proximity. As an example, consider an N-monolayer quantum well centered on $z=0$. The material properties in this case are proportional to

$$\Pi_B(z)=\theta_B(z+Na/2)-\theta_B(z-Na/2), \tag{3.9}$$

which has the Fourier transform

$$\Pi_B(k)=\frac{a}{\sqrt{2\pi}}B(k)\,\frac{\sin(kNa/2)}{\sin(ka/2)}. \tag{3.10}$$

The magnitude of this function is plotted in Fig. 1 for quantum wells of 4, 8, and 16 ML (i.e., well widths of 11.3, 22.6, and 45.2 Å in GaAs). Even for $N=4$ it is clear that the dominant part of the spectrum lies within the inner half of the BZ. As N increases, the accuracy of the approximation improves.

In the most general case of two media with greatly different basis functions, the GSVAA will break down for $N<4$. However, in most heterostructures of common interest, the change in basis functions at an interface is small[48] (say 10%), so in addition to the spatially varying part described above, there is a large constant term whose contribution to the Fourier spectrum is proportional to $\delta(\mathbf{k})$. In such cases, the basis functions $u_n(\mathbf{r},\mathbf{R})$ can be treated as GSV functions of **R** even in the limit of a single-monolayer quantum well.

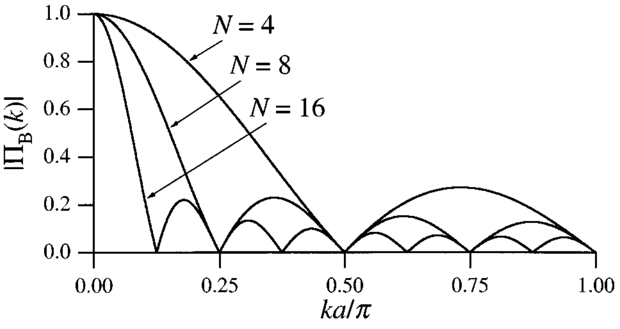

FIG. 1. Fourier spectrum $\Pi_B(k)$ for three different quantum-well widths.

Much the same conclusions are reached by examining the properties of the envelope functions. A sample envelope in a 10-ML quantum well is shown in Fig. 2, along with its Fourier transform. This function is clearly slowly varying by the **k**-space criterion, despite the 2:1 amplitude discontinuity at the interface. Of course, even in a wide quantum well, rapidly varying envelopes can always be generated at sufficiently high quantum numbers. One should always check after a given calculation that the envelopes are indeed small in the outer half of the Brillouin zone.

In a very thin (single monolayer) quantum well, the envelopes will usually be GSV unless the change in basis functions is large, since in the single-monolayer limit the envelopes consist almost entirely of long exponential tails in the barrier regions.[18] The only structure where the GSVAA is questionable even in the ground state is a superlattice in

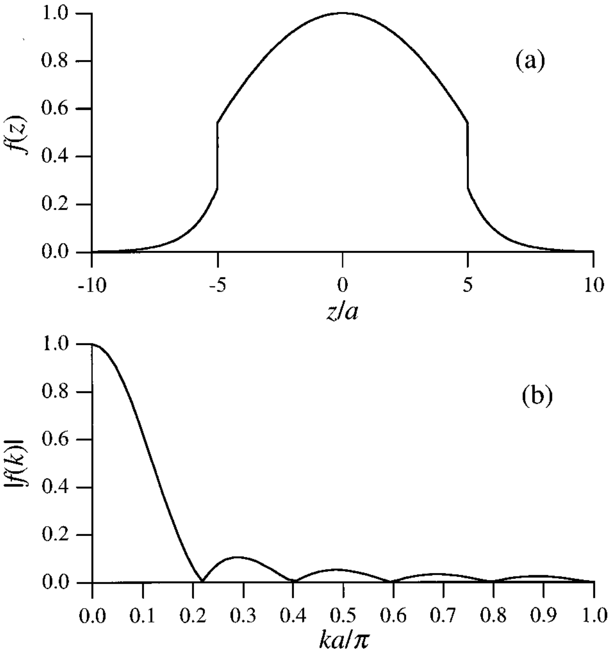

FIG. 2. (a) One-dimensional example of an envelope (not a solution to any equation in the text) with a 2:1 amplitude discontinuity in a 10-ML quantum well. (b) Fourier transform of (a).

which both the wells and barriers are very thin. In this case there will be large-**k** components in both the material properties and the envelope functions, so the magnitude of these terms should be examined critically before the results of a GSV calculation are accepted.

We thus have a clear, simple criterion for establishing which terms in the Hamiltonian (2.10) are physically significant. Section IV begins by considering the potential energy.

IV. ENVELOPE-FUNCTION HAMILTONIAN

A. Potential energy

To simplify the analysis, the potential energy $V(\mathbf{r})$ will be separated into two parts: a ‘‘flat-band’’ term $V_{\mathrm{FB}}(\mathbf{r})$, which describes how the bulk potential varies with material composition, and an additional term $\Phi(\mathbf{r})$ representing space-charge effects and externally applied fields:

$$V(\mathbf{r})=V_{\mathrm{FB}}(\mathbf{r})+\Phi(\mathbf{r}). \tag{4.1}$$

The external potential $\Phi(\mathbf{r})$ is usually assumed to be slowly varying, but the flat-band potential is certainly not. It will be represented here by an expansion similar to that used for the envelopes (2.11):

$$V_{\mathrm{FB}}(\mathbf{r})=v\sum_{\mathbf{R}} V_0(\mathbf{r},\mathbf{R})\,\delta_B(\mathbf{r}-\mathbf{R}), \tag{4.2}$$

where $V_0(\mathbf{r},\mathbf{R})$ is the bulk potential of the medium in the unit cell at $\mathbf{R}$. Because $\delta_B(\mathbf{r}-\mathbf{R})$ is not strictly zero outside the unit cell at $\mathbf{R}$, Eq. (4.2) generates a small overlap of the bulk potentials at an interface. This is physically reasonable since some overlap always occurs, although it may not be precisely of the form (4.2). Any deviations can be taken up in the potential $\Phi(\mathbf{r})$.

The flat-band part of the potential-energy operator in the Hamiltonian (2.10) is

$$V_{nn'}^{\mathrm{FB}}(\mathbf{R},\mathbf{R}')=\int w_n^*(\mathbf{r},\mathbf{R})V_{\mathrm{FB}}(\mathbf{r})w_{n'}(\mathbf{r},\mathbf{R}')d^3r. \tag{4.3}$$

It is shown in Appendix C that for GSV envelopes and material properties, this matrix element reduces to the simple expression

$$V_{nn'}^{\mathrm{FB}}(\mathbf{R},\mathbf{R}')=V_{nn'}^{0}(\mathbf{R})\,\delta_{\mathbf{R},\mathbf{R}'}, \tag{4.4}$$

where

$$V_{nn'}^{0}(\mathbf{R})=\int_v u_n^*(\mathbf{r},\mathbf{R})V_0(\mathbf{r},\mathbf{R})u_{n'}(\mathbf{r},\mathbf{R})d^3r \tag{4.5}$$

is the bulk matrix element for the medium at $\mathbf{R}$. Equation (4.4) is the same as the usual result for locally slowly varying functions, but we see here that it is valid at abrupt junctions as well. Hence, contrary to popular opinion, there is no need to fear that the abrupt change in microscopic potential at an ideal interface will lead to gross deviations from bulk envelope-function theory. This fear arises from the assumption that the flat-band potential must be treated within the confines of traditional effective-mass theory,[27,28] i.e., as an arbitrary rapidly varying perturbation. However, the flat-band potential is not arbitrary, nor is it treated here as a perturbation, nor is it rapidly varying in $\mathbf{R}$ (at least not in the global sense). In a material-dependent basis, it is only natural that the potential which generates the basis functions should have simple matrix elements.

The same procedure can be used to calculate the matrix elements of $\Phi(\mathbf{r})$. If $\Phi(\mathbf{k})$ meets the slowly varying criterion, then these reduce to

$$\Phi_{nn'}(\mathbf{R},\mathbf{R}')=\Phi(\mathbf{R})\,\delta_{nn'}\,\delta_{\mathbf{R},\mathbf{R}'}, \tag{4.6}$$

which is the same as the well-known result from bulk effective-mass theory.

Therefore, if the microscopic potential can be separated into flat-band and slowly varying components, the potential-energy matrix element is no different from the bulk. Such a separation is not always possible, since in general there will exist rapidly varying interface terms that cannot be expressed as a flat-band potential.[10] These terms give rise to additional interband couplings not included here. However, the flat-band potential is by far the dominant effect of the abrupt transition between two media, so no corrections to this approximation will be considered here. (Many microscopic theories, especially those based on the pseudopotential method,[9–12] also make use of this assumption.)

Since the potential energy has such a simple structure, the true interface effects must arise from the kinetic energy.

B. Kinetic energy

The kinetic-energy part of the Hamiltonian (2.10) is given by

$$T_{nn'}(\mathbf{R},\mathbf{R}')=\frac{1}{2m}\int w_n^*(\mathbf{r},\mathbf{R})\mathbf{p}^2 w_{n'}(\mathbf{r},\mathbf{R}')d^3r. \tag{4.7}$$

This can be rewritten using the completeness condition (2.7) as

$$T_{nn'}(\mathbf{R},\mathbf{R}')=\frac{1}{2m}\sum_{n'',\mathbf{R}''}\mathbf{p}_{nn''}^{\dagger}(\mathbf{R},\mathbf{R}'')\cdot\mathbf{p}_{n''n'}(\mathbf{R}'',\mathbf{R}'), \tag{4.8}$$

in which

$$\mathbf{p}_{nn'}(\mathbf{R},\mathbf{R}')=\int w_n^*(\mathbf{r},\mathbf{R})\mathbf{p}w_{n'}(\mathbf{r},\mathbf{R}')d^3r \tag{4.9}$$

is the (Hermitian) momentum matrix in the Wannier basis. It is shown in Appendix D that this matrix takes the following form within the GSVAA:

$$\mathbf{p}_{nn'}(\mathbf{R},\mathbf{R}')=[\delta_{nn'}\mathbf{P}+\boldsymbol{\mathcal{P}}_{nn'}(\mathbf{R})]\,\delta_{\mathbf{R},\mathbf{R}'}. \tag{4.10}$$

Here $\mathbf{P}=-i\hbar\nabla_{\mathbf{R}}$ is the discrete momentum operator,[49] and $\boldsymbol{\mathcal{P}}_{nn'}$ is defined by

$$\boldsymbol{\mathcal{P}}_{nn'}(\mathbf{R})=\mathbf{p}_{nn'}(\mathbf{R})+\mathbf{P}_{nn'}(\mathbf{R}). \tag{4.11}$$

The first term in $\boldsymbol{\mathcal{P}}_{\mathrm{nn}'}$ is just the bulk momentum matrix

$$\mathbf{p}_{nn'}(\mathbf{R})=\int_v u_n^*(\mathbf{r},\mathbf{R})\mathbf{p}u_{n'}(\mathbf{r},\mathbf{R})d^3r, \tag{4.12}$$

but the second is an interface term that is not included in conventional envelope-function theories:

$$\mathbf{P}_{nn'}(\mathbf{R}) = \int_v u_n^*(\mathbf{r},\mathbf{R})\mathbf{P}u_{n'}(\mathbf{r},\mathbf{R})d^3r. \tag{4.13}$$

This matrix describes the strength of the interband coupling induced by the change in basis functions at an interface; it was first derived by Karavaev and Tikhodeev[20] (in a slightly different form appropriate for slowly varying functions). The main difference between $\mathbf{P}_{nn'}$ and $\mathbf{p}_{nn'}$ is the fact that $\mathbf{P}$ operates only on $\mathbf{R}$, not on $\mathbf{r}$. Thus, when we are in the bulk regions of a heterostructure away from any interface, the basis states $u_n(\mathbf{r},\mathbf{R})$ are independent of $\mathbf{R}$, and $\mathbf{P}_{nn'}$ vanishes. Near an interface, however, the basis states have a step-function dependence, so $\mathbf{P}_{nn'}$ behaves like δ_B in this region. This generates interband mixing across the junction, hence $\mathbf{P}_{nn'}$ may lead to qualitative changes in the interface connection rules (which will be discussed further below).

Although the bulk matrix $\mathbf{p}_{nn'}$ is strictly Hermitian, the interface matrix $\mathbf{P}_{nn'}$ is only approximately so. It is shown in Appendix E that $\mathbf{P}_{nn'}$ is not Hermitian in general, but becomes Hermitian when the GSVAA is invoked. The remainder of this paper considers only GSV functions, so any deviations from Hermiticity will be ignored in what follows.

Because the operator $\mathbf{P}$ does not act on $\mathbf{r}$, $\mathbf{P}_{nn'}$ couples only bands of the same symmetry. In general it will have diagonal matrix elements, but for some important cases these terms vanish. For example, at the Γ point of a zinc-blende crystal, the basis functions u_n can always be chosen real because of time-reversal symmetry,[6] which means that $\mathbf{P}_{nn'}$ is purely imaginary. Since this matrix is also Hermitian (within the GSVAA), its diagonal elements must be zero.

C. Approximate Hamiltonian

The above results for the GSV kinetic and potential energy operators may now be inserted into the Hamiltonian (2.10) and written in continuum form using (2.11). This leads to the approximate envelope-function equations

$$\sum_{n'} H_{nn'}(\mathbf{r},\mathbf{p})f_{n'}(\mathbf{r}) = i\hbar\frac{\partial}{\partial t}f_n(\mathbf{r}), \tag{4.14}$$

where

$$\begin{aligned} H_{nn'}(\mathbf{r},\mathbf{p}) = {} & \left[\frac{\mathbf{p}^2}{2m} + E_n(\mathbf{r}) + \Phi(\mathbf{r})\right]\delta_{nn'} \\ & + \frac{1}{2m}[\boldsymbol{\mathcal{P}}_{nn'}(\mathbf{r})\cdot\mathbf{p} + \mathbf{p}\cdot\boldsymbol{\mathcal{P}}_{nn'}(\mathbf{r})] \\ & + K_{nn'}(\mathbf{r}) + Q_{nn'}(\mathbf{r}). \end{aligned} \tag{4.15}$$

This Hamiltonian is very similar in appearance to the bulk $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian.[6] The three diagonal terms, in particular, are exactly the same, except that the band-edge energy E_n is now spatially varying. The terms linear in $\mathbf{p}$ are also spatially varying, and the explicit symmetry of these terms ensures their Hermiticity. This symmetry came directly from the microscopic Schrödinger equation (and the approximate Hermiticity of $\mathbf{P}_{nn'}$); it was not artificially imposed, as is frequently done in phenomenological theories.[2,3,50]

The primary difference from bulk $\mathbf{k}\cdot\mathbf{p}$ theory lies in those terms which depend explicitly on the change in basis functions at the interface, namely, $\mathbf{P}_{nn'}$, $K_{nn'}$, and $Q_{nn'}$. These last two terms are defined as follows:

$$\begin{aligned} K_{nn'} &= \frac{1}{2m}\sum_{n''}(\mathbf{p}_{nn''}\cdot\mathbf{P}_{n''n'} + \mathbf{P}_{nn''}\cdot\mathbf{p}_{n''n'}), \\ Q_{nn'} &= \frac{1}{2m}\sum_{n''}\mathbf{P}_{nn''}\cdot\mathbf{P}_{n''n'}. \end{aligned} \tag{4.16}$$

The matrix $K_{nn'}$ is responsible for the zone-center HH-LH coupling in zinc-blende structures. The bulk momentum matrix $\mathbf{p}_{nn''}$ couples the Γ_{15} valence bands to other Γ_{15} bands (such as those directly above the Γ_1 conduction band) via matrix elements of the type $\langle x|p_y|z''\rangle$; the interface matrix $\mathbf{P}_{n''n'}$ then couples back to the Γ_{15} valence bands through terms such as $\langle z''|P_z|z'\rangle$. Since $\mathbf{P}_{n''n'}$ behaves like δ_B at an abrupt interface, $K_{nn'}$ and $Q_{nn'}$ have this character as well. However, $Q_{nn'}$, unlike $K_{nn'}$, couples only bands of the same symmetry.

The Hamiltonian (4.15) is exactly the same as that derived by Karavaev and Tikhodeev[20] under more restrictive conditions, namely, envelopes and material properties that are locally slowly varying. The above derivation shows that their analysis is actually valid for a much broader range of problems, and that no corrections to their Hamiltonian are required until the Fourier spectrum becomes significant in the outer half of the BZ.

The equations presented here involve sums over an infinite number of bands, which obviously cannot be implemented in practical calculations. The following section therefore considers how we may reduce the number of bands using perturbation theory.[29]

V. SINGLE-BAND EFFECTIVE-MASS MODEL

This section focuses primarily on deriving an effective-mass Hamiltonian and boundary conditions for the nondegenerate Γ_1 conduction band of a zinc-blende heterostructure, although some general results for degenerate bands are also presented. The derivation initially follows that given by Karavaev and Tikhodeev,[20] but extends their results by developing explicit interface connection rules for the envelopes at an abrupt junction. It will be assumed that we are considering a system (e.g., GaAs/$\mathrm{Al_{0.3}Ga_{0.7}As}$) in which, for energies close to the conduction-band edge $E_s(\mathbf{r})$, the dominant envelope function throughout the structure is the conduction-band envelope $f_s(\mathbf{r})$. All other bands $E_j(\mathbf{r})$ are sufficiently remote in energy that the corresponding envelopes $f_j(\mathbf{r})$ are small compared to $f_s(\mathbf{r})$; they will thus be treated as first-order perturbations.

We are interested here in calculating the energy eigenfunctions of (4.14), which are the solutions to

$$\sum_{n'} H_{nn'}(\mathbf{r},\mathbf{p})f_{n'}(\mathbf{r}) = Ef_n(\mathbf{r}). \tag{5.1}$$

The equation for the remote band $n=j$ can be used to find a first-order solution for f_j as a function of f_s:[18,20]

$$f_j = G_j[\tfrac{1}{2}(\boldsymbol{\nabla}\cdot\tilde{\mathbf{p}}_{js} + \tilde{\mathbf{p}}_{js}\cdot\boldsymbol{\nabla}) + K_{js}]f_s, \tag{5.2}$$

where $G_j = (E - E_j - \Phi)^{-1}$, and $\tilde{\mathbf{p}}_{nn'} = (\hbar/im)\mathbf{p}_{nn'}$ is the Kane matrix,[6] which is real and antisymmetric. Since s is a Γ_1 band, symmetry considerations[51] show that (5.2) will be nonzero only if j is a Γ_{15} band. Substitution of (5.2) into (5.1) with $nn' = sj$ yields a single-band effective-mass equation of the form

$$H(\mathbf{r},\mathbf{p},E) f(\mathbf{r}) = E f(\mathbf{r}), \tag{5.3}$$

in which $f(\mathbf{r}) \equiv f_s(\mathbf{r})$ and

$$H = E_s + \Phi + Q_{ss} - \frac{\hbar^2}{2m}\nabla^2 + \sum_j^{\Gamma_{15}} [\tfrac{1}{2}(\boldsymbol{\nabla}\cdot\tilde{\mathbf{p}}_{sj} + \tilde{\mathbf{p}}_{sj}\cdot\boldsymbol{\nabla}) + K_{sj}]$$
$$\times G_j[\tfrac{1}{2}(\boldsymbol{\nabla}\cdot\tilde{\mathbf{p}}_{js} + \tilde{\mathbf{p}}_{js}\cdot\boldsymbol{\nabla}) + K_{js}]. \tag{5.4}$$

Consider now the special case of a layered medium grown along the [001] crystal axis, and let Φ be a function of z alone. Then (5.4) reduces to[20]

$$\begin{aligned} H = {} & E_s(z) + \Phi(z) - \nabla\cdot\frac{\hbar^2}{2m^*(z)}\nabla + Q_{ss}(z) \\ & + \sum_j^{\Gamma_{15}(z)} G_j[K_{sj} - \tfrac{1}{2}(\partial_z \tilde{p}^z_{sj})]^2 \\ & + \sum_j^{\Gamma_{15}(z)} \partial_z\{\tilde{p}^z_{sj} G_j[K_{sj} - \tfrac{1}{2}(\partial_z \tilde{p}^z_{sj})]\}, \end{aligned} \tag{5.5}$$

in which

$$\frac{m}{m^*(z,E)} = 1 + \frac{2}{m}\sum_j^{\Gamma_{15}(z)} \frac{|p^z_{sj}(z)|^2}{E - E_j(z) - \Phi(z)} \tag{5.6}$$

is the energy-dependent effective mass, and $p^z_{sj}(z) = \langle s|p_z|j\rangle$, where j is restricted to states $|z\rangle$ that are of Γ_{15} symmetry. In both sums in (5.5), the operator ∂_z acts only on the function immediately to its right; it does not act upon the envelope f.

The most interesting terms in (5.5) are $Q_{ss}(z)$ and the two sums, since it is these terms that represent the deviation from conventional theory. It is tempting at this stage to approximate the material parameters as abruptly varying, in which case $\tilde{p}^z_{sj}$ is a Heaviside step function and K_{sj} is a Dirac δ function. However, this causes immediate problems, since the penultimate term in (5.5) is then the square of a δ function, i.e., a δ function of infinite weight. The interface is hence an impenetrable barrier, and the only solutions to (5.3) are those in which f vanishes at the interface. This is obviously not a physically reasonable model for a heterostructure. The problem is that the true material properties are not mathematically abrupt; all functions in (5.5) are smoothly varying and finite, due to the quasicontinuum restriction on allowed wave vectors. Therefore, if we wish to approximate the junction as mathematically abrupt, we must multiply the quasicontinuum functions in (5.5) *before* taking the abrupt limit.[52]

Let us take the interface to be at $z=0$, and assume that the material parameters are of the form

$$Q_{ss}(z) = Q^0_{ss}\delta_B(z), \quad K_{sj}(z) = K^0_{sj}\delta_B(z),$$
$$\partial_z \tilde{p}^z_{sj}(z) = \Delta\tilde{p}^z_{sj}\delta_B(z), \tag{5.7}$$

where $\delta_B(z) = \sin(\pi z/a)/\pi z$, $Q^0_{ss} = \int Q_{ss}(z)dz$, and $\Delta\tilde{p}^z_{sj}$ is the change in $\tilde{p}^z_{sj}(z)$ across the interface. Then in the abrupt limit of (5.5) we have simply $\delta_B(z) \to \delta(z)$ and $\delta^2_B(z) \to a^{-1}\delta(z)$. The abrupt approximation to (5.5) can therefore be written as

$$\begin{aligned} H = {} & E_s(z) + \Phi(z) - \nabla\cdot\frac{\hbar^2}{2m^*(z)}\nabla + \Gamma\,\delta(z) \\ & + \partial_z\left[\frac{\hbar^2 C(z)}{2m^*(z)}\,\delta(z)\right], \end{aligned} \tag{5.8}$$

in which

$$\Gamma = Q^0_{ss} + a^{-1}\sum_j^{\Gamma_{15}(z)} G_j(0)(K^0_{sj} - \tfrac{1}{2}\Delta\tilde{p}^z_{sj})^2 \tag{5.9}$$

collects all of the δ-function coefficients, and

$$C(z) = \frac{2m^*(z)}{\hbar^2}\sum_j^{\Gamma_{15}(z)} \tilde{p}^z_{sj}(z) G_j(z)(K^0_{sj} - \tfrac{1}{2}\Delta\tilde{p}^z_{sj}). \tag{5.10}$$

In (5.8) the functions $E_s(z)$, $m^*(z)$, and $C(z)$ have step discontinuities at the interface (but no δ functions). Again, in the last term, ∂_z acts only on the function in brackets, not on the envelope f.

We are now in a position to derive boundary conditions from the effective-mass equation (5.3). Since the Hamiltonian (5.8) has translational invariance in the xy plane, we can choose the envelopes to be of the form $f(\mathbf{r}) = f(z)e^{i\mathbf{k}_\parallel\cdot\mathbf{r}}$, where $\mathbf{k}_\parallel = \hat{\mathbf{x}}k_x + \hat{\mathbf{y}}k_y$. The boundary conditions will be sought in the form of a connection rule between the functions $f(z)$ and $g(z) \equiv f'(z)/m^*(z)$ on opposite sides of the interface. If we define $f_\pm$ to be the value of f just to the right (left) of the interface, then the connection rule can be expressed in terms of the transfer matrix[16]

$$\begin{bmatrix} f_+ \\ g_+ \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}\begin{bmatrix} f_- \\ g_- \end{bmatrix}. \tag{5.11}$$

In order to conserve current density across the interface, the determinant of T must be 1 (to within an overall complex phase factor).[16,25] In the conventional effective-mass theory, this condition is externally imposed by choosing T to be the unit matrix. We will see below that this is not true in general.

The Hamiltonian (5.8) is valid for all z, so the desired boundary conditions can be obtained by integrating the effective-mass equation across the interface.[50,53] This yields the transfer matrix (see Appendix F)

$$T_{11} = \frac{1 + C_0/2}{1 - C_0/2} \qquad T_{12} = 0,$$
$$T_{21} = \frac{2\Gamma}{\hbar^2(1 - C_0^2/4),} \quad T_{22} = \frac{1 - C_0/2}{1 + C_0/2}, \tag{5.12}$$

in which $C_0 = C(z=0)$. This matrix obviously conserves current density, but it differs from the unit matrix in two significant ways. First, the parameter C_0 generates a discontinuity in the envelope function, which comes from the δ-function derivative in (5.8). If C_0 is small, then the connection rule reduces to $f_+ \cong (1+C_0) f_-$, hence C_0 gives the approximate fractional change in f at the interface. Second, the δ-function coefficient Γ modifies the connection rule for the slope of the envelope.

This agrees with the transfer matrices calculated from other microscopic models,[10,16] but here the origin of these terms is clear. They come from the change in basis functions across the interface, which appears in $\mathbf{P}_{nn'}$ and in the spatial dependence of $\mathbf{p}_{nn'}$. Note that the off-diagonal element T_{12} vanishes, as suggested by Laikhtman.[21] This term is treated as generally nonzero in the theory of Ando *et al.*,[16] but their numerical calculations show that it is always negligible unless T is used to connect two bands of different symmetry (e.g., Γ and X). This is not the case treated here, so (5.12) agrees with their results.

The specific formalism used by Ando *et al.*[16] has been criticized by Cuypers and van Haeringen[10] because it is based on a current conservation law that assumes parabolic bands, which the latter authors have shown to be invalid in real semiconductors. For example, in a GaAs/Al$_x$Ga$_{1-x}$As structure, the (complex) dispersion of Al$_x$Ga$_{1-x}$As may be significantly nonparabolic at the band edge of GaAs. The same point was raised by the present author in connection with the envelope-function theory of optical phonons.[37] It is therefore important to include the energy dependence of the effective mass (5.6) in most calculations.

The above derivation was limited to a single nondegenerate band. It can, however, easily be extended to degenerate or quasidegenerate bands. Consider a model in which a finite set of bands n is treated directly, with the effects of other remote bands j included as second-order perturbations. The perturbations renormalize Eq. (4.15) as follows:

$$\tilde{H}_{nn'} = H_{nn'} + \sum_j \left[\frac{1}{2m} (\mathbf{p}\cdot\mathbf{p}_{nj} + \mathbf{p}_{nj}\cdot\mathbf{p}) + K_{nj} \right] \times G_j \left[\frac{1}{2m} (\mathbf{p}\cdot\mathbf{p}_{jn'} + \mathbf{p}_{jn'}\cdot\mathbf{p}) + K_{jn'} \right]. \tag{5.13}$$

This generates effects similar to those found in the nondegenerate case (viz., discontinuities and δ functions), but the main qualitative differences from conventional theory are those interface effects already included in (4.15), such as the HH-LH coupling due to $K_{nn'}$. A more detailed analysis and discussion of (5.13) will be presented elsewhere.

VI. COMPARISON WITH OTHER MODELS

If the zone-center Bloch functions are the same on both sides of a heterojunction, all of the interface effects described in the previous section vanish. The transfer matrix (5.12) consequently reduces to the unit matrix, and the model derived above reduces to the conventional envelope-function model proposed by Bastard[4] and others on the grounds of current conservation. The present derivation therefore supports the use of this model in situations where the difference in basis functions is small enough to be neglected—provided one maintains the correct operator ordering[5,18] given by (5.13). Note, however, that a small change in basis functions is not always negligible, since it can lead to qualitative differences such as the HH-LH coupling discussed above.

The present theory is very similar to the exact envelope-function theory of Burt,[18] which uses a material-independent Luttinger-Kohn basis $U_n(\mathbf{r})$ to represent the entire heterostructure. This basis is related to the modified Wannier basis (2.6) by the unitary transformation

$$u_n(\mathbf{r},\mathbf{R}) = \sum_{n'} A_{n'n}(\mathbf{R}) U_{n'}(\mathbf{r}),$$

$$U_n(\mathbf{r}) = \sum_{n'} A^*_{nn'}(\mathbf{R}) u_{n'}(\mathbf{r},\mathbf{R}), \tag{6.1}$$

in which [cf. (C3)]

$$A_{nn'}(\mathbf{R}) = \int_v U^*_n(\mathbf{r}) u_{n'}(\mathbf{r},\mathbf{R}) d^3r. \tag{6.2}$$

The Burt envelopes F_n are therefore given in terms of the f_n by

$$f_n(\mathbf{R}) = \sum_{n'} A^*_{n'n}(\mathbf{R}) F_{n'}(\mathbf{R}),$$

$$F_n(\mathbf{R}) = \sum_{n'} A_{nn'}(\mathbf{R}) f_{n'}(\mathbf{R}). \tag{6.3}$$

One can in fact derive the Burt formalism from the present theory without resorting to such mathematics, since $U_n(\mathbf{r})$ is merely the special case in which $u_n(\mathbf{r},\mathbf{R})$ is chosen to be independent of $\mathbf{R}$. The bulk momentum matrix $\mathbf{p}_{nn'}$ is thus a constant, while the interface matrices $\mathbf{6}_{nn'}$, $K_{nn'}$, and $Q_{nn'}$ are zero. This enormously simplifies the description of the interface. However, $U_n(\mathbf{r})$ no longer diagonalizes the bulk Hamiltonian, so in (4.15) the diagonal matrix $E_n(\mathbf{r})\delta_{nn'}$ must be replaced by a nondiagonal matrix $W_{nn'}(\mathbf{r})$; the off-diagonal terms are responsible for the differences between the Burt theory and conventional theory. Their effect may be seen by replacing K_{nj} in (5.13) with W_{nj}; this generates HH-LH coupling as before, but now as a *bulk* effect due to linear-$\mathbf{k}$ terms in the bulk dispersion (as in the perturbation theory of Smith and Mailhiot[9]). These linear terms cannot be neglected in general because their coefficients are not the same as those in the standard $\mathbf{k}\cdot\mathbf{p}$ theory.[6,54] Hence a treatment of such effects in the Burt formalism requires one to tamper with the form of the bulk dispersion, and the simplicity of the interface is achieved at the cost of added complexity in the bulk. The present theory has the opposite effect. Which model is chosen to describe a given problem is thus primarily a matter of aesthetics.

As mentioned in Sec. IV, the approximate Hamiltonian (4.15) is identical to one derived previously by Karavaev and Tikhodeev.[20] The unitary transformation (6.1) shows the connection between the two theories. Karavaev and Tikhodeev derived their model from just such a transformation, but one that was continuous rather than discrete. As a consequence, their formalism was not unique, and to maintain the uniqueness of their model they were forced to limit the analysis to cases in which the material properties and envelope functions were slowly varying in $\mathbf{r}$ space. The present formalism, by starting from an exact representation, is able to demonstrate explicitly what corrections to this model are needed and when they arise, thus proving that their approximations remain valid even at abrupt interfaces. Questions of uniqueness aside, their derivation offers a quick, easily understandable route to the Hamiltonian (4.15), and the author highly recommends a perusal of this paper[20] for an alternative view of the problem.

Indeed, the discussion of the GSVAA in Sec. III opens up the possibility that other effective-mass derivations in the literature, which were previously thought to apply only to smoothly graded media, may in fact be valid for abrupt heterostructures as well. Most such derivations,[39,41–44,46] however, cannot be generalized in this manner, because they invoke further approximations that do not correctly account for the spatial variation of the material properties. Two notable exceptions are the theories of Leibler[40] and Young,[45] both of whom obtained effective-mass equations similar to (5.5). Leibler used a material-independent Luttinger-Kohn basis and treated the $\mathbf{k}\cdot\mathbf{p}$ interaction and the inhomogeneous potential as perturbations, much the same as Smith and Mailhiot's approach to abrupt junctions.[9] Young used perturbation theory in a material-independent Wannier basis, arriving at essentially the same result as Leibler (with minor differences due to a small error in Leibler's solution).

von Roos and Mavromatis[44] and Elçi[55] have given derivations almost identical to Leibler's, but they incorrectly concluded that the effective mass is independent of position, because they failed to include the third-order perturbations necessary to describe its position dependence. This highlights a common feature of all derivations that treat the material dependence of the basis functions as a perturbation, namely, that they yield only an approximation (good if applied correctly,[9,18] poor if not[44,55]) to the true spatial variation of the material properties. In contrast, the material-dependent basis functions used in the present paper reproduce the correct $\mathbf{k}\cdot\mathbf{p}$ equations in every bulk medium in the heterostructure, regardless of how much the microscopic potential changes between media.

Material-dependent Wannier functions have been used previously in several envelope-function models based on the tight-binding method.[15–17] However, the complexity of this approach normally limits its application to nearest-neighbor models with a small set of basis functions, and completeness and orthogonality issues are not usually addressed. As discussed in Sec. V, the present theory reproduces all qualitative features of the interface connection rules derived from such models, but it substantially improves upon these results by providing a straightforward extension to a complete orthonormal basis set.

The most rigorous envelope-function connection rules thus far have been those derived from large-basis pseudopotential calculations.[9–11] This technique, although highly accurate, requires specific knowledge of the microscopic basis functions throughout the calculation, and it is rather difficult to glean much intuitive information from such an approach. In contrast, the present theory condenses all microscopic information into the interface matrix $\mathbf{P}_{nn'}$. The envelope-function equations can thus be manipulated analytically, and the qualitative effects of the change in basis functions at an interface can be determined without the need for extensive numerical work.

VII. SUMMARY AND CONCLUSIONS

This paper has presented an envelope-function model that takes explicit account of the change in band-edge Bloch functions at an interface between two semiconductors. The formalism was derived by choosing a set of material-dependent, spatially localized basis functions that diagonalize the bulk zone-center Hamiltonian in each unit cell of the crystal. This modified Wannier basis is complete and orthonormal, thus providing an exact envelope-function representation of the Schrödinger equation. An approximate model was developed by assuming that all terms arising exclusively from wave vectors in the outer half of the Brillouin zone could be neglected; this approximation was shown to be valid even for highly discontinuous functions. Differences from conventional envelope-function theory appeared in the kinetic energy (but not the potential energy), where the material dependence of the basis functions led to new interband mixings, such as zone-center HH-LH coupling.

Perturbation theory was used to derive a single-band effective-mass equation. This equation contained kinetic-energy terms proportional (in the abrupt limit) to a Dirac δ function and its derivative. Connection rules were derived for the envelope function, showing that the δ' term gives rise to an amplitude discontinuity, while the δ term modifies the connection rule for the slope. These results were shown to be in accord with the transfer matrix obtained by other methods.

The derivation presented in this paper has the advantage of showing clearly the underlying unity of the subject matter. Although the interface effects described here are well known, they could not previously be included in a rigorous envelope-function model without intensive numerical calculations. The present theory offers a simple intuitive description of how such effects arise, and it can be used to predict the behavior of the envelopes in general qualitative terms without resorting to microscopic calculations. Furthermore, it demonstrates conclusively that envelope-function techniques are indeed valid at an abrupt interface. Thus the phenomenological approach of including the interface region in an effective-mass equation[50] is fully justified, and the strong criticism[19] leveled at such theories (as unjustifiable in principle) now seems overly pessimistic. If the interface is treated correctly, the interface effective-mass equation is completely equivalent to the transfer-matrix approach, so there is no reason to favor one technique over the other.

ACKNOWLEDGMENTS

I wish to thank Mike Burt for valuable discussions and advice. This material is based upon work supported by the North Atlantic Treaty Organization under a grant awarded in 1995.

APPENDIX A: COMPLETENESS AND ORTHOGONALITY

Consider first the Luttinger-Kohn functions $u_n(\mathbf{r},\mathbf{R})$. Because these functions are periodic in $\mathbf{r}$, we can expand them in a Fourier series:

$$u_n(\mathbf{r},\mathbf{R}) = v^{-1/2} \sum_{\mathbf{G}} u_{n\mathbf{G}}(\mathbf{R}) e^{i\mathbf{G}\cdot\mathbf{r}},$$

$$u_{n\mathbf{G}}(\mathbf{R}) = v^{-1/2} \int_v u_n(\mathbf{r},\mathbf{R}) e^{-i\mathbf{G}\cdot\mathbf{r}}\, d^3r, \tag{A1}$$

where the sum is over all reciprocal lattice vectors **G**, and the integral is over any primitive cell volume v. The Fourier coefficients $u_{n\mathbf{G}}(\mathbf{R})$ are complete and orthonormal[18,30] with respect to the plane waves $e^{i\mathbf{G}\cdot\mathbf{r}}$:

$$\sum_n u_{n\mathbf{G}}(\mathbf{R}) u^*_{n\mathbf{G}'}(\mathbf{R}) = \delta_{\mathbf{G},\mathbf{G}'},$$

$$\sum_{\mathbf{G}} u^*_{n\mathbf{G}}(\mathbf{R}) u_{n'\mathbf{G}}(\mathbf{R}) = \delta_{nn'}. \tag{A2}$$

These equations are independent of **R**, and are equivalent to the following relations for $u_n(\mathbf{r},\mathbf{R})$:

$$\sum_n u_n(\mathbf{r},\mathbf{R}) u^*_n(\mathbf{r}',\mathbf{R}) = v^{-1} \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot(\mathbf{r}-\mathbf{r}')}$$

$$= \sum_{\mathbf{R}'} \delta(\mathbf{r}-\mathbf{r}'-\mathbf{R}'),$$

$$\int_v u^*_n(\mathbf{r},\mathbf{R}) u_{n'}(\mathbf{r},\mathbf{R}) d^3r = \delta_{nn'}. \tag{A3}$$

The former equation, which is to be integrated over any cell v, expresses the completeness of the basis $u_n(\mathbf{r},\mathbf{R})$ with respect to cell-periodic functions. (This was first noted by Luttinger and Kohn.[28]) The latter specifies the normalization convention used in this paper.

The completeness and orthogonality of the modified Wannier basis $w_n(\mathbf{r},\mathbf{R})$ may be proved easily from the above relations. One simply inserts the definition (2.6) into the left-hand side of (2.7) and expands the functions $\delta_B(\mathbf{r}-\mathbf{R})$ and $u_n(\mathbf{r},\mathbf{R})$ in Fourier space using (2.3) and (A1). The right-hand side of (2.7) then follows directly from Eqs. (A2) upon use of the identities[30]

$$\sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} = \frac{(2\pi)^3}{v} \sum_{\mathbf{G}} \delta(\mathbf{k}-\mathbf{G}) \tag{A4}$$

and

$$B(\mathbf{k}) B(\mathbf{k}+\mathbf{G}) = B(\mathbf{k}) \delta_{\mathbf{G},\mathbf{0}}. \tag{A5}$$

Note that the above proof of (2.7) is not limited to basis functions taken from the center of the Brillouin zone. A trivial extension is to replace $u_n(\mathbf{r},\mathbf{R})$ in (2.6) with $u_{n\mathbf{k}_0}(\mathbf{r},\mathbf{R})$, where $\mathbf{k}_0$ is an arbitrary function of **R**. One could also replace $u_n(\mathbf{r},\mathbf{R})$ with $\psi_{n\mathbf{k}_0}(\mathbf{r},\mathbf{R})$, but then (2.7) is valid only if $\mathbf{k}_0$ is independent of **R**.

APPENDIX B: $\delta_B(\mathbf{r})$

The behavior of $\delta_B(\mathbf{r})$ is easiest to see from a specific example. Let the primitive translations in the direct and reciprocal lattices be denoted by $\mathbf{a}_i$ and $\mathbf{b}_i$, respectively, with $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$, and let the coordinate systems be defined by $\mathbf{r} = \Sigma_i x_i \mathbf{a}_i$ and $\mathbf{k} = \Sigma_i k_i \mathbf{b}_i$, where lattice sites occur at integer values of x_i and k_i. Then the simplest possible Brillouin zone is the parallelepiped $|k_i| < \frac{1}{2}$, which yields for δ_B the three-dimensional sinc function

$$\delta_B(\mathbf{r}) = \frac{1}{v} \prod_{i=1}^{3} \frac{\sin(\pi x_i)}{\pi x_i}. \tag{B1}$$

Other choices of BZ (such as the conventional Wigner-Seitz cell) will lead to a more complicated functional form for $\delta_B(\mathbf{r})$, but Eq. (2.5) will be satisfied at the lattice sites for any such choice.

APPENDIX C: POTENTIAL ENERGY

In this appendix, Eq. (4.4) is derived by applying the GSVAA to the flat-band matrix element

$$V^{\mathrm{FB}}_{nn'}(\mathbf{R},\mathbf{R}') = v \sum_{\mathbf{R}''} \int w^*_n(\mathbf{r},\mathbf{R}) \delta_B(\mathbf{r}-\mathbf{R}'')$$

$$\times V_0(\mathbf{r},\mathbf{R}'') w_{n'}(\mathbf{r},\mathbf{R}') d^3r. \tag{C1}$$

The first step is to move the right-hand basis function to the location $\mathbf{R}''$ by applying the transformation

$$u_{n'}(\mathbf{r},\mathbf{R}') = \sum_{n'''} A_{n'''n'}(\mathbf{R}'',\mathbf{R}') u_{n'''}(\mathbf{r},\mathbf{R}''), \tag{C2}$$

where

$$A_{nn'}(\mathbf{R},\mathbf{R}') = \int_v u^*_n(\mathbf{r},\mathbf{R}) u_{n'}(\mathbf{r},\mathbf{R}') d^3r \tag{C3}$$

is a unitary matrix describing the degree of orthogonality of the basis functions in different unit cells [cf. (A3)]. The transformation (C2) leaves us with a product $V_0(\mathbf{r},\mathbf{R}'') u_{n'''}(\mathbf{r},\mathbf{R}'')$. Since this is a periodic function of **r**, we can expand it in terms of the complete periodic basis $u_n(\mathbf{r},\mathbf{R}'')$, namely, as

$$V_0(\mathbf{r},\mathbf{R}'') u_{n'''}(\mathbf{r},\mathbf{R}'') = \sum_{n''} V^0_{n''n'''}(\mathbf{R}'') u_{n''}(\mathbf{r},\mathbf{R}''), \tag{C4}$$

where the expansion coefficients are the bulk matrix elements (4.5). The flat-band operator is therefore of the form

$$V^{\mathrm{FB}}_{nn'}(\mathbf{R},\mathbf{R}') = \sum_{n''n'''} \sum_{\mathbf{R}''} \Delta_{nn''}(\mathbf{R},\mathbf{R}',\mathbf{R}'')$$

$$\times V^0_{n''n'''}(\mathbf{R}'') A_{n'''n'}(\mathbf{R}'',\mathbf{R}'), \tag{C5}$$

where the only quantity not yet defined is the integral

$$\Delta_{nn''}(\mathbf{R},\mathbf{R}',\mathbf{R}'') = v \int w^*_n(\mathbf{r},\mathbf{R}) \delta_B(\mathbf{r}-\mathbf{R}') w_{n''}(\mathbf{r},\mathbf{R}'') d^3r. \tag{C6}$$

This gives us an expression for $V^{\mathrm{FB}}_{nn'}(\mathbf{R},\mathbf{R}')$ in terms of $V^0_{nn'}(\mathbf{R})$, but it is in general very complex, with many nonlocal and interband couplings. Let us see if we can simplify it by invoking the GSVAA. We can start by inserting into (C6) the Fourier expansions (2.3) and (A1). Upon performing the integral over **r**, one finds

$$\Delta_{nn''}(\mathbf{R},\mathbf{R}',\mathbf{R}'')=\frac{v^2}{(2\pi)^6}\int d^3k\int d^3k' \times\int d^3k''\sum_{\mathbf{G},\mathbf{G}''}u^*_{n\mathbf{G}}(\mathbf{R})u_{n''\mathbf{G}''}(\mathbf{R}'') \times\delta(\mathbf{k}+\mathbf{k}'-\mathbf{k}''+\mathbf{G}-\mathbf{G}'')B(\mathbf{k})B(\mathbf{k}') \times B(\mathbf{k}'')e^{i\mathbf{k}\cdot\mathbf{R}}e^{i\mathbf{k}'\cdot\mathbf{R}'}e^{-i\mathbf{k}''\cdot\mathbf{R}''}. \tag{C7}$$

Integrating over $\mathbf{k}''$, this becomes

$$\Delta_{nn''}(\mathbf{R},\mathbf{R}',\mathbf{R}'')=\frac{v^2}{(2\pi)^6}\int d^3k\int d^3k'\sum_{\mathbf{G},\mathbf{G}''}u^*_{n\mathbf{G}}(\mathbf{R})u_{n''\mathbf{G}''} \times(\mathbf{R}'')B(\mathbf{k})B(\mathbf{k}')B(\mathbf{k}+\mathbf{k}'+\mathbf{G}-\mathbf{G}'') \times e^{i\mathbf{k}\cdot(\mathbf{R}-\mathbf{R}'')}e^{i\mathbf{k}'\cdot(\mathbf{R}'-\mathbf{R}'')}. \tag{C8}$$

Now if $\mathbf{k}$ and $\mathbf{k}'$ were limited to the inner half of the BZ, the third B function would be nonzero only for $\mathbf{G}=\mathbf{G}''$. Let us assume that this remains true for all $\mathbf{k}$ and $\mathbf{k}'$. Then (C8) can be evaluated immediately using (A2) and (2.3); the result is

$$\Delta_{nn''}(\mathbf{R},\mathbf{R}',\mathbf{R}'')=\delta_{nn''}\delta_{\mathbf{R},\mathbf{R}''}\delta_{\mathbf{R}',\mathbf{R}''}. \tag{C9}$$

Upon inserting this in (C5) and using $A_{n'''n'}(\mathbf{R}',\mathbf{R}')=\delta_{n'''n'}$ [from (C3) and (A3)], the flat-band operator reduces to the strikingly simple expression (4.4). Hence all of the complications in (C5) drop out, and we are left with nothing but the local bulk matrix element.

The justification for the above steps is as follows. In general we must allow for $\mathbf{G}\neq\mathbf{G}''$ in (C8), and there will consequently be corrections to (C9) arising from integrals over the regions where $\mathbf{k}+\mathbf{k}'$ is out of the BZ. However, if we insert (C5) and (C7) into the variational expression for the energy [cf. (2.9)]

$$E=v\sum_{n,n'}\sum_{\mathbf{R},\mathbf{R}'}f^*_n(\mathbf{R})H_{nn'}(\mathbf{R},\mathbf{R}')f_{n'}(\mathbf{R}'), \tag{C10}$$

then it becomes clear that these corrections will be negligible if the functions of $\mathbf{R}$ and $\mathbf{R}'$ which couple to the $e^{i\mathbf{k}\cdot\mathbf{R}}$ and $e^{i\mathbf{k}'\cdot\mathbf{R}'}$ terms in (C7) are GSV. The relevant function of $\mathbf{R}$ is $f^*_n(\mathbf{R})u^*_{nG}(\mathbf{R})$, while that of $\mathbf{R}'$ is $A_{n'''n'}(\mathbf{R}'',\mathbf{R}')f_{n'}(\mathbf{R}')$. Thus the solution (4.4) is valid if the envelopes and material properties are small in the outer half of the BZ.

APPENDIX D: MOMENTUM MATRIX

The momentum matrix (4.9) can be represented exactly in the form

$$\mathbf{p}_{nn'}(\mathbf{R},\mathbf{R}')=[\delta_{nn'}\mathbf{P}+\boldsymbol{\mathcal{P}}_{nn'}(\mathbf{R})]\delta_{R,R'}+\mathbf{D}_{nn'}(\mathbf{R},\mathbf{R}'). \tag{D1}$$

This differs from (4.10) only in the extra term $\mathbf{D}_{nn'}(\mathbf{R},\mathbf{R}')$, which is negligible within the GSVAA. To derive (D1), we can start by inserting the definition (2.6) of the right-hand Wannier function into (4.9):

$$\mathbf{p}_{nn'}(\mathbf{R},\mathbf{R}')=v\int w^*_n(\mathbf{r},\mathbf{R})\delta_B(\mathbf{r}-\mathbf{R}')\mathbf{p}u_{n'}(\mathbf{r},\mathbf{R}')d^3r +v\int w^*_n(\mathbf{r},\mathbf{R})u_{n'}(\mathbf{r},\mathbf{R}')\mathbf{p}\delta_B(\mathbf{r}-\mathbf{R}')d^3r. \tag{D2}$$

In the first term, the function $\mathbf{p}u_{n'}(\mathbf{r},\mathbf{R}')$ is periodic in $\mathbf{r}$, so it can be expanded over the complete periodic basis $u_n(\mathbf{r},\mathbf{R})$ as

$$\mathbf{p}u_{n'}(\mathbf{r},\mathbf{R}')=\sum_{n''}\mathbf{p}_{n''n'}(\mathbf{R}')u_{n''}(\mathbf{r},\mathbf{R}'), \tag{D3}$$

in which $\mathbf{p}_{nn'}(\mathbf{R})$ is the bulk matrix element (4.12). In the second term, we can use the quasicontinuum envelope definition (2.11) to obtain the expansion

$$\mathbf{p}\delta_B(\mathbf{r}-\mathbf{R}')=\sum_{\mathbf{R}''}\delta_B(\mathbf{r}-\mathbf{R}'')\mathbf{P}''\delta_{\mathbf{R}'',\mathbf{R}'}. \tag{D4}$$

To put the results back into Wannier-function form, we can use the unitary transformation (C2). The integrals in (D2) can then be performed using the orthogonality relation (2.7), with the result

$$\mathbf{p}_{nn'}(\mathbf{R},\mathbf{R}')=\mathbf{p}_{nn'}(\mathbf{R})\delta_{\mathbf{R},\mathbf{R}'}+A_{nn'}(\mathbf{R},\mathbf{R}')\mathbf{P}\delta_{\mathbf{R},\mathbf{R}'}. \tag{D5}$$

The next step is to work the second term in (D5) into a more convenient form. To see more clearly which terms can be neglected under the GSVAA, it helps to couple this operator directly to an envelope function [cf. (2.9)]:

$$\mathbf{c}_n(\mathbf{R})=\sum_{n',\mathbf{R}'}[A_{nn'}(\mathbf{R},\mathbf{R}')\mathbf{P}\delta_{\mathbf{R},\mathbf{R}'}]f_{n'}(\mathbf{R}'). \tag{D6}$$

We can use Fourier transforms to rewrite this function as

$$\mathbf{c}_n(\mathbf{R})=\frac{v}{(2\pi)^6}\sum_{n',\mathbf{R}'}\int\int\int B(\mathbf{k})B(\mathbf{k}') \times B(\mathbf{k}'')\hbar\mathbf{k}''e^{i\mathbf{k}''\cdot\mathbf{R}}e^{i(\mathbf{k}+\mathbf{k}'-\mathbf{k}'')\cdot\mathbf{R}'} \times A_{nn'}(\mathbf{R},\mathbf{k})f_{n'}(\mathbf{k}')d^3k\ d^3k'\ d^3k''. \tag{D7}$$

The sum over $\mathbf{R}'$ can be performed using the identity (A4); then integrating over $\mathbf{k}''$, this becomes

$$\mathbf{c}_n(\mathbf{R})=(2\pi)^{-3}\sum_{n'}\int\int B(\mathbf{k})B(\mathbf{k}') \times\hbar[\mathbf{k}+\mathbf{k}'+\mathbf{G}(\mathbf{k}+\mathbf{k}')]e^{i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{R}} \times A_{nn'}(\mathbf{R},\mathbf{k})f_{n'}(\mathbf{k}')d^3k\ d^3k', \tag{D8}$$

where $\mathbf{G}(\mathbf{k}+\mathbf{k}')$ is the reciprocal lattice vector that brings the sum $\mathbf{k}+\mathbf{k}'+\mathbf{G}$ into the first BZ. The terms in (D8) proportional to $\mathbf{k}$ and $\mathbf{k}'$ are easily identifiable as spatial derivatives of $A_{nn'}$ and $f_{n'}$. We can therefore transform (D8) back to $\mathbf{R}$ space and find

$$\mathbf{c}_n(\mathbf{R}) = \sum_{n'} \mathbf{P}_{nn'}(\mathbf{R}) f_{n'}(\mathbf{R}) + \mathbf{P} f_n(\mathbf{R}) + \sum_{n'\mathbf{R}'} \mathbf{D}_{nn'}(\mathbf{R},\mathbf{R}') f_{n'}(\mathbf{R}'), \tag{D9}$$

which completes the derivation of (D1). The nonlocal operator $\mathbf{D}_{nn'}(\mathbf{R},\mathbf{R}')$ is defined by the integral

$$\mathbf{D}_{nn'}(\mathbf{R},\mathbf{R}') = \frac{v}{(2\pi)^{9/2}} \int\int \hbar \mathbf{G}(\mathbf{k}+\mathbf{k}') \times A_{nn'}(\mathbf{R},\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{R}} e^{i\mathbf{k}'\cdot(\mathbf{R}-\mathbf{R}')} d^3k\, d^3k', \tag{D10}$$

which can be calculated explicitly:

$$D_{nn'}(\mathbf{R},\mathbf{R}') = [A_{nn'}(\mathbf{R},\mathbf{R}') - \delta_{nn'}]\mathbf{P}\delta_{\mathbf{R},\mathbf{R}'} - \mathbf{P}_{nn'}(\mathbf{R})\delta_{\mathbf{R},\mathbf{R}'}. \tag{D11}$$

This operator satisfies

$$\sum_{\mathbf{R}'} \mathbf{D}_{nn'}(\mathbf{R},\mathbf{R}') = \mathbf{0}, \tag{D12}$$

so it will couple weakly to slowly varying envelopes.[18,35] Indeed, since $\mathbf{G}(\mathbf{k}+\mathbf{k}')$ is nonzero only where $\mathbf{k}+\mathbf{k}'$ is outside the BZ, it is clear from (D8) that $\mathbf{D}_{nn'}$ can be discarded if $A_{nn'}$ and $f_{n'}$ are GSV. Hence Eq. (4.10) will be an excellent approximation to (D1) under most circumstances.

This result can be used to derive a simple expression for the probability current density. The microscopic current density is defined by

$$\mathbf{J}(\mathbf{r}) = \frac{1}{m} \operatorname{Re}[\psi^*(\mathbf{r})\mathbf{p}\psi(\mathbf{r})]. \tag{D13}$$

Upon expanding $\psi(\mathbf{r})$ in the modified Wannier basis and applying the GSVAA, one finds that the first-zone part of $\mathbf{J}(\mathbf{r})$ is approximately

$$\mathbf{J}_B(\mathbf{r}) = \frac{1}{m} \sum_{nn'} \operatorname{Re}\{f_n^*(\mathbf{r})[\delta_{nn'}\mathbf{p} + \boldsymbol{\mathcal{P}}_{nn'}(\mathbf{r})] f_{n'}(\mathbf{r})\}, \tag{D14}$$

where $\mathbf{J}_B(\mathbf{r})$ is defined in (3.1).

APPENDIX E: HERMITICITY OF THE INTERFACE MATRIX

To establish whether $\mathbf{P}_{nn'}(\mathbf{R})$ is Hermitian, we must be able to calculate discrete derivatives[49] of the basis functions $u_n(\mathbf{r},\mathbf{R})$. The first step is to define the quasicontinuum basis functions [cf. (2.11)]

$$u_n(\mathbf{r},\mathbf{r}') = v \sum_{\mathbf{R}} u_n(\mathbf{r},\mathbf{R})\delta_B(\mathbf{R}-\mathbf{r}'). \tag{E1}$$

Now consider the integral

$$I_{nn'}(\mathbf{r}) = \int_v u_n^*(\mathbf{r}',\mathbf{r}) u_{n'}(\mathbf{r}',\mathbf{r}) d^3r'. \tag{E2}$$

At the lattice sites $\mathbf{r}=\mathbf{R}$, we have

$$I_{nn'}(\mathbf{R}) = \delta_{nn'}, \tag{E3}$$

which follows from the orthogonality (A3) of the discrete basis functions. However, for $\mathbf{r}\neq\mathbf{R}$, the quasicontinuum functions (E1) are not orthogonal:

$$I_{nn'}(\mathbf{r}) = v^2 \sum_{\mathbf{R}',\mathbf{R}''} \delta_B(\mathbf{r}-\mathbf{R}') A_{nn'}(\mathbf{R}',\mathbf{R}'') \delta_B(\mathbf{R}''-\mathbf{r}), \tag{E4}$$

where $A_{nn'}$ is the unitary matrix (C3). This lack of orthogonality means that $\mathbf{P}_{nn'}(\mathbf{R})$ is not generally Hermitian. If we take the gradient of Eq. (E2), multiply by $\hbar/i$, and evaluate the resulting function at a lattice site, we find the relation

$$\mathbf{P}_{nn'}(\mathbf{R}) - \mathbf{P}_{n'n}^*(\mathbf{R}) = [\mathbf{p} I_{nn'}(\mathbf{r})]|_{\mathbf{r}=\mathbf{R}}, \tag{E5}$$

which does not vanish in general. However, if the basis functions $u_n(\mathbf{r},\mathbf{R})$ are GSV—that is, if $u_n(\mathbf{r},\mathbf{k})$ is negligible in the outer half of the BZ—then $I_{nn'}(\mathbf{r})$ is a quasicontinuum function, which implies that $I_{nn'}(\mathbf{r})=\delta_{nn'}$ everywhere [see (E3) and (2.11)], and the right-hand side of (E5) vanishes. Therefore, even though $\mathbf{P}_{nn'}(\mathbf{R})$ is not strictly Hermitian, it may be treated as such within the context of the GSVAA.

APPENDIX F: INTERFACE TRANSFER MATRIX

Multiplying (5.3) by $2/\hbar^2$ and integrating from $z_0<0$ to z, one finds

$$g(z) - g(z_0) = \frac{2}{\hbar^2}\int_{z_0}^{z} R(z')dz' + \frac{C(z)}{m^*(z)} f(z)\delta(z) - C_0 g_0 \theta(z), \tag{F1}$$

in which $\theta(z)$ is the Heaviside step function, $g_0 = g(z=0) = \frac{1}{2}(g_+ + g_-)$, and

$$R(z) = [E_s(z) + \Phi(z) + \Gamma\delta(z) + \hbar^2 k_\parallel^2/2m^*(z) - E] f(z). \tag{F2}$$

If we now multiply (F1) by $m^*(z)$, integrate between $-\varepsilon$ and ε, and let $\varepsilon\to 0$, the result is

$$f_+ - f_- = C_0 f_0, \tag{F3}$$

which specifies T_{11} and T_{12} in (5.12). The boundary condition on g is obtained from (F1) by setting $z_0=-\varepsilon$, $z=\varepsilon$, and taking the limit $\varepsilon\to 0$, which yields

$$g_+ - g_- = (2/\hbar^2)\Gamma f_0 - C_0 g_0. \tag{F4}$$

This gives T_{21} and T_{22} after a rearrangement of terms.

[1] M. Altarelli, in *Heterojunctions and Semiconductor Superlattices*, edited by G. Allan *et al.* (Springer, Berlin, 1986), p. 12.

[2] W. Pötz and D. K. Ferry, Superlatt. Microstruct. **3**, 57 (1987).

[3] R. Eppenga, M. F. H. Schuurmans, and S. Colak, Phys. Rev. B **36**, 1554 (1987).

[4] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures* (Wiley, New York, 1988).

[5] B. A. Foreman, Phys. Rev. B **48**, 4964 (1993).

[6] E. O. Kane, in *Physics of III-V Compounds*, edited by R. K. Willardson and A. C. Beer, Semiconductors and Semimetals Vol. 1 (Academic, New York, 1966), p. 75.

[7] P. Enders, A. Bärwolff, M. Woerner, and D. Suisky, Phys. Rev. B **51**, 16 695 (1995).

[8] J. N. Schulman and Y. C. Chang, Phys. Rev. B **31**, 2056 (1985).

[9] D. L. Smith and C. Mailhiot, Phys. Rev. B **33**, 8345 (1986); **33**, 8360 (1986); J. Vac. Sci. Technol. B **8**, 793 (1990); Rev. Mod. Phys. **62**, 173 (1990).

[10] J. P. Cuypers and W. van Haeringen, J. Phys. Condens. Matter **4**, 2587 (1992); Phys. Rev. B **47**, 10 310 (1993); **48**, 11 469 (1993).

[11] C. Aversa and J. E. Sipe, Phys. Rev. B **47**, 6590 (1993); Appl. Phys. Lett. **63**, 1975 (1993); Phys. Rev. B **49**, 14 542 (1994).

[12] G. Edwards and J. C. Inkson, Semicond. Sci. Technol. **9**, 178 (1994); **9**, 310 (1994); Solid State Commun. **89**, 595 (1994); **91**, 84(E) (1994).

[13] R. Sooryakumar, D. S. Chemla, A. Pinczuk, A. Gossard, W. Wiegmann, and L. J. Sham, Solid State Commun. **54**, 859 (1985).

[14] T. B. Norris, N. Vodjdani, B. Vinter, E. Costard, and E. Böckenhoff, Phys. Rev. B **43**, 1867 (1991).

[15] P. J. Price, in *Proceedings of the International Conference on the Physics of Semiconductors, Exeter, 1962,* edited by A. C. Stickland (Adlard & Son, Dorking, England, 1962), p. 99.

[16] T. Ando and S. Mori, Surf. Sci. **113**, 124 (1982); T. Ando, S. Wakahara, and H. Akera, Phys. Rev. B **40**, 11 609 (1989); T. Ando, *ibid.* **47**, 9621 (1993).

[17] Q.-G. Zhu and H. Kroemer, Phys. Rev. B **27**, 3519 (1983).

[18] M. G. Burt, Semicond. Sci. Technol. **3**, 739 (1988); J. Phys. Condens. Matter **4**, 6651 (1992); Phys. Rev. B **50**, 7518 (1994).

[19] W. Trzeciakowski, Phys. Rev. B **38**, 12 493 (1988); Semicond. Sci. Technol. **10**, 768 (1995).

[20] G. F. Karavaev and Yu. S. Tikhodeev, Fiz. Tekh. Poluprovodn. **25**, 1237 (1991) [Sov. Phys. Semicond. **25**, 745 (1991)].

[21] B. Laikhtman, Phys. Rev. B **46**, 4769 (1992).

[22] P. von Allmen, Phys. Rev. B **46**, 15 377 (1992).

[23] T. Yamanaka, H. Kamada, Y. Yoshikuni, W. W. Lui, S. Seki, and K. Yokoyama, J. Appl. Phys. **76**, 2347 (1994).

[24] S. N. Grinyaev, G. F. Karavaev, and V. N. Chernyshov, Fiz. Tekh. Poluprovodn. **28**, 1393 (1994) [Semicond. **28**, 784 (1994)].

[25] R. Balian, D. Bessis, and G. A. Mezincescu, Phys. Rev. B **51**, 17 624 (1995).

[26] B. A. Foreman, Phys. Rev. B **52**, 12 241 (1995).

[27] G. H. Wannier, Phys. Rev. **52**, 191 (1937); J. C. Slater, *ibid.* **76**, 1592 (1949); E. N. Adams, *ibid.* **85**, 41 (1952).

[28] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).

[29] P.-O. Löwdin, J. Chem. Phys. **19**, 1396 (1951).

[30] J. Callaway, *Quantum Theory of the Solid State*, 2nd ed. (Academic, New York, 1991).

[31] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders, Philadelphia, 1976).

[32] V. Heine, Proc. Phys. Soc. London **81**, 300 (1963).

[33] E. T. Whittaker, Proc. R. Soc. Edinburgh **35**, 181 (1915); C. E. Shannon, Proc. IRE **37**, 10 (1949).

[34] J. A. Krumhansl, in *Lattice Dynamics*, Proceedings of the International Conference at Copenhagen, 1963, edited by R. F. Wallis (Pergamon, Oxford, 1965), p. 627.

[35] I. A. Kunin, *Elastic Media with Microstructure* (Springer, Berlin, 1982).

[36] W. Garczynski and J. Stelmach, J. Math. Phys. **22**, 1106 (1981); **23**, 1168 (1982).

[37] B. A. Foreman, Phys. Rev. B **52**, 12 260 (1995).

[38] E. I. Blount, in *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic, New York, 1962), Vol. 13, pp. 321–322.

[39] T. Gora and F. Williams, in *II-VI Semiconducting Compounds*, edited by D. G. Thomas (Benjamin, New York, 1967), p. 639; Phys. Rev. **177**, 1179 (1969).

[40] L. Leibler, Phys. Rev. B **12**, 4443 (1975).

[41] C. M. van Vliet and A. H. Marshak, Phys. Rev. B **26**, 6734 (1982); **29**, 5960 (1984).

[42] O. von Roos, Phys. Rev. B **27**, 7547 (1983).

[43] R. A. Morrow, Phys. Rev. B **29**, 5962 (1984).

[44] O. von Roos and H. Mavromatis, Phys. Rev. B **31**, 2294 (1985).

[45] K. Young, Phys. Rev. B **39**, 13 434 (1989).

[46] M. R. Geller and W. Kohn, Phys. Rev. Lett. **70**, 3103 (1993); Phys. Rev. B **48**, 14 085 (1993); W. Kohn, Physica B **212**, 305 (1995).

[47] R. E. Ziemer, W. H. Tranter, and D. R. Fannin, *Signals and Systems: Continuous and Discrete* (Macmillan, New York, 1983).

[48] This is strictly true only within the context of a pseudopotential model. The present paper treats only local pseudopotentials [M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789 (1966)], but the same methods can be applied to the nonlocal case (for details, see Ref. 18). For the true potential including all core effects, the basis functions in different media will usually be strongly dissimilar, and may be close to orthogonal.

[49] The derivative of a discrete function is defined to be the derivative of the corresponding quasicontinuum function evaluated at a lattice site. For example, $\mathbf{P}\delta_{\mathbf{R},\mathbf{R}'}=v[\mathbf{p}\delta_B(\mathbf{r}-\mathbf{R}')]|_{\mathbf{r}=\mathbf{R}}$, which is nonlocal.

[50] R. A. Morrow and K. R. Brownstein, Phys. Rev. B **30**, 678 (1984); R. A. Morrow, *ibid.* **36**, 4836 (1987); J. Thomsen, G. T. Einevoll, and P. C. Hemmer, *ibid.* **39**, 12 783 (1989).

[51] R. Parmenter, Phys. Rev. **100**, 573 (1995); G. Dresselhaus, *ibid.* **100**, 580 (1955).

[52] This shows that the argument used in Ref. 50 to rule out certain Hamiltonians is mathematically correct but physically meaningless, since it depends critically on the assumption of abruptly varying material properties.

[53] B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1956), pp. 174–177.

[54] M. Cardona, N. E. Christensen, and G. Fasol, Phys. Rev. B **38**, 1806 (1988).

[55] A. Elçi, Phys. Rev. B **49**, 7432 (1994).