First-principles envelope-function theory for lattice-matched semiconductor heterostructures

Bradley A. Foreman*

Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
(Received 17 June 2005; revised manuscript received 26 August 2005; published 28 October 2005)

In this paper a multiband envelope-function Hamiltonian for lattice-matched semiconductor heterostructures is derived from first-principles self-consistent norm-conserving pseudopotentials. The theory is applicable to isovalent or heterovalent heterostructures with macroscopically neutral interfaces and no spontaneous bulk polarization. The key assumption—proved in earlier numerical studies—is that the heterostructure can be treated as a weak perturbation with respect to some periodic reference crystal, with the nonlinear response small in comparison to the linear response. Quadratic response theory is then used in conjunction with $\mathbf{k} \cdot \mathbf{p}$ perturbation theory to develop a multiband effective-mass Hamiltonian (for slowly varying envelope functions) in which all interface band-mixing effects are determined by the linear response. To within terms of the same order as the position dependence of the effective mass, the quadratic response contributes only a bulk band offset term and an interface dipole term, both of which are diagonal in the effective-mass Hamiltonian. The interface band mixing is therefore described by a set of bulklike parameters modulated by a structure factor that determines the distribution of atoms in the heterostructure. The same linear parameters determine the interface band-mixing Hamiltonian for slowly varying and (sufficiently large) abrupt heterostructures of arbitrary shape and orientation. Long-range multipole Coulomb fields arise in quantum wires or dots, but have no qualitative effect in two-dimensional systems beyond a dipole contribution to the band offsets. The method of invariants is used to determine the explicit form of the Hamiltonian for Γ_6 and Γ_8 states in semiconductors with the zinc-blende structure, and for intervalley mixing of Γ and X electrons in (001) GaAs/AlAs heterostructures.

I. INTRODUCTION

A. Background and motivation

Envelope-function models continue to play a key role in the design and interpretation of experiments on semiconductor heterostructures. The canonical "envelope-function approximation," which by definition makes use of only bulk effective-mass parameters and heterojunction band offsets, has been very successful in explaining a wide range of experiments.[1,2] However, recent studies have increasingly emphasized interface-related effects lying outside the scope of conventional envelope-function theory, such as optical[3–5] and electrical[6–8] anisotropy, intervalley mixing,[9–15] and spin polarization phenomena.[16–27] These effects are generated (wholly or in part) by interface band-mixing terms in the heterostructure Hamiltonian, such as valence-band mixing of light and heavy holes,[28–30] Γ-X coupling,[10,11] and the conduction or valence band Rashba coupling.[31]

A great deal of progress in the modeling of these effects can be made on the basis of symmetry information alone. The method of invariants, which was originally developed for bulk semiconductors,[32–35] has proved a powerful tool in the study of heterostructures as well.[6,15–31,36–38] The standard method of invariants uses symmetry information to construct an explicit interface Hamiltonian, but the same information (supplemented by hermiticity or current-conservation requirements) may also be used to construct connection rules for the envelope functions on opposite sides of the interface.[39–46] In either case, one obtains a phenomenological model containing some interface parameters whose value is determined by comparison with experiment.

Useful as this approach may be, it does not provide any information about the magnitude of the interface parameters,

some of which (for reasons unrelated to symmetry) may happen to be zero or negligibly small. To obtain this information (along with a deeper understanding of the physical origin of the interface phenomena), one must turn to a more detailed microscopic model of the interface. Thus, numerous envelope-function models have been derived directly from the microscopic potential energy; these include interface Hamiltonians[47–71] and connection rules[72–82] as well as numerical approaches based on solving the $\mathbf{k} \cdot \mathbf{p}$ equations in momentum space.[83,84]

However, all of the cited derivations are based on empirical pseudopotentials, in which the potential energy is determined by the choice of some specific model for the interface. This leads to ambiguity in the results, as different choices may yield conflicting predictions as to which interface parameters are important,[12–14,61,64,69] or whether interface band-mixing effects are directly related to band offsets.[70,71,85,86] Such conflicts can only be resolved by deriving the envelope-function Hamiltonian from an *ab initio* self-consistent potential, as suggested by Sham and Lu.[87] The purpose of this paper is to present such a derivation and examine its implications for interface band-mixing effects. Numerical applications of the theory are not considered here. A preliminary account of these results has been presented elsewhere.[88]

B. Basic assumptions and limitations

The current "standard model" for condensed-matter physics is based on density-functional theory in the local density approximation (LDA).[89–91] However, it is well known that this ground-state formalism does not accurately predict en-

ergy gaps in semiconductors. Thus, the present state-of-the-art in band-structure theory involves calculations of the electron self-energy in the shielded-interaction[92,93] or *GW* approximation.[94–97] This field is not yet fully mature, as apparent early successes[98–100] have been questioned in light of recent developments.[101–105] Nevertheless, any future refinements in approximation techniques for calculating the quasiparticle band structure will continue to be based on the self-energy in Dyson's equation. For this reason, the present work relies only upon general properties of the self-energy operator, not upon any specific approximation to this quantity.

For simplicity, however, this paper assumes that the ionic cores are described in terms of norm-conserving pseudopotentials.[106–110] The projector augmented wave (PAW) method of Blöchl,[103,105,111] which does not involve this approximation, has become increasingly popular in recent years. A generalization of the present theory to incorporate the PAW formalism should be possible, but to avoid undue complexity it is not considered here. In addition, this paper considers only lattice-matched systems (with no bulk or interface strain).

The key assumption used in the present work is that the heterostructure is a small perturbation[112] with respect to some periodic virtual reference crystal (such as $Al_{0.5}Ga_{0.5}As$ for a GaAs/AlAs heterostructure). This assumption has been verified within LDA for both isovalent and heterovalent systems, including GaAs/AlAs,[113,114] Ge/GaAs,[114,115] $In_{0.53}Ga_{0.47}As$/InP,[114,116] GaAs/Si/AlAs and GaAs/Ge/AlAs,[117,118] Si/Ge,[119,120] and InAs/GaSb.[121,122] Within the framework of pseudopotential theory, one would expect it to be valid for other similar heterostructures also. This assumption makes it possible to describe the self-consistent heterostructure perturbation in terms of nonlinear response theory, with the linear response predominant and the quadratic response providing a weak correction.[113–122]

Such an approach has been used by Sham[123] to derive an effective-mass equation for shallow impurity states in semiconductors. Sham's work was extended in the preceding paper[124] to obtain expressions for the self-energy at small values of the crystal momentum in lattice-matched heterostructures described by spin-dependent nonlocal pseudopotentials. These expressions are used here to construct a multi-band effective-mass theory for lattice-matched heterostructures.

Although the basic formalism developed here is quite general, to keep the paper to a reasonable length it is necessary to impose some restrictions on the material systems that are treated in detail. (This allows the power series in Sec. V to be terminated at a reasonably low order.) Since only lattice-matched systems are considered here, it is assumed that the symmetry of the reference crystal does not support a spontaneous polarization, because that would generate macroscopic electric fields and piezoelectric strain fields inconsistent with the lattice-matching assumption. Therefore, the present theory is not directly applicable to wurtzite materials.

In addition, it is assumed that the interfaces in a heterovalent system such as Ge/GaAs are macroscopically neutral,[114,125] so that the atoms can be grouped together into neutral clusters (of fractional atoms) that carry a dipole moment only at the interface.[124,126] Thus, the main application

of the present theory would be to isovalent or heterovalent heterostructures of semiconductors with the zinc-blende or diamond structure. However, the only examples treated explicitly here are from isovalent systems.

Lattice mismatch could not be included in this theory without fundamental changes to account for atomic relaxation. Nevertheless, the present results provide a solid foundation for subsequent extensions of the theory to include this effect. Other restrictions, such as macroscopic neutrality and the lack of bulk dipole terms, could be lifted merely by extending some power series expansions to higher order (although, depending on the accuracy that is desired,[126] this may require the inclusion of certain nonanalytic terms neglected in Ref. 124).

The fact that (within the pseudopotential approximation) a typical heterostructure is a weak perturbation[112–122] makes possible the existence of energy eigenstates $|\psi\rangle$ whose wave function $\langle n\mathbf{k}|\psi\rangle$ in the Luttinger-Kohn representation[127] $|n\mathbf{k}\rangle$ of the reference crystal is negligible outside a small region in $\mathbf{k}$ space near the high-symmetry points of the Brillouin zone.[128–131] These eigenstates are just the low-energy excitations that are of greatest interest experimentally. Their existence, which has been verified by extensive numerical work on empirical pseudopotential models,[128–131] is precisely the condition needed for the validity of an effective-mass approximation for slowly varying envelope functions $F_n(\mathbf{x})$,[47–52,127] since $F_n(\mathbf{k}) \equiv \langle n\mathbf{k}|\psi\rangle$. The existence of slowly varying envelopes[50,51,128–131] provides the foundation for the effective-mass theory derived here.

The present approach, based on quadratic response theory,[124] fits well with the $\mathbf{k}\cdot\mathbf{p}$ perturbation formalism of Leibler,[47,48,65–69] in which the heterostructure perturbation is treated as small in comparison to the energy separating the bands of interest from other remote bands. Leibler's theory is used here to develop a multi-band effective-mass theory that includes all terms of the same order as the position dependence of the effective mass. This includes cubic and quartic dispersion terms in the Hamiltonian of the reference crystal, as well as the leading contributions from the quadratic response. As shown in Sec. II, higher-order terms cannot consistently be described in terms of local differential equations.[65–69]

C. Summary of key results

In this paper, it is shown that the dominant interface band-mixing terms are those arising from the linear response. Indeed, the linear response is the *only* contribution to band mixing that fits within the framework of the perturbation theory defined above, and the only contribution that can consistently be described in terms of local differential equations. To this level of accuracy, the quadratic response contributes only a bulk band offset term and an interface dipole term, neither of which produces any band mixing in the effective-mass Hamiltonian.

This represents a major simplification, since it implies that the interface band-mixing terms in the Hamiltonian are just a superposition of parameters derived from the linear response to individual ionic perturbations (or neutral cluster

perturbations[124,126] in the case of heterovalent substitutions). These parameters are calculated once and for all for a given material system; they then appear as coefficients in front of structure factors describing the distribution of atoms in the heterostructure. The Hamiltonian contains spatial derivatives of (and differences between) these structure factors, which generate δ-like functions at an interface. The atomic distribution functions change for heterostructures of different shape (wells, wires, or dots) and orientation, but their coefficients do not. Thus, a single set of interface parameters governs the band mixing for any type of heterostructure[132] in a given material system.

If desired, one can express these linear parameters as a *difference* between the properties of the various *bulk* materials that make up the heterostructure. (A similar result was obtained for linear band offsets in Refs. 113–122.) However, in a no-common-atom system such as InAs/GaSb, this must include the (lattice-matched) "interface" materials GaAs and InSb. Also, these bulk-like linear response parameters (which in principle do not require a supercell calculation[133]) cannot be determined from experiments on bulk semiconductors.

For some terms in the linear response (derived from the analytic part of the self-energy[124]), the interface Hamiltonian itself has the form of a macroscopic average of the Dirac δ function or its derivatives. For the remaining terms (derived from the nonanalytic part of the self-energy[124]), the effective charge density has this localized form, but the potential energy is not as well localized, having the form of a long-range multipole potential. However, in a quantum well or any other heterostructure with two-dimensional translation symmetry, all terms are well localized except for the interface dipole terms (which merely modify the band offsets).

The band-mixing Hamiltonian has the same general form for slowly graded structures (within the virtual crystal approximation) and abrupt heterojunctions, the only difference being the rate of change of the structure factor. That is, in contrast to previous theories based on model potentials or empirical pseudopotentials, no new band-mixing parameters appear at an abrupt junction. Indeed, the localized interface terms in the Hamiltonian derived here are qualitatively identical to the Hamiltonian derived by Leibler[47,48] for slowly graded heterostructures.

These results shed light on recent suggestions[12–14,64,69] that the Γ_1-X_{1z} and X_{1x}-X_{1y} intervalley mixing potentials at an ideal (001) GaAs/AlAs heterojunction should be proportional to $\delta(z)$. It is shown here that such mixing arises only from the quadratic response and is therefore negligible in comparison to Γ_1-X_{3z}, X_{1z}-X_{3z}, and X_{3x}-X_{3y} mixing. There is, however, a linear contribution to the Γ_1-X_{1z} and X_{1x}-X_{1y} mixing that is proportional to $\delta'(z)$, as well as a linear $\delta(z)$ mixing for nonideal interfaces or when spin-orbit coupling is included. These contributions may help to explain the experimental observations in Refs. 12–14.

D. Outline of the paper

The paper begins in Sec. II with a discussion of which terms are to be retained in $\mathbf{k} \cdot \mathbf{p}$ perturbation theory, based upon a review of important recent work by Takhtamirov and

Volkov.[65–69] The main results of the quadratic response theory developed in the preceding paper[124] are presented (in modified form) in Sec. III. The basic envelope-function formalism is developed in Sec. IV, while Sec. V describes the power series expansions that are used to obtain approximate expressions valid for slowly varying envelopes. Perturbation theory is used to eliminate the coupling to "remote" bands in Sec. VI, yielding the basic expression for the multi-band envelope-function Hamiltonian. Modifications to the Hamiltonian that are necessary if one wishes to describe the material parameters of an abrupt heterojunction as piecewise constant are discussed in Sec. VII. Symmetry properties of the Hamiltonian are discussed in Sec. VIII, where explicit matrix representations of the material properties are given for semiconductors with the zinc-blende structure. Finally, the significance of the results and their relation to previous work in the literature are discussed in Sec. IX.

II. ORDER OF TERMS INCLUDED

Takhtamirov and Volkov[65–69] have recently demonstrated (using an instructive analogy to the leading relativistic corrections to the Schrödinger equation) that most derivations of heterostructure effective-mass equations in the literature (including those of the present author) do not consistently include all perturbative corrections of the same order. This section presents a review and discussion of their results, with the objective of establishing which terms are to be retained in subsequent perturbative approximations.

The first case to be considered is a wide-gap system satisfying

$$\Delta \bar{V} \ll \bar{E}_g, \qquad (2.1)$$

where $\Delta \bar{V}$ is a typical heterojunction band offset, and $\bar{E}_g$ is a typical energy gap (for the virtual bulk reference crystal) separating the band in question from all other "remote" bands. It is assumed that in the reference crystal this band is describable by an effective-mass equation with effective mass m^*. This will be the case if the $\mathbf{k} \cdot \mathbf{p}$ interaction with remote bands

$$\frac{\hbar \bar{k} \bar{p}}{m} \equiv (\bar{k}\lambda)\bar{E}_g \qquad (2.2)$$

can be treated as a small perturbation. Here m is the free-electron mass, $\bar{k} \sim 2\pi/L$ is a typical envelope function wavenumber for a quantum well of width L, $\bar{p} \sim 2\pi\hbar/a$ is a typical interband momentum matrix element (where a is the lattice constant), and it is assumed that $L \gg a$. Equation (2.2) defines a length parameter $\lambda = \hbar \bar{p}/m\bar{E}_g$, which (if the free-electron contribution to m^* is negligible) may also be written as $\lambda \approx \hbar(2m^*\bar{E}_g)^{-1/2}$. For GaAs, both expressions for λ give $\lambda \approx 6$ Å, thus $\lambda \sim a$ and $\bar{k}\lambda \ll 1$ in a wide quantum well. For the remainder of this paper the parameters λ and a are used interchangeably, although in general error estimates should be based on the larger of the two quantities.

In general one is interested in cases where the kinetic energy is comparable to the band offset:

$$\frac{\hbar^2 \bar{k}^2}{2m^*} \sim \Delta \bar{V} \sim (\bar{k}\lambda)^2 \bar{E}_{\mathrm{g}}. \tag{2.3}$$

This is the order of terms included in ordinary effective-mass theory,[127] which may be compared with the nonrelativistic (Schrödinger) approximation to the Dirac equation. The position dependence of the effective mass can be calculated by treating $\Delta \bar{V}/\bar{E}_{\mathrm{g}}$ as a perturbation,[47,48] yielding a correction of order

$$\frac{\hbar^2 \bar{k}^2}{2m^*} \left(\frac{\Delta \bar{V}}{\bar{E}_{\mathrm{g}}} \right) \sim \frac{(\Delta \bar{V})^2}{\bar{E}_{\mathrm{g}}} \sim (\bar{k}\lambda)^2 \Delta \bar{V}, \tag{2.4}$$

which is of the same order as the nonparabolic k^4 terms in the kinetic energy of the reference crystal. Hence, these terms (analogous to the relativistic mass corrections to the Schrödinger equation) must be included if one is to retain all terms of the same order as the position dependence of the effective mass.

As will be seen below, the interface band-mixing terms proportional to a δ function are of order

$$A \langle \delta(z) \rangle \sim (\bar{k}a) \Delta \bar{V}, \tag{2.5}$$

while those proportional to the derivative of a δ function are of order

$$B \langle \delta'(z) \rangle \sim (\bar{k}a)^2 \Delta \bar{V}, \tag{2.6}$$

in which A and B are constants. Hence, the latter terms are also comparable to the position dependence of the effective mass.

For a step-function discontinuity ΔV in the potential energy, the envelope function F has a discontinuity in its second derivative given by

$$\frac{\Delta F''}{F} = \frac{2m^*}{\hbar^2} \Delta V, \tag{2.7}$$

where $F \sim L^{-1/2}$. This gives rise to an asymptotic behavior in k space of $F(k) \sim (\Delta F'')k^{-3}$. But in the exact Luttinger-Kohn envelope-function representation,[127] the envelope functions are limited to wave vectors inside the first Brillouin zone. Hence, the use of a local differential effective-mass equation, which gives rise to nonvanishing $F(k)$ outside the Brillouin zone, generates an error in the kinetic energy of order

$$(\bar{k}a)^3 \left(\frac{\hbar^2 \bar{k}^2}{2m^*} \right) \sim (\bar{k}a)^3 \Delta \bar{V}. \tag{2.8}$$

Therefore, in the local approximation, the accuracy is limited by Eq. (2.8), and contributions beyond the level of (2.4) and (2.6) should for consistency be omitted.

Of course, this is not a fundamental limitation of the envelope-function method, as one can always choose to work in **k** space (which is a common choice for numerical work[134–136]). However, there is another similar source of error that arises from anticrossings of the bands explicitly included in the envelope-function model with those treated as remote perturbations. This occurs, for example, in the Γ_6 conduction band of GaAs at about a third of the distance to

the Brillouin zone boundary in the Δ direction. Beyond such an anticrossing the model is no longer valid and an error of order (2.8) arises even in a **k**-space formalism. To eliminate such errors one must enlarge the Hamiltonian by treating these remote bands explicitly, thus obtaining a full-zone **k·p** model.[74,83,137,138]

Such a multiband envelope-function model is also needed for medium-gap ($\Delta \bar{V} \sim E_{\mathrm{g}}^{(0)}$) and narrow-gap ($E_{\mathrm{g}}^{(0)} \ll \Delta \bar{V}$) systems, where $E_{\mathrm{g}}^{(0)}$ is the energy gap of the reference crystal. The perturbative approach based on Eq. (2.1) can still be used, provided that $\bar{E}_{\mathrm{g}}$ refers to the energy gap between the bands of interest and those treated as remote perturbations. Takhtamirov and Volkov[69] have considered the extreme narrow-gap limit for the case in which the dispersion is dominated by linear-k terms for all energy ranges of interest, and showed that in this case the error generated by the local approximation is of order $(\bar{k}a)^2 \Delta \bar{V}$. In such cases the terms (2.4) and (2.6) should be omitted for consistency. However, in practical situations one is more likely to encounter cases in which $\Delta \bar{V} \sim E_{\mathrm{g}}^{(0)}$ and the k^2 dispersion terms are comparable to the linear-k terms. Thus, in this paper all terms of order $(\bar{k}a)^2 \Delta \bar{V}$ are retained, while those at the level of the local approximation (2.8) are omitted.

It should be noted that a fully self-consistent perturbation scheme is not always desirable. It would of course be impossible to achieve an accurate numerical prediction for the fine structure of the hydrogen atom without including the relativistic-mass correction and the Darwin term in addition to the spin-orbit splitting. However, for qualitative considerations one is often interested primarily in symmetry-breaking effects, in which case the former two contributions may justifiably be omitted. Likewise, although the present work retains all terms of order $(\bar{k}a)^2 \Delta \bar{V}$, certain of these (such as the k^4 bulk dispersion terms) may possibly be omitted for applications in which the primary focus is on symmetry-breaking interface effects. [As an example, Takhtamirov and Volkov have proposed a model[6] based on certain terms of order $(\bar{k}a)^3 \Delta \bar{V}$, while neglecting larger terms of order $(\bar{k}a) \Delta \bar{V}$ and $(\bar{k}a)^2 \Delta \bar{V}$.] However, the neglect of these terms can only be justified for specific individual applications and should not be presumed to hold in general.

The discussion here has focused on wide quantum wells, but this should not be taken to imply that effective-mass theory is inapplicable in other cases. For example, in a narrow quantum well,[50,68] an effective-mass approximation can be developed along the same lines as for shallow impurities.[123,127] In this case, however, the above estimates of the interface terms and local approximation are no longer valid; see Ref. 68 for further details.

The remainder of this paper uses atomic units with $\hbar = m = e = 1$.

III. QUADRATIC RESPONSE THEORY

A. Basic definitions

The problem of interest is the Dyson eigenvalue equation (see Ref. 124 and references therein)

$$-\tfrac{1}{2}\nabla^2\psi(\mathbf{x},\omega) + \int V(\mathbf{x},\mathbf{x}',\omega)\psi(\mathbf{x}',\omega)d^3x' = E(\omega)\psi(\mathbf{x},\omega),$$

$$(3.1)$$

in which ω is a complex energy parameter, E is the complex eigenvalue, and the self-consistent potential V is the sum of a fixed norm-conserving ionic pseudopotential V_{ion} and the non-Hermitian self-energy operator Σ:

$$V(\mathbf{x},\mathbf{x}',\omega) = V_{\text{ion}}(\mathbf{x},\mathbf{x}') + \Sigma(\mathbf{x},\mathbf{x}',\omega). \qquad (3.2)$$

For notational simplicity, the ω dependence will be suppressed in most of the equations that follow. In Eq. (3.1), ψ and V are spinors, with V having the form

$$V = \mathbb{1}\,V_{\text{sc}} + \boldsymbol{\sigma}\cdot\mathbf{V}, \qquad (3.3)$$

in which $\mathbb{1}$ is the 2×2 spinor unit matrix, V_{sc} is the scalar relativistic part of V, $\boldsymbol{\sigma}$ is the Pauli matrix, and $\mathbf{V}$ accounts for spin-orbit coupling.[139–142] Note that Eq. (3.1) incorporates all relativistic corrections of order $Z^2\alpha^2$ (where Z is the atomic number and α is the fine-structure constant), but neglects terms of order α^2, such as spin-orbit coupling outside the atomic cores.[107,108]

In a heterostructure, it is convenient to partition the ionic pseudopotential as

$$V_{\text{ion}}(\mathbf{x},\mathbf{x}') = V_{\text{ion}}^{(0)}(\mathbf{x},\mathbf{x}') + \Delta V_{\text{ion}}(\mathbf{x},\mathbf{x}'), \qquad (3.4)$$

where $V_{\text{ion}}^{(0)}$ is the ionic pseudopotential of some periodic reference crystal (which may be a virtual crystal). This is defined as

$$V_{\text{ion}}^{(0)}(\mathbf{x},\mathbf{x}') = \sum_{a,j,\mathbf{R}} f^{aj} v_{\text{ion}}^a(\mathbf{x}-\mathbf{R}_j, \mathbf{x}'-\mathbf{R}_j), \qquad (3.5)$$

in which $v_{\text{ion}}^a(\mathbf{x},\mathbf{x}')$ is the ionic pseudopotential for atomic species a, f^{aj} is the fractional weight associated with atom[143] a on site j at position $\boldsymbol{\tau}_j$ in the unit cell of the reference crystal, and $\mathbf{R}_j = \mathbf{R}+\boldsymbol{\tau}_j$, where $\mathbf{R}$ is a Bravais lattice vector of the reference crystal. f^{aj} must satisfy

$$0 \leq f^{aj} \leq 1, \quad \sum_a f^{aj} = 1, \qquad (3.6)$$

although the former constraint need not be strictly enforced.

The term ΔV_{ion} is the perturbation due to the heterostructure:

$$\Delta V_{\text{ion}}(\mathbf{x},\mathbf{x}') = \sum_{a,j,\mathbf{R}} \theta_{\mathbf{R}}^{aj} v_{\text{ion}}^a(\mathbf{x}-\mathbf{R}_j, \mathbf{x}'-\mathbf{R}_j). \qquad (3.7)$$

Here $\theta_{\mathbf{R}}^{aj}$ is the change in fractional weight (relative to the reference crystal) of atom a at position $\mathbf{R}_j$ in the heterostructure, which must satisfy

$$0 \leq f^{aj} + \theta_{\mathbf{R}}^{aj} \leq 1, \quad \sum_a \theta_{\mathbf{R}}^{aj} = 0. \qquad (3.8)$$

The constraint (3.8) permits one to rewrite ΔV_{ion} as

$$\Delta V_{\text{ion}}(\mathbf{x},\mathbf{x}') = \sum_{\alpha,\mathbf{R}}{}' \theta_{\mathbf{R}}^{\alpha} \Delta v_{\text{ion}}^{\alpha}(\mathbf{x}-\mathbf{R}_\alpha, \mathbf{x}'-\mathbf{R}_\alpha), \qquad (3.9)$$

in which

$$\Delta v_{\text{ion}}^{\alpha}(\mathbf{x},\mathbf{x}') \equiv v_{\text{ion}}^a(\mathbf{x},\mathbf{x}') - v_{\text{ion}}^{\bar{a}_j}(\mathbf{x},\mathbf{x}'). \qquad (3.10)$$

Here $\alpha=(a,j)$ is a composite index, $\bar{a}_j$ is the label of some given atom on site j, and the prime on the summation symbol indicates that the values $\bar{\alpha}=(\bar{a}_j,j)$ are excluded. To simplify the interpretation of (3.9), the atom $\bar{a}_j$ is chosen to be the same as the (virtual) atom on site j in the reference crystal.

In this paper the self-consistent potential $V(\mathbf{x},\mathbf{x}')$ is treated using nonlinear response theory.[124] The fundamental assumption is that V can be expressed as a power series in the variables $\theta_{\mathbf{R}}^{\alpha}$:

$$V(\mathbf{x},\mathbf{x}') = V^{(0)}(\mathbf{x},\mathbf{x}') + V^{(1)}(\mathbf{x},\mathbf{x}') + V^{(2)}(\mathbf{x},\mathbf{x}') + \cdots.$$

$$(3.11)$$

Here $V^{(0)}(\mathbf{x},\mathbf{x}')$ is the self-consistent potential of the reference crystal, which has the same periodicity as the ionic potential (3.5):

$$V^{(0)}(\mathbf{x},\mathbf{x}') = V^{(0)}(\mathbf{x}+\mathbf{R}, \mathbf{x}'+\mathbf{R}). \qquad (3.12)$$

The linear response to the heterostructure perturbation has the form

$$V^{(1)}(\mathbf{x},\mathbf{x}') = \sum_{\alpha,\mathbf{R}}{}' \theta_{\mathbf{R}}^{\alpha} \Delta v_{\mathbf{R}}^{\alpha}(\mathbf{x},\mathbf{x}'), \qquad (3.13)$$

while the quadratic response is

$$V^{(2)}(\mathbf{x},\mathbf{x}') = \sum_{\alpha,\mathbf{R}}{}' \sum_{\alpha',\mathbf{R}'}{}' \theta_{\mathbf{R}}^{\alpha} \theta_{\mathbf{R}'}^{\alpha'} \Delta v_{\mathbf{R}\mathbf{R}'}^{\alpha\alpha'}(\mathbf{x},\mathbf{x}'). \qquad (3.14)$$

Here the expansion coefficients in the power series are defined by

$$\Delta v_{\mathbf{R}}^{\alpha}(\mathbf{x},\mathbf{x}') = \frac{\partial V(\mathbf{x},\mathbf{x}')}{\partial \theta_{\mathbf{R}}^{\alpha}},$$

$$\Delta v_{\mathbf{R}\mathbf{R}'}^{\alpha\alpha'}(\mathbf{x},\mathbf{x}') = \frac{1}{2}\frac{\partial^2 V(\mathbf{x},\mathbf{x}')}{\partial \theta_{\mathbf{R}}^{\alpha}\,\partial \theta_{\mathbf{R}'}^{\alpha'}}, \qquad (3.15)$$

where the derivatives are evaluated with respect to the reference crystal. These derivatives can be evaluated numerically by applying perturbations (3.9) with small values of $\theta_{\mathbf{R}}^{\alpha}$.

The leading source of error in quadratic response theory lies in the neglected cubic response term $V^{(3)}$. This error can be reduced somewhat by a suitable choice of reference potential. For example, in a GaAs/AlAs heterostructure, choosing $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}$ as a reference crystal instead of GaAs would reduce the cubic error in AlAs by a factor of 8. However, it would also create a comparable error in GaAs. Thus, the cubic error in the interface Hamiltonian for an $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}$ reference potential would be about $\tfrac{1}{4}$ that of a GaAs or AlAs reference potential. This is certainly an improvement, but since the quadratic response is already quite small (see Sec. III B), it is unlikely to make much practical difference.

In the momentum representation, the Dyson equation (3.1) has the form

$$\frac{1}{2}k^2\psi(\mathbf{k}) + \sum_{\mathbf{k}'} V(\mathbf{k},\mathbf{k}')\psi(\mathbf{k}') = E\psi(\mathbf{k}). \qquad (3.16)$$

Making use of the translation symmetry of the reference crystal, one can write the Fourier transforms of the potentials (3.15) as

$$\Delta v_{\mathbf{R}}^{\alpha}(\mathbf{k},\mathbf{k}') \equiv e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{\alpha}}\Delta v^{\alpha}(\mathbf{k},\mathbf{k}'),$$

$$\Delta v_{\mathbf{RR}'}^{\alpha\alpha'}(\mathbf{k},\mathbf{k}') \equiv e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{\alpha}}\Delta v^{\alpha\alpha'\mathbf{R}''}(\mathbf{k},\mathbf{k}'), \quad (3.17)$$

in which $\mathbf{R}'' = \mathbf{R}' - \mathbf{R}$, and the coordinate origin is $\mathbf{R}_{\alpha}$ for the modified functions on the right-hand side. The Fourier transform of the linear response (3.13) can therefore be written as

$$V^{(1)}(\mathbf{k},\mathbf{k}') = N\sum_{\alpha}{}' \theta^{\alpha}(\mathbf{k}-\mathbf{k}')\Delta v^{\alpha}(\mathbf{k},\mathbf{k}'). \qquad (3.18)$$

Here $N = \Omega/\Omega_0$ is the number of unit cells (of volume Ω_0) in the reference crystal (of volume Ω), $\theta^{\alpha}(\mathbf{k})$ is the Fourier transform of $\theta_{\mathbf{R}}^{\alpha}$:

$$\theta_{\mathbf{R}}^{\alpha} = \sum_{\mathbf{k}\in\Omega_0^*} \theta^{\alpha}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{R}_{\alpha}}, \qquad (3.19a)$$

$$\theta^{\alpha}(\mathbf{k}) = \frac{1}{N}\sum_{\mathbf{R}\in\Omega} \theta_{\mathbf{R}}^{\alpha}e^{-i\mathbf{k}\cdot\mathbf{R}_{\alpha}}, \qquad (3.19b)$$

and $\Omega_0^* = (2\pi)^3/\Omega_0$ is the volume of a unit cell in the reciprocal lattice. Note from (3.19b) that $\theta^{\alpha}(\mathbf{k})$ is quasiperiodic:

$$\theta^{\alpha}(\mathbf{k}+\mathbf{G}) = \theta^{\alpha}(\mathbf{k})e^{-i\mathbf{G}\cdot\tau_{\alpha}}, \qquad (3.20)$$

where $\mathbf{G}$ is a reciprocal lattice vector of the reference crystal.

In a similar fashion, inserting (3.17) into the Fourier transform of (3.14) and relabeling $\mathbf{R}'$ gives

$$V^{(2)}(\mathbf{k},\mathbf{k}') = N\sum_{\alpha,\alpha',\mathbf{R}'}{}' \theta^{\alpha\alpha'\mathbf{R}'}(\mathbf{k}-\mathbf{k}')\Delta v^{\alpha\alpha'\mathbf{R}'}(\mathbf{k},\mathbf{k}'),$$

$$(3.21)$$

in which $\theta^{\alpha\alpha'\mathbf{R}'}(\mathbf{k})$ is the Fourier transform of the pair distribution function $\theta^{\alpha\alpha'\mathbf{R}'}(\mathbf{R}) \equiv \theta_{\mathbf{R}}^{\alpha}\theta_{\mathbf{R}+\mathbf{R}'}^{\alpha'}$.

B. Estimation of magnitude

A crude estimate of the relative magnitudes of the linear and quadratic response can be obtained from a simple non-linear Thomas-Fermi model,[89] with the result[144]

$$V^{(1)} \sim \Delta\bar{V}, \qquad \frac{V^{(2)}}{V^{(1)}} \sim \frac{\Delta\bar{V}}{4\epsilon_{\mathrm{F}}}, \qquad (3.22)$$

where $\Delta\bar{V}$ is the typical band offset defined in Sec. II, and ϵ_{F} is the Fermi energy. Now the difference in screened pseudopotentials between typical III-V semiconductors is roughly $\Delta\bar{V} \sim 0.02$–0.05 Ry,[145] which yields the estimate $|V^{(2)}/V^{(1)}| \sim 0.01$. This suggests that the quadratic response is indeed very small.

This estimate is supported by the LDA calculations of Wang and Zunger[130] for GaAs/AlAs heterostructures, in which they found (see Table I of Ref. 130) that the interface band-mixing terms arising from the linear response were on average about 1000 times larger than those arising from the quadratic response (with the ratio between the smallest linear and largest quadratic terms being about 100, in agreement with the estimate obtained above). However, the LDA calculations for GaAs/AlAs and $In_{0.53}Ga_{0.47}As/InP$ presented in Refs. 114 and 116 indicate that the quadratic density response is only about 10 times smaller than the linear density response (the linear and quadratic potentials were not given in these papers).

Nevertheless, this is still a sufficiently large ratio to establish the validity of the quadratic approximation used here. For the purposes of the perturbation scheme of Sec. II, the factor $4\epsilon_{\mathrm{F}}$ in the denominator of Eq. (3.22) will be treated formally as of order $\bar{E}_{\mathrm{g}}$, so that $\Delta V^{(2)}$ is considered to be of the same order $[(\Delta\bar{V})^2/\bar{E}_{\mathrm{g}} \sim (\bar{k}a)^2\Delta\bar{V}]$ as the smallest terms retained in Sec. II.

C. Functional form

As shown in Ref. 124, the Coulomb interaction gives rise to singularities in the linear and quadratic potentials $\Delta v(\mathbf{k},\mathbf{k}')$ when $\mathbf{k}-\mathbf{k}'$ is equal to a reciprocal lattice vector. The explicit form of the linear potential is[124]

$$\Delta v_{\mathbf{R}}^{\alpha}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}') = w_{\mathbf{R}}^{\alpha}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}')$$
$$+ \Lambda_{\mathbf{GG}'}(\mathbf{k},\mathbf{k}')\varphi_{\mathbf{R}}^{\alpha}(\mathbf{k}-\mathbf{k}'),$$

$$(3.23)$$

where the potential $w_{\mathbf{R}}^{\alpha}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}')$ is an analytic function of $\mathbf{k}$ and $\mathbf{k}'$ for wave vectors inside the first Brillouin zone of the reference crystal. For small values of $\mathbf{k}$ and $\mathbf{k}'$ it can be represented as a Taylor series, which is the basis for the effective-mass theory developed in the following section. The effective vertex function $\Lambda_{\mathbf{GG}'}(\mathbf{k},\mathbf{k}')$ is also an analytic function of $\mathbf{k}$ and $\mathbf{k}'$.

The singular contributions come from the screened potential φ, which is a spin scalar of the form[124,146]

$$\varphi_{\mathbf{R}}^{\alpha}(\mathbf{q}) = \frac{v_c(\mathbf{q})n_{\mathbf{R}}^{\alpha}(\mathbf{q})}{\epsilon(\mathbf{q})}, \qquad (3.24)$$

where $\epsilon(\mathbf{q})$ is the static electronic dielectric function of the reference crystal and

$$v_c(\mathbf{q}) = \begin{cases} 4\pi/q^2 & \text{if } q \neq 0, \\ 0 & \text{if } q = 0. \end{cases} \qquad (3.25)$$

The function $n_{\mathbf{R}}^{\alpha}(\mathbf{q})$ is an effective electron density containing partial contributions from the bare ionic pseudocharge and the screening charge. This is an analytic function of $\mathbf{q}$ (for small q) with the symmetry of site $\mathbf{R}_{\alpha}$ in the reference crystal. For isovalent substitutions in zinc-blende crystals, the leading nonanalytic terms in Eq. (3.24) are octopole and hexadecapole potentials proportional to $q_x q_y q_z/q^2$ and $(q_x^4 + q_y^4 + q_z^4)/q^2$.[124]

The quadratic potential has a similar form:

$$\Delta v^{\alpha\alpha'}_{\mathbf{RR}'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}') = w^{\alpha\alpha'}_{\mathbf{RR}'}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}')$$
$$+ \Lambda_{\mathbf{GG}'}(\mathbf{k},\mathbf{k}')\varphi^{\alpha\alpha'}_{\mathbf{RR}'}(\mathbf{k}-\mathbf{k}'),$$
$$(3.26)$$

in which Λ is the same vertex function as above. In this case, however, the potential $w^{\alpha\alpha'}_{\mathbf{RR}'}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}')$ is not an analytic function of $\mathbf{k}$ and $\mathbf{k}'$. Nevertheless, to within the accuracy required here (namely, zeroth order in $\mathbf{k}$ and $\mathbf{k}'$), the nonanalytic part can be neglected for neutral perturbations. Note that the analytic part of the quadratic potential is analytic over a smaller region in $\mathbf{k}$ space than the linear potential. Whereas $w^{\alpha}_{\mathbf{R}}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}')$ was analytic for $\mathbf{k}$ and $\mathbf{k}'$ inside the first Brillouin zone, the analytic part of $w^{\alpha\alpha'}_{\mathbf{RR}'}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}')$ is analytic only over the inner "half" of the Brillouin zone (i.e., over the Brillouin zone of a crystal whose lattice constants are double those of the reference crystal).

The quadratic screened potential is defined by

$$\varphi^{\alpha\alpha'}_{\mathbf{RR}'}(\mathbf{q}) = \frac{v_c(\mathbf{q})n^{\alpha\alpha'}_{\mathbf{RR}'}(\mathbf{q})}{\epsilon(\mathbf{q})}, \qquad (3.27)$$

in which the effective density $n^{\alpha\alpha'}_{\mathbf{RR}'}(\mathbf{q})$ is likewise not analytic, but can be approximated as such. The leading terms here are dipole and quadrupole potentials (since the monopole term vanishes for an insulator at zero temperature). In zinc-blende crystals, only the interface dipole term is non-negligible under the approximation scheme used in this paper.

For use in Eqs. (3.18) and (3.21), one requires also expressions for the modified linear and quadratic potentials defined in Eq. (3.17). These are given by

$$\Delta v^{\alpha}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}') = w^{\alpha}_{\mathbf{GG}'}(\mathbf{k},\mathbf{k}')$$
$$+ e^{i(\mathbf{G}-\mathbf{G}')\cdot\tau_{\alpha}}\Lambda_{\mathbf{GG}'}(\mathbf{k},\mathbf{k}')\varphi^{\alpha}(\mathbf{k}-\mathbf{k}'),$$
$$(3.28a)$$

$$\Delta v^{\alpha\alpha'\mathbf{R}'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}') = w^{\alpha\alpha'\mathbf{R}'}_{\mathbf{GG}'}(\mathbf{k},\mathbf{k}') + e^{i(\mathbf{G}-\mathbf{G}')\cdot\tau_{\alpha}}$$
$$\times \Lambda_{\mathbf{GG}'}(\mathbf{k},\mathbf{k}')\varphi^{\alpha\alpha'\mathbf{R}'}(\mathbf{k}-\mathbf{k}'),$$
$$(3.28b)$$

in which the terms on the right-hand side are defined by expressions similar to (3.17).

IV. ENVELOPE-FUNCTION EQUATIONS

In this section, the Dyson equation (3.16) for a heterostructure is written in the Luttinger-Kohn representation,[127] in which the basis functions are defined to be plane waves multiplied by the zone-center Bloch functions U_n of the periodic reference crystal.[47,49] Here the set $\{U_n\}$ is chosen to be a complete set of solutions to the equation [cf. Eq. (3.1)]

$$-\frac{1}{2}\nabla^2 U_n(\mathbf{x},\omega) + \int V^{(\mathrm{r},0)}(\mathbf{x},\mathbf{x}',\omega)U_n(\mathbf{x}',\omega)d^3x'$$
$$= E_n(\omega)U_n(\mathbf{x},\omega) \qquad (4.1)$$

that satisfy the periodic boundary conditions

$$U_n(\mathbf{x},\omega) = U_n(\mathbf{x}+\mathbf{R},\omega), \qquad (4.2)$$

in which $V^{(\mathrm{r},0)}$ is the Hermitian part of the reference potential $V^{(0)}$. For a general operator A, the "real" (Hermitian) and "imaginary" (anti-Hermitian) parts are defined here as

$$A^{(\mathrm{r})} = \frac{1}{2}(A^{\dagger}+A), \quad A^{(\mathrm{i})} = \frac{i}{2}(A^{\dagger}-A). \qquad (4.3)$$

The solutions to Eq. (4.1) form a complete orthonormal set of periodic functions[127] for any value of ω. In the Fourier series representation

$$U_n(\mathbf{x},\omega) = \sum_{\mathbf{G}} U_{n\mathbf{G}}(\omega)e^{i\mathbf{G}\cdot\mathbf{x}}, \qquad (4.4)$$

the orthogonality and completeness relations are

$$\sum_{\mathbf{G}} U^{\dagger}_{n\mathbf{G}}(\omega)U_{n'\mathbf{G}}(\omega) = \delta_{nn'}, \qquad (4.5a)$$

$$\sum_n U_{n\mathbf{G}}(\omega)U^{\dagger}_{n\mathbf{G}'}(\omega) = \mathbb{1}\delta_{\mathbf{GG}'}, \qquad (4.5b)$$

in which $U^{\dagger}_{n\mathbf{G}}$ denotes the hermitian conjugate of the spinor $U_{n\mathbf{G}}$.

Note that the Bravais lattice chosen for the periodic boundary conditions in Eq. (4.2) need not be the same as that in (3.12); for certain applications it may be preferable to impose periodicity with respect to some (mathematically defined) supercell instead. For example, in treating intervalley Γ-X coupling in semiconductors with the zinc-blende structure, it is convenient[50,61] to choose a nonprimitive simple cubic unit cell of volume $\Omega_0 = a^3$ (where a is the conventional cubic lattice constant), which encompasses four primitive fcc cells.[147] This folds the X valleys onto the Brillouin zone center of the supercell, thereby permitting intervalley Γ-X mixing to be described in the same notation as that for ordinary Γ states (although the Γ and X states are of course not coupled by the $\mathbf{k}\cdot\mathbf{p}$ interaction).

The wave function in the Luttinger-Kohn representation (or envelope function) $F_n(\mathbf{k})$ is defined as

$$F_n(\mathbf{k}) = \sum_{\mathbf{G}} U^{\dagger}_{n\mathbf{G}}\psi(\mathbf{k}+\mathbf{G}), \qquad (4.6a)$$

$$\psi(\mathbf{k}+\mathbf{G}) = \sum_n F_n(\mathbf{k})U_{n\mathbf{G}}. \qquad (4.6b)$$

This definition is valid for any value of $\mathbf{k}$, but since only those values from one unit cell Ω_0^* are needed to determine $\psi(\mathbf{x})$, it is convenient to set $F_n(\mathbf{k}) \equiv 0$ when $\mathbf{k} \notin \Omega_0^*$.[49,148] The Fourier transform of (4.6b) is then the usual exact envelope-function expansion[49,127]

$$\psi(\mathbf{x}) = \sum_n F_n(\mathbf{x})U_n(\mathbf{x}). \qquad (4.7)$$

The Dyson equation in the Luttinger-Kohn representation is given by Eqs. (3.16), (4.5), and (4.6) as

$$E_n F_n(\mathbf{k}) + \sum_{n'} L_{nn'}(\mathbf{k}) F_{n'}(\mathbf{k}) + \sum_{n'} \sum_{\mathbf{k}' \in \Omega_0^*} \Delta V_{nn'}(\mathbf{k}, \mathbf{k}') F_{n'}(\mathbf{k}')$$

$$= E F_n(\mathbf{k}). \tag{4.8}$$

Here $\Delta V = V - V^{(0)}$ is the perturbation due to the heterostructure, the matrix elements of which are

$$\Delta V_{nn'}(\mathbf{k}, \mathbf{k}') = \sum_{\mathbf{G}, \mathbf{G}'} U_{n\mathbf{G}}^\dagger \Delta V(\mathbf{k} + \mathbf{G}, \mathbf{k}' + \mathbf{G}') U_{n'\mathbf{G}'}. \tag{4.9}$$

The term $L_{nn'}(\mathbf{k})$ groups together all contributions from the bulk reference crystal Hamiltonian except E_n; i.e.,

$$L_{nn'}(\mathbf{k}) = V_{nn'}^{(0)}(\mathbf{k}) - V_{nn'}^{(r,0)}(\mathbf{0}) + \mathbf{k} \cdot \mathbf{p}_{nn'} + \tfrac{1}{2} k^2 \delta_{nn'}, \tag{4.10}$$

in which $V_{nn'}^{(0)}(\mathbf{k}) = V_{nn'}^{(0)}(\mathbf{k}, \mathbf{k})$ is the potential energy of the reference crystal and

$$\mathbf{p}_{nn'} = \sum_{\mathbf{G}} \mathbf{G} U_{n\mathbf{G}}^\dagger U_{n'\mathbf{G}} \tag{4.11}$$

is the momentum matrix of the reference crystal.

Within quadratic response theory, the perturbation $\Delta V_{nn'}(\mathbf{k}, \mathbf{k}')$ is obtained by substituting Eqs. (3.18), (3.21), and (3.28) into Eq. (4.9). The result is

$$\Delta V_{nn'}(\mathbf{k}, \mathbf{k}') = W_{nn'}(\mathbf{k}, \mathbf{k}') + \Lambda_{nn'}(\mathbf{k}, \mathbf{k}') \varphi(\mathbf{k} - \mathbf{k}'), \tag{4.12}$$

where the vertex function

$$\Lambda_{nn'}(\mathbf{k}, \mathbf{k}') = \sum_{\mathbf{G}, \mathbf{G}'} U_{n\mathbf{G}}^\dagger \Lambda_{\mathbf{G}\mathbf{G}'}(\mathbf{k}, \mathbf{k}') U_{n'\mathbf{G}'} \tag{4.13}$$

is again an analytic function of $\mathbf{k}$ and $\mathbf{k}'$. The W term is defined by $W = W^{(1)} + W^{(2)}$, where

$$W_{nn'}^{(1)}(\mathbf{k}, \mathbf{k}') = \sum_\alpha{}' \theta^\alpha(\mathbf{k} - \mathbf{k}') W_{nn'}^\alpha(\mathbf{k}, \mathbf{k}'), \tag{4.14}$$

$$W_{nn'}^\alpha(\mathbf{k}, \mathbf{k}') = N \sum_{\mathbf{G}, \mathbf{G}'} U_{n\mathbf{G}}^\dagger w_{\mathbf{G}\mathbf{G}'}^\alpha(\mathbf{k}, \mathbf{k}') U_{n'\mathbf{G}'} e^{i(\mathbf{G}' - \mathbf{G}) \cdot \tau_\alpha}. \tag{4.15}$$

The screened potential φ is defined by

$$\varphi(\mathbf{q}) = \frac{v_c(\mathbf{q}) n(\mathbf{q})}{\epsilon(\mathbf{q})}, \tag{4.16}$$

where $n = n^{(1)} + n^{(2)}$ is an effective electron density for the heterostructure perturbation:

$$n^{(1)}(\mathbf{q}) = N \sum_\alpha{}' \theta^\alpha(\mathbf{q}) n^\alpha(\mathbf{q}). \tag{4.17}$$

The quadratic contributions $W^{(2)}$ and $n^{(2)}$ are given by obvious generalizations [see Eq. (3.28)] of the above expressions.

The physical interpretation of these results is considered below.

V. POWER SERIES EXPANSIONS

In this section, power series expansions are used to obtain approximate expressions for the Hamiltonian matrix elements in the envelope-function equations (4.8). This approximation is justified by the existence[50,51,128–131] of slowly varying envelope functions $F_n(\mathbf{x})$, for which $F_n(\mathbf{k})$ is negligible unless $\mathbf{k}$ is small. This expansion provides a starting point for the development of an approximate effective-mass theory, and also assists in the physical interpretation of the various terms in the Hamiltonian.

A. Reference crystal Hamiltonian

The leading terms in the bulk Hamiltonian $L_{nn'}(\mathbf{k})$ of Eq. (4.10) are derived from a Taylor series expansion of $V_{nn'}^{(0)}(\mathbf{k})$, and are given through terms of the fourth order in k by

$$L_{nn'}(\mathbf{k}) = k_\lambda \pi_{nn'}^\lambda + k_\lambda k_\mu \widetilde{D}_{nn'}^{\lambda\mu} + k_\lambda k_\mu k_\kappa \widetilde{C}_{nn'}^{\lambda\mu\kappa} + k_\lambda k_\mu k_\kappa k_\nu \widetilde{Q}_{nn'}^{\lambda\mu\kappa\nu}. \tag{5.1}$$

Here a summation with respect to the Cartesian indices $\lambda, \mu, \kappa,$ and ν is implicit, and the coefficients are

$$\pi_{nn'}^\lambda = p_{nn'}^\lambda + \left(\frac{\partial V_{nn'}^{(r,0)}(\mathbf{k})}{\partial k_\lambda} \right)_{\mathbf{k}=0},$$

$$\widetilde{D}_{nn'}^{\lambda\mu} = \frac{1}{2} \left(\delta_{\lambda\mu} \delta_{nn'} + \frac{\partial^2 V_{nn'}^{(r,0)}(\mathbf{k})}{\partial k_\lambda \, \partial k_\mu} \right)_{\mathbf{k}=0},$$

$$\widetilde{C}_{nn'}^{\lambda\mu\kappa} = \frac{1}{3!} \left(\frac{\partial^3 V_{nn'}^{(r,0)}(\mathbf{k})}{\partial k_\lambda \, \partial k_\mu \, \partial k_\kappa} \right)_{\mathbf{k}=0},$$

$$\widetilde{Q}_{nn'}^{\lambda\mu\kappa\nu} = \frac{1}{4!} \left(\frac{\partial^4 V_{nn'}^{(r,0)}(\mathbf{k})}{\partial k_\lambda \, \partial k_\mu \, \partial k_\kappa \, \partial k_\nu} \right)_{\mathbf{k}=0}. \tag{5.2}$$

The various derivatives of $V_{nn'}^{(r,0)}(\mathbf{k})$ account for contributions from the nonlocal part of the potential energy to the dispersion relation of the reference crystal. The quantity $\pi_{nn'}^\lambda$ is the kinetic momentum matrix of the reference crystal, whereas the other terms give partial contributions (see Sec. VI for the remaining contributions) to the effective-mass tensor and the cubic and quartic dispersion terms of the reference crystal.

In principle, $L_{nn'}(\mathbf{k})$ should include contributions from the antihermitian part of the self-energy $\Sigma_{nn'}^{(i,0)}(\mathbf{k}, \omega) = V_{nn'}^{(i,0)}(\mathbf{k}, \omega)$. However, in Appendix A it is shown that, for energies ω near the band gap of the reference crystal, the contributions from $\Sigma^{(i)}$ are much smaller than the smallest terms retained in the present approximation scheme. Such contributions were therefore neglected in the above expressions, and are likewise neglected in subsequent analysis of the heterostructure perturbation. However, in any calculation where it is desired

to include the effects of a finite quasiparticle lifetime, the dominant terms may be restored to leading order by replacing E_n with

$$E_n(\omega) \rightarrow E_n(\omega) + i\Sigma_{nn}^{(i,0)}(\mathbf{k}=\mathbf{0},\omega), \qquad (5.3)$$

and then retaining the imaginary part only to first order in perturbation theory in all subsequent analysis.

B. Linear heterostructure potential

A similar Taylor series expansion technique is useful for the terms $W_{nn'}(\mathbf{k},\mathbf{k}')$ and $\Lambda_{nn'}(\mathbf{k},\mathbf{k}')$ in the heterostructure perturbation (4.12). This subsection begins by considering the simple special case in which the screened atomic pseudopotentials $w_{\mathbf{GG}'}^{\alpha}(\mathbf{k},\mathbf{k}')$ in Eqs. (3.28a) and (4.15) have the form of a local potential; i.e.,

$$w_{\mathbf{GG}'}^{\alpha}(\mathbf{k},\mathbf{k}') = w^{\alpha}(\mathbf{k}-\mathbf{k}'+\mathbf{G}-\mathbf{G}'). \qquad (5.4)$$

Such would be the case, for example, in an LDA calculation based on local ionic pseudopotentials. In this case, $W_{nn'}^{(1)} \times (\mathbf{k},\mathbf{k}')$ is also a local potential of the form $W_{nn'}^{(1)}(\mathbf{k},\mathbf{k}') = W_{nn'}^{(1)}(\mathbf{k}-\mathbf{k}')$. This simplification makes it easier to grasp the physical significance of the power series expansion, and also facilitates a comparison between the present theory and earlier envelope-function models based on local empirical pseudopotentials.

1. Local analytic terms

The approximation technique used here for $W_{nn'}^{(1)}(\mathbf{k}-\mathbf{k}')$ is to expand the term $W_{nn'}^{\alpha}(\mathbf{k}-\mathbf{k}')$ on the right-hand side of Eq. (4.14) in a Taylor series in $\mathbf{q} \equiv \mathbf{k}-\mathbf{k}'$:

$$W_{nn'}^{(1)}(\mathbf{q}) = \sum_{\alpha}{}' \theta^{\alpha}(\mathbf{q})[W_{nn'}^{\alpha} + iq_{\lambda}(Z_{nn'}^{\lambda})^{\alpha} - q_{\lambda}q_{\mu}(Y_{nn'}^{\lambda\mu})^{\alpha}].$$

$$(5.5)$$

Here the terms $W_{nn'}^{\alpha}$, $(Z_{nn'}^{\lambda})^{\alpha}$, and $(Y_{nn'}^{\lambda\mu})^{\alpha}$ are $\mathbf{q}$-independent expansion coefficients. The series (5.5) has been truncated at the second order in q because such terms are of order $(\bar{k}a)^2\Delta\bar{V}$, which are the smallest corrections permitted in the present perturbation scheme.

The result (5.5) may be used directly in a $\mathbf{k}$-space envelope-function calculation based on Eq. (4.8), in which the $\mathbf{k}$ values are expressly limited to the unit cell Ω_0^*. However, to obtain a local differential equation, one must allow $\mathbf{k}$ to range over all possible values. Since $\theta^{\alpha}(\mathbf{k})$ is quasiperiodic [see Eq. (3.20)], this local approximation will generate large-$\mathbf{k}$ terms in the envelope functions unless a $\mathbf{k}$-space cutoff is introduced. If this is done, the Fourier transform of (5.5) is the local potential

$$W_{nn'}^{(1)}(\mathbf{x}) = \sum_{\alpha}{}' [W_{nn'}^{\alpha} + (Z_{nn'}^{\lambda})^{\alpha}\partial_{\lambda} + (Y_{nn'}^{\lambda\mu})^{\alpha}\partial_{\lambda}\partial_{\mu}]\theta^{\alpha}(\mathbf{x}),$$

$$(5.6)$$

in which $\partial_{\lambda} = \partial/\partial x_{\lambda}$ and [cf. Eq. (3.19a)]

$$\theta^{\alpha}(\mathbf{x}) = \sum_{\mathbf{k}} B(\mathbf{k})\theta^{\alpha}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}, \qquad (5.7)$$

where the cutoff function $B(\mathbf{k})$ is defined in Appendix B. From Eq. (3.8), the constraint

$$\sum_a \theta^{aj}(\mathbf{x}) = 0 \qquad (5.8)$$

is satisfied for any choice of $B(\mathbf{k})$.

The physical significance of the result (5.6) can be appreciated by considering a specific example such as a (001) GaAs/AlAs heterojunction. In this case, as will be shown in Sec. VII, the function $\theta^{\alpha}(\mathbf{x})$ depends only on the z coordinate and behaves like a smooth step function at the interface. The spatial derivatives in Eq. (5.6) therefore generate finite-width δ-like terms at the interface, with the Z term proportional to $\delta(z)$ and the Y term proportional to $\delta'(z)$. Hence, Eq. (5.6) provides a first example of the interface band-mixing terms alluded to previously in Eqs. (2.5) and (2.6).

The physical origin of these terms can be understood by going back one step further in the derivation. From Eq. (4.15), it is clear that the Taylor series expansion of $W_{nn'}^{\alpha}(\mathbf{q})$ in Eq. (5.5) is equivalent to a Taylor series expansion of the screened atomic pseudopotential (5.4) with respect to $\mathbf{q} = \mathbf{k}-\mathbf{k}'$. Hence, the physical origin of the linear and quadratic (in q) terms in Eq. (5.5) is simply the finite slope and curvature of the screened atomic pseudopotentials in momentum space.

This demonstrates that the Hamiltonian of a heterostructure depends not just on the values of the atomic pseudopotentials $w^{\alpha}(\mathbf{k})$ at the reciprocal lattice vectors $\mathbf{G}$, but also at a *range* of $\mathbf{k}$ values in a finite neighborhood of each $\mathbf{G}$ (the size of the neighborhood depending on how rapidly varying the envelope function is). The necessity for an accurate fitting of empirical pseudopotentials over a range of $\mathbf{k}$ values has been emphasized particularly in the work of Mäder and Zunger.[149] The truncated expansion (5.5) shows that the present perturbation scheme relies for its accuracy upon the validity of a quadratic extrapolation of $w^{\alpha}(\mathbf{k})$ in the neighborhood of each $\mathbf{G}$. Inspection of the form of typical atomic pseudopotentials[145,150] shows this to be a good approximation.

The definition (5.5) of the interface band-mixing parameters Z and Y highlights a significant difference between the present approach and previous derivations of envelope-function Hamiltonians. In previous derivations, the heterostructure potential was chosen to have the form

$$V(\mathbf{x}) = \sum_l V^l(\mathbf{x})\theta^l(\mathbf{x}), \qquad (5.9)$$

in which $V^l(\mathbf{x})$ is the periodic potential for the bulk material l, and $\theta^l(\mathbf{x})$ is a form factor determining the composition of the heterostructure. Within this model, $V(\mathbf{x})$ depends only upon the atomic pseudopotentials at the reciprocal lattice vectors $\mathbf{G}$, with the magnitude of the interface δ-like terms determined by $\theta^l(\mathbf{x})$. In the limit of slowly varying $\theta^l(\mathbf{x})$, the interface terms vanish.

The present results show that this behavior is an unphysical artifact of the model (5.9). In the present theory, the strength of the interface terms is the same (in the virtual crystal approximation) for slowly graded and abrupt heterostructures, with the rate of change of $\theta^\alpha(\mathbf{x})$ affecting only the width of the interface terms.

2. Nonlocal analytic terms

The more general case of a nonlocal potential involves only a straightforward extension of these results. The matrix element $W^\alpha_{nn'}(\mathbf{k},\mathbf{k}')$ of Eq. (4.15) is expanded to second order in $\mathbf{k}$ and $\mathbf{k}'$ as follows:

$$W^\alpha_{nn'}(\mathbf{k},\mathbf{k}') = W^\alpha_{nn'} + k_\lambda(\widetilde{J}^\lambda_{nn'})^\alpha + k'_\lambda[(\widetilde{J}^\lambda_{n'n})^\alpha]^* + k_\lambda k_\mu(\widetilde{M}^{\lambda\mu}_{nn'})^\alpha$$
$$+ k'_\lambda k'_\mu[(\widetilde{M}^{\mu\lambda}_{n'n})^\alpha]^* + k_\lambda k'_\mu(\widetilde{R}^{\lambda\mu}_{nn'})^\alpha. \tag{5.10}$$

Here the anti-Hermitian part of the self-energy was neglected, as discussed above Eq. (5.3) and in Appendix A. This approximation simplifies the expansion by providing a relationship between (for example) the coefficients of k_λ and k'_λ. The physical interpretation of the various expansion coefficients is discussed below (in Sec. VI), after perturbation theory has been used to eliminate the coupling to remote bands.

When this expansion is substituted into Eq. (4.14) for $W^{(1)}_{nn'}(\mathbf{k},\mathbf{k}')$, the result can be written as

$$W^{(1)}_{nn'}(\mathbf{k},\mathbf{k}') = W^{(1)}_{nn'}(\mathbf{q}) + k_\lambda\widetilde{J}^\lambda_{nn'}(\mathbf{q}) + k'_\lambda[\widetilde{J}^\lambda_{n'n}(-\mathbf{q})]^*$$
$$+ k_\lambda k_\mu\widetilde{M}^{\lambda\mu}_{nn'}(\mathbf{q}) + k'_\lambda k'_\mu[\widetilde{M}^{\mu\lambda}_{n'n}(-\mathbf{q})]^*$$
$$+ k_\lambda k'_\mu\widetilde{R}^{\lambda\mu}_{nn'}(\mathbf{q}), \tag{5.11}$$

in which $\mathbf{q}=\mathbf{k}-\mathbf{k}'$ as before, and the various $\mathbf{q}$-dependent functions are defined by expressions of the form

$$\widetilde{R}^{\lambda\mu}_{nn'}(\mathbf{q}) = \sum_\alpha{}' \theta^\alpha(\mathbf{q})(\widetilde{R}^{\lambda\mu}_{nn'})^\alpha. \tag{5.12}$$

These functions all have a step-function-like behavior in $\mathbf{x}$ space at a heterojunction (see Sec. VIII). Since $\theta^\alpha_\mathbf{R}$ is real, they also have the hermiticity and symmetry properties

$$[W^{(1)}_{nn'}(\mathbf{q})]^* = W^{(1)}_{n'n}(-\mathbf{q}),$$

$$[\widetilde{R}^{\lambda\mu}_{nn'}(\mathbf{q})]^* = \widetilde{R}^{\mu\lambda}_{n'n}(-\mathbf{q}),$$

$$\widetilde{M}^{\lambda\mu}_{nn'}(\mathbf{q}) = \widetilde{M}^{\mu\lambda}_{nn'}(\mathbf{q}). \tag{5.13}$$

3. Nonanalytic terms

The term $\Lambda_{nn'}(\mathbf{k},\mathbf{k}')\varphi^{(1)}(\mathbf{q})$ describing the nonanalytic contributions to Eq. (4.12) is handled in much the same way. The vertex function $\Lambda_{nn'}(\mathbf{k},\mathbf{k}')$ is expanded in a Taylor series of the form (5.10), while the linear electron density in Eq. (4.17) is expanded as

$$n^\alpha(\mathbf{q}) = q_\lambda n^\alpha_\lambda + q_\lambda q_\mu n^\alpha_{\lambda\mu} + q_\lambda q_\mu q_\kappa n^\alpha_{\lambda\mu\kappa} + q_\lambda q_\mu q_\kappa q_\nu n^\alpha_{\lambda\mu\kappa\nu}, \tag{5.14}$$

which is valid for neutral perturbations. In a heterovalent zinc-blende heterostructure described by neutral perturbations,[124,126] the dipole and traceless quadrupole terms n^α_λ and $n^\alpha_{\lambda\mu}-\frac{1}{3}n^\alpha_{\nu\nu}\delta_{\lambda\mu}$ are nonvanishing only at interfaces. For isovalent substitutions in zinc-blende, $n^\alpha_\lambda=0$ and $n^\alpha_{\lambda\mu}=\frac{1}{3}n^\alpha_{\nu\nu}\delta_{\lambda\mu}$ everywhere, so the contribution from the latter term is analytic and can be absorbed into the definition of the analytic potential (5.10). The remaining terms $n^\alpha_{\lambda\mu\kappa}$ and $n^\alpha_{\lambda\mu\kappa\nu}$ describe octopole and hexadecapole moments.[124]

The inverse dielectric function $\epsilon^{-1}(\mathbf{q})$ in Eq. (4.16) can also be expanded in a power series. For isovalent perturbations in zinc-blende this has no qualitative significance, since the leading correction merely renormalizes the hexadecapole term $n^\alpha_{\lambda\mu\kappa\nu}$.[124] However, the wave vector dependence of $\epsilon^{-1}(\mathbf{q})$ does generate a qualitatively new contribution from the interface dipole term in Eq. (5.14); see Eq. (6.23) of Ref. 124 for the explicit form of this term.

With the expansion (5.14), the effective density (4.17) has the same form as that derived above for the local potential in Eq. (5.5). In coordinate space, it involves a series of derivatives of $\theta^\alpha(\mathbf{x})$, similar to the result shown in Eq. (5.6). This is the same as the usual multipole expansion of the macroscopic charge density in classical electromagnetism.[151] In particular, note that for a perturbation consisting of a single impurity atom, the function $\theta^\alpha(\mathbf{x})$ is just the macroscopic average of a Dirac δ function, in complete agreement with the expressions given in Ref. 151.

Since the Taylor series expansion for $\Lambda_{nn'}(\mathbf{k},\mathbf{k}')$ is identical in form to that for $W^\alpha_{nn'}(\mathbf{k},\mathbf{k}')$, the contribution from this term will not be written out explicitly here. All subsequent perturbation theory analysis for the two terms is formally identical, except that the vertex function is multiplied by an extra factor of $\varphi(\mathbf{q})$.

C. Quadratic heterostructure potential

1. Analytic terms

Because the quadratic response is already of order $(\Delta\bar{V})^2/\bar{E}_\mathrm{g}$, the quadratic version of Eq. (4.14) can be replaced by the zeroth-order approximation

$$W^{(2)}_{nn'}(\mathbf{k},\mathbf{k}') = \sum_{\alpha,\alpha',\mathbf{R}'}{}' \theta^{\alpha\alpha'\mathbf{R}'}(\mathbf{q})W^{\alpha\alpha'\mathbf{R}'}_{nn'} \equiv \widetilde{W}^{(2)}_{nn'}(\mathbf{q}), \tag{5.15}$$

in which

$$W^{\alpha\alpha'\mathbf{R}'}_{nn'} = \lim_{k,k'\to 0} W^{\alpha\alpha'\mathbf{R}'}_{nn'}(\mathbf{k},\mathbf{k}'), \tag{5.16}$$

where the limit is well defined for neutral perturbations.[124] This contributes a local potential-energy term similar to that given by $W^{(1)}_{nn'}(\mathbf{q})$ in Eq. (5.11). At a heterojunction, the functional dependence in $\mathbf{x}$ space is similar to that of a smooth (macroscopically averaged) step function, with possible deviations in the vicinity of the junction.

Now since $W_{nn'}^{\alpha\alpha'\mathbf{R}'}$ depends only on the part of $W_{nn'}^{\alpha\alpha'\mathbf{R}'}(\mathbf{k},\mathbf{k}')$ that is analytic in $\mathbf{k}$ and $\mathbf{k}'$, it is a short-range quantity that is significant only when $\mathbf{R}'+\boldsymbol{\tau}_{\alpha'}-\boldsymbol{\tau}_{\alpha}$ is comparable to the lattice constant a. Therefore, at a heterojunction, replacing $\widetilde{W}_{nn'}^{(2)}(\mathbf{q})$ with an ideal step function would generate an error of order $\bar{k}a$ in a term of order $(\bar{k}a)^2\Delta\bar{V}$. Hence, the error is of the same order as the local approximation (2.8) and can be neglected.

Thus, within the present perturbation scheme there is no interface contribution from $\widetilde{W}_{nn'}^{(2)}(\mathbf{q})$. This is an important result, as it simplifies the analysis of interface effects in later sections of this paper.

Within the context of an empirical pseudopotential model,[149] the nonlinear bulk term $\widetilde{W}_{nn'}^{(2)}(\mathbf{q})$ derived here can be represented as an environment dependence of the screened empirical pseudopotential, in which (for example) the pseudopotential for an As atom in GaAs is different from that for an As atom in AlAs. The importance of accounting for such effects has been emphasized by Mäder and Zunger.[149]

2. Nonanalytic terms

The leading terms in the quadratic density are the dipole and quadrupole terms

$$n^{\alpha\alpha'\mathbf{R}'}(\mathbf{q}) = q_\lambda n_\lambda^{\alpha\alpha'\mathbf{R}'} + q_\lambda q_\mu n_{\lambda\mu}^{\alpha\alpha'\mathbf{R}'}, \qquad (5.17)$$

which again is valid for neutral perturbations. (Here the constant term vanishes even for charged perturbations as long as the system is insulating,[124] but in this case there is an additional nonanalytic term of order q^2.[124]) In zinc-blende materials, the net contributions to $n^{(2)}(\mathbf{q})$ from $n_\lambda^{\alpha\alpha'\mathbf{R}'}$ and the traceless part of $n_{\lambda\mu}^{\alpha\alpha'\mathbf{R}'}$ both vanish in the bulk regions of a heterostructure. The interface part of $n_{\lambda\mu}^{\alpha\alpha'\mathbf{R}'}$ is negligible under the current approximation scheme, while the bulk part can be absorbed into the definition of the analytic potential $\widetilde{W}_{nn'}^{(2)}(\mathbf{q})$. Therefore, only the interface dipole term remains under the current approximation scheme.

D. Two-dimensional systems

In a heterostructure (such as a quantum well) with two-dimensional translation symmetry, the nonanalytic terms arising from $\varphi(\mathbf{q})$ have a particularly simple form. Let the dimensionless coordinates x_i and k_j be defined by $\mathbf{x}=x_i\mathbf{a}_i$ and $\mathbf{k}=k_j\mathbf{b}_j$, where $\mathbf{a}_i$ and $\mathbf{b}_j$ are basis vectors for the direct and reciprocal lattices of the reference crystal, with $\mathbf{a}_i\cdot\mathbf{b}_j=2\pi\delta_{ij}$. In these coordinates, the lattice sites are defined by $x_i=R_i$ and $k_j=G_j$, where R_i and G_j are integers. In such a two-dimensional system, one can choose $\mathbf{a}_1$ and $\mathbf{a}_2$ to lie parallel to the junction plane, so that the atomic distribution function $\theta_\mathbf{R}^\alpha$ is independent of R_1 and R_2. The Fourier transform $\theta^\alpha(\mathbf{k})$ then has the form

$$\theta^\alpha(\mathbf{k}) = \theta^\alpha(k_3)\sum_{\mathbf{G}_\parallel} \delta_{\mathbf{k}_\parallel\mathbf{G}_\parallel}e^{-i\mathbf{G}_\parallel\cdot\boldsymbol{\tau}_\alpha}, \qquad (5.18)$$

where $\mathbf{k}_\parallel=k_1\mathbf{b}_1+k_2\mathbf{b}_2$. For small $\mathbf{k}_\parallel$, the $\mathbf{k}_\parallel$ dependence is simply $\delta_{\mathbf{k}_\parallel 0}$. The same conclusion holds for the pair distribution function $\theta^{\alpha\alpha'\mathbf{R}'}(\mathbf{k})$.

Therefore, the nonanalytic potential (4.16) has the form $\varphi(\mathbf{q})=\varphi(q_3)\delta_{\mathbf{q}_\parallel 0}$, which is independent of the direction of $\mathbf{q}$. As a result, all terms in $\varphi(\mathbf{q})$ except the monopole and dipole terms reduce to analytic functions of q_3, which can be absorbed into the definition of the analytic potential W. In regard to the monopole and dipole terms, this paper considers only neutral perturbations in crystals with no bulk dipole moment. Thus the only nonanalytic contributions are the $1/q_3$ terms generated by the interface dipoles in Eqs. (5.14) and (5.17).

These have the same q_3 dependence as the Fourier transform of a step function, and merely add extra terms to the band offsets at a heterojunction. Hence, the interface dipole contributions can be absorbed into the definition of $W_{nn'}^{(1)}(\mathbf{q})$ and $\widetilde{W}_{nn'}^{(2)}(\mathbf{q})$. Note that in a no-common-atom system such as InAs/GaSb, the contribution from the quadratic interface dipole to the band offset has a different value for GaAs-like and InSb-like interfaces.[152]

In summary, for the material systems considered in this paper, the nonanalytic potential $\varphi(\mathbf{q})$ does not contribute anything qualitatively new in a heterostructure with two-dimensional translation symmetry. Only the interface dipole term is truly nonanalytic, and that can be absorbed into the definition of the band offsets (although this contribution does depend on the microscopic structure of the interface[152]). Long-range potentials arising from the direction dependence of $\varphi(\mathbf{q})$ appear only in structures with lower translation symmetry, such as quantum wires and dots.

VI. ELIMINATION OF INTERBAND COUPLING

In this section, perturbation theory[31,34,127] is used to derive a multi-band effective-mass Hamiltonian from the infinite-dimensional matrix equations (4.8). The method is outlined briefly here; for further details, see Refs. 34 and 31.

The zone-center states n of the reference crystal are divided into a class A containing the states of interest, and a class B containing all other states. The total Hamiltonian is written as $H=H_0+H'$, where H_0 has matrix elements $(H_0)_{mm'}=E_m\delta_{mm'}$, and $m=(n,\mathbf{k})$ is a composite index. The unperturbed Hamiltonian H_0 is assumed to be Hermitian, but the perturbation H' need not be. A similarity transformation $\bar{H}=e^{-S}He^S$ is used to eliminate the coupling between sets A and B to any desired order in the perturbation H'. (The transformation is unitary if H' is Hermitian, as is approximately the case here.) This yields a finite-dimensional effective Hamiltonian for states $m,m'\in A$, which is given explicitly (to third order in H') in Eq. (C1) of Appendix C.

For the case considered here, the matrix elements of H_0 and H' are given by Eq. (4.8) as $\langle n\mathbf{k}|H_0|n'\mathbf{k}'\rangle=E_n\delta_{nn'}\delta_{\mathbf{kk}'}$ and

$$\langle n\mathbf{k}|H'|n'\mathbf{k}'\rangle = L_{nn'}(\mathbf{k})\delta_{\mathbf{kk}'} + W_{nn'}^{(1)}(\mathbf{k},\mathbf{k}') + \widetilde{W}_{nn'}^{(2)}(\mathbf{q})$$
$$+ \Lambda_{nn'}(\mathbf{k},\mathbf{k}')\varphi(\mathbf{q}), \tag{6.1}$$

in which $L_{nn'}(\mathbf{k})$ is defined in Eq. (5.1), $W_{nn'}^{(1)}(\mathbf{k},\mathbf{k}')$ in Eq. (5.11), and $\widetilde{W}_{nn'}^{(2)}(\mathbf{q})$ in Eq. (5.15). Upon inserting these matrix elements into Eq. (C1), one obtains the effective-mass Hamiltonian (for $n, n' \in A$)

$$\langle n\mathbf{k}|\bar{H}|n'\mathbf{k}'\rangle = (E_n\delta_{nn'} + k_\lambda \pi_{nn'}^\lambda + k_\lambda k_\mu D_{nn'}^{\lambda\mu} + k_\lambda k_\mu k_\kappa C_{nn'}^{\lambda\mu\kappa}$$
$$+ k_\lambda k_\mu k_\kappa k_\nu Q_{nn'}^{\lambda\mu\kappa\nu})\delta_{\mathbf{kk}'} + W_{nn'}^{(1)}(\mathbf{q}) + W_{nn'}^{(2)}(\mathbf{q})$$
$$+ k_\lambda J_{nn'}^\lambda(\mathbf{q}) + k_\lambda'[J_{n'n}^\lambda(-\mathbf{q})]^* + k_\lambda k_\mu M_{nn'}^{\lambda\mu}(\mathbf{q})$$
$$+ k_\lambda' k_\mu'[M_{n'n}^{\mu\lambda}(-\mathbf{q})]^* + k_\lambda k_\mu' R_{nn'}^{\lambda\mu}(\mathbf{q}) + [\Lambda_{nn'}$$
$$+ k_\lambda \hat{J}_{nn'}^\lambda + k_\lambda'(\hat{J}_{n'n}^\lambda)^* + k_\lambda k_\mu \hat{M}_{nn'}^{\lambda\mu} + k_\lambda' k_\mu'(\hat{M}_{n'n}^{\mu\lambda})^*$$
$$+ k_\lambda k_\mu' \hat{R}_{nn'}^{\lambda\mu}]\varphi(\mathbf{q}). \tag{6.2}$$

In this expression, $\pi_{nn'}^\lambda$ is the kinetic momentum matrix (5.2) of the reference crystal, $2D_{nn'}^{\lambda\mu}$ is the inverse effective mass tensor of the reference crystal [see Eq. (C2)], and $C_{nn'}^{\lambda\mu\kappa}$ and $Q_{nn'}^{\lambda\mu\kappa\nu}$ are the coefficients of the cubic and quartic dispersion terms in the reference crystal [see Eqs. (C3) and (C4)]. These are just renormalized versions of the quantities $\widetilde{D}_{nn'}^{\lambda\mu}, \widetilde{C}_{nn'}^{\lambda\mu\kappa}$, and $\widetilde{Q}_{nn'}^{\lambda\mu\kappa\nu}$ defined previously in Eq. (5.2).

Likewise, the functions $W_{nn'}^{(2)}(\mathbf{q}), J_{nn'}^\lambda(\mathbf{q}), M_{nn'}^{\lambda\mu}(\mathbf{q})$, and $R_{nn'}^{\lambda\mu}(\mathbf{q})$ are all renormalized versions of quantities defined previously. The renormalized functions are given explicitly in Appendix C.

The terms multiplying the screened nonanalytic potential $\varphi(\mathbf{q})$ are derived from the Taylor series expansion and $\mathbf{k}\cdot\boldsymbol{\pi}$ renormalization of the vertex function $\Lambda_{nn'}(\mathbf{k},\mathbf{k}')$. The various constant coefficients (e.g., $\hat{R}_{nn'}^{\lambda\mu}$) are defined in the same way as the analogous $\mathbf{q}$-dependent functions [e.g., $R_{nn'}^{\lambda\mu}(\mathbf{q})$] given in Appendix C, but with $W_{nn'}^{(1)}(\mathbf{q})$ replaced by the constant $\Lambda_{nn'} = \Lambda_{nn'}(\mathbf{0},\mathbf{0})$.

Equation (6.2) can now be rearranged[47,48] and Fourier transformed to obtain the effective-mass Hamiltonian

$$\bar{H}_{nn'}(\mathbf{x},\mathbf{p}) = E_n\delta_{nn'} + \{p_\lambda, \pi_{nn'}^\lambda(\mathbf{x})\} + \{p_\lambda p_\mu, D_{nn'}^{\{\lambda\mu\}}(\mathbf{x})\}$$
$$+ p_\lambda p_\mu p_\kappa C_{nn'}^{\lambda\mu\kappa} + p_\lambda p_\mu p_\kappa p_\nu Q_{nn'}^{\lambda\mu\kappa\nu} + W_{nn'}^{(1)}(\mathbf{x})$$
$$+ W_{nn'}^{(2)}(\mathbf{x}) + \partial_\lambda Z_{nn'}^\lambda(\mathbf{x}) + \partial_\lambda \partial_\mu Y_{nn'}^{\lambda\mu}(\mathbf{x})$$
$$+ \{p_\lambda, \partial_\mu \Gamma_{nn'}^{\lambda\mu}(\mathbf{x})\} + [\partial_\mu \Phi_{nn'}^{\lambda\mu}(\mathbf{x})]p_\lambda + \Lambda_{nn'}\varphi(\mathbf{x})$$
$$+ \hat{\pi}_{nn'}^\lambda\{p_\lambda, \varphi(\mathbf{x})\} + \hat{D}_{nn'}^{\{\lambda\mu\}}\{p_\lambda p_\mu, \varphi(\mathbf{x})\}$$
$$+ \hat{Z}_{nn'}^\lambda \partial_\lambda \varphi(\mathbf{x}) + \hat{Y}_{nn'}^{\lambda\mu} \partial_\lambda \partial_\mu \varphi(\mathbf{x}) + \hat{\Gamma}_{nn'}^{\lambda\mu}\{p_\lambda, \partial_\mu \varphi(\mathbf{x})\}$$
$$+ \hat{\Phi}_{nn'}^{\lambda\mu}[\partial_\mu \varphi(\mathbf{x})]p_\lambda, \tag{6.3}$$

in which $\{A,B\} = \{AB\} = \frac{1}{2}(AB+BA)$ is the symmetrized product, $\mathbf{p} = -i\boldsymbol{\nabla}$ is the momentum operator, and $\partial_\lambda = \partial/\partial x_\lambda$ acts

only on the function immediately to its right.

All of the $\mathbf{x}$-dependent functions are defined in terms of the $\mathbf{k}$-space cutoff (5.7). The quantity

$$\pi_{nn'}^\lambda(\mathbf{x}) = \pi_{nn'}^\lambda + J_{nn'}^\lambda(\mathbf{x}) + [J_{n'n}^\lambda(\mathbf{x})]^* \tag{6.4}$$

is the material-dependent kinetic momentum matrix for the heterostructure, while

$$D_{nn'}^{\lambda\mu}(\mathbf{x}) = D_{nn'}^{\lambda\mu} + M_{nn'}^{\lambda\mu}(\mathbf{x}) + [M_{n'n}^{\mu\lambda}(\mathbf{x})]^* + R_{nn'}^{\lambda\mu}(\mathbf{x}) \tag{6.5}$$

is half the inverse effective mass tensor for the heterostructure. This has symmetric and antisymmetric parts:

$$D_{nn'}^{\{\lambda\mu\}}(\mathbf{x}) = \frac{1}{2}[D_{nn'}^{\lambda\mu}(\mathbf{x}) + D_{nn'}^{\mu\lambda}(\mathbf{x})],$$

$$D_{nn'}^{[\lambda\mu]}(\mathbf{x}) = \frac{1}{2}[D_{nn'}^{\lambda\mu}(\mathbf{x}) - D_{nn'}^{\mu\lambda}(\mathbf{x})], \tag{6.6}$$

although the antisymmetric part has a nonvanishing contribution only in the presence of a magnetic field.

The functions Z, Y, Γ, and Φ in Eq. (6.3) are all interface terms. The first two terms

$$Z_{nn'}^\lambda(\mathbf{x}) = -i\frac{1}{2}(J_{nn'}^\lambda(\mathbf{x}) - [J_{n'n}^\lambda(\mathbf{x})]^*), \tag{6.7}$$

$$Y_{nn'}^{\lambda\mu}(\mathbf{x}) = \frac{1}{2}R_{nn'}^{\{\lambda\mu\}}(\mathbf{x}), \tag{6.8}$$

are renormalized versions of the δ and δ' mixing potentials considered previously in Eqs. (5.5) and (5.6). However, the other two terms

$$\Gamma_{nn'}^{\lambda\mu}(\mathbf{x}) = -i(M_{nn'}^{\{\lambda\mu\}}(\mathbf{x}) - [M_{n'n}^{\{\lambda\mu\}}(\mathbf{x})]^*), \tag{6.9}$$

$$\Phi_{nn'}^{\lambda\mu}(\mathbf{x}) = iR_{nn'}^{[\lambda\mu]}(\mathbf{x}), \tag{6.10}$$

were not present in Eqs. (5.5) and (5.6). The term $\Phi_{nn'}^{\lambda\mu}(\mathbf{x})$, which is antisymmetric in λ and μ, is just a generalized Rashba coefficient[31,153–156] for multiband Hamiltonians. However, the term $\Gamma_{nn'}^{\lambda\mu}(\mathbf{x})$, which is symmetric in λ and μ,[157] has received little attention in the literature. Its physical interpretation will be discussed below in Sec. VIII.

Since the functions Z, Y, Γ, and Φ behave to lowest order as step functions at an abrupt junction, the Hamiltonian (6.3) shows explicitly that these functions produce interface terms proportional to δ or δ'.

The remaining terms ($\hat{\pi}, \hat{Z}$, etc.) in Eq. (6.3) that appear in front of $\varphi(\mathbf{x})$ are defined by the obvious generalizations of Eqs. (6.4)–(6.10). The interpretation of these terms parallels that of the terms already discussed, except that the contributions from φ are not as well localized at the interface.

Here it is worth noting that in the GW approximation,[94,95] one has $\Lambda_{nn'} = \delta_{nn'}$ and $\hat{\pi}_{nn'}^\lambda = \hat{D}_{nn'}^{\lambda\mu} = 0$, so the potential $\varphi(\mathbf{x})$ does not contribute to the momentum matrix or the effective-mass tensor. In LDA, these simplifications are also valid, and one has in addition $\hat{Z}_{nn'}^\lambda = 0$, since the exchange-correlation potential is short-ranged.

For isovalent zinc-blende systems, the leading term in $\varphi^{(1)}$ is an octopole potential, so the contributions from the second-rank tensors $\hat{D}, \hat{Y}, \hat{\Gamma}$, and $\hat{\Phi}$ are negligible. These

terms are non-negligible only for the linear interface dipole term in a heterovalent zinc-blende system (or for a slowly varying external potential, which is not considered explicitly here). Likewise, $\varphi^{(2)}$ is negligible in all terms except $\Lambda_{nn'}$.

Equation (6.3) is the main result obtained in this paper. The qualitative form of this Hamiltonian is very similar to the Leibler Hamiltonian[47,48] for slowly graded heterostructures, as amended by Takhtamirov and Volkov.[65,66,68] The differences are primarily due to the use of atomic pseudopotentials (rather than a model based on periodic bulk potentials), the inclusion of long-range Coulomb potentials, and the use of linear response theory to simplify the interface Hamiltonian, as discussed in Sec. IX.

The explicit form of the various material parameters in Eq. (6.3) for semiconductors with the zinc-blende structure is given below in Sec. VIII. First, however, the possibility of representing the material parameters as piecewise constant is considered.

VII. SIMPLIFIED MATHEMATICAL REPRESENTATION OF HETEROSTRUCTURE MATERIAL PROPERTIES

This section discusses several ways in which the mathematical description of material properties can be simplified. The first is to label the materials in terms of bulk compounds (e.g., GaAs) rather than atoms; the second is to approximate the properties of an abrupt junction using piecewise constant material parameters with δ functions and their derivatives at interfaces. For simplicity, only two-dimensional isovalent systems are considered here (see Sec. V D).

A. Transformation from atomic to bulk-crystal description

All of the linear-response terms can be transformed immediately to a bulk-crystal representation similar to that described above in Eq. (5.9):

$$W_{nn'}^{(1)}(\mathbf{x}) = \sum_{\alpha}{}' \theta^\alpha(\mathbf{x}) W_{nn'}^\alpha, \qquad (7.1a)$$

$$= \sum_{l}{}' \theta^l(\mathbf{x}) W_{nn'}^l, \qquad (7.1b)$$

in which l labels the different bulk materials, and [cf. Eq. (5.8)]

$$\sum_l \theta^l(\mathbf{x}) = 0. \qquad (7.1c)$$

For example, in a GaAs/AlAs heterostructure, the linear potentials for the two bulk media are defined by

$$W_{nn'}^{GaAs} = W_{nn'}^{Ga} + W_{nn'}^{As},$$

$$W_{nn'}^{AlAs} = W_{nn'}^{Al} + W_{nn'}^{As}. \qquad (7.2)$$

If the reference crystal is chosen to be GaAs, then $W_{nn'}^{Ga} = W_{nn'}^{As} = 0$. [This choice is made for clarity of exposition; the final results given in Eqs. (7.4), (7.9), (7.14) and (7.16) do

not depend on the choice of reference crystal.] The transformation from (7.1a) and (7.1b) is then simply

$$W_{nn'}^{(1)}(\mathbf{x}) = \theta^{Al}(\mathbf{x}) W_{nn'}^{Al} = \theta^{Al}(\mathbf{x}) W_{nn'}^{AlAs}. \qquad (7.3)$$

Hence

$$\theta^{AlAs} = \theta^{Al}, \qquad \theta^{GaAs} = \theta^{Ga}. \qquad (7.4)$$

For a no-common-atom system such as InAs/GaSb, four different bulk potentials can be defined:

$$W_{nn'}^{InAs} = W_{nn'}^{In} + W_{nn'}^{As},$$

$$W_{nn'}^{InSb} = W_{nn'}^{In} + W_{nn'}^{Sb},$$

$$W_{nn'}^{GaAs} = W_{nn'}^{Ga} + W_{nn'}^{As},$$

$$W_{nn'}^{GaSb} = W_{nn'}^{Ga} + W_{nn'}^{Sb}. \qquad (7.5)$$

If the reference crystal is chosen to be InAs, then $W_{nn'}^{In} = W_{nn'}^{As} = 0$. In the atomic description, the linear response potential is

$$W_{nn'}^{(1)}(\mathbf{x}) = \theta^{Ga}(\mathbf{x}) W_{nn'}^{Ga} + \theta^{Sb}(\mathbf{x}) W_{nn'}^{Sb}, \qquad (7.6)$$

which can be rewritten (using $\theta^{In} + \theta^{Ga} = 0$ and $\theta^{As} + \theta^{Sb} = 0$) in several different ways, two of which are

$$W_{nn'}^{(1)} = \theta^{Ga} W_{nn'}^{GaSb} + (\theta^{In} + \theta^{Sb}) W_{nn'}^{InSb} \qquad (7.7a)$$

and

$$W_{nn'}^{(1)} = \theta^{Sb} W_{nn'}^{GaSb} + (\theta^{Ga} + \theta^{As}) W_{nn'}^{GaAs}. \qquad (7.7b)$$

Equation (7.7a) is useful for describing a (001) heterojunction with an InSb-like interface:

$$\cdots\text{-As-In-As-In-Sb-Ga-Sb-Ga-}\cdots, \qquad (7.8a)$$

while Eq. (7.7b) is useful for describing a heterojunction with a GaAs-like interface:

$$\cdots\text{-In-As-In-As-Ga-Sb-Ga-Sb-}\cdots. \qquad (7.8b)$$

In the first case one can identify the bulk functions

$$\theta^{GaSb} = \theta^{Ga}, \qquad \theta^{InAs} = \theta^{As},$$

$$\theta^{InSb} = \theta^{In} + \theta^{Sb}, \qquad \theta^{GaAs} = 0, \qquad (7.9a)$$

while in the second case

$$\theta^{GaSb} = \theta^{Sb}, \qquad \theta^{InAs} = \theta^{In},$$

$$\theta^{GaAs} = \theta^{Ga} + \theta^{As}, \qquad \theta^{InSb} = 0. \qquad (7.9b)$$

This type of transformation can be used for any term in the linear response. For the quadratic response, such a description is not appropriate, but all contributions (including the interface dipole potential) can be approximated as abrupt step functions, as discussed in Secs. V C 1 and V D.

B. Piecewise constant material parameters

The next step in simplifying the description of the material properties is to approximate an ideal heterostructure as

piecewise constant. As a specific example, the case of a (001) heterojunction between semiconductors with the zinc-blende structure is considered here.

For this case, a convenient slab-adapted[158,159] unit cell is defined by the basis vectors

$$\mathbf{a}_1 = \frac{a}{2}(1,-1,0), \quad \mathbf{a}_2 = \frac{a}{2}(1,1,0), \quad \mathbf{a}_3 = \frac{a}{2}(1,0,1).$$
(7.10)

Periodic boundary conditions are applied over the crystal volume $\Omega = \mathbf{L}_1 \cdot (\mathbf{L}_2 \times \mathbf{L}_3)$, where $\mathbf{L}_i = N_i \mathbf{a}_i$ and N_i is an integer (thus $\Omega = N\Omega_0$, where $N = N_1 N_2 N_3$).

The bulk properties of the heterojunction are to be represented in terms of the periodic step function

$$\Theta(\mathbf{x}) = \Theta(z) = \begin{cases} 1, & 0 < z < \frac{1}{2}L_z, \\ 0, & -\frac{1}{2}L_z < z < 0, \end{cases}$$
(7.11)

in which $L_z = \frac{1}{2}N_3 a$ is the period in the z direction, and $\Theta(\mathbf{x}) = \Theta(\mathbf{x}+\mathbf{L}_i)$. The interface properties are to be represented by the derivatives $\Theta_z(\mathbf{x}) = \partial\Theta/\partial z$ and $\Theta_{zz}(\mathbf{x}) = \partial^2\Theta/\partial z^2$, which are periodic arrays of δ and δ' functions.

The first example of an actual (001) heterojunction to be considered is a common-atom GaAs/AlAs junction. The coordinate origin is chosen to be an interface As atom, with unit-cell basis vectors $\boldsymbol{\tau}_a = \mathbf{0}$ and $\boldsymbol{\tau}_c = \frac{1}{4}a(1,1,1)$ for anions and cations, respectively. The discrete function $\theta_\mathbf{R}^\alpha$ for $\alpha = \mathrm{Al}$ is therefore

$$\theta_\mathbf{R}^{\mathrm{Al}} = \begin{cases} 1, & 0 \leq R_3 \leq \frac{1}{2}N_3 - 1, \\ 0, & -\frac{1}{2}N_3 \leq R_3 \leq -1, \end{cases}$$
(7.12)

where the dimensionless coordinate R_3 is an integer (see Sec. V D). From the Fourier transform of this function, one finds that for small k,

$$\theta^{\mathrm{Al}}(\mathbf{k}) \simeq \Theta(\mathbf{k}) - \frac{a^2}{96}\Theta_{zz}(\mathbf{k}).$$
(7.13)

Upon inserting this result into Eq. (7.3), one finds that for $|z| < \frac{1}{2}L_z$, the linear band offset at a GaAs/AlAs heterojunction can be represented as

$$W_{nn'}^{(1)}(\mathbf{x}) = (W_{nn'}^{\mathrm{AlAs}} - W_{nn'}^{\mathrm{GaAs}})[\Theta(z) - x]$$
$$- \frac{a^2}{96}(W_{nn'}^{\mathrm{AlAs}} - W_{nn'}^{\mathrm{GaAs}})\delta'(z),$$
(7.14)

which has been written in a general form suitable for a virtual reference crystal $(\mathrm{GaAs})_{1-x}(\mathrm{AlAs})_x$. Thus, if the band offset is treated as piecewise constant, there is an additional correction proportional to $\delta'(z)$ that has the effect of renormalizing the matrix Y in the Hamiltonian (6.3). This merely reflects the fact that $\theta^{\mathrm{Al}}(z)$ is a smooth (macroscopically averaged) step function, and the difference between a smooth step function and an abrupt step function is, to leading order, proportional to the macroscopic average of $\delta'(z)$.

The remaining material-dependent parameters in Eq. (6.3) can be treated using the same approach, but for these parameters the term proportional to Θ_{zz} yields a correction of order

$(\bar{k}a)^3\Delta\bar{V}$ or higher, which can be neglected. Thus for an ideal GaAs/AlAs junction, all of the material parameters in (6.3) except $W_{nn'}^{(1)}(\mathbf{x})$ can be replaced by abrupt step functions.

For the case of a no-common-atom InAs/GaSb junction, the situation is more complicated. The example considered here is the GaAs-like junction in Eq. (7.8b), where the origin is the midpoint of an As-Ga bond and $\boldsymbol{\tau}_c = -\boldsymbol{\tau}_a = \frac{1}{8}a(1,1,1)$. In this case, the small-k behavior of the atomic distribution functions is

$$\theta^{\mathrm{Ga}}(\mathbf{k}) \simeq \Theta(\mathbf{k}) + \frac{a}{8}\Theta_z(\mathbf{k}) - \frac{a^2}{384}\Theta_{zz}(\mathbf{k}),$$

$$\theta^{\mathrm{Sb}}(\mathbf{k}) \simeq \Theta(\mathbf{k}) - \frac{a}{8}\Theta_z(\mathbf{k}) - \frac{a^2}{384}\Theta_{zz}(\mathbf{k}).$$
(7.15)

Upon inserting these results into Eq. (7.7b), one finds that for $|z| < \frac{1}{2}L_z$, the linear band offset may be approximated by

$$W_{nn'}^{(1)}(\mathbf{x}) = (W_{nn'}^{\mathrm{GaSb}} - W_{nn'}^{\mathrm{InAs}})[\Theta(z) - x] + \frac{a}{8}(2W_{nn'}^{\mathrm{GaAs}} - W_{nn'}^{\mathrm{InAs}}$$
$$- W_{nn'}^{\mathrm{GaSb}})\delta(z) - \frac{a^2}{384}(W_{nn'}^{\mathrm{GaSb}} - W_{nn'}^{\mathrm{InAs}})\delta'(z),$$
(7.16a)

which is written in a form suitable for the virtual reference crystal $(\mathrm{InAs})_{1-x}(\mathrm{GaSb})_x$. Note that this has the form

$$W_{nn'}^{(1)}(\mathbf{x}) = (W_{nn'}^{(+)} - W_{nn'}^{(-)})[\Theta(z) - x] + \frac{a}{8}(2W_{nn'}^{(i)} - W_{nn'}^{(+)}$$
$$- W_{nn'}^{(-)})\delta(z) - \frac{a^2}{384}(W_{nn'}^{(+)} - W_{nn'}^{(-)})\delta'(z), \quad (7.16b)$$

in which (+) and (−) label the bulk materials to the right and left of the junction, (i) labels the "interface" material, and the reference crystal is $(+)_{1-x}(-)_x$. This result is valid for any no-common-atom (001) junction, regardless of the convention chosen for the $\boldsymbol{\tau}$ vectors. Thus for a no-common-atom junction, $W_{nn'}^{(1)}(\mathbf{x})$ contains terms that renormalize both the Z and Y matrices.

The $\delta(z)$ term in (7.16) has a very simple physical interpretation in which half a monolayer $(a/4)$ at the interface is occupied by GaAs instead of either InAs or GaSb. This concept has been used previously in the construction of envelope-function models for electrons[160] and phonons.[161]

Note that if one replaces Sb with As in Eq. (7.16a), the result does not reduce directly to an expression of the form (7.14). However, the difference is merely due to the different choice of coordinate origin in the two cases. To leading order, $\Theta(z - \frac{1}{8}a) \simeq \Theta(z) - \frac{1}{8}a\delta(z) + \frac{1}{128}a^2\delta'(z)$, $\delta(z - \frac{1}{8}a) \simeq \delta(z) - \frac{1}{8}a\delta'(z)$, and $\delta'(z - \frac{1}{8}a) \simeq \delta'(z)$. Thus, replacing Sb with As and shifting the origin by $\frac{1}{8}a(1,1,1)$ does indeed reduce (7.14), (7.15), and (7.16a).

The result (7.16b) can be applied to all other linear-response terms in the Hamiltonian (6.3), but since the $\Delta\pi(\mathbf{x})$ and $Z(\mathbf{x})$ terms are already of order $(\bar{k}a)\Delta\bar{V}$, the $\delta'(z)$ term in (7.16b) yields a negligible correction of order $(\bar{k}a)^3\Delta\bar{V}$. Thus

$$Z_{nn'}^{\lambda}(\mathbf{x}) = (Z_{nn'}^{\lambda(+)} - Z_{nn'}^{\lambda(-)})[\Theta(z) - x]$$
$$+ \frac{a}{8}(2Z_{nn'}^{\lambda(i)} - Z_{nn'}^{\lambda(+)} - Z_{nn'}^{\lambda(-)})\delta(z), \qquad (7.17)$$

and we see that Z generates a correction to Y, while π generates a correction to Γ. However, the position dependence of all remaining terms in (6.3) (including the quadratic response) can be represented as a simple step function:

$$Y_{nn'}^{\lambda\mu}(\mathbf{x}) = (Y_{nn'}^{\lambda\mu(+)} - Y_{nn'}^{\lambda\mu(-)})[\Theta(z) - x], \qquad (7.18)$$

since these terms are already of order $(\bar{k}a)^2\Delta\bar{V}$.

VIII. THEORY OF INVARIANTS

In this section the method of invariants[32–36] is used to construct the explicit form of the Hamiltonian (6.3) for Γ electrons in semiconductors with the zinc-blende structure.[162,163] The results are then compared with the interface Hamiltonian given by the method of invariants for a common-atom (001) junction, including all (linear and nonlinear) interface terms of order $(\bar{k}a)^2\Delta\bar{V}$ that are permitted by symmetry. Finally, the interface terms of order $(\bar{k}a)\Delta\bar{V}$ are considered for the case of Γ-X coupling in GaAs/AlAs. The results in this section make use of the time reversal and crystal symmetry properties of the self-energy that were derived in Ref. 124.

A. Generalized Leibler Hamiltonian

As demonstrated above, to within terms of order $(\bar{k}a)^2\Delta\bar{V}$, the position dependence of the interface parameters can be calculated entirely in terms of *bulk*-like matrix elements modulated by the atomic form factors $\theta^{\alpha}(\mathbf{x})$. Thus, the Hamiltonian for a zinc-blende heterostructure of arbitrary composition can be determined by constructing invariants transforming as Γ_1 under the symmetry operations of the T_d group. This amounts to treating $\theta^{\alpha}(\mathbf{x})$ as a "slowly varying" function that is an invariant of T_d.[48]

The relevant basis functions for the representations of T_d are given in Table I. The specific example to be considered here is that of the Γ_8 valence band, with Γ_{15} and Γ_6 derivable as special cases of the Γ_8 results. The extension of these results to multi-band Hamiltonians (e.g., $\Gamma_6 \oplus \Gamma_7 \oplus \Gamma_8$) can be handled using the methods of Refs. 33–35, but is not considered explicitly here.

The momentum matrix $\pi^{\lambda}(\mathbf{x})$ must transform as a vector (Γ_{15}) that is odd under time reversal, which leaves only one possibility:

$$\pi^{\lambda}(\mathbf{x}) = V_{\lambda}\xi(\mathbf{x}), \qquad (8.1)$$

in which $V_x \equiv \{J_x(J_y^2 - J_z^2)\}$ is a product of angular momentum matrices J_{ν} for a particle with spin $\frac{3}{2}$, and $\xi(\mathbf{x})$ is some linear

TABLE I. Basis functions for constructing invariants of T_d. Here **P** is a vector operator whose components need not commute, while $\boldsymbol{\sigma}, \mathbf{I}$, and $\mathbf{J}$ are angular momentum (pseu- dovector) operators corresponding to angular momentum $\frac{1}{2}$, 1, and $\frac{3}{2}$, respectively. Cyclic permutations of x, y, and z also yield acceptable basis functions.

Rep.	Basis functions
Γ_1	$1, P^2$
Γ_2	$J_x J_y J_z + J_z J_y J_x$
Γ_{12}	$P_x^2 - \frac{1}{3}P^2, I_x^2 - \frac{1}{3}I^2, J_x^2 - \frac{1}{3}J^2$
Γ_{15}	$P_x, \{P_x P_y\}, \{I_x I_y\}, \{J_x J_y\}, V_x \equiv \{J_x(J_y^2 - J_z^2)\}$
Γ_{25}	$i[P_x, P_y], \sigma_x, I_x, J_x, J_x^3$

combination of $\theta^l(\mathbf{x})$ functions. The corresponding results for Γ_{15} and Γ_6 states are obtained by replacing $J_{\nu} \to I_{\nu}$ (spin 1) and $J_{\nu} \to \sigma_{\nu}$ (spin $\frac{1}{2}$), respectively. This yields zero in both cases, hence $\pi^{\lambda}(\mathbf{x})$ occurs only for Γ_8 and is a relativistic effect.

For the effective-mass tensor $D^{\lambda\mu}(\mathbf{x})$ one obtains the well-known result[32]

$$-D^{\lambda\mu}(\mathbf{x}) = \frac{1}{2}\gamma_1(\mathbf{x})\delta_{\lambda\mu}1 - \gamma_2(\mathbf{x})\delta_{\lambda\mu}\left(J_{\lambda}^2 - \frac{1}{3}J^2\right)$$
$$- \gamma_3(\mathbf{x})(1 - \delta_{\lambda\mu})\{J_{\lambda}J_{\mu}\}$$
$$+ i\epsilon_{\lambda\mu\nu}[\kappa(\mathbf{x})J_{\nu} + q(\mathbf{x})J_{\nu}^3], \qquad (8.2)$$

in which the Luttinger parameters $\gamma_1(\mathbf{x}), \gamma_2(\mathbf{x}), \gamma_3(\mathbf{x}), \kappa(\mathbf{x})$, and $q(\mathbf{x})$ are all linear combinations of $\theta^l(\mathbf{x})$ functions, and $\epsilon_{\lambda\mu\nu}$ is the antisymmetric unit tensor. For Γ_{15} the parameter q is not independent (since $I_{\nu}^3 = I_{\nu}$), whereas for Γ_6 only γ_1 and κ are independent.

The Z matrix has the same symmetry as the coordinate operator (i.e., a vector that is even under time reversal), so it has the form[30]

$$Z^{\lambda}(\mathbf{x}) = |\epsilon_{\lambda\mu\nu}|\{J_{\mu}J_{\nu}\}\zeta(\mathbf{x}), \qquad (8.3)$$

where $\zeta(\mathbf{x})$ is a linear combination of $\theta^l(\mathbf{x})$ functions. This result has the same form for Γ_{15}, but vanishes for Γ_6. The coupling (8.3) generates a zone-center mixing of heavy and light holes, and was proposed independently in Refs. 3, 28, 59, and 60.

The Y interface matrix has the same symmetry as the symmetric part of D, so it can be written

$$Y^{\lambda\mu}(\mathbf{x}) = \frac{1}{2}\eta_1(\mathbf{x})\delta_{\lambda\mu}1 - \eta_2(\mathbf{x})\delta_{\lambda\mu}\left(J_{\lambda}^2 - \frac{1}{3}J^2\right)$$
$$- \eta_3(\mathbf{x})(1 - \delta_{\lambda\mu})\{J_{\lambda}J_{\mu}\}. \qquad (8.4)$$

This term has not been studied previously for Γ_8 states, although the corresponding term for Γ_6 electrons (a direct analog of the Darwin term from the Dirac equation) is well known.[47,48,53]

The Γ matrix has the form

$$\Gamma^{\lambda\mu}(\mathbf{x}) = |\epsilon_{\lambda\mu\nu}|V_{\nu}\chi(\mathbf{x}), \qquad (8.5)$$

in which $\chi(\mathbf{x})$ is a linear combination of $\theta^l(\mathbf{x})$ functions. The closely related coupling $\hat{\Gamma}^{\lambda\mu}$ arising from an external electric field was proposed recently in Ref. 18 as a possible mecha-

TABLE II. Basis function for D_{2d}. The components of $\boldsymbol{\sigma}, \mathbf{I}$, and $\mathbf{B} = ic(\mathbf{P} \times \mathbf{P})$ transform as those of $\mathbf{J}$.

Rep.	Basis functions
X_1	$1,$ $P_x^2 + P_y^2, P_z^2, J_x^2 + J_y^2, J_z^2, P_z\{J_xJ_y\}, P_zV_z,$ $P_xJ_x - P_yJ_y, P_xJ_x^3 - P_yJ_y^3, P_xV_x + P_yV_y,$ $P_x\{J_yJ_z\} + P_y\{J_zJ_x\}$
X_2	$P_x^2 - P_y^2, J_x^2 - J_y^2, P_zJ_z, P_zJ_z^3, P_xJ_x + P_yJ_y,$ $P_xJ_x^3 + P_yJ_y^3, P_xV_x - P_yV_y, P_x\{J_yJ_z\} - P_y\{J_zJ_x\}$
X_3	$P_z, \{P_xP_y\}, \{J_xJ_y\}, V_z, P_zJ_z^2, P_xJ_y - P_yJ_x,$ $P_xJ_y^3 - P_yJ_x^3, P_xV_y + P_yV_x, P_x\{J_xJ_z\} + P_y\{J_yJ_z\}$
X_4	$J_z, J_z^3, P_z(J_x^2 - J_y^2), P_xJ_y + P_yJ_x,$ $P_xJ_y^3 + P_yJ_x^3, P_xV_y - P_yV_x, P_x\{J_xJ_z\} - P_y\{J_yJ_z\}$
X_5	$(P_x, P_y), (\{P_yP_z\}, \{P_zP_x\}), (J_x, -J_y), (J_x^3, -J_y^3),$ $(V_x, V_y), (\{J_yJ_z\}, \{J_zJ_x\}), (P_zJ_y, -P_zJ_x),$ $(P_zJ_y^3, -P_zJ_x^3), (P_z\{J_zJ_x\}, P_z\{J_zJ_y\}), (P_zV_y, P_zV_x),$ $(P_yJ_z, -P_xJ_z), (P_xJ_z^2, P_yJ_z^2), (P_yJ_z^3, -P_xJ_z^3),$ $(P_yV_z, P_xV_z), (P_y\{J_xJ_y\}, P_x\{J_xJ_y\})$

TABLE IV. Terms in the (001) GaAs/AlAs interface Hamiltonian constructed from invariants of C_{2v}.

Term	D_{2d}	$\Gamma_1(T_d)$	Origin
$1\,\delta(z)$	X_1	no	
$(J_z^2 - \tfrac{1}{3}J^2)\delta(z)$	X_1	no	
$\{J_xJ_y\}\delta(z)$	X_3	yes	Z^λ
$1\,\delta'(z)$	X_3	yes	$Y^{\lambda\mu}$
$(J_z^2 - \tfrac{1}{3}J^2)\delta'(z)$	X_3	yes	$Y^{\lambda\mu}$
$\{J_xJ_y\}\delta'(z)$	X_1	no	
$V_z\{P_z\delta(z)\}$	X_1	no	
$(P_xV_x + P_yV_y)\delta(z)$	X_1	no	
$(P_xV_y + P_yV_x)\delta(z)$	X_3	yes	$\Gamma^{\lambda\mu}$
$(P_xJ_x - P_yJ_y)\delta(z)$	X_1	no	
$(P_xJ_y - P_yJ_x)\delta(z)$	X_3	yes	$\Phi^{\lambda\mu}$
$(P_xJ_x^3 - P_yJ_y^3)\delta(z)$	X_1	no	
$(P_xJ_y^3 - P_yJ_x^3)\delta(z)$	X_3	yes	$\Phi^{\lambda\mu}$

nism for the linear and circular photogalvanic effects. This term is relativistic in origin and does not occur for Γ_{15} or Γ_6.

The matrix Φ has the same symmetry as the antisymmetric part of D, except that it is odd rather than even under time reversal. It can therefore be written as

$$\Phi^{\lambda\mu}(\mathbf{x}) = \epsilon_{\lambda\mu\nu}[\kappa'(\mathbf{x})J_\nu + q'(\mathbf{x})J_\nu^3]. \quad (8.6)$$

For Γ_6 electrons (where q' is not independent) this is the analog of the Rashba spin-splitting effect for low-symmetry bulk semiconductors,[153] which occurs also in heterostructures of cubic semiconductors[48,154,164,165] due to the reduced symmetry at a surface or interface. The corresponding Hamiltonian for Γ_8 states was proposed in Ref. 58, and has received renewed attention as the valence-band Rashba coupling[27,31,155,156] in recent years.

The coefficients of the various terms involving φ in Eq. (6.3) have the same form as those already given. The symmetry restrictions on the Taylor series expansions for $n^{(1)}$ and $n^{(2)}$ were given in Ref. 124.

B. GaAs/AlAs (001) heterojunction

The significance of these results is now investigated by comparing them with all interface terms of order $(\bar{k}a)^2 \Delta\bar{V}$

TABLE III. Basis functions for C_{2v}. The components of $\boldsymbol{\sigma}, \mathbf{I}$, and $\mathbf{B} = ic(\mathbf{P} \times \mathbf{P})$ transform as those of $\mathbf{J}$.

Rep.	Basis functions
Δ_1	$1, P_z, \{P_xP_y\}, P_x^2 + P_y^2, P_z^2, \{J_xJ_y\}, J_x^2 + J_y^2, J_z^2, J_z^3,$ V_z
Δ_2	$P_x^2 - P_y^2, J_z, J_x^2 - J_y^2$
Δ_3	$P_x + P_y, \{P_z(P_x + P_y)\}, J_x - J_y, \{J_z(J_x + J_y)\},$ $J_x^3 - J_y^3, V_x + V_y$
Δ_4	$P_x - P_y, \{P_z(P_x - P_y)\}, J_x + J_y, \{J_z(J_x - J_y)\},$ $J_x^3 + J_y^3, V_x - V_y$

allowed by symmetry for a GaAs/AlAs (001) heterojunction, which has the point group C_{2v}. To better understand this comparison, it is helpful to begin by studying the symmetry properties of the linear and quadratic response in this system.

The starting point is the observation that although the point group of the heterojunction is C_{2v} (with the coordinate origin at an interface As atom), if the reference crystal is chosen to be the virtual crystal $Al_{0.5}Ga_{0.5}As$, then the perturbation due to the ionic pseudopotentials has a higher symmetry: it transforms according to the representation X_3 of the D_{2d} group. In other words, it transforms according to the identity representation Δ_1 of the C_{2v} group (none of whose operations change the z coordinate), but in addition it has odd parity with respect to those elements of D_{2d} that change the sign of z. This occurs because the atomic mole fraction $\theta_{\mathbf{R}}^\alpha$ itself transforms as $X_3(D_{2d})$, whereas the atomic perturbation $\Delta v_{\rm ion}^\alpha(\mathbf{x}, \mathbf{x}')$ has the site symmetry $\Gamma_1(T_d)$ or $X_1(D_{2d})$. Therefore, $\Delta V_{\rm ion}(\mathbf{x}, \mathbf{x}')$ transforms as $X_1 \otimes X_3 = X_3$ under the operations of D_{2d}.

To determine the behavior of the screened potential $\Delta V(\mathbf{x}, \mathbf{x}')$, note that since the vertex functions[124] $\Gamma^{(1)}$ and $\Gamma^{(2)}$ of the reference crystal transform as $\Gamma_1(T_d)$, the linear response $V^{(1)}(\mathbf{x}, \mathbf{x}')$ transforms as $X_1 \otimes X_3 = X_3$, while the quadratic response $V^{(2)}(\mathbf{x}, \mathbf{x}')$ transforms as $X_1 \otimes X_3 \otimes X_3 = X_1$. (In general, all odd-order terms in the response transform as X_3, while all even-order terms transform as X_1.) Thus the total response $\Delta V(\mathbf{x}, \mathbf{x}') = V^{(1)}(\mathbf{x}, \mathbf{x}') + V^{(2)}(\mathbf{x}, \mathbf{x}')$ transforms as neither X_3 nor X_1, but as $\Delta_1(C_{2v})$.

If the quadratic response is of the same order as the linear response, then one must include all possible invariants of C_{2v} that are of order $(\bar{k}a)^2 \Delta\bar{V}$ when constructing the interface Hamiltonian. Basis functions for D_{2d} and C_{2v} are given in Tables II and III, respectively; note that both X_1 and X_3 are compatible with Δ_1. All interface invariants of C_{2v} that are (1) Hermitian, (2) time-reversal invariant, and (3) of order $(\bar{k}a)^2 \Delta\bar{V}$ or less are listed in Table IV.

The second column in this table lists the symmetry of each term under the operations of D_{2d}. The third column

indicates whether the term can be constructed as an invariant of T_d, treating $\Theta(z)$ as an invariant of T_d. (As discussed above, the latter approach can be used for all linear-response interface terms.) Note that in the T_d symmetry analysis, $\delta(z) = d\Theta/dz$ transforms as z (Γ_{15}) and $\delta'(z)$ transforms as z^2 (Γ_1 or Γ_{12}), whereas in the D_{2d} symmetry analysis, $\delta(z)$ transforms as z^2 (X_1) and $\delta'(z)$ transforms as z (X_3).

In agreement with the general symmetry properties of $V^{(1)}$ and $V^{(2)}$ derived earlier, Table IV shows that all linear inter-face terms [derivable from $\Gamma_1(T_d)$] transform as $X_3(D_{2d})$. The remaining terms that transform as X_1 all originate in the quadratic or higher-order response. Thus, if one accounts for the smallness of the quadratic response (Sec. III B), seven out of the thirteen possible interface invariants for (001) GaAs/AlAs heterojunctions can be omitted because they are actually of order $(\bar{k}a)^3 \Delta \bar{V}$ or higher. This represents a major simplification over the general case in which all response terms are of the same order.

For a no-common-atom heterojunction, the results of Sec. VIII A remain valid (as they depend only on the bulk symmetry), but the linear response now has only the symmetry $\Delta_1(C_{2v})$. Thus the terms labeled X_1 in Table IV now arise (in general) even in linear-response theory, due to the renormalization effects described above in Eqs. (7.16) and (7.17). For example, the $\delta(z)$ term in Eq. (7.16) generates the term in the first row of Table IV. However, the term in the second row of Table IV is still zero (within linear response), because $W_{nn'}^{(1)}(\mathbf{x})$ in Eq. (7.16) couples only states of the same symmetry, with $W_{nn'}^{(1)}(\mathbf{x}) = \delta_{nn'} W_{nn}^{(1)}(\mathbf{x})$ and $W_{nn}^{(1)}(\mathbf{x}) = W_{n'n'}^{(1)}(\mathbf{x})$ whenever $E_n = E_{n'}$.

C. Γ-X coupling

If one extends the above analysis to the case of intervalley Γ-X coupling at a (001) GaAs/AlAs junction, it is immediately apparent (because the Z matrix has the symmetry of a coordinate matrix) that the linear response produces $\delta(z)$ coupling between the bands

$$\Gamma_1 - X_{3z}, \quad X_{1z} - X_{3z}, \quad X_{3x} - X_{3y}, \tag{8.7}$$

but no $\delta(z)$ coupling between the bands

$$\Gamma_1 - X_{1z}, \quad X_{1x} - X_{1y}. \tag{8.8}$$

Coupling of the latter type only occurs in the quadratic response, or in the linear response from terms proportional to $\delta'(z)$ or $i\{P_z \delta(z)\}$. These conclusions agree with those of Ref. 61, which were derived from a model potential constructed from a linear superposition of atomic-like pseudopotentials.

Note, however, that these results hold only for an ideal heterojunction. If interdiffusion of Ga and Al atoms breaks the X_3 symmetry of the linear response, then there will be $\delta(z)$ coupling between the bands (8.8) due to terms similar to those derived in Eq. (7.16).

IX. DISCUSSION

Since a summary of the main results of this paper has already been given in Sec. I C, this section will be limited to a discussion of the differences between the present theory and previous envelope-function theories that have appeared in the literature. The principal differences are that (1) the present Hamiltonian is constructed from atomic pseudopotentials rather than the periodic potential of a bulk crystal; (2) the present theory is self-consistent, accounting fully for electron-electron Coulomb interactions; and (3) the present approach uses linear and quadratic response theory to simplify the functional form of the heterostructure Hamiltonian.

Starting with point (1), previous envelope-function theories[47–82] were based on model potentials of the form (5.9), which is a linear combination of periodic bulk potentials multiplied by steplike functions $\theta^l(\mathbf{x})$. The form of $\theta^l(x)$ near an interface was either specified as part of the model (e.g., abrupt step function) or treated as unknown—and in principle unknowable, at least within the model potential approach.

Although the present theory led to a similar expression (7.1b), this was a *derived* result based on linear response theory, in which the functions $\theta^l(\mathbf{x})$ have a known form determined by the distribution $\theta_{\mathbf{R}}^{\alpha}$ of atoms in the heterostructure. The potential (7.1b) includes contributions from the "interface" materials (e.g., GaAs and InSb in an InAs/GaSb heterostructure), which are often omitted in the model-potential approach. Furthermore, for the model potential, the coefficients of the δ terms in the Hamiltonian depend on (i) the values of the *bulk* potentials *at* the reciprocal lattice vectors $\mathbf{G}$; and (ii) whether the functions $\theta^l(\mathbf{x})$ are smooth (zero outside the Brillouin zone) or sharp, with some δ terms vanishing in the smooth case.[65–69] In the present theory these coefficients are determined not by $\theta^l(\mathbf{x})$, but instead by the properties of the linear *atomic* pseudopotentials in a *finite neighborhood* of each $\mathbf{G}$.

For point (2), the present theory includes long-range multipole Coulomb potentials (because the self-energy is nonanalytic at $\mathbf{q} = \mathbf{G}$), which were not considered in previous studies. These terms have no qualitative effect in two-dimensional systems (other than to renormalize the band offsets), but they are present in quantum wires and dots, even for isovalent heterostructures.

Finally, for point (3), if the functions $\theta^l(\mathbf{x})$ are treated as unknown in the model potential approach, one has no way of knowing how large the various interface terms are, and all terms permitted by symmetry should in principle be included.[65–69] However, since the linear response is dominant (for typical heterostructures) in a first-principles theory, one can eliminate all nonlinear interface band-mixing terms, thereby simplifying the Hamiltonian considerably.

Aside from these differences, the qualitative form of the Hamiltonian derived here is very similar to those derived by Leibler[47,48] for slowly graded heterostructures and Takhtamirov and Volkov[65–69] for abrupt heterostructures. The basic structure of these Hamiltonians is similar because the same type of perturbation theory[31,34,127] was used in their derivation.

Other differences arise in theories using different approximation techniques. The well-known theory of Burt[50] is based on the Luttinger-Kohn representation[47,127] with an energy-dependent approximation similar to Löwdin perturbation

theory[166] used to eliminate the coupling to remote bands. Estimation of the error involved in this approximation shows that Burt's general theory[50] includes all terms of order $(\bar{k}a)\Delta\bar{V}$, but retains some terms of order $(\bar{k}a)^2\Delta\bar{V}$ while omitting others. In particular, his results give an essentially correct (within the model potential approach) description of the δ-function term Z and the Rashba term Φ, but omit the contributions from remote-band coupling to the terms Y and Γ.

In Ref. 58, I used Burt's theory, but introduced an additional approximation (also adopted in Refs. 52 and 62) whereby the energy eigenvalue in the denominator of the energy-dependent effective mass was replaced with the position-dependent energy of the bulk valence band maximum. This has the effect of replacing the Rashba coefficient in Eq. (6.10) with $iD_{nn'}^{[\lambda\mu]}(\mathbf{x})$. In other words, the Rashba parameters κ' and q' of Eq. (8.6) were replaced by the Luttinger parameters κ and q of Eq. (8.2). This simplifies applications of the theory because κ and q are known from bulk magnetoabsorption measurements,[167,168] and κ can also be estimated from γ_1, γ_2, and γ_3.[58,167] However, since the true Rashba parameters κ' and q' cannot be determined from bulk measurements, this simplification will produce values that have the correct order of magnitude but are quantitatively incorrect.

A similar approximation was used in Refs. 3, 70, and 71 to estimate the δ-function mixing parameter for light and heavy holes [see Table IV and Eq. (8.3)] from the valence-band offset. The original proposal was based on the H_{BF} model,[3] in which the zone-center Γ_{15} Hamiltonian is written as a sum of operators $B = \frac{1}{2} + \{I_x I_y\}$ and $F = \frac{1}{2} - \{I_x I_y\}$ multiplied by position-dependent bulk valence-band energies E_v. For a (001) GaAs/AlAs heterojunction this yields the Γ_{15} interface Hamiltonian

$$\frac{a}{4}\{I_x I_y\}(E_v^{\text{GaAs}} - E_v^{\text{AlAs}})\delta(z), \tag{9.1}$$

whereas for an InAs/GaSb junction with a GaAs-like interface one obtains

$$\frac{a}{4}(1 - 2\{I_x I_y\})\Big[E_v^{\text{GaAs}} - \tfrac{1}{2}(E_v^{\text{InAs}} + E_v^{\text{GaSb}})\Big]\delta(z). \tag{9.2}$$

The corresponding results for Γ_8 are given by the substitution $\{I_x I_y\} \rightarrow \frac{1}{3}\{J_x J_y\}$.[32] Now the diagonal term [proportional to $1\delta(z)$] in Eq. (9.2) is identical (within linear response) to the $\delta(z)$ term derived above in Eq. (7.16). However, the valence-band *mixing* term derived in Eqs. (6.7) and (8.3) is not related to the valence-band offset. Therefore, although the model of Refs. 3, 70, and 71 yields a coupling of the correct symmetry and order of magnitude, its numerical value is not reliable.[85] The same conclusion was reached in Ref. 86.

In Ref. 60, I proposed a smooth model potential[169] for abrupt heterostructures that yields a valence-band mixing determined chiefly by the change in Bloch functions at an interface. This result is of course not valid for general model potentials[60,61] (despite claims to the contrary[56,57]), and it is not supported by the present work. However, in Ref. 68 it

was argued that the valence-band mixing generated by Bloch-function differences is negligible in comparison to that generated by a discontinuous model potential. Although the latter does not appear in the present first-principles theory, its analog is the Z term shown in Eq. (5.6). The former contribution is given by the term proportional to $W_{in'}^{(1)}(\mathbf{q})$ in Eqs. (6.7) and (C6). Inspection of these terms shows that both are of order $(\bar{k}a)\Delta\bar{V}$. Therefore, neither the model of Ref. 60 [which omits the interface contributions from Eq. (5.5)] nor the model of Refs. 66–68 (which omits the contribution from Bloch-function differences) gives a correct value for the valence-band mixing coefficient.

In Refs. 12–14, 64, and 69 it was proposed that intervalley Γ_1-X_{1z} and X_{1x}-X_{1y} coupling should be proportional to $\delta(z)$ and of the same order of magnitude as Γ_1-X_{3z}, X_{1z}-X_{3z}, and X_{3x}-X_{3y} coupling. However, as shown above in Sec. VIII C, the former types of coupling should generally be substantially weaker than the latter for ideal interfaces. The same conclusion was reached in Ref. 130, where it was shown that despite the very small magnitude of the *direct* Γ_1-X_{1z} coupling, there is a net *effective* Γ_1-X_{1z} coupling (as revealed by the anticrossing of superlattice subbands in Fig. 1 of Ref. 130) generated by Γ_1-X_{3z}, X_{1z}-X_{3z}, and $\mathbf{k}\cdot\mathbf{p}$ couplings that is only a factor of two or three smaller than the net $\Gamma_1 - X_{3z}$ coupling.

However, in Refs. 12–14 it was found that X_x-X_y mixing experiments could not adequately be explained without including δ-function X_{1x}-X_{1y} coupling of comparable magnitude to X_{3x}-X_{3y} coupling. Nevertheless, since Refs. 12–14 did not test the effect of $\delta'(z)X_{1x}$-X_{1y} coupling, it is not yet clear that their experiments require such a large $\delta(z)$ coupling. If $\delta'(z)$ or $i\{P_z\delta(z)\}$ coupling is unable to explain their results, then their data may indicate the presence of nonideal interfaces, as discussed in Sec. VIII C. Another possible explanation would be spin-orbit coupling, which generates a direct $\delta(z)$ coupling between X_{6x} and X_{6y} even for ideal interfaces.

Two important physical effects not considered in this paper are the contributions from alloying (beyond the virtual crystal approximation) and from strain (in the bulk or at an interface). The former effect is in principle already encompassed by the present formalism (for lattice-matched systems in which linear response is dominant), although for practical applications one would need to perform a detailed analysis of the significance of different intervalley mixing effects.[130] The contribution from strain would require a nontrivial extension of this theory, but since an *ab initio* linear-response approach to strain has already been successfully applied to $In_{0.53}Ga_{0.47}As/InP$,[114,116] Si/Ge,[119,120] and $InAs/GaSb$[122] within LDA, the general perturbation scheme used in this paper should work for lattice-mismatched systems also.

APPENDIX A: IMAGINARY PART OF THE SELF-ENERGY

Luttinger[170] has shown for homogeneous systems that $\Sigma^{(i)}$ vanishes at the Fermi surface to all orders in perturbation

theory. The generalization of his result to the case of inhomogeneous systems is $\Sigma^{(i)}(\omega \pm i0^+) = \mp J(\omega)$, where ω is real and[171,172]

$$J(\omega) = \begin{cases} C_+(\omega - \mu_+)^2, & \omega \geqslant \mu_+, \\ 0, & \mu_- \leqslant \omega \leqslant \mu_+, \\ C_-(\omega - \mu_-)^2, & \omega \leqslant \mu_-, \end{cases} \qquad \text{(A1)}$$

for ω near $\mu_\pm \equiv \mu \pm \frac{1}{2} E_g$. Here μ is the chemical potential (in the limit of zero temperature), E_g is the energy gap, and $C_\pm \geqslant 0$ is independent of ω. The magnitude of $C_\pm$ may be estimated from the calculations of Quinn and Ferrell[173] for a degenerate homogeneous electron gas. Their results may be expressed as $\langle C_\pm \rangle = 1/E_i$, where[173,174]

$$E_i = \frac{256}{\sqrt{3}\pi^2} \frac{\epsilon_F^2}{\omega_p}, \qquad \text{(A2)}$$

in which ϵ_F is the Fermi energy and ω_p is the plasma frequency. If one treats semiconductors with the diamond or zinc-blende structure as a homogeneous gas with a density of eight valence electrons per primitive unit cell, one obtains values of $E_i \simeq 126$ eV for GaAs and $E_i \simeq 140$ eV for Si. This is very large in comparison to the energy gap and suggests that $\Sigma^{(i)}$ will have a negligible influence on the band structure in typical semiconductor heterostructures (although it has an important qualitative effect in producing a finite quasiparticle lifetime).

Since a real semiconductor is not a degenerate homogeneous electron gas, one may question whether this estimate is reliable. However, the calculations of Fleszar and Hanke[175] for Si yield $E_i \gtrsim 100$ eV (estimated from Fig. 3 of Ref. 175), with some asymmetry between electrons and holes. Thus it is reasonable in heterostructures to assume that $E_i \gg \bar{E}_g$. Since $(\omega - \mu_\pm)^2$ is of order $(\Delta \bar{V})^2$, this means that $\Sigma^{(i)}$ is negligible under the present perturbation scheme. Thus it will, for the most part, not be considered explicitly in the envelope-function equations derived here. However, the leading contribution from $\Sigma^{(i)}$ is noted in Eq. (5.3) so that it may be included if desired.

APPENDIX B: CUTOFF FUNCTION

This appendix considers possible definitions of the Brillouin zone cutoff function $B(\mathbf{k})$ introduced in Eq. (5.7). The most obvious choice [from Eq. (3.19a)] would be

$$B(\mathbf{k}) = \begin{cases} 1, & \mathbf{k} \in \Omega_0^*, \\ 0, & \mathbf{k} \notin \Omega_0^*, \end{cases} \qquad \text{(B1)}$$

which was used previously in Refs. 49 and 50. However, the sharp cutoff at the boundary of Ω_0^* has the undesirable effect of producing Gibbs oscillations in $\mathbf{x}$ space. Other possibilities that eliminate this problem are $B(\mathbf{k}) = [1 + \beta(ka)^2] \times \exp[-\beta(ka)^2]$ and $B(\mathbf{k}) = \exp[-\beta(ka)^4]$, in which β is some number of order 1. These choices of $B(\mathbf{k})$ are smooth, spherically symmetric, and introduce a negligible error of order $(\bar{k}a)^4 \Delta \bar{V}$.

For some applications one need not introduce any cutoff function at all, as an alternative power-series approximation for $\theta^\alpha(\mathbf{k})$ is often more convenient. This approximation is discussed in Sec. VII B.

APPENDIX C: RENORMALIZED EFFECTIVE-MASS PARAMETERS

This appendix presents some details of the perturbation theory used in Sec. VI. For states $m, m' \in A$, the effective Hamiltonian $\bar{H}$ is given by[31,34]

$$\begin{aligned} \bar{H}_{mm'} = {} & H_{mm'} + \frac{1}{2}\sum_i^B H'_{mi}H'_{im'}\left(\frac{1}{E_{mi}} + \frac{1}{E_{m'i}}\right) \\ & + \frac{1}{2}\sum_i^B \sum_j^B H'_{mi}H'_{ij}H'_{jm'}\left(\frac{1}{E_{mi}E_{mj}} + \frac{1}{E_{m'i}E_{m'j}}\right) \\ & - \frac{1}{2}\sum_i^B \sum_{m''}^A \left(\frac{H'_{mm''}H'_{m''i}H'_{im'}}{E_{mi}E_{m''i}} + \frac{H'_{mi}H'_{im''}H'_{m''m'}}{E_{m'i}E_{m''i}}\right) \\ & + O((H')^4), \end{aligned} \qquad \text{(C1)}$$

in which $E_{mi} = E_m - E_i$. This expression can be used to derive most of the terms in the effective-mass Hamiltonian (6.2). The coefficients of the quadratic, cubic, and quartic dispersion terms of the reference crystal are given by

$$D_{nn'}^{\lambda\mu} = \tilde{D}_{nn'}^{\lambda\mu} + \frac{1}{2}\sum_i^B \pi_{ni}^\lambda \pi_{in'}^\mu \left(\frac{1}{E_{ni}} + \frac{1}{E_{n'i}}\right), \qquad \text{(C2)}$$

$$\begin{aligned} C_{nn'}^{\lambda\mu\kappa} = {} & \tilde{C}_{nn'}^{\lambda\mu\kappa} + \frac{1}{2}\sum_i^B (\pi_{ni}^\lambda \tilde{D}_{in'}^{\mu\kappa} + \tilde{D}_{ni}^{\lambda\mu}\pi_{in'}^\kappa)\left(\frac{1}{E_{ni}} + \frac{1}{E_{n'i}}\right) + \frac{1}{2}\sum_i^B \sum_j^B \pi_{ni}^\lambda \pi_{ij}^\mu \pi_{jn'}^\kappa \left(\frac{1}{E_{ni}E_{nj}} + \frac{1}{E_{n'i}E_{n'j}}\right) \\ & - \frac{1}{2}\sum_i^B \sum_{n''}^A \left(\frac{\pi_{nn''}^\lambda \pi_{n''i}^\mu \pi_{in'}^\kappa}{E_{ni}E_{n''i}} + \frac{\pi_{ni}^\lambda \pi_{in''}^\mu \pi_{n''n'}^\kappa}{E_{n'i}E_{n''i}}\right), \end{aligned} \qquad \text{(C3)}$$

$$Q_{nn'}^{\lambda\mu\kappa\nu} = \widetilde{Q}_{nn'}^{\lambda\mu\kappa\nu} + \frac{1}{2}\sum_i^B (\widetilde{D}_{ni}^{\lambda\mu}\widetilde{D}_{in'}^{\kappa\nu} + \pi_{ni}^{\lambda}\widetilde{C}_{in'}^{\mu\kappa\nu} + \widetilde{C}_{ni}^{\lambda\mu\kappa}\pi_{in'}^{\nu})\left(\frac{1}{E_{ni}} + \frac{1}{E_{n'i}}\right)$$

$$+ \frac{1}{2}\sum_i^B \sum_j^B (\widetilde{D}_{ni}^{\lambda\mu}\pi_{ij}^{\kappa}\pi_{jn'}^{\nu} + \pi_{ni}^{\lambda}\widetilde{D}_{ij}^{\mu\kappa}\pi_{jn'}^{\nu} + \pi_{ni}^{\lambda}\pi_{ij}^{\mu}\widetilde{D}_{jn'}^{\kappa\nu})\left(\frac{1}{E_{ni}E_{nj}} + \frac{1}{E_{n'i}E_{n'j}}\right)$$

$$- \frac{1}{2}\sum_i^B \sum_{n''}^A \left(\frac{\widetilde{D}_{nn''}^{\lambda\mu}\pi_{n''i}^{\kappa}\pi_{in'}^{\nu} + \pi_{nn''}^{\lambda}\widetilde{D}_{n''i}^{\mu\kappa}\pi_{in'}^{\nu} + \pi_{nn''}^{\lambda}\pi_{n''i}^{\mu}\widetilde{D}_{in'}^{\kappa\nu}}{E_{ni}E_{n''i}} + \frac{\widetilde{D}_{ni}^{\lambda\mu}\pi_{in''}^{\kappa}\pi_{n''n'}^{\nu} + \pi_{ni}^{\lambda}\widetilde{D}_{in''}^{\mu\kappa}\pi_{n''n'}^{\nu} + \pi_{ni}^{\lambda}\pi_{in''}^{\mu}\widetilde{D}_{n''n'}^{\kappa\nu}}{E_{n'i}E_{n''i}}\right) + O(\pi^4).$$

$$\text{(C4)}$$

In the last expression, the symbol $O(\pi^4)$ denotes terms of the fourth order in the kinetic momentum π. These cannot be obtained from Eq. (C1), but they can be derived easily from Eq. (B15e) on p. 205 of Ref. 31.

The term $W_{nn'}^{(2)}(\mathbf{q})$ is a renormalized second-order contribution to the band offsets, given by

$$W_{nn'}^{(2)}(\mathbf{x}) = \widetilde{W}_{nn'}^{(2)}(\mathbf{x}) + \frac{1}{2}\sum_i^B \left(\frac{1}{E_{ni}} + \frac{1}{E_{n'i}}\right)[W_{ni}^{(1)}(\mathbf{x}) + \Lambda_{ni}\varphi^{(1)}(\mathbf{x})]$$

$$\times [W_{in'}^{(1)}(\mathbf{x}) + \Lambda_{in'}\varphi^{(1)}(\mathbf{x})], \quad \text{(C5)}$$

where $W_{nn'}^{(1)}(\mathbf{q})$ and $\widetilde{W}_{nn'}^{(2)}(\mathbf{q})$ were defined in Eqs. (5.11) and (5.15). The only contribution from $\varphi^{(1)}$ to be included here is the linear interface dipole term (5.14) for heterovalent systems; but note that this contribution is zero in the *GW* approximation[94,95] and in density-functional theory[89,90] (since $\Lambda_{nn'} = \delta_{nn'}$ in these cases).

The remaining terms in (6.2) are renormalized versions of the functions $\widetilde{J}_{nn'}^{\lambda}(\mathbf{q})$, $\widetilde{M}_{nn'}^{\lambda\mu}(\mathbf{q})$, and $\widetilde{R}_{nn'}^{\lambda\mu}(\mathbf{q})$ defined in Eqs. (5.11) and (5.12):

$$J_{nn'}^{\lambda}(\mathbf{q}) = \widetilde{J}_{nn'}^{\lambda}(\mathbf{q}) + \frac{1}{2}\sum_i^B \pi_{ni}^{\lambda}W_{in'}^{(1)}(\mathbf{q})\left(\frac{1}{E_{ni}} + \frac{1}{E_{n'i}}\right), \quad \text{(C6)}$$

$$M_{nn'}^{\lambda\mu}(\mathbf{q}) = \widetilde{M}_{nn'}^{\lambda\mu}(\mathbf{q}) + \frac{1}{2}\sum_i^B \widetilde{D}_{ni}^{\lambda\mu}W_{in'}^{(1)}(\mathbf{q})\left(\frac{1}{E_{ni}} + \frac{1}{E_{n'i}}\right)$$

$$+ \frac{1}{2}\sum_i^B \pi_{ni}^{\lambda}\widetilde{J}_{in'}^{\mu}(\mathbf{q})\left(\frac{1}{E_{ni}} + \frac{1}{E_{n'i}}\right)$$

$$+ \frac{1}{2}\sum_i^B \sum_j^B \pi_{ni}^{\lambda}\pi_{ij}^{\mu}W_{jn'}^{(1)}(\mathbf{q})\left(\frac{1}{E_{ni}E_{nj}} + \frac{1}{E_{n'i}E_{n'j}}\right)$$

$$- \frac{1}{2}\sum_i^B \sum_{n''}^A \left(\frac{\pi_{nn''}^{\lambda}\pi_{n''i}^{\mu}W_{in'}^{(1)}(\mathbf{q})}{E_{ni}E_{n''i}} + \frac{\pi_{ni}^{\lambda}\pi_{in''}^{\mu}W_{n''n'}^{(1)}(\mathbf{q})}{E_{n'i}E_{n''i}}\right),$$

$$\text{(C7)}$$

$$R_{nn'}^{\lambda\mu}(\mathbf{q}) = \widetilde{R}_{nn'}^{\lambda\mu}(\mathbf{q}) + \frac{1}{2}\sum_i^B \left(\pi_{ni}^{\lambda}[\widetilde{J}_{n'i}^{\mu}(-\mathbf{q})]^* + \widetilde{J}_{ni}^{\lambda}(\mathbf{q})\pi_{in'}^{\mu}\right)$$

$$\times \left(\frac{1}{E_{ni}} + \frac{1}{E_{n'i}}\right)$$

$$+ \frac{1}{2}\sum_i^B \sum_j^B \pi_{ni}^{\lambda}W_{ij}^{(1)}(\mathbf{q})\pi_{jn'}^{\mu}\left(\frac{1}{E_{ni}E_{nj}} + \frac{1}{E_{n'i}E_{n'j}}\right)$$

$$- \frac{1}{2}\sum_i^B \sum_{n''}^A \left(\frac{\pi_{nn''}^{\lambda}W_{n''i}^{(1)}(\mathbf{q})\pi_{in'}^{\mu}}{E_{ni}E_{n''i}} + \frac{\pi_{ni}^{\lambda}W_{in''}^{(1)}(\mathbf{q})\pi_{n''n'}^{\mu}}{E_{n'i}E_{n''i}}\right).$$

$$\text{(C8)}$$

Note that $M_{nn'}^{\lambda\mu}(\mathbf{q}) \neq M_{nn'}^{\mu\lambda}(\mathbf{q})$, in contrast to Eq. (5.13).

*Electronic address: phbaf@ust.hk

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

[2] G. Bastard, J. A. Brum, and R. Ferreira, in *Solid State Physics*, edited by H. Ehrenreich and D. Turnbull (Academic, Boston, 1991), Vol. 44, pp. 229–415.

[3] O. Krebs and P. Voisin, Phys. Rev. Lett. **77**, 1829 (1996).

[4] O. Krebs, D. Rondi, J. L. Gentner, L. Goldstein, and P. Voisin, Phys. Rev. Lett. **80**, 5770 (1998).

[5] A. V. Platonov, V. P. Kochereshko, E. L. Ivchenko, G. V. Mikhailov, D. R. Yakovlev, M. Keim, W. Ossau, A. Waag, and G. Landwehr, Phys. Rev. Lett. **83**, 3546 (1999).

[6] É. E. Takhtamirov and V. A. Volkov, JETP Lett. **71**, 422 (2000).

[7] T. Reker, H. Im, L. E. Bremme, H. Choi, Y. Chung, P. C. Klipstein, and H. Shtrikman, Phys. Rev. Lett. **88**, 056403 (2002).

[8] U. Rössler and J. Kainz, Solid State Commun. **121**, 313 (2002).

[9] N. J. Pulsford, R. J. Nicholas, P. Dawson, K. J. Moore, G. Duggan, and C. T. B. Foxon, Phys. Rev. Lett. **63**, 2284 (1989).

[10] I. L. Aleĭner and E. L. Ivchenko, Semiconductors **27**, 330 (1993).

[11] Y. Fu, M. Willander, E. L. Ivchenko, and A. A. Kiselev, Phys. Rev. B **47**, 13498 (1993).

[12] H. Im, P. C. Klipstein, R. Grey, and G. Hill, Phys. Rev. Lett. **83**, 3693 (1999).

[13] H. Im, L. E. Bremme, Y. C. Chung, P. C. Klipstein, R. Grey, and G. Hill, Physica E (Amsterdam) **6**, 214 (2000).

[14] P. C. Klipstein, Phys. Status Solidi B **223**, 87 (2001).

[15] G. F. Glinskii, V. A. Lakisov, A. G. Dolmatov, and K. O. Kravchenko, Nanotechnology **11**, 233 (2000).

[16] S. D. Ganichev, E. L. Ivchenko, S. N. Danilov, J. Eroms, W. Wegscheider, D. Weiss, and W. Prettl, Phys. Rev. Lett. **86**, 4358 (2001).

[17] S. D. Ganichev, E. L. Ivchenko, V. V. Bel'kov, S. A. Tarasenko, M. Sollinger, D. Weiss, W. Wegscheider, and W. Prettl, Nature (London) **417**, 153 (2002).

[18] S. D. Ganichev, U. Rössler, W. Prettl, E. L. Ivchenko, V. V. Bel'kov, R. Neumann, K. Brunner, and G. Abstreiter, Phys. Rev. B **66**, 075328 (2002).

[19] L. E. Golub, Phys. Rev. B **67**, 235320 (2003).

[20] S. D. Ganichev, P. Schneider, V. V. Bel'kov, E. L. Ivchenko, S. A. Tarasenko, W. Wegscheider, D. Weiss, D. Schuh, B. N. Murdin, P. J. Phillips, C. R. Pidgeon, D. G. Clarke, M. Merrick, P. Murzyn, E. V. Beregulin, and W. Prettl, Phys. Rev. B **68**, 081302(R) (2003).

[21] V. V. Bel'kov, S. D. Ganichev, P. Schneider, C. Back, M. Oestreich, J. Rudolph, D. Hägele, L. E. Golub, W. Wegscheider, and W. Prettl, Solid State Commun. **128**, 283 (2003).

[22] S. D. Ganichev and W. Prettl, J. Phys.: Condens. Matter **15**, R935 (2003).

[23] S. D. Ganichev, V. V. Bel'kov, L. E. Golub, E. L. Ivchenko, P. Schneider, S. Giglberger, J. Eroms, J. De Boeck, G. Borghs, W. Wegscheider, D. Weiss, and W. Prettl, Phys. Rev. Lett. **92**, 256601 (2004).

[24] J. Sinova, D. Culcer, Q. Niu, N. A. Sinitsyn, T. Jungwirth, and A. H. MacDonald, Phys. Rev. Lett. **92**, 126603 (2004).

[25] S.-Q. Shen, M. Ma, X. C. Xie, and F. C. Zhang, Phys. Rev. Lett. **92**, 256603 (2004).

[26] S. Zhang and Z. Yang, Phys. Rev. Lett. **94**, 066602 (2005).

[27] B. A. Bernevig and S.-C. Zhang, Phys. Rev. Lett. **95**, 016801 (2005).

[28] I. L. Aleĭner and E. L. Ivchenko, JETP Lett. **55**, 692 (1992).

[29] E. L. Ivchenko, A. Yu. Kaminskiĭ, and I. L. Aleĭner, JETP **77**, 609 (1993).

[30] E. L. Ivchenko, A. Yu. Kaminski, and U. Rössler, Phys. Rev. B **54**, 5852 (1996).

[31] R. Winkler, *Spin-Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems*, Springer Tracts in Modern Physics Vol. 191 (Springer, Berlin, 2003).

[32] J. M. Luttinger, Phys. Rev. **102**, 1030 (1956).

[33] K. Suzuki and J. C. Hensel, Phys. Rev. B **9**, 4184 (1974).

[34] G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors* (Wiley, New York, 1974).

[35] H.-R. Trebin, U. Rössler, and R. Ranvaud, Phys. Rev. B **20**, 686 (1979).

[36] E. L. Ivchenko and G. E. Pikus, *Superlattices and Other Heterostructures: Symmetry and Optical Phenomena*, 2nd ed. (Springer, Berlin, 1997).

[37] G. F. Glinskii and K. O. Kravchenko, cond-mat/9808174 (unpublished).

[38] G. F. Glinskiĭ and K. O. Kravchenko, Phys. Solid State **40**, 803 (1998).

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

[40] S. de-Leon, B. Laikhtman, and L. D. Shvartsman, J. Phys.: Condens. Matter **10**, 8715 (1998).

[41] M. V. Kisin, Semiconductors **27**, 274 (1993).

[42] M. V. Kisin, Semiconductors **28**, 1143 (1994).

[43] M. V. Kisin, B. L. Gelmont, and S. Luryi, Phys. Rev. B **58**, 4605 (1998).

[44] I. V. Tokatly, A. G. Tsibizov, and A. A. Gorbatsevich, Phys. Rev. B **65**, 165328 (2002).

[45] A. V. Rodina, A. Yu. Alekseev, A. L. Efros, M. Rosen, and B. K. Meyer, Phys. Rev. B **65**, 125302 (2002).

[46] A. V. Rodina, Al. L. Efros, and A. Yu. Alekseev, Phys. Rev. B **67**, 155312 (2003).

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

[48] L. Leibler, Phys. Rev. B **16**, 863 (1977).

[49] M. G. Burt, Semicond. Sci. Technol. **3**, 739 (1988).

[50] M. G. Burt, J. Phys.: Condens. Matter **4**, 6651 (1992).

[51] M. G. Burt, Phys. Rev. B **50**, 7518 (1994).

[52] M. G. Burt, J. Phys.: Condens. Matter **11**, R53 (1999).

[53] K. Young, Phys. Rev. B **39**, 13434 (1989).

[54] J. P. Cuypers and W. van Haeringen, Physica B **168**, 58 (1991).

[55] G. F. Karavaev and Yu. S. Tikhodeev, Sov. Phys. Semicond. **25**, 745 (1991).

[56] G. F. Karavaev and I. N. Krivorotov, Semiconductors **30**, 102 (1996).

[57] G. F. Karavaev and I. N. Krivorotov, Russ. Phys. J. **39**, 719 (1996).

[58] B. A. Foreman, Phys. Rev. B **48**, R4964 (1993).

[59] B. A. Foreman, Phys. Rev. B **52**, 12241 (1995).

[60] B. A. Foreman, Phys. Rev. B **54**, 1909 (1996).

[61] B. A. Foreman, Phys. Rev. Lett. **81**, 425 (1998).

[62] B. A. Foreman, Phys. Rev. Lett. **86**, 2641 (2001).

[63] M. G. Burt and B. A. Foreman, in *Proceedings of the 24th International Conference on the Physics of Semiconductors, Jerusalem, 1998*, edited by D. Gershoni (World Scientific, Singapore, 1999), CD-ROM file 119.pdf.

[64] P. C. Klipstein, in *Proceedings of the 24th International Conference on the Physics of Semiconductors, Jerusalem, 1998*, edited by D. Gershoni (World Scientific, Singapore, 1999), CD-ROM file 1265.pdf.

[65] É. E. Takhtamirov and V. A. Volkov, Semicond. Sci. Technol. **12**, 77 (1997).

[66] V. A. Volkov and É. E. Takhtamirov, Phys. Usp. **40**, 1071 (1997).

[67] É. E. Takhtamirov and V. A. Volkov, Phys. Low-Dimens. Semicond. Struct. **3/4**, 203 (1999).

[68] É. E. Takhtamirov and V. A. Volkov, JETP **89**, 1000 (1999).

[69] É. E. Takhtamirov and V. A. Volkov, JETP **90**, 1063 (2000).

[70] S. Cortez, O. Krebs, and P. Voisin, J. Vac. Sci. Technol. B **18**, 2232 (2000).

[71] S. Cortez, O. Krebs, and P. Voisin, Eur. Phys. J. B **21**, 241 (2001).

[72] D. L. Smith and C. Mailhiot, Phys. Rev. B **33**, 8345 (1986).

[73] D. L. Smith and C. Mailhiot, Rev. Mod. Phys. **62**, 173 (1990).

[74] C. Mailhiot and D. L. Smith, J. Vac. Sci. Technol. B **8**, 793 (1990).

[75] W. Trzeciakowski, Phys. Rev. B **38**, 12493 (1988).

[76] T. Ando, S. Wakahara, and H. Akera, Phys. Rev. B **40**, 11609 (1989).

[77] J. P. Cuypers and W. van Haeringen, J. Phys.: Condens. Matter **4**,

2587 (1992).

[78] J. P. Cuypers and W. van Haeringen, Phys. Rev. B **47**, 10310 (1993).

[79] J. P. Cuypers and W. van Haeringen, Phys. Rev. B **48**, R11469 (1993).

[80] C. Aversa and J. E. Sipe, Phys. Rev. B **47**, 6590 (1993).

[81] C. Aversa and J. E. Sipe, Phys. Rev. B **49**, 14542 (1994).

[82] S. N. Grinyaev, G. F. Karavaev, and V. N. Chernyshov, Semiconductors **28**, 784 (1994).

[83] L.-W. Wang and A. Zunger, Phys. Rev. B **54**, 11417 (1996).

[84] H. Yi and M. Razeghi, Phys. Rev. B **56**, 3933 (1997).

[85] B. A. Foreman, Phys. Rev. Lett. **82**, 1339 (1999).

[86] R. Magri and A. Zunger, Phys. Rev. B **62**, 10364 (2000).

[87] L. J. Sham and Y.-T. Lu, J. Lumin. **44**, 207 (1989).

[88] B. A. Foreman, in *Proceedings of the 27th International Conference on the Physics of Semiconductors, Flagstaff, Arizona, 2004*, edited by J. Menéndez and C. G. Van de Walle (American Institute of Physics, Melville, NY, 2005), pp. 413–414.

[89] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[90] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[91] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).

[92] G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961).

[93] G. Baym, Phys. Rev. **127**, 1391 (1962).

[94] L. Hedin, Phys. Rev. **139**, A796 (1965).

[95] L. Hedin and S. Lundqvist, in *Solid State Physics*, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 1969), Vol. 23, pp. 1–181.

[96] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, in *Solid State Physics*, edited by H. Ehrenreich and F. Spaepen (Academic, San Diego, 2000), Vol. 54, pp. 1–218.

[97] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).

[98] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[99] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).

[100] X. Zhu and S. G. Louie, Phys. Rev. B **43**, 14142 (1991).

[101] R. T. M. Ummels, P. A. Bobbert, and W. van Haeringen, Phys. Rev. B **57**, 11962 (1998).

[102] W.-D. Schöne and A. G. Eguiluz, Phys. Rev. Lett. **81**, 1662 (1998).

[103] B. Arnaud and M. Alouani, Phys. Rev. B **62**, 4464 (2000).

[104] W. Ku and A. G. Eguiluz, Phys. Rev. Lett. **89**, 126401 (2002).

[105] S. Lebègue, B. Arnaud, M. Alouani, and P. E. Bloechl, Phys. Rev. B **67**, 155208 (2003).

[106] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).

[107] L. Kleinman, Phys. Rev. B **21**, 2630 (1980).

[108] G. B. Bachelet and M. Schlüter, Phys. Rev. B **25**, 2103 (1982).

[109] G. B. Bachelet, D. R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).

[110] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).

[111] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[112] W. Andreoni, A. Baldereschi, and R. Car, Solid State Commun. **27**, 821 (1978).

[113] R. Resta, S. Baroni, and A. Baldereschi, Superlattices Microstruct. **6**, 31 (1989).

[114] S. Baroni, R. Resta, A. Baldereschi, and M. Peressi, in *Spectroscopy of Semiconductor Microstructures*, NATO Advanced Study Institute, Series B: Physics, Vol. 206, edited by G. Fasol, A.

Fasolino, and P. Lugli (Plenum, New York, 1989), pp. 251–272.

[115] S. Baroni, R. Resta, and A. Baldereschi, in *Proceedings of the 19th International Conference on the Physics of Semiconductors, Warsaw, 1988*, edited by W. Zawadzki (Institute of Physics, Polish Academy of Sciences, Warsaw, 1988), pp. 525–528.

[116] M. Peressi, S. Baroni, A. Baldereschi, and R. Resta, Phys. Rev. B **41**, 12106 (1990).

[117] M. Peressi, S. Baroni, R. Resta, and A. Baldereschi, Phys. Rev. B **43**, 7347 (1991).

[118] S. Baroni, M. Peressi, R. Resta, and A. Baldereschi, in *Proceedings of the 21st International Conference on the Physics of Semiconductors, Beijing, 1992*, edited by P. Jiang and H.-Z. Zheng (World Scientific, Singapore, 1992), pp. 689–696.

[119] L. Colombo, R. Resta, and S. Baroni, Phys. Rev. B **44**, 5572 (1991).

[120] M. Peressi and S. Baroni, Phys. Rev. B **49**, 7490 (1994).

[121] M. Peressi, B. Montanari, S. Baroni, and E. Molinari, in *Proceedings of the 23rd International Conference on the Physics of Semiconductors, Berlin, 1996*, edited by M. Scheffler and R. Zimmermann (World Scientific, Singapore, 1996), pp. 943–946.

[122] B. Montanari, M. Peressi, S. Baroni, and E. Molinari, Appl. Phys. Lett. **69**, 3218 (1996).

[123] L. J. Sham, Phys. Rev. **150**, 720 (1966).

[124] B. A. Foreman, preceding paper, Phys. Rev. B **72**, 165344 (2005).

[125] W. A. Harrison, E. A. Kraut, J. R. Waldrop, and R. W. Grant, Phys. Rev. B **18**, 4402 (1978).

[126] There are three possible ways of dealing with heterovalent systems. (i) Treating the monopole terms directly, at the level of the approximation scheme defined in Ref. 124. This is the same level of accuracy as ordinary effective-mass theory (Ref. 127), which means that the position dependence of the effective mass (2.4) and the interface terms (2.5) and (2.6) are neglected. (ii) Treating the monopole terms directly, but extending the analysis in Ref. 124 to higher order, including nonanalytic terms in $W^{(2)}$ and $n^{(2)}$. (iii) Grouping the monopole terms into neutral clusters of fractional atoms (Ref. 124), which are then treated as the fundamental unit of perturbation. This is the approach adopted here, because it allows heterovalent and isovalent systems to be treated as formally the same. However, since the cluster size is several nearest-neighbor spacings, this approach places more stringent limits on what constitutes a "slowly varying" envelope.

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

[128] J.-B. Xia, Phys. Rev. B **39**, 3310 (1989).

[129] L.-W. Wang, A. Franceschetti, and A. Zunger, Phys. Rev. Lett. **78**, 2819 (1997).

[130] L.-W. Wang and A. Zunger, Phys. Rev. B **56**, 12395 (1997).

[131] L.-W. Wang and A. Zunger, Phys. Rev. B **59**, 15806 (1999).

[132] This statement refers, of course, only to the class of heterostructures for which the slowly varying envelope approximation and the error analysis in Sec. II remain valid.

[133] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, Rev. Mod. Phys. **73**, 515 (2001).

[134] G. A. Baraff and D. Gershoni, Phys. Rev. B **43**, 4011 (1991).

[135] D. Gershoni, C. H. Henry, and G. A. Baraff, IEEE J. Quantum Electron. **29**, 2433 (1993).

[136] R. Winkler and U. Rössler, Phys. Rev. B **48**, 8918 (1993).

[137] M. Cardona and F. H. Pollak, Phys. Rev. **142**, 530 (1966).

[138] F. H. Pollak, C. W. Higginbotham, and M. Cardona, J. Phys. Soc. Jpn. **21**, 20 (1966).

[139] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 2920 (1986).

[140] M. P. Surh, M.-F. Li, and S. G. Louie, Phys. Rev. B **43**, 4286 (1991).

[141] L. A. Hemstreet, C. Y. Fong, and J. S. Nelson, Phys. Rev. B **47**, 4238 (1993).

[142] G. Theurich and N. A. Hill, Phys. Rev. B **64**, 073106 (2001).

[143] Although the language used here (and throughout this paper) refers to atoms, in heterovalent systems it is convenient (Ref. 126) to use these symbols as labels for the neutral clusters of (fractional) atoms defined in Ref. 124.

[144] B. A. Foreman (unpublished).

[145] M. L. Cohen and V. Heine, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1970), Vol. 24, pp. 37–248.

[146] In Eq. (3.24), an analytic contribution to $\varphi_{\mathbf{R}}^{\alpha\tau}(\mathbf{k}-\mathbf{k}')$ from Ref. 124 has been absorbed into the definition of $w_{\mathbf{R}}^{\alpha\tau}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}')$.

[147] This choice of basis was criticized in Ref. 69 as being nonorthogonal and overcomplete. However, since (for example) the X_z states change sign under any fcc lattice translation whose z component is $\pm a/2$, the basis functions are orthogonal with respect to the simple cubic unit cell. Also, although there are four times as many zone-center basis functions in the supercell, a primitive unit cell in the reciprocal lattice has a volume $\Omega_0^* = (2\pi)^3/\Omega_0$ only one-fourth the original volume, so the basis is complete but not overcomplete.

[148] It was shown in Refs. 54 and 68 that when Ω_0^* is chosen to be the conventional 3D Brillouin zone of the reference crystal, difficulties involving "umklapp"-like processes generated by $V_{nn'}(\mathbf{k},\mathbf{k}')$ may arise in certain applications for heterostructures with 2D translation symmetry. However, as noted by Burt (Ref. 50), these difficulties are easily avoided by choosing a "slab-adapted" unit cell.[158,159] It is assumed here that this choice has been made.

[149] K. A. Mäder and A. Zunger, Phys. Rev. B **50**, 17393 (1994).

[150] M. L. Cohen and J. R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors*, 2nd ed. (Springer, Berlin, 1989).

[151] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, New York, 1999), pp. 253–254.

[152] R. G. Dandrea, C. B. Duke, and A. Zunger, J. Vac. Sci. Technol. B **10**, 1744 (1992).

[153] É. I. Rashba, Sov. Phys. Solid State **2**, 1109 (1960).

[154] F. T. Vas'ko, JETP Lett. **30**, 541 (1979).

[155] R. Winkler, Phys. Rev. B **62**, 4245 (2000).

[156] R. Winkler, H. Noh, E. Tutuc, and M. Shayegan, Phys. Rev. B **65**, 155303 (2002).

[157] The expression given by Leibler in Eq. (30) of Ref. 48 is antisymmetric in the Cartesian indices, but this is obviously incorrect, as the net contribution from this term would then be identically zero.

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

[159] L. Kleinman, Phys. Rev. B **24**, 7412 (1981).

[160] J. R. Waterman, B. V. Shanabrook, R. J. Wagner, M. J. Yang, J. L. Davis, and J. P. Omaggio, Semicond. Sci. Technol. **8**, S106 (1993).

[161] B. A. Foreman, Phys. Rev. B **52**, 12260 (1995).

[162] R. H. Parmenter, Phys. Rev. **100**, 573 (1955).

[163] G. Dresselhaus, Phys. Rev. **100**, 580 (1955).

[164] G. Bastard, Phys. Rev. B **24**, 5693 (1981).

[165] G. Bastard, Phys. Rev. B **25**, 7584 (1982).

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

[167] C. R. Pidgeon and R. N. Brown, Phys. Rev. **146**, 575 (1966).

[168] C. R. Pidgeon and S. H. Groves, Phys. Rev. **186**, 824 (1969).

[169] This potential was of the form (5.9), but with $\theta^l(\mathbf{q})=0$ outside the first Brillouin zone.

[170] J. M. Luttinger, Phys. Rev. **121**, 942 (1961).

[171] J. W. Negele and H. Orland, *Quantum Many-Particle Systems* (Westview, Oxford, 1998), p. 254.

[172] B. Farid, in *Electron Correlation in the Solid State*, edited by N. H. March (Imperial College, London, 1999), pp. 103–261.

[173] J. J. Quinn and R. A. Ferrell, Phys. Rev. **112**, 812 (1958).

[174] J. Callaway, *Quantum Theory of the Solid State*, 2nd ed. (Academic, San Diego, 1991), p. 820.

[175] A. Fleszar and W. Hanke, Phys. Rev. B **56**, 10228 (1997).