

The justification for applying the effective-mass approximation to microstructures

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1992 J. Phys.: Condens. Matter 4 6651

(<http://iopscience.iop.org/0953-8984/4/32/003>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.96

The article was downloaded on 11/05/2010 at 00:22

Please note that [terms and conditions apply](#).

INVITED PAPER

The justification for applying the effective-mass approximation to microstructures

M G Burt

BT Laboratories, Martlesham Heath, Ipswich, Suffolk IP5 7RE, UK

Received 9 March 1992

Abstract. The assumptions of conventional effective-mass theory, especially the one of continuity of the envelope function at an abrupt interface, are reviewed critically so that the need for a fresh approach becomes apparent. A new envelope-function method, developed by the author over the past few years, is reviewed. This new method is based on both a generalization and a novel application to microstructures of the Luttinger–Kohn envelope-function expansion. The differences between this new method and the conventional envelope-function method are emphasized. An alternative derivation of the new envelope-function equations, which are exact, to that already published is provided. A new and improved derivation of the author's effective-mass equation is given, in which the differences in the zone-centre eigenstates of the constituent crystals are taken into account. This derivation is valid for abrupt interfaces and relies only on the slow variation of the envelope function(s). Unambiguous boundary conditions that automatically conserve probability current are derived. The cause of the kinks in the conventional effective-mass envelope function, at abrupt effective-mass changes, is identified. The formal results are extensively illustrated with a numerical example. A plot of the author's exact envelope function shows that it has a soft kink at an effective-mass discontinuity. This soft kink is also seen in the exact wavefunction. It is this feature which is approximated in conventional models by the effective-mass-related kink. The reasons why conventional effective-mass theory works so well become clear. The distinction between Luttinger–Kohn and Wannier–Slater envelope functions is highlighted. It is demonstrated that the author's generalization and new application of the Luttinger–Kohn envelope-function expansion can also be carried out for the Wannier–Slater case. However, it is found that the derivation of effective-mass equations for the generalized Wannier–Slater envelope functions is not as straightforward as for the Luttinger–Kohn case. Some problems uncovered recently in trying to extend effective-mass theory to the non-parabolic regime are resolved.

1. Introduction

The use of the effective-mass approximation, in various guises, is almost as old as solid-state theory itself, and its formal justification for microstructures has long provided a challenge for the theorist. The approximation has been used extensively to describe electronic motion near band extrema in the presence of slowly varying weak perturbations, such as applied magnetic and electric fields, as well as for the shallow-impurity problem in semiconductors; see e.g. Luttinger and Kohn (1955) especially for earlier references, Kohn (1957) and Altarelli and Bassani (1982). With the advent of semiconductor quantum wells (Dingle *et al* 1974), it was natural to try to apply an effective-mass approximation, the famous 'particle in a box' model,

to the calculation of the low-lying electronic states, because the calculation is simple and at least one expects the envelope of the wavefunction to be slowly varying on the atomic scale, one of the conditions needed to derive effective-mass equations for homogeneous crystals. However, the other condition for the validity of the effective-mass approximation, that the perturbing potential is weak, is clearly violated for quantum wells, because the difference between the potentials in the well and barrier layer is not small, neither is it slowly varying. Yet the 'particle in a box' model seemed to work (Dingle *et al* 1974) and has been in extensive use ever since. So the advent of quantum wells brought about a hiatus in the development of effective-mass theory from the well understood regime of weak perturbations in bulk crystals to the little-understood strong perturbation regime in microstructures.

Why the effective-mass theory works so well for microstructures has been an enigma, and progress in resolving it has been mixed. Much of this work has been on the boundary-condition problem, i.e. how to connect the solutions on either side of an atomically abrupt interface (see e.g. Altarelli 1983a,b,c, 1986, Smith and Mailhot 1986, Eppenga *et al* 1987, Baraff and Gershoni 1991) or equivalently trying to determine the Hamiltonian phenomenologically (see e.g. Morrow and Brownstein 1984, Morrow 1987a, b, Einevoll and Hemmer 1988, Thomsen *et al* 1989, Einevoll *et al* 1990, Einevoll 1990). What has hampered progress has been the lack of a derivation of the effective-mass equations from first principles, without the use of heuristic arguments, i.e. a derivation that starts from the microscopic Schrödinger equation and at each stage provides a means of estimating the error in any approximation made. The main problem has been the heuristic nature of the conventional envelope-function method (for a recent review of the applications of this method see Bastard *et al* (1991)). A few years ago, the author managed to develop a new envelope-function method, derive the corresponding exact envelope-function equations and demonstrate how to arrive at an effective-mass equation (Burt 1987, 1988a, b, 1989). This paper is aimed at showing how this approach can resolve the above-mentioned problems.

In the next section the problems associated with conventional approaches to the justification of the effective-mass approximation are reviewed. While the problem of finding the correct boundary condition for the derivative of the envelope function has received extensive attention, continuity of the envelope function is widely assumed (see e.g. Bastard 1981, Bastard *et al* 1991), though there are exceptions (e.g. Ando *et al* 1989, Tikhodeev 1991, Cuypers and van Haeringen 1992). The potential errors in this assumption are highlighted and the need for a definition of envelope functions that builds in continuity from the start becomes clear. Such a definition has been the basis of the author's approach, and this approach is reviewed along with the derivation of the exact equations for these new envelope functions (Burt 1987, 1988a, 1989) in section 3. The differences between the author's new envelope-function method and the conventional method are emphasized. In section 4 an alternative and, as yet, unpublished derivation of these exact envelope-function equations is given, a derivation that emphasizes that the new exact envelope-function equations are related to the original Schrödinger equation by unitary transformations; such a derivation may be more appealing to some readers. In particular, this derivation demonstrates the power of this new envelope-function method. It can handle non-local potentials, for instance, and can easily be adapted to include spin-orbit interaction (Burt 1989). And, indeed, on first applying a change in independent variable, it could be applied to strained structures, as has already been done in earlier work (Burt 1989). The next step on the road to the derivation of an effective-mass equation is to justify the

dropping of the non-local terms in the author's exact envelope-function equations. This is sketched in section 5 along with a review of the other simplifications and their justification to get approximate envelope-function equations with piecewise-constant coefficients for microstructures with atomically abrupt boundaries. In section 6 the author's approximate envelope-function equations are used to derive effective-mass equations. *In this derivation, the change in each and every zone-centre eigenfunction as one moves across an interface is taken into account as a matter of course, even though the basis functions remain the same.* Also, provided the right level of approximation is retained (Burt 1988b), the author's envelope function and its derivative are continuous everywhere, even at atomically abrupt interfaces. Returning to the author's original envelope-function equations, the question of boundary conditions at abrupt interfaces under more severe approximations is dealt with in section 7. In the author's formalism, this question only arises when a piecewise-continuous approximation is made to the coefficients in the envelope-function equations. It is shown that all the solutions of these equations and their derivatives must be continuous. However, if, as a further approximation, some of the second-derivative terms are neglected, then kinks in the author's envelope functions at effective-mass discontinuities may arise. In this same section, the relation between the conservation of the probability current and the above-mentioned boundary conditions is discussed. As a separate issue, the problem of boundary conditions at an impenetrable barrier is examined, pointing out that the apparently obvious answer, though approximately correct, is not quite so easy to justify as one might suppose.

Section 8 gives numerical results for a model superlattice. The model, being one-dimensional, is not intended, of course, to represent a real physical system. Its primary purpose is to illustrate the new mathematical method and this it does very well with minimal complexity. On solving the Schrödinger equation for this model, the ground-state wavefunction is found to have a pronounced soft kink in its envelope at the interfaces. The author's new exact envelope-function equations are also solved exactly and a rapid variation in the derivative of the dominant envelope function corresponding to the effective-mass change is found as predicted earlier (Burt 1988b). This rapid variation in the author's envelope function corresponds to the effective-mass-related kink generated by conventional treatments. A number of plots are given for the various functions that arise in the author's formulation.

So far, the paper is concerned with the author's generalization and novel application of the envelope functions introduced by Luttinger and Kohn (1955). However, there is an alternative approach to envelope-function theory using the definition of Wannier (1937) and Slater (1949) based on Wannier orbitals. In section 9, it is shown how the method used in section 4 to derive exact equations for the author's generalization of Luttinger-Kohn envelope functions can be used to do the same for the Wannier-Slater case. In the same section we extend a derivation by Young (1989) of the effective-mass equation based on Wannier orbitals. It is pointed out that it is not necessary to restrict this derivation, as Young did, to structures with slowly varying and small overall composition changes. The derivation turns out to be considerably more complicated than the derivation presented in section 6, and still does not cover the possibility of substantial changes in composition dealt with there.

The extension of effective-mass theory to the non-parabolic regime is a topic of considerable importance for cases of strong quantum confinement. In the author's approach, and in the $k \cdot p$ method generally, the use of an energy-dependent effective mass seems to emerge naturally. This approach has, nevertheless, come under

criticism recently. Section 10 is devoted, in part, to clarifying the situations in which the energy-dependent effective-mass approach is a useful one. It also gives an opportunity to show the power of an *ab initio* method, such as the author has developed, to resolve the type of questions raised by the above-mentioned criticism; it makes a fitting point at which to finish by giving a summary and some perspective (section 11).

Finally, a note on nomenclature. It is important to keep in mind the distinction between conventional envelope functions and those introduced by the author. It would, however, be clumsy to make this distinction fully and explicitly at absolutely every relevant point in the text. So, as a shorthand, the envelope functions introduced by the author will often be referred to simply as the 'new' envelope functions, even though it is some five years since they were first introduced (Burt 1987). And in the absence of any qualification it is to be assumed that the envelope functions concerned are 'new'. The exception is in the next section, dealing with conventional effective-mass theory, where the envelope function referred to is the conventional one and denoted by f . The 'new' envelope functions introduced and used in subsequent sections will be denoted by F_n .

2. The problems with conventional effective-mass theory

The problems with conventional effective-mass theory can be highlighted by considering electronic motion in a quantum well in a one-dimensional crystal system. For both the well and the barrier layers, one can write the wavefunction Ψ approximately as

$$\Psi(z) = f(z)U(z) \quad (2.1)$$

where $U(z)$ is the wavefunction for the band-edge eigenstate and $f(z)$ is a slowly varying envelope function. To find the eigenstates of the structure one has to match the wavefunction in the well and barrier crystals at the interfaces. For an interface at $z = 0$ we have

$$\Psi^{(w)}(0) = \Psi^{(b)}(0) \quad (2.2)$$

so

$$f^{(w)}(0)U^{(w)}(0) = f^{(b)}(0)U^{(b)}(0). \quad (2.3)$$

It is almost universally assumed that the well and barrier crystals are so similar in their chemical nature that the U 's are approximately equal (see e.g. Bastard *et al* 1991) and one obtains the approximate continuity of the envelope function

$$f^{(w)}(0) = f^{(b)}(0) \quad (2.4)$$

provided the U 's are not zero at $z = 0$. Let us take a look at this assumption that the U 's are equal. Suppose the error in this assumption is 5%, a typical figure for semiconductor quantum wells. The error in the energy is of the order of the square of the error in Ψ , i.e. 0.25%. This would appear to be a highly satisfactory state of affairs if this error referred to the size quantization energy. However, it does not. Rather, it refers to the total energy, and this is more realistically measured from the

kinetic energy zero of the 'empty' lattice. This total energy is typically about 10 eV in semiconductors and hence the error in the energy is of the order of 25 meV! So the similarities in the U 's does not explain why the effective-mass method works so well. There must be a more subtle reason, to which we will return.

One could object to the above reasoning concerning the errors inherent in the approximation (2.4) by saying that the effective-mass equation incorporating (2.4) is of the Sturm–Liouville type and so errors in the eigenvalue (the size quantization energy) are of the order of the square of the errors in f . However, for this argument to be acceptable, one has to show that an effective-mass-type equation of the Sturm–Liouville form exists which incorporates the true boundary condition (2.3).

Another way of viewing the problem of assuming that the U 's are equal is to focus on the size of the terms one is throwing out of the Hamiltonian when this approximation is made. An idea of the size can be obtained from the work of Pollak and Cardona (1966) and Pollak *et al* (1966), who produced $k \cdot p$ band structures for III–V compounds using zone-centre wavefunctions of elemental semiconductors as a basis. The matrix elements that arise because the zone-centre eigenfunctions for, say, Ge do not diagonalize the GaAs Hamiltonian are in excess of 1.5 eV! This example, of course, overstates the case for systems of common interest such as GaAlAs/GaAs and GaInAs/InP, but even here the terms one throws away are of the order of several hundred meV (Burt 1988c), which is about the size of the ' $k \cdot p$ ' type terms that are conventionally retained.

Now consider the boundary condition on the derivative. The usual way of justifying the continuity of $(1/m^*)df/dz$ at an interface is to appeal to conservation of the current, $\text{Re}[(1/m^*)f^*(-i\hbar df/dz)]$. If continuity of f is assumed, then continuity of $(1/m^*)df/dz$ will ensure current conservation, though this is not the only choice. The puzzling aspect of this conclusion is that it puts a kink in the approximate wavefunction, a kink that cannot literally be present in the true wavefunction if its second derivative is to be finite. On the other hand, if one omits the effective mass in the derivative boundary condition, then one has problems with current conservation. This is not a problem with the bound states, which do not carry current, but it is for the extended states, and if a boundary condition is to be universal, then it must apply to both types of states.

It is sometimes said that this boundary-condition problem is not an important one because the eigenvalues are insensitive to whether or not the effective mass is included. This is indeed the case when the confining potential is weak and the states are extended, so that the derivative of the envelope function is small and plays little role in determining the energy. For cases in which the confinement is not weak, the inclusion or otherwise of the effective mass in the derivative boundary condition can lead to large differences, as will be demonstrated later in section 8.

Another approach for justifying the application of the effective-mass method to microstructures (see e.g. Pötz *et al* (1985) for the equivalent arguments for the $k \cdot p$ case) is to start with the effective-mass equation for a bulk crystal,

$$-\frac{\hbar^2}{2m^*} \frac{d^2 f}{dz^2} + E_0 f = E f \quad (2.5)$$

where E_0 is the band-edge energy, and generalize it to the case of slowly varying composition changes by allowing m^* and E_0 to depend on z . The kinetic energy operator, $-(\hbar^2/2m^*)d^2/dz^2$, becomes non-Hermitian because m^* now depends on z . To overcome this objection the operator is replaced by a Hermitian

one. There are an infinite number of choices at this point, but the usual one is $-(\hbar^2/2)(d/dz)(1/m^*)(d/dz)$. So one ends up with the equation

$$-(\hbar^2/2)(d/dz)((1/m^*)df/dz) + E_0f = Ef \quad (2.6)$$

which one then assumes, without further justification, will also be valid for the case of rapidly varying compositions, including atomically abrupt interfaces. Integrating this equation across an interface gives the boundary condition $(1/m^*)df/dz$ continuous, which is consistent with current conservation.

There have been a number of attempts to determine which symmetrization procedure is the correct one, including an elegant method to fix it using exactly solvable models, which support the symmetrization given above (Einevoll and Hemmer 1988, Thomsen *et al* 1989, Einevoll *et al* 1990). The limitations of such approaches are that they assume that there exists a unique symmetrization procedure, that the effective potential energy can be represented by the band edge, E_0 , and that they leave open the problem of the relation of the envelope function to the wavefunction.

Both the approaches described here, the matching of bulk solutions across an interface, and the generalization of the bulk equation to the case of spatially varying composition, have a common weakness: they provide no way of estimating the errors used in their use except against microscopic calculations, which can only be *ad hoc*. The only way to get around this problem is to provide a derivation of the effective-mass equation starting from the Schrödinger equation so that the effect of discarded terms can be estimated and any symmetrization of the kinetic energy operator emerges naturally from the derivation. The next section outlines how the first step of such a derivation has been achieved.

3. The new envelope-function method: a review

A prerequisite for a satisfactory effective-mass theory is a well formulated envelope-function method, and that in turn relies on suitably defined envelope functions, F_n . The author (Burt 1987, 1988a, b, c, 1989) has provided such an envelope-function method using the following definition: Consider a complete set of functions $U_n(\mathbf{R})$ that are periodic in the position variable \mathbf{R} , with respect to a Bravais lattice. The new envelope functions, $F_n(\mathbf{R})$, for an arbitrary wavefunction $\Psi(\mathbf{R})$, are defined as the coefficients in the expansion

$$\Psi(\mathbf{R}) = \sum_n F_n(\mathbf{R})U_n(\mathbf{R}) \quad (3.1)$$

with the proviso that the plane-wave expansion of the $F_n(\mathbf{R})$ is restricted to the first Brillouin zone (for a discussion on the choice of Brillouin zones, see appendix 1) corresponding to the Bravais lattice used to define the $U_n(\mathbf{R})$. This definition of the envelope functions is, in essence, the one given by Luttinger and Kohn (1955). In that paper, it was applied to bulk semiconductors with shallow acceptor levels or applied magnetic fields. The envelope-function expansion was defined, from the outset, with the $U_n(\mathbf{R})$ as the zone-centre eigenfunctions of the semiconductor, though the extension to arbitrary periodic functions is implicit in their work. In subsequent applications to microstructures by almost all other authors, the Luttinger and Kohn choice

(which will be referred to as the conventional choice) of the zone-centre eigenfunctions has been made in each layer, i.e. a different expansion is made in each layer. In contrast, the author (Burt 1988a, 1989) has elected to apply one such expansion to an entire microstructure; the same $U_n(\mathbf{R})$ are used throughout even though there are regions where they are not eigenstates of the local Hamiltonian. There are two advantages. First, the $F_n(\mathbf{R})$ are unique: once $\Psi(\mathbf{R})$ is given, then the $F_n(\mathbf{R})$ are given by an inversion formula without further constraints such as obeying the $\mathbf{k} \cdot \mathbf{p}$ band structure equations, though they do in fact obey related equations (see (3.12)) that also provide the band-structure, far from interfaces when $\Psi(\mathbf{R})$ is an energy eigenstate. Secondly, the boundary-condition problem is completely circumvented: the $F_n(\mathbf{R})$ and all their derivatives are smooth and continuous everywhere for well behaved Ψ , even at abrupt interfaces, because of the limited range of wavevectors used in their plane expansion.

It is well worth elaborating on the difference between the definition of the new envelope functions given above and the definition of the conventional envelope functions (see e.g. Altarelli 1983a,b,c, 1986). There are two differences. The first lies in the choice of the periodic basis functions. The conventional envelope functions are defined with respect to zone-centre eigenfunctions, i.e. periodic functions that diagonalize the local Hamiltonian while the new expansion (3.1) employs no such restriction. This generalization is not new. It was used by Pollak *et al* (1966) and Pollak and Cardona (1966) to obtain the $\mathbf{k} \cdot \mathbf{p}$ band-structure of III-V compounds using group IV zone-centre eigenfunctions as a basis; in this case the envelope functions are propagating plane waves. Smith and Mailhot (1986) also used the same periodic basis functions in each layer in their $\mathbf{k} \cdot \mathbf{p}$ theory for superlattices though here they use a combination of propagating and evanescent plane waves. The second and much more important difference lies in the scope of the expansion. The conventional envelope-function expansion in each homogeneous material region is limited to a solution of the time-independent Schrödinger equation for the local Hamiltonian and this also applies to the work of Smith and Mailhot (1986). On the other hand, the new expansion introduced above can represent *any* well behaved wavefunction regardless of whether or not it is an eigenstate of a particular Hamiltonian. It is just this increased scope that makes this new expansion so powerful.

Armed with this new exact and unique expansion we are in a position to derive exact equations for these envelope functions, $F_n(\mathbf{R})$, for the case in which $\Psi(\mathbf{R})$ is an eigenstate of a microstructure; the existence of such equations is in no doubt, given the uniqueness of this envelope-function expansion. The basic idea is to insert the envelope-function expansion (3.1) for $\Psi(\mathbf{R})$ into both sides of the Schrödinger equation, rearrange the expressions into envelope expansion form and then equate coefficients of $U_n(\mathbf{R})$ on both sides. For the kinetic energy term we get

$$T\Psi = -\frac{\hbar^2}{2m} \sum_n [(\nabla^2 F_n)U_n + 2(\nabla F_n) \cdot (\nabla U_n) + F_n \nabla^2 U_n]. \quad (3.2)$$

Fortunately, this expression is already close to the envelope-function expansion form (3.1) because each term is a product of a periodic function and a function with a plane-wave expansion restricted to the first Brillouin zone. All we have to do is express the periodic functions, ∇U_n and $\nabla^2 U_n$, in terms of the U_n . This is easily done in terms of the matrix elements of T and \mathbf{p} , the kinetic energy and momentum,

with respect to the U_n . We obtain

$$T\Psi = \sum_n \left(-\frac{\hbar^2}{2m} \nabla^2 F_n + \sum_{n'} \frac{-i\hbar}{m} \mathbf{p}_{nn'} \cdot \nabla F_{n'} + \sum_{n'} T_{nn'} F_{n'} \right) U_n \quad (3.3)$$

where

$$\mathbf{p}_{nn'} = \frac{1}{V_c} \int U_n^* \mathbf{p} U_{n'} d^3 R \quad (3.4)$$

$$T_{nn'} = \frac{1}{V_c} \int U_n^* T U_{n'} d^3 R \quad (3.5)$$

and the normalization is

$$\frac{1}{V_c} \int U_n^* U_{n'} d^3 R = \delta_{nn'} \quad (3.6)$$

with the integrals taken over a unit cell (volume V_c) of the Bravais lattice used to define the $U_n(\mathbf{R})$. The U_n have been chosen orthonormal for convenience. Expressions for $\mathbf{p}_{nn'}$ and $T_{nn'}$ for the general case in which the U_n are not orthogonal, but still independent, are given in section 4.

More manipulation is needed to bring the potential energy term, $V\Psi$, into envelope-function expansion form (3.1), because V for a microstructure contains plane-wave components with wavevectors both inside and outside the Brillouin zone. The manipulation is achieved by expanding all the factors in plane waves, expressing each product of plane waves into a single plane wave of the form $\exp[i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{R}]$ with \mathbf{G} a reciprocal-lattice vector and \mathbf{k} inside the Brillouin zone, and then expressing the periodic plane waves, $\exp(i\mathbf{G} \cdot \mathbf{R})$, in terms of the periodic functions U_n . The end result is of the form

$$\sum_n \left(\sum_{n'} \int V_{nn'}(\mathbf{R}, \mathbf{R}') F_{n'}(\mathbf{R}') d^3 \mathbf{R}' \right) U_n(\mathbf{R}) \quad (3.7)$$

where the integral is over the large volume over which cyclic boundary conditions have been applied and the dependence of $V_{nn'}$ on both \mathbf{R} and \mathbf{R}' can each be expressed as a plane-wave expansion restricted to the Brillouin zone.

The RHS of the Schrödinger equation $H\Psi = (T + V)\Psi = E\Psi$ is already in envelope expansion form (3.1). Equating coefficients of U_n on both sides gives the exact equations

$$-\frac{\hbar^2}{2m} \nabla^2 F_n(\mathbf{R}) + \sum_{n'} \frac{-i\hbar}{m} \mathbf{p}_{nn'} \cdot \nabla F_{n'}(\mathbf{R}) + \sum_{n'} \int H_{nn'}(\mathbf{R}, \mathbf{R}') F_{n'}(\mathbf{R}') d^3 \mathbf{R}' = E F_n \quad (3.8)$$

where

$$H_{nn'}(\mathbf{R}, \mathbf{R}') = T_{nn'} \Delta(\mathbf{R} - \mathbf{R}') + V_{nn'}(\mathbf{R}, \mathbf{R}') \quad (3.9)$$

and

$$\Delta(\mathbf{R} - \mathbf{R}') = \frac{1}{\Omega} \sum_{\mathbf{k}} \exp[i\mathbf{k} \cdot (\mathbf{R} - \mathbf{R}')] \quad (3.10)$$

is the plane-wave expansion of the delta function restricted to wavevectors in the first Brillouin zone (see e.g. Stoneham 1975).

Away from interfaces, $H_{nn'}(\mathbf{R}, \mathbf{R}')$ reduces to $H_{nn'}\Delta(\mathbf{R} - \mathbf{R}')$ where

$$H_{nn'} = \frac{1}{V_c} \int U_n^* H U_n d^3R \quad (3.11)$$

(see appendix 3) provided the U_n have been chosen to have the same period as the crystal Hamiltonian at \mathbf{R} . $H_{nn'}$ depends only on the bulk material since H represents the bulk Hamiltonian at large distances from any interface. The equation (3.8) becomes

$$-\frac{\hbar^2}{2m} \nabla^2 F_n(\mathbf{R}) + \sum_{n'} \frac{-i\hbar}{m} \mathbf{p}_{nn'} \cdot \nabla F_{n'}(\mathbf{R}) + \sum_{n'} H_{nn'} F_{n'}(\mathbf{R}) = E F_n. \quad (3.12)$$

These equations have constant coefficients and therefore have plane-wave solutions leading to an eigenvalue problem for the band structure of the generalized $\mathbf{k} \cdot \mathbf{p}$ type introduced by Pollak *et al* (1966) and Pollak and Cardona (1966). When the microscopic potential changes abruptly at interfaces and there are no built-in or external fields, equation (3.12) can be used as the basis for a $\mathbf{k} \cdot \mathbf{p}$ theory of superlattices (Smith and Mailhiot 1986). In each layer, in the spirit of the conventional envelope-function method, one writes the wavefunction as $\sum_n f_n U_n$, where the f_n are a trial combination of solutions (both propagating and evanescent waves) of (3.12). These f_n do not coincide with the F_n near interfaces; they still obey (3.12) *not* the more general (3.8); they also do not have any restriction on the wavevectors of the plane waves included in their Fourier transform. The trial solutions $\sum_n f_n U_n$ in adjoining layers then have to be matched across their common interface. The boundary conditions cannot be obtained simply by integrating (3.12) across the interface (Smith and Mailhiot (1986), quite rightly, eschew this approach) since it has not been established that (3.12) holds there; in general it does not (see, for instance, the interface terms discussed in section 5). On the other hand, the F_n obey the more complicated equation (3.8), which is valid *everywhere* even at abrupt interfaces and their continuity is built in from the start. More discussion on this point is to be found in Burt (1988a).

Finally, it is worth emphasizing the general applicability of this transformation into the new envelope-function representation. It is not restricted to a particular type of Hamiltonian. It can treat spin-orbit interaction, non-local potentials such as one comes across in pseudopotential and many-electron theory, as well as non-lattice-matched structures (Burt 1989). Because the envelope-function expansion (3.1) does not rely on the wavefunction being an eigenstate, one can also treat time-dependent problems. The only changes one has to make to the above is to recognize that Ψ can be time-dependent if the F_n are. The manipulation of the LHS is the same. The manipulation of the RHS is still trivial; $i\hbar \partial / \partial t$ just replaces the energy, E .

4. An alternative derivation of the new envelope-function equations

The approach used in this section emphasizes the fact that the exact envelope-function equations (3.8) are related to the original Schrödinger equation by a unitary transformation. It is also the approach needed in section 9 to derive the corresponding equations for the Wannier-Slater type envelope functions and clarify the distinction between those envelope functions and the Luttinger-Kohn type envelope functions introduced in the previous section.

Our starting point is the Schrödinger equation in the plane-wave representation

$$\frac{\hbar^2}{2m}(\mathbf{k} + \mathbf{G})^2 \tilde{\Psi}_{\mathbf{G}}(\mathbf{k}) + \sum_{\mathbf{k}', \mathbf{G}'} \langle \mathbf{k} + \mathbf{G} | V | \mathbf{k}' + \mathbf{G}' \rangle \tilde{\Psi}_{\mathbf{G}'}(\mathbf{k}') = E \tilde{\Psi}_{\mathbf{G}}(\mathbf{k}) \quad (4.1)$$

where the wavefunction $\Psi(\mathbf{R})$ and its Fourier transform $\tilde{\Psi}_{\mathbf{G}}(\mathbf{k})$ are related by

$$\tilde{\Psi}_{\mathbf{G}}(\mathbf{k}) = \frac{1}{\Omega} \int \Psi(\mathbf{R}) \exp[-i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{R}] d^3R \quad (4.2a)$$

$$\Psi(\mathbf{R}) = \sum_{\mathbf{k}, \mathbf{G}} \tilde{\Psi}_{\mathbf{G}}(\mathbf{k}) \exp[i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{R}]. \quad (4.2b)$$

Here Ω is the large volume completely enclosing the microstructure over which the usual cyclic boundary conditions are applied. Wavevectors have been expressed in the form $\mathbf{k} + \mathbf{G}$ where \mathbf{G} is a reciprocal-lattice vector of a Bravais lattice and \mathbf{k} lies within a suitably chosen Brillouin zone (see appendix 1). We now introduce a new complete set of periodic functions $U_n(\mathbf{R})$ with plane-wave expansion

$$U_n(\mathbf{R}) = \sum_{\mathbf{G}} U_{n\mathbf{G}} \exp(i\mathbf{G} \cdot \mathbf{R}). \quad (4.3a)$$

Because the $U_n(\mathbf{R})$ are complete we may expand the plane waves in terms of them

$$\exp(i\mathbf{G} \cdot \mathbf{R}) = \sum_n (U^{-1})_{\mathbf{G}n} U_n(\mathbf{R}) \quad (4.3b)$$

the completeness being expressed by

$$\sum_n (U^{-1})_{\mathbf{G}n} U_{n\mathbf{G}'} = \delta_{\mathbf{G}\mathbf{G}'}. \quad (4.3c)$$

Substitution for $\exp(i\mathbf{G} \cdot \mathbf{R})$ in (4.2b) using (4.3b) gives the envelope-function expansion

$$\Psi(\mathbf{R}) = \sum_n F_n(\mathbf{R}) U_n(\mathbf{R}) \quad (4.4)$$

with

$$F_n(\mathbf{R}) = \sum_{\mathbf{k}} \tilde{F}_n(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{R}) \quad (4.5a)$$

$$\tilde{F}_n(\mathbf{k}) = \frac{1}{\Omega} \int F_n(\mathbf{R}) \exp(-i\mathbf{k} \cdot \mathbf{R}) d^3R \quad (4.5b)$$

$$\tilde{F}_n(\mathbf{k}) = \sum_{\mathbf{G}} \tilde{\Psi}_{\mathbf{G}}(\mathbf{k}) (U^{-1})_{\mathbf{G}n}. \quad (4.6)$$

Premultiplying (4.1) by $(U^{-1})_{Gn}$, summing over G and using the inverse of (4.6), i.e.

$$\tilde{\Psi}_{G'}(k') = \sum_n U_{nG'} \tilde{F}_n(k') \tag{4.7}$$

we get

$$\frac{\hbar^2 k^2}{2m} \tilde{F}_n(k) + \frac{\hbar}{m} \sum_{n'} k \cdot p_{nn'} \tilde{F}_{n'}(k) + \sum_{n'} \sum_{k'} \tilde{H}_{nn'}(k, k') \tilde{F}_{n'}(k') = E \tilde{F}_n(k) \tag{4.8}$$

where

$$p_{nn'} = \sum_G \hbar G (U^{-1})_{Gn} U_{n'G} \tag{4.9}$$

$$\tilde{H}_{nn'}(k, k') = T_{nn'} \delta_{kk'} + \tilde{V}_{nn'}(k, k') \tag{4.10}$$

$$T_{nn'} = \sum_G \frac{\hbar^2 G^2}{2m} (U^{-1})_{Gn} U_{n'G} \tag{4.11}$$

$$\tilde{V}_{nn'}(k, k') = \sum_{G, G'} (U^{-1})_{Gn} (k + G | V | k' + G') U_{n'G'} \tag{4.12}$$

Transforming into real space gives the exact envelope-function equations (3.8) with

$$H_{nn'}(R, R') = \frac{1}{\Omega} \sum_{k, k'} \exp(+ik \cdot R) \tilde{H}_{nn'}(k, k') \exp(-ik' \cdot R') \tag{4.13}$$

giving $H_{nn'}(R, R')$ in terms of the potential V , via (4.10), (4.11) and (4.12). The definitions of $p_{nn'}$, and $T_{nn'}$ coincide with those of the previous section when the transformation U is unitary, i.e. $(U^{-1})_{Gn} = U_{nG}^*$.

This derivation of equation (3.8) is not restricted to local potentials, as no assumption has been made concerning the dependence of $\langle k + G | V | k' + G' \rangle$ on the wavevectors. Indeed, one can apply the above derivation to a truncated plane-wave Hamiltonian (thereby creating an effective potential that is non-local even though the original potential was local), one in which the number of reciprocal-lattice vectors G is finite; the eigenvalues of the truncated plane-wave Hamiltonian and the exact envelope-function equations will be identical. The numerical example treated in section 8 is based on such a truncated plane-wave Hamiltonian. The derivation could also be generalized with little effort to include spin-orbit interaction.

5. Approximations to the new envelope-function equations

The new envelope-function equations (3.8) can be solved exactly for systems with periodic boundary conditions using a plane-wave representation; this technique is used later in this paper in the numerical example (section 8). This approach is unattractive, in general, because it involves diagonalizing large matrices. The attraction of envelope-function and effective-mass methods has been their algebraic and

computational simplicity. We now survey the approximations we must make to (3.8) to get more convenient equations in real space.

We first need to justify the dropping of the non-local parts of $V_{nn'}(\mathbf{R}, \mathbf{R}')$, for this will convert our integrodifferential equations into differential equations. Fortunately, this is not difficult to justify for local potentials and slowly varying envelope functions; the details are given in appendix 2. The main point is that $V_{nn'}(\mathbf{R}, \mathbf{R}')$ can be replaced by $V_{nn'}(\mathbf{R})\Delta(\mathbf{R} - \mathbf{R}')$ where $V_{nn'}(\mathbf{R})$ is a new function distinguished notationally from $V_{nn'}(\mathbf{R}, \mathbf{R}')$ by the different number of arguments. The new envelope-function equations (3.8) become

$$-\frac{\hbar^2}{2m}\nabla^2 F_n(\mathbf{R}) + \sum_{n'} \frac{-i\hbar}{m} \mathbf{p}_{nn'} \cdot \nabla F_{n'}(\mathbf{R}) + \sum_{n'} H_{nn'}(\mathbf{R}) F_{n'}(\mathbf{R}) = E F_n(\mathbf{R}) \quad (5.1)$$

where $H_{nn'}(\mathbf{R}) = T_{nn'} + V_{nn'}(\mathbf{R})$. In a region far from any interface, $H_{nn'}(\mathbf{R})$ becomes $H_{nn'}$ for the appropriate bulk crystal (see appendix 3) and we see that if the U_n are chosen to diagonalize $H_{nn'}$ for that bulk crystal, then the envelope-function equations (5.1) in that region become the real-space equivalent of the familiar $k \cdot p$ equations (see e.g. Kane 1966). Note that far from any interface the non-local part of $H_{nn'}(\mathbf{R}, \mathbf{R}')$ vanishes without making any approximation (appendix 3). The non-local part of $H_{nn'}(\mathbf{R}, \mathbf{R}')$ is only non-zero in the region of an interface.

In going from one bulk region to another, $H_{nn'}(\mathbf{R})$ changes from one bulk value of $H_{nn'}$ to another. It turns out that for an abrupt planar interface or superlattice $H_{nn'}(\mathbf{R})$ can be resolved into two functions (appendix 4 and also Burt (1988a)),

$$H_{nn'}(\mathbf{R}) = H_{nn'}^{(b)}(\mathbf{R}) + H_{nn'}^{(i)}(\mathbf{R}) \quad (5.2)$$

where $H_{nn'}^{(b)}(\mathbf{R})$ depends only on the bulk values of $H_{nn'}$ and the position of the interfaces and $H_{nn'}^{(i)}(\mathbf{R})$ depends on the details of how the potential changes from one bulk value to the other. One can describe the dependence of $H_{nn'}^{(b)}$ on \mathbf{R} as an abrupt step from one value of $H_{nn'}$ to the next at the interface, but softened by Gibbs oscillations due to the restriction of the plane-wave expansion of $H_{nn'}^{(b)}$ to the Brillouin zone. For the case in which the potential changes abruptly at the interface, each term in the Fourier transform, $\tilde{H}_{nn'}^{(i)}(\mathbf{k})$, of $H_{nn'}^{(i)}$ has a large wavevector denominator while $\tilde{H}_{nn'}^{(b)}(\mathbf{k})$ has a small wavevector denominator. This makes $H_{nn'}^{(i)}$ small compared with $H_{nn'}^{(b)}$. This fact combined with the fact that $H_{nn'}^{(i)}$ is only non-zero in the neighbourhood of an interface means that $H_{nn'}^{(i)}$ has little influence on the energy eigenvalues. To a good approximation it can be neglected and we can replace $H_{nn'}(\mathbf{R})$ with $H_{nn'}^{(b)}$. One can go even further and drop the Gibbs oscillations, i.e. replace them with the abrupt step approximation on the grounds that this only alters the large wavevector components and the energy of states with slowly varying envelope functions will not be sensitive to such a change.

Of course, the general form of the resolution (5.2) is not restricted just to the case of an abrupt change in potential. Any deviations from an abrupt change due to charge rearrangement at an interface would contribute to $H_{nn'}^{(i)}$ and its contribution determined using (4.13).

6. Derivation of effective-mass equations

The derivation of effective-mass-type equations from envelope-function equations (conventional or otherwise) consists of eliminating small envelope functions in favour of the dominant ones. The derivation given here, however, will take into account the off-diagonal components of the zone-centre Hamiltonian $H_{nn'}$, and this will give it wider applicability than previous work; it will be able to handle the case of microstructures in which the component crystals cannot be considered to have the same zone-centre eigenfunctions. It will give valuable insight into why the effective-mass equation works so well.

One divides the slowly varying envelope functions into two groups S and R ; those in the S group, denoted by $F_s, F_{s'}$, etc., are dominant, while those in group R , denoted by $F_r, F_{r'}$, etc., are small and are eliminated approximately in favour of the members of group S . The envelope-function equations neglecting the non-local terms are

$$-\frac{\hbar^2}{2m}\nabla^2 F_n(\mathbf{R}) + \sum_{n'} \frac{-i\hbar}{m} \mathbf{p}_{nn'} \cdot \nabla F_{n'}(\mathbf{R}) + \sum_{n'} H_{nn'}(\mathbf{R}) F_{n'}(\mathbf{R}) = E F_n(\mathbf{R}). \tag{6.1}$$

For slowly varying envelope functions, using the equation for $n = r$, we have approximately

$$F_r = (E - H_{rr})^{-1} \sum_{s'} \left(\frac{-i\hbar}{m} \mathbf{p}_{rs'} \cdot \nabla F_{s'} + H_{rs'} F_{s'} \right). \tag{6.2}$$

If one substitutes for F_r in equation (6.1) with $n = s$ using equation (6.2), one obtains

$$\begin{aligned} & -\frac{\hbar^2}{2m} \sum_{s'} \nabla \cdot \left[\gamma_{ss'}^{(r)} \cdot \nabla F_{s'}(\mathbf{R}) \right] + \sum_{s'} \frac{-i\hbar}{m} \mathbf{p}_{ss'} \cdot \nabla F_{s'}(\mathbf{R}) \\ & + \sum_{s'} H_{ss'}^{(2)}(E, \mathbf{R}) F_{s'}(\mathbf{R}) + \sum_{s',r} \frac{-i\hbar}{m} \mathbf{p}_{sr} \cdot \nabla [(E - H_{rr})^{-1} H_{rs'}] F_{s'}(\mathbf{R}) \\ & + \sum_{s',r} \frac{-i\hbar}{m} (E - H_{rr})^{-1} (\mathbf{p}_{sr} H_{rs'} + H_{sr} \mathbf{p}_{rs'}) \cdot \nabla F_{s'}(\mathbf{R}) = E F_s(\mathbf{R}) \end{aligned} \tag{6.3}$$

where the contribution $\gamma_{ss'}^{(r)}$ from remote bands (i.e. group R) to the generalized effective-mass tensor is

$$\gamma_{ss'}^{(r)}(E, \mathbf{R}) = I \delta_{ss'} + \frac{2}{m} \sum_r \mathbf{p}_{sr} [E - H_{rr}(\mathbf{R})]^{-1} \mathbf{p}_{rs'} \tag{6.4}$$

and

$$H_{ss'}^{(2)}(E, \mathbf{R}) = H_{ss'}(\mathbf{R}) + \sum_r H_{sr}(\mathbf{R}) [E - H_{rr}(\mathbf{R})]^{-1} H_{rs'}(\mathbf{R}) \tag{6.5}$$

is the effective Hamiltonian folded down to second order onto the manifold s, s', s'' , etc.

To proceed further we note that the fourth term on the RHS of (6.3) namely

$$\sum_{s',r} \frac{-i\hbar}{m} p_{sr} \cdot \nabla [(E - H_{rr})^{-1} H_{rs'}] F_{s'}(\mathbf{R}) \quad (6.6)$$

is zero in the bulk because H_{rr} and $H_{rs'}$ are constant and the gradient of $(E - H_{rr})^{-1} H_{rs'}$ vanishes. Since this fourth term is only non-zero near interfaces and involves large energy denominators, it can usually be safely neglected. In a similar spirit we also neglect the interface contributions to the effective Hamiltonian $H_{nn'}(\mathbf{R})$, though here it is the wavevector denominators that are large.

Consider the simplest case of conduction band states in a type one zincblende microstructure. Here the energy is much closer to the conduction band edge than to any other band edge, so we expect $F_c(\mathbf{R})$, the envelope function corresponding to the conduction band-like basis state $U_c(\mathbf{R})$, to be dominant, all other envelope functions being small. For such an approximation to work, of course, one must choose U_c to be a reasonable approximation to the conduction band minimum wavefunction of each of the constituent materials. The momentum matrix $p_{ss'}$ has only one element p_{cc} which is zero because $U_c(\mathbf{R})$ corresponds to a band extremum. The fifth term on the RHS of equation (6.3), i.e.

$$\sum_{s',r} \frac{-i\hbar}{m} (E - H_{rr})^{-1} (p_{sr} H_{rs'} + H_{sr} p_{rs'}) \cdot \nabla F_{s'}(\mathbf{R}) \quad (6.7)$$

becomes

$$\sum_r \frac{-i\hbar}{m} (E - H_{rr})^{-1} (p_{cr} H_{rc} + H_{cr} p_{rc}) \cdot \nabla F_c(\mathbf{R}) \quad (6.8)$$

and in the bulk will lead to a term in the band structure linear in wavevector. Ignoring such small effects, which are due to the bulk parts of H_{rc} and H_{cr} , we conclude that this term is dominated by the interface parts of H_{rc} and H_{cr} , and in keeping with our earlier approximations we neglect it. Hence we arrive at our version of the effective-mass equation

$$-\frac{\hbar^2}{2m} \nabla \cdot \left(\frac{1}{m_c(E, \mathbf{R})} \nabla F_c(\mathbf{R}) \right) + H_{cc}^{(2)}(E, \mathbf{R}) F_c(\mathbf{R}) = E F_c(\mathbf{R}) \quad (6.9)$$

where the isotropy of the effective-mass tensor, $\gamma_{cc}(E, \mathbf{R})$, has been put in explicitly using the effective-mass function, $m_c(E, \mathbf{R})$, defined by

$$\frac{m}{m_c(E, \mathbf{R})} = 1 + \frac{2}{m} \sum_r p_{cr} [E - H_{rr}(\mathbf{R})]^{-1} p_{rc} \quad (6.10)$$

with p standing for the component of the momentum along one of the cubic axes.

The most important thing to note here is that $H_{cc}^{(2)}$ in the bulk represents the conduction band edge to second order, hence providing an accurate description of the change in conduction band edge as one moves from one material to another

without any need to introduce a change in the basis function. The basis function $U_c(\mathbf{R})$ is the same everywhere and the change in zone-centre eigenfunctions and the conduction band minimum comes about naturally because off-diagonal elements of the zone-centre Hamiltonian have been included.

As another example consider the application of equation (6.3) to the case of valence band states for the same type one zincblende microstructures, in particular, cases in which only the envelope functions for the light- and heavy-hole basis functions are large. Again we neglect terms that are appreciable only near interfaces and those that lead to contributions to the bulk band structure linear in wavevector. Noting also that $p_{ss'}$ is zero for these band-edge valence states, one obtains

$$-\frac{\hbar^2}{2m} \sum_{s'} \nabla \cdot [\gamma_{ss'} \cdot \nabla F_{s'}(\mathbf{R})] + H_{ss'}^{(2)}(E, \mathbf{R}) F_s(\mathbf{R}) = E F_s(\mathbf{R}) \quad (6.11)$$

where we have made use of the fact that $H_{ss'}^{(2)}(E, \mathbf{R})$ is diagonal from tetrahedral symmetry if interface terms are neglected. Again $H_{ss'}^{(2)}(E, \mathbf{R})$ approximates the profile of the relevant band edge to second order without any need to resort to changes in the basis function on crossing an interface. One might object to the above derivation on the grounds that it is unrealistic to assume that the envelope function for the spin split-off band is small because in many cases the spin split-off band is close in energy to the light- and heavy-hole bands. This is equivalent to saying that the contribution corresponding to $r = \text{spin split-off band}$ in the term

$$\sum_{s', r} \frac{-i\hbar}{m} (E - H_{rr})^{-1} (p_{sr} H_{rs'} + H_{sr} p_{rs'}) \cdot \nabla F_{s'}(\mathbf{R}) \quad (6.12)$$

is not negligible. However, p_{sr} vanishes by symmetry when r corresponds to the spin split-off band, showing that it is symmetry and not energy separation that keeps the spin split-off band envelope function small and makes the approximation of considering light- and heavy-hole bands alone a good one provided the spin split-off band is not so close as to make the light-hole band non-parabolic.

The procedure used in this section is equivalent to 'folding down' secular equations (see e.g. Heine and Cohen 1970) working to second order in small quantities. One could proceed to derive more accurate equations, as is needed in some cases to treat non-parabolicity for example, if desired, being mindful of the possibility that the 'non-local' terms one has jettisoned (section 5) may eventually become important.

7. Boundary conditions

7.1. Normal interface

The great strength of the author's approach has been the circumvention of the boundary-condition problem; the new envelope functions and all their derivatives from their very definition (3.1) are continuous everywhere even at abrupt interfaces. However, when one makes approximations to the exact equations (3.8) for these envelope functions and solves them in the coordinate representation, there is no longer any restriction on the range of wavevectors in the plane-wave expansion of the solutions. So the possibility of discontinuities in the approximate envelope functions and their derivatives arises.

Consider a planar abrupt interface and take this to occupy the plane $z = 0$. With translational symmetry in the (x, y) plane we can find a complete set of solutions in the form

$$F_n(\mathbf{R}) = (1/S^{1/2}) \exp(ik_{\parallel} \cdot \mathbf{r}) F_n(z)$$

where $\mathbf{r} = (x, y)$ and cyclic boundary conditions are taken over an area S , which in the abrupt step approximation obey the equation

$$\begin{aligned} -\frac{\hbar^2}{2m} \frac{d^2 F_n(z)}{dz^2} + \frac{-i\hbar}{m} \sum_{n'} p_{nn'}^z \frac{dF_{n'}(z)}{dz} + \frac{\hbar^2 k_{\parallel}^2}{2m} F_n(z) \\ + \sum_{n'} \frac{\hbar}{m} k_{\parallel} \cdot p_{nn'} F_{n'}(z) + \sum_{n'} H_{nn'} F_{n'}(z) = E F_n(z). \end{aligned} \quad (7.1)$$

If the envelope-function expansion (3.1) is truncated to a finite number of terms, as is necessarily the case in any numerical work, then one sees that the solutions F_n of (7.1) must be continuous at the interface $z = 0$ (it is hoped that the reader will not be confused by the use of the same symbols for the exact envelope functions and the solutions to (7.1) that are approximations to them); if F_n were to have a discontinuity there, then the term involving the second derivative would behave as the derivative of a delta function with no other term to neutralize it. Integrating over an infinitesimal interval about $z = 0$ (this is a valid procedure because we have derived (7.1) for all z including interfaces) one finds that

$$-i\hbar \frac{dF_n}{dz} + 2 \sum_{n'} p_{nn'}^z F_{n'} \quad (7.2)$$

is continuous at $z = 0$. Noting that the continuity of the $F_{n'}$ has already been established and that $p_{nn'}^z$ is a constant, one sees that the derivatives must also be continuous across the interface. Note that the presence of the second derivative is crucial; its omission can lead to entirely different boundary conditions, including the loss even of the continuity of some of the envelope functions (Taylor and Burt 1987).

So we have seen that the abrupt step approximation does not alter the continuity of the new envelope functions and their derivatives at an abrupt planar interface (provided, of course, all the solutions including out-of-zone solutions are included). However, if we use the effective-mass approximation as well as the abrupt step approximation, then an effective-mass-related discontinuity in the derivative arises. For example, take the effective-mass equation (6.9) for the conduction band quantum well. Carrying out the same integration used to derive (7.2) gives $(1/m_c) dF_c/dz$ continuous. Because m_c is discontinuous in the abrupt step approximation, then dF_c/dz is necessarily so as well. The author has already explained how this kink approximates a rapid change in the derivative in the region of the interface in the true envelope function by considering the effective-mass equation derived with the Gibbs oscillations retained (Burt 1988b); the reader will appreciate that the omission of the off-diagonal elements of the Hamiltonian in this reference does not alter the conclusions. An example of the rapid change in the derivative of the envelope function due to a change in effective mass is given in the section on numerical results (next section).

The existence of a kink in the effective-mass envelope function in the abrupt step approximation comes about because the approximation used for the small envelope functions neglects their second derivatives (one also loses the out-of-zone solutions) and hence the proof of their continuity and that of the derivative of the dominant envelope function given above fails; the lack of continuity in the small envelope functions $F_{n'}$, means that the continuity of

$$-i\hbar \frac{dF_c}{dz} + 2 \sum_{n'} p_{cn'}^z F_{n'} \tag{7.3}$$

requires a discontinuity in dF_c/dz .

7.2. Current conservation

It is appropriate to demonstrate that these boundary conditions are consistent with conservation of probability current. Using the standard textbook technique on equation (7.1) (i.e. multiplying by F_n^* , taking the imaginary part and rearranging) one finds that

$$\frac{1}{m} \text{Re} \sum_n \left[F_n^* \left(-i\hbar \frac{dF_n}{dz} \right) + \sum_{n'} F_n^* p_{nn'}^z F_{n'} \right] \tag{7.4}$$

is independent of z . Using the technique described by the author (Burt 1987) one can show, with an appropriate choice of the three-dimensional Brillouin zone (see appendix 1), that (7.4) is the z component of the probability current crossing the (x, y) plane, provided that the envelope functions are sufficiently slowly varying for one to be able to restrict the plane-wave expansion to the inner half of the Brillouin zone, i.e. bilinear products such as $F_n^* F_{n'}$, have a plane-wave expansion restricted to the Brillouin zone. This is a mild restriction. One finds that the conservation of (7.4) is entirely in keeping with the boundary condition, i.e. continuity of (7.2) derived from (7.1) as one would expect because they are based on the same approximations. Because both the solutions of (7.1) and their derivatives are continuous across the interface, the current (7.4) is automatically conserved. One can readily confirm, following Burt (1988b), that, to first order in small quantities, the effective-mass approximation for the current (7.4) is consistent with the boundary condition (7.3). This result even extends to the non-parabolic case, in which the non-parabolicity can be described by an energy-dependent effective mass (section 10).

7.3. Interface with an impenetrable barrier

It is often assumed that the correct boundary condition at an interface with an impenetrable barrier is that the envelope function should be zero there. (It is assumed that the wavefunction is being approximated by just one term of the envelope-function expansion, i.e. $\Psi(z) = F(z)U(z)$.) However, this is not necessarily so if the periodic function $U(z)$ has a node at the interface. (I was fortunate to become aware of this point some years ago when Professor Heine raised it with a seminar speaker at the Cavendish.) The problem of justifying this boundary condition can be solved by inverting the envelope-function expansion to give the envelope function in terms of the wavefunction. This can be done using elementary manipulations: start with (4.6)

with unitary U , substitute for $\tilde{\Psi}_G(k)$ using (4.2a), then use (4.3a) to eliminate the sum over G , and finally take the Fourier transform with respect to k . One obtains

$$F_n(\mathbf{R}) = \int \Delta(\mathbf{R} - \mathbf{R}') U_n^*(\mathbf{R}') \Psi(\mathbf{R}') d^3R' \quad (7.5)$$

where $\Delta(\mathbf{R} - \mathbf{R}')$ is defined by (3.10). $\Delta(\mathbf{R} - \mathbf{R}')$ is only appreciably different from zero near $\mathbf{R} = \mathbf{R}'$. It follows that $F_n(\mathbf{R})$ will fall rapidly to zero as one moves into any region in which Ψ is zero, regardless of the detailed properties of the U_n . Putting the $F_n(\mathbf{R})$ to zero at the interface is a good approximation if they are slowly varying.

8. Numerical results: the new envelope-function method

The model microstructure used here, to illustrate the algebraic results derived in earlier sections, is a superlattice constructed from one-dimensional 'Mathieu' crystals, so called because with their sinusoidal potential the corresponding Schrödinger equation becomes the Mathieu equation. The band structure of the crystal forming the well is shown in figure 1 and the potential for the superlattice in figure 2. The wavefunction for one of the zero-wavevector eigenstates of the superlattice, computed by diagonalizing the plane-wave Hamiltonian, is given in figure 3. One can see the pronounced and rapid change in the derivative of the envelope at the interfaces related to the effective-mass change.

To apply the author's envelope-function method to the computation of this eigenstate, the wavefunction $\Psi(z)$ is expanded as

$$\Psi(z) = \sum_n F_n(z) U_n(z) \quad (8.1)$$

where the $U_n(z)$ have been chosen as the zero-wavevector eigenstates of the virtual crystal. The envelope functions $F_n(z)$ have plane-wave expansions limited to the range $-\pi/a < k \leq \pi/a$. The periodicity of the superlattice restricts the number of terms in the plane-wave expansion to the number of unit cells in the superlattice, so that the exact envelope functions

$$-\frac{\hbar^2}{2m} \frac{d^2 F_n}{dz^2} + \sum_{n'} \frac{-i\hbar}{m} p_{nn'} \frac{dF_{n'}}{dz} + \sum_{n'} \int H_{nn'}(z, z') F_{n'}(z') dz' = E F_n \quad (8.2)$$

can be solved to machine accuracy in k -space as a matrix eigenvalue problem. The two dominant envelope functions are given in figure 4. There is a smoothed-out kink in the dominant envelope function at the interfaces. The derivative changes by a factor of about 2.5 within a distance of a lattice spacing or so compared with an effective-mass ratio of 2.25 at that energy.

To derive the corresponding effective-mass equation, as illustrated in a previous section, one must start from the local approximation to the exact envelope-function equations, namely from

$$-\frac{\hbar^2}{2m} \frac{d^2 F_n}{dz^2} + \sum_{n'} \frac{-i\hbar}{m} p_{nn'} \frac{dF_{n'}}{dz} + \sum_{n'} H_{nn'}(z) F_{n'}(z) = E F_n. \quad (8.3)$$

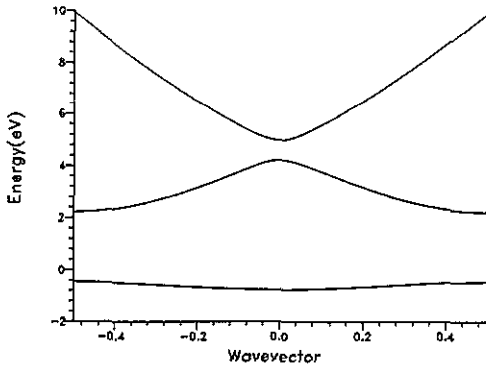


Figure 1. The band structure of the one-dimensional crystal that constitutes the well in the superlattice results presented in this paper. The potential is $V_s \cos(2\pi z/a)$ with $V_s = 2.7512$ eV and the lattice period $a = 5.86$ Å. The wavevector is given in units of $2\pi/a$. In a crude way the band structure mimics, in part, that of a direct-gap III-V semiconductor and it will be convenient to use this analogy further for ease of notation and refer to the lower two bands as valence bands and the third band (the highest shown) as the conduction band. The lattice constant and band gap (0.750 eV) at zero wavevector have been chosen to correspond to the ternary semiconductor alloy GaInAs lattice-matched to InP. Even though the band gap (0.750 eV) is relatively small compared to the mid-gap energy (4.612 eV), this is not a two-band model crystal as can be seen from the large zone-edge gap between the valence bands. The results (figure 3 onwards) of the superlattice calculations are for the lowest conduction band eigenstate.

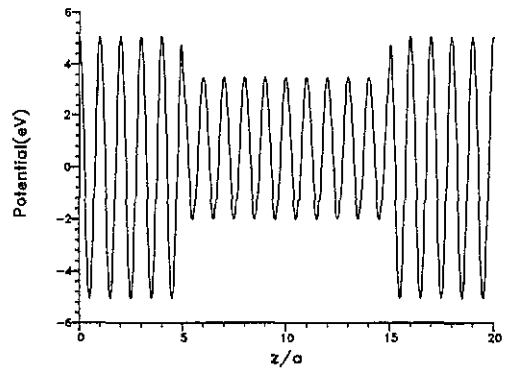


Figure 2. One period of the superlattice potential from the middle of one barrier layer to the middle of the next. The well crystal that occupies the 10 lattice periods $5a < z < 15a$ has potential $V_0 + V_s \cos(2\pi z/a)$ with $V_0 = 0.7460$ eV and $V_s = 2.7512$ eV. The barrier crystal, which occupies the 10 lattice periods $0 < z < 5a$ and $15a < z < 20a$, has a potential of the same form with $V_0 = 0.0$ eV and $V_s = 5.0596$ eV. The band gap of the barrier crystal is 2.060 eV and the conduction band-edge discontinuity, i.e. the well depth, is 0.2301 eV. The transition from barrier potential to well potential is not abrupt because the superlattice potential is represented here by its plane-wave expansion as used in the calculations. The plane-wave expansion is restricted to plane waves with wavevectors (in units of $2\pi/a$) $k + G$, with $|G| \leq 5$ and with $-\frac{1}{2} \leq k < +\frac{1}{2}$, i.e. k is inside the Brillouin zone of the underlying Bravais lattice.

The accuracy of this approximation has been tested for the case depicted in figure 4 and found to produce errors in the eigenvalue of less than 0.1 meV and plots of the envelope functions indistinguishable from those shown. The reasons for this small error are (i) that the plane-wave components \tilde{F}_k are small for $|k| > \pi/(2a)$, where a is the lattice period, i.e. for wavevectors in the 'outer' half of the Brillouin zone and (ii) the expectation value of the non-local parts of the Hamiltonian (Burt 1988a) evaluated in k -space involves terms containing the product $\tilde{F}_k^* \tilde{F}_{k'}$, with a wavevector denominator that is large except when k and k' are near opposite edges of the Brillouin zone, in which case the product $\tilde{F}_k^* \tilde{F}_{k'}$ is extremely small.

In figures 4 to 8 various aspects of the effective Hamiltonian function $H_{nn'}(z)$ are illustrated. In general, the $H_{nn'}(z)$ can be decomposed into both 'bulk' and 'interface' terms,

$$H_{nn'}(z) = H_{nn'}^{(b)}(z) + H_{nn'}^{(i)}(z). \tag{8.4}$$

For this Mathieu superlattice, with barrier and well widths L_B and L_W respectively,

$$H_{nn'}^{(b)}(z) = \bar{H}_{nn'} + (H_{nn'}^B - H_{nn'}^W) \sum_{k \neq 0} \frac{\sin(kL_B/2)}{k(L_B + L_W)/2} \exp(ikz) \tag{8.5}$$

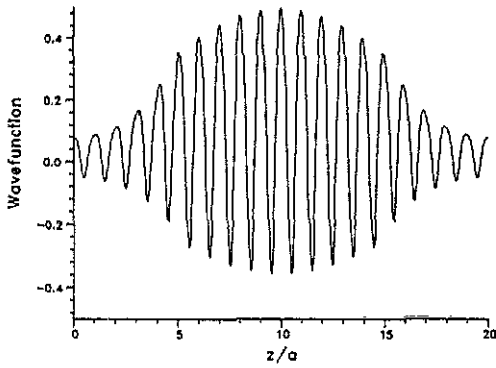


Figure 3. The wavefunction for the lowest-lying conduction band state of the superlattice described in figure 2. The energy of this state is 5.7961 eV. The conduction band edge for the well is at 5.7333 eV so the size quantization energy is 62.8 meV. Note how the rate of change of the peak heights in going from peak to peak (the slope of the envelope) changes rapidly in the region of the interfaces (at $z = 5a$ and $z = 15a$).

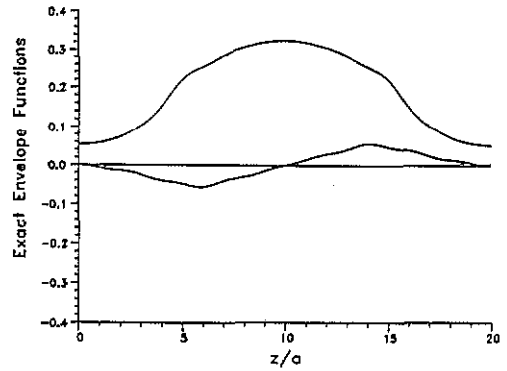


Figure 4. The new (as opposed to the conventional) envelope functions for the wavefunction shown in figure 3 defined with respect to periodic functions that are the zero-wavevector eigenstates of the virtual crystal, i.e. the crystal with potential $\bar{V}_0 + \bar{V}_s \cos(2\pi z/a)$ where $\bar{V}_0 = (V_0^{\text{well}} + V_0^{\text{barrier}})/2$ and $\bar{V}_s = (V_s^{\text{well}} + V_s^{\text{barrier}})/2$. The symmetric positive function is the envelope function (which is real) for the conduction band. The antisymmetric function is the imaginary part of the envelope function (the real part is zero) for the upper valence band. The other envelope functions are too small to show clearly on this scale. Although the constituent crystals are not simple two-band model crystals, their potentials being too strong for the band-edge wavefunctions to be composed merely of two plane waves, the wavefunction in this case can be represented quite well by a two-band envelope-function expansion. The relatively rapid change in the derivative of the conduction band envelope function at the interfaces is related to the change in conduction band effective mass.

where $\bar{H}_{nn'}$ is the molar average $(L_B H_{nn'}^B + L_W H_{nn'}^W)/(L_B + L_W)$, $H_{nn'}^B$, and $H_{nn'}^W$ are the bulk values for $H_{nn'}$, i.e. the zone-centre Hamiltonian in the barrier and well crystals, and

$$H_{nn'}^{(i)}(z) = \sum_{k \neq 0} 2 \sin(kL_B/2) \Delta \tilde{V}_{nn'}(k) \exp(ikz) \quad (8.6)$$

with

$$\Delta \tilde{V}_{nn'}(k) = \sum_{G, G'} U_{nG}^* \left(\sum_{G'' \neq G-G'} \frac{\tilde{V}_{G''}^B - \tilde{V}_{G''}^W}{(k + G - G' - G'')(L_B + L_W)} \right) U_{n'G'}. \quad (8.7)$$

The $H_{nn}(z)$ corresponding to the dominant envelope function of figure 4 is shown in figure 5 along with the abrupt step approximation to this function determined by the bulk values, i.e. the piecewise-constant function equal to H_{nn}^B for $0 < z < 5a$ and $15a < z < 20a$ and equal to H_{nn}^W for $5a < z < 10a$. The plots of $H_{nn}^{(b)}(z)$

and $H_{nn}^{(i)}(z)$ are given in figures 6 and 7 for the dominant envelope function. In figure 6 the abrupt step approximation is also given as in figure 5, emphasizing that $H_{nn}^{(b)}(z)$ differs from that piecewise-continuous function only by the softening effect of the Gibbs oscillations. The interface term $H_{nn}^{(i)}(z)$ is plotted in figure 7 on a much magnified scale; it is small because the wavevector denominators in $\Delta\tilde{V}_{nn'}(k)$ are all large, i.e. greater than π/a . Omission of these interface terms only leads one to overestimate the eigenvalue by 1.1 meV.

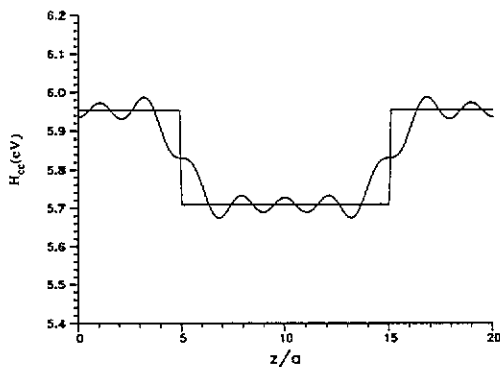


Figure 5. The function $H_{nn}(z)$ for the case of $n =$ conduction band, defined with respect to the same virtual-crystal periodic basis functions as the envelope functions (see caption to figure 4). When the interband coupling due to the superlattice potential (i.e. the zone-centre Hamiltonian as distinct from the conventional ' $k \cdot p$ ' coupling) is sufficiently weak, then this function is the effective potential for conduction band electrons. The square well gives the profile of the bulk value.

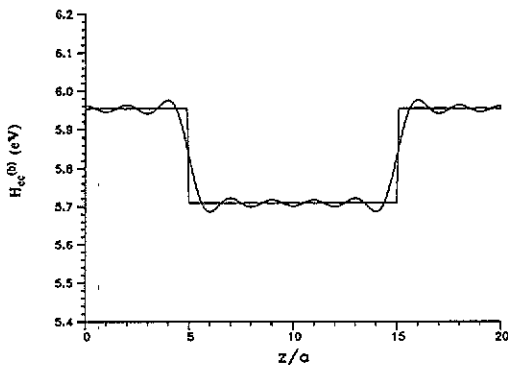


Figure 6. The 'bulk' part of $H_{cc}(z)$, i.e. that part that is independent of the way in which the superlattice potential changes over at the interfaces from one bulk crystal potential to the next. It depends only on the bulk values of H_{cc} and the structure factor for the superlattice. Except for the softening of discontinuities due to Gibbs oscillations, the function is just the piecewise-continuous function (also shown) that takes the corresponding bulk value for H_{cc} in well and barrier regions.

In figure 8 the $H_{nn'}(z)$ coupling the two envelope functions shown in figure 4 is plotted. Such off-diagonal terms are implicitly ignored in conventional envelope-function methods. This particular term would vanish in bulk crystals because the corresponding periodic basis functions are of opposite symmetry and, indeed, $H_{nn}^{(b)}(z)$ defined above vanishes identically. Only the interface term $H_{nn'}^{(i)}(z)$ remains and one sees this clearly in figure 8 with $H_{nn'}(z)$ only being appreciable near the interfaces where the reflection symmetry about the centre of the unit cell in the bulk is broken.

The differences in the functional form between the two 'interface' terms in figures 7 and 8 can be explained using the symmetry properties of the basis functions ($U_{n(\pm G)} = \pm U_{nG}$) and the Taylor expansion (using the binomial expansion of $(k + G - G' - G'')^{-1}$ in k of the summand of $\Delta\tilde{V}_{nn'}(k)$ in (8.7)). Such an expansion is justified because k is always inside the first Brillouin zone while $G - G' - G''$ is a non-zero reciprocal-lattice vector. The first term is just a constant and the coefficients in the plane-wave expansion vary as $\sin(kL_B/2)$, which corresponds to the differential of the band-edge profile. Since the latter in our case is a square wave, we expect to see delta functions, softened of course by Gibbs oscillations, of alternating sign at the successive interfaces. This is indeed a good description of

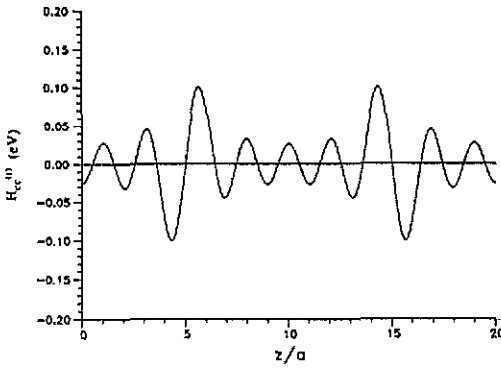


Figure 7. The 'interface' part of $H_{cc}(z)$, i.e. that part that depends on the way, abruptly in the case treated in this paper, in which the superlattice potential changes over at the interfaces from one bulk crystal potential to the next. It is the difference between the full function $H_{cc}(z)$ shown in figure 5 and the 'bulk' part of $H_{cc}(z)$ shown in figure 6. This 'interface' part is implicitly ignored in the conventional envelope-function method.

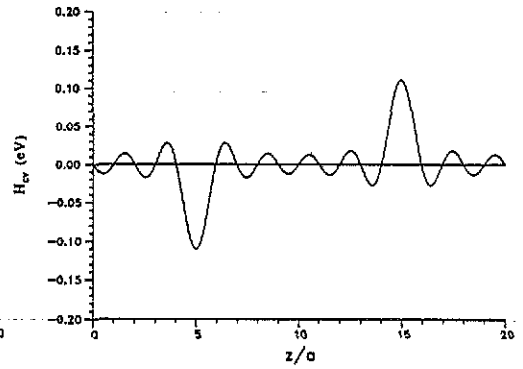


Figure 8. The imaginary part of the function $H_{nn'}(z)$ (the real part is zero) for the case of $n =$ conduction band, $n' =$ upper valence band representing the interband coupling due to the superlattice potential. For bulk crystals this function would be zero by symmetry and the function is only appreciable where the bulk symmetry (a reflection symmetry in this case) is broken, i.e. near an interface. This function is implicitly ignored in the conventional envelope-function method.

what one sees in figure 8. The reason why the first term in the binomial expansion of $(k + G - G' - G'')^{-1}$ works reasonably well is that the term proportional to k vanishes because the periodic basis functions are of opposite symmetry and the next non-vanishing term varies as $(k/G)^2$. In figure 7 the indices n and n' are identical and the basis functions are necessarily of the same symmetry. The binomial expansion leads to a Taylor series for $\Delta \tilde{V}_{nn'}(k)$ that only contains odd powers in k . Taking just the first term, the plane-wave expansion coefficients vary as $k \sin(kL_B/2)$, which explains why the z variation of figure 7 looks like the differential of that in figure 8.

Figures 9, 10 and 11 are concerned with the effective-mass approximation to the state depicted in figures 3 and 4. Neglecting interface terms the generalized effective-mass equation for the eigenstate in the region of the s th band edge is

$$-(\hbar^2/2)(d/dz)\{[1/m_s(z, E)] dF_s/dz\} + H_{ss}^{(2)}F_s(z) = EF_s \quad (8.8)$$

where

$$H_{ss}^{(2)} = H_{ss}(z) + \sum_{r \neq s} \frac{H_{sr}(z)H_{rs}(z)}{E - H_{rr}(z)} \quad (8.9)$$

is the effective band-edge profile and $m/m_s(z, E)$ is the dimensionless effective-mass function

$$\frac{m}{m_s(z, E)} = 1 + \frac{2}{m} \sum_{r \neq s} \frac{|p_{sr}|^2}{E - H_{rr}(z)}. \quad (8.10)$$

The effective-mass function is plotted in figure 9. The piecewise-continuous profile of the bulk effective-mass values is also shown.

In figure 10 the effective-mass approximation (8.8) to the dominant envelope function shown in figure 4 is plotted. The rapid change of the derivative at the interfaces caused by the effective-mass change is very marked and highlighted by plotting the derivative directly in figure 11. It should be emphasized that the change in derivative is not entirely dominated by the change in the bulk effective masses because the effective-mass function only takes on the bulk values asymptotically at large distances from the interfaces; the derivation of the connection rule is also only approximate.

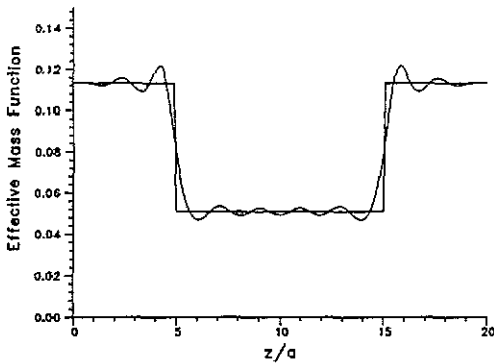


Figure 9. The conduction band effective-mass function calculated using the 'bulk' parts of the quantities $H_{nn}(z)$. This function is a smoothed-out version of the piecewise-constant variations in effective mass through the superlattice (m_c/m is 0.05091 in the well and 0.1132 in the barrier). It depends on the energy and is evaluated at the energy of the state shown in figure 3, i.e. 5.7932 eV.

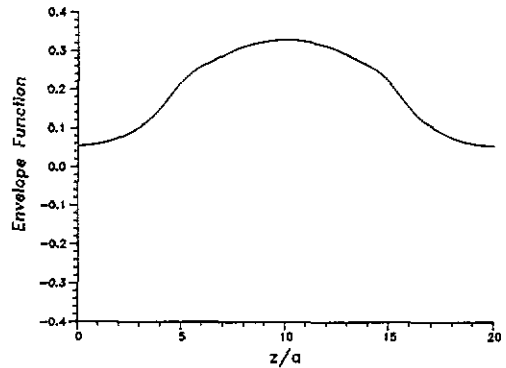


Figure 10. The conduction band envelope function F computed using the new (as opposed to the conventional) effective-mass equation (8.8) with the 'interface' terms in the $H_{nn}(z)$ omitted; it is an approximation to the dominant envelope function in figure 4. The effective-mass function used is, therefore, that given in figure 9. The rapid change in the derivative at the interfaces is approximated by a kink in conventional formulations of the effective-mass method.

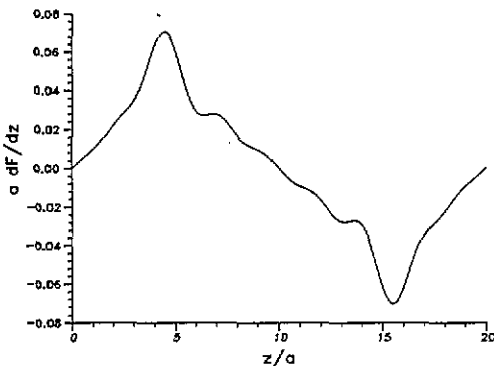


Figure 11. The derivative of the envelope function shown in figure 10 multiplied by the lattice constant to give a dimensionless quantity. There is a rapid change in the derivative at the interfaces due to the change in effective mass; the derivative falls by a factor of 2.5 between $z = 4.5a$ and $z = 5.5a$ compared with the ratio of the bulk effective masses of 2.25.

The rapid change in derivative at the interface displayed in figures 10 and 11 demonstrates that the kink introduced into the wavefunction envelope by including the effective mass in the conventional derivative boundary condition is not artificial,

but does approximate the actual behaviour of the envelope near the interface. The kink is necessary to get the energy right. Using an energy-dependent effective mass in the standard 'particle in a box' (Kronig-Penney) model, with the effective mass included in the derivative boundary condition, gives a size quantization energy of 60.1 meV (using effective masses 0.050 91 and 0.1132) compared to the exact value of 62.8 meV. If, however, the effective mass is omitted from the derivative boundary condition, and the kink thereby removed, then the size quantization energy rises to 94.3 meV, an overestimate of about 50%!!

At one time (Burt 1989) the author had reservations concerning the inclusion of the effective mass in the derivative boundary condition, beyond those mentioned in section 2, because the ground-state energy had an unexpected behaviour: when the barrier effective mass decreases, all other parameters being held constant, the size quantization energy is predicted to increase. At first sight, one might expect the size quantization energy to fall because the localization has decreased and this decreases the kinetic energy, via the uncertainty principle, and indeed this is the prediction of the 'particle in a box' model when the effective mass is omitted from the derivative boundary condition. However, calculations on the Mathieu superlattice used here have supported the inclusion of the effective mass in the derivative boundary condition. For example, the barrier potential form factor, V_s , was changed to 3.8801 eV to give a lower gap and lower effective-mass barrier crystal, and the constant potential, V_0 , in the well was decreased to 0.2299 eV to keep the band offset at 0.2301 eV. So from the phenomenological point of view nothing had been changed save the barrier effective mass had been decreased. The size quantization energy turned out to be 67.6 meV compared with the earlier value of 62.8 meV (see caption to figure 3), a rise of 4.8 meV. The prediction of the 'particle in a box' (Kronig-Penney) model with the effective mass included in the derivative boundary condition is a size quantization energy of 67.2 meV (using the appropriate energy-dependent effective masses of 0.051 08 and 0.072 68) compared with the previous value (preceding paragraph) of 60.1 meV. It approximates the exact results very well. It appears that the fall in energy due to decreasing localization must be more than counteracted by the potential energy gained due to extra penetration of the barrier.

Finally it should be mentioned that the results presented here are not inconsistent with those of Cunningham *et al* (1988). These authors solved the Schrödinger equation for a model one-dimensional quantum-well system algebraically and compared the results with the 'particle in a box' model predictions. Taking the non-parabolicity into account, they found that the omission of the effective mass in the derivative boundary condition gave the most accurate result; inclusion of the band-edge effective mass in the derivative boundary condition gave significant error; they did not consider the use of an energy-dependent effective mass in the derivative boundary condition. However, the energy-dependent effective mass in their model is almost the same in the well and barrier. (It is essentially a two-band model with no valence band offset so that, for the conduction band states that they treat, m_c is proportional to $E - E_v$, which is the same in both the well and the barrier.) Hence inclusion of the energy-dependent effective mass in the derivative boundary condition, the result the author has derived, would produce almost the same result as continuity of the derivative of the envelope function in their special model and would be a good approximation to their exact results.

9. Wannier–Slater envelope functions

9.1. Exact equations

The difference between Wannier–Slater (WS) envelope functions and the Luttinger–Kohn (LK) envelope functions is most readily seen in their Fourier transforms. To construct the wavefunction from these Fourier transforms, one uses Bloch functions for zero wavevector in the Luttinger–Kohn case but wavevector-dependent Bloch functions in the Wannier–Slater case. So, while the wavefunction $\Psi(\mathbf{R})$ is related to the Fourier transform $\tilde{F}_n(\mathbf{k})$ of the Luttinger–Kohn envelope functions by

$$\Psi(\mathbf{R}) = \sum_{\mathbf{k}} \sum_n \tilde{F}_n(\mathbf{k}) U_n(\mathbf{R}) \exp(i\mathbf{k} \cdot \mathbf{R}) \quad (9.1)$$

where the $U_n(\mathbf{R})$ are the zero-wavevector eigenstates of a periodic reference Hamiltonian H_0 , the corresponding expression for the Wannier–Slater envelope functions $F_n^{\text{WS}}(\mathbf{R})$ is

$$\Psi(\mathbf{R}) = \sum_{\mathbf{k}} \sum_n \tilde{F}_n^{\text{WS}}(\mathbf{k}) U_{n\mathbf{k}}(\mathbf{R}) \exp(i\mathbf{k} \cdot \mathbf{R}) \quad (9.2)$$

where $U_{n\mathbf{k}}(\mathbf{R})$ is the periodic Bloch function at wavevector \mathbf{k} for H_0 . Since there is still a unique relation between the wavefunction and the envelope function in the Wannier–Slater case, it is still possible to derive exact envelope-function equations, as we will now show using a method analogous to that used in section 4.

We start with the Schrödinger equation

$$(H_0 + \Delta V)|\Psi\rangle = E|\Psi\rangle \quad (9.3)$$

where ΔV is the difference between the potential experienced by the electron in the microstructure and the periodic reference potential in H_0 , and use the Bloch states $|n\mathbf{k}\rangle$ defined by

$$H_0|n\mathbf{k}\rangle = E_n^{(0)}(\mathbf{k})|n\mathbf{k}\rangle \quad (9.4)$$

as the basis for expanding the state vector. We obtain

$$\sum_{n',\mathbf{k}'} [E_n^{(0)}(\mathbf{k})\delta_{\mathbf{k}\mathbf{k}'}\delta_{nn'} + \langle n\mathbf{k}|\Delta V|n'\mathbf{k}'\rangle]\langle n'\mathbf{k}'|\Psi\rangle = E\langle n\mathbf{k}|\Psi\rangle. \quad (9.5)$$

From the definition of $\tilde{F}_n^{\text{WS}}(\mathbf{k})$ we find that

$$F_n^{\text{WS}}(\mathbf{R}) = \frac{1}{\Omega^{1/2}} \sum_{\mathbf{k}} \langle n\mathbf{k}|\Psi\rangle \exp(i\mathbf{k} \cdot \mathbf{R}) \quad (9.6)$$

where Ω is the normalizing volume. Transforming into real space and using the definition of $F_n^{\text{WS}}(\mathbf{R})$ one obtains the exact equations

$$E_n^{(0)}(-i\nabla)F_n^{\text{WS}}(\mathbf{R}) + \sum_{n'} \int \Delta V_{n\mathbf{n}'}(\mathbf{R}, \mathbf{R}') F_{n'}^{\text{WS}}(\mathbf{R}') d^3\mathbf{R}' = E F_n^{\text{WS}}(\mathbf{R}) \quad (9.7)$$

where

$$\Delta V_{nn'}(\mathbf{R}, \mathbf{R}') = \frac{1}{\Omega} \sum_{\mathbf{k}, \mathbf{k}'} \exp(i\mathbf{k} \cdot \mathbf{R}) \langle n\mathbf{k} | \Delta V | n'\mathbf{k}' \rangle \exp(-i\mathbf{k}' \cdot \mathbf{R}'). \quad (9.8)$$

The Wannier–Slater envelope functions $F_n^{\text{WS}}(\mathbf{R})$ used here are closely related to the coefficients in the expansion of the wavefunction in terms of the Wannier functions defined with respect to the reference crystal. The Wannier basis vector $|n\mathbf{R}_m\rangle$, for the band n and the lattice site \mathbf{R}_m , is related to the Bloch vector $|n\mathbf{k}\rangle$ by

$$|n\mathbf{R}_m\rangle = \frac{1}{N^{1/2}} \sum_{\mathbf{k}} \exp(-i\mathbf{k} \cdot \mathbf{R}_m) |n\mathbf{k}\rangle \quad (9.9)$$

where N is the number of unit cells in the normalizing volume. Elementary manipulations give

$$F_n^{\text{WS}}(\mathbf{R}) = (V_c)^{1/2} \sum_{\mathbf{R}_m} \Delta(\mathbf{R} - \mathbf{R}_m) \langle n\mathbf{R}_m | \Psi \rangle \quad (9.10)$$

where V_c is the volume of a unit cell in the reference crystal and Δ is the delta function with limited plane-wave expansion introduced in section 3. In (9.10), $\langle n\mathbf{R}_m | \Psi \rangle$ is the amplitude for finding the electron in the Wannier orbital $\langle \mathbf{R} | n\mathbf{R}_m \rangle$ and is often referred to as the envelope function, the continuation of the function between successive lattice points being given scant attention. The above relation between $F_n^{\text{WS}}(\mathbf{R})$ and $\langle n\mathbf{R}_m | \Psi \rangle$ gives the continuation explicitly.

9.2. Effective-mass equation

Subsequent to the author's derivation of a new effective-mass equation including the order of the differential operators in the kinetic energy term based on the LK envelope-function expansion, a derivation has been given by Young (1989) based on the WS envelope-function expansion. Young's discussion is restricted to the case in which ΔV is of the form LU where U is a periodic potential and L is a small and slowly varying function on the scale of the lattice constant, i.e. restricted to a system composed of alloys of two components. The discussion is also restricted to first order in L , i.e. to first order in ΔV . The purpose here is to show that the approach can be modified to include abrupt changes in composition profile, but the inclusion of large changes in composition is much more difficult than in the LK approach.

Because the derivation is complex, it will be presented here explicitly only for a two-band model of a one-dimensional microstructure. We start from the k -space version of (9.7) and for definiteness we consider conduction band states so that $\tilde{F}_c^{\text{WS}}(k)$, the conduction band envelope function in k -space, is dominant. To first order in ΔV , the term involving the valence band envelope function in k -space, $\tilde{F}_v^{\text{WS}}(k)$, in the exact equation for $\tilde{F}_c^{\text{WS}}(k)$ (i.e. the equation for $n = c$) can be neglected, so that the equation for $\tilde{F}_c^{\text{WS}}(k)$ is simply

$$E_c^{(0)}(k) \tilde{F}_c^{\text{WS}}(k) + \sum_{k'} \langle ck | \Delta V | ck' \rangle \tilde{F}_c^{\text{WS}}(k') = E \tilde{F}_c^{\text{WS}}(k). \quad (9.11)$$

Now we are interested in slowly varying solutions in real space. These can be present even if the difference potential ΔV is changing abruptly from one cell to the next, in which case, only the values of $E_c^{(0)}(k)$ and $\langle ck|\Delta V|ck'\rangle$ at small k and k' will substantially determine the eigenvalue. We can, on this understanding, still follow Young and expand $\langle ck|\Delta V|ck'\rangle$ to second order in the wavevectors, though we will take a less formal approach and use explicit expansions based on the two-band model. The Bloch basis states for the conduction band are expressed in terms of the $k = 0$ eigenstates, $U_n(z)$, of the reference Hamiltonian

$$\langle z|ck\rangle = \frac{e^{ikz}}{L^{1/2}} \sum_n A_n(k) U_n(z) \quad (9.12)$$

where the sum is over the conduction and valence bands and to second order in k

$$A_c(k) = 1 - \frac{1}{2}(\hbar|p_{vc}|/mE_g)^2 k^2 \quad (9.13)$$

$$A_v(k) = (\hbar p_{vc}/mE_g)k. \quad (9.14)$$

To evaluate $\langle ck|\Delta V|ck'\rangle$ we need the plane-wave expansion

$$\Delta V(z) = \sum_{k,G} \Delta \tilde{V}_G(k) \exp[i(k+G)z]. \quad (9.15)$$

For $k - k'$ inside the Brillouin zone

$$\begin{aligned} \langle ck|\Delta V|ck'\rangle &= \sum_{n,n'} A_n(k)^* A_{n'}(k') \sum_{G,G'} U_{nG}^* \Delta \tilde{V}_{G-G'}(k-k') U_{n'G'} \\ &= \sum_{n,n'} A_n(k)^* A_{n'}(k') \Delta \tilde{V}_{nn'}(k-k'). \end{aligned} \quad (9.16)$$

There are four terms on the RHS. The easiest to deal with is the one corresponding to $n = n' = v$. This is

$$(\hbar p_{vc}/mE_g)^* (\hbar p_{vc}/mE_g) k \Delta \tilde{V}_{vv}(k-k') k' \quad (9.17)$$

and, when equation (9.11) is transformed into real space, gives a term

$$- \left(\frac{\hbar|p_{vc}|}{mE_g} \right)^2 \frac{d}{dz} \left(\Delta V_{vv}(z) \frac{dF_c^{WS}}{dz} \right). \quad (9.18)$$

The $n = n' = c$ term, which is more complicated, is

$$\left[1 - \frac{1}{2}(\hbar|p_{vc}|/mE_g)^2 (k^2 + k'^2) \right] \Delta \tilde{V}_{cc}(k-k') \quad (9.19)$$

which, on using $k^2 + k'^2 = 2kk' + (k-k')^2$, leads to a contribution

$$\begin{aligned} \Delta V_{cc}(z) F_c^{WS} + (\hbar|p_{vc}|/mE_g)^2 \left[(d/dz)(\Delta V_{cc}(z) dF_c^{WS}/dz) \right. \\ \left. + \frac{1}{2}(d^2 \Delta V_{cc}(z)/dz^2) F_c^{WS} \right] \end{aligned} \quad (9.20)$$

to the real-space equation. The last term, involving the second derivative of $\Delta V_{cc}(z)$, corresponds to the term involving the second derivative of L , the composition profile in Young's equation (26). For abrupt interfaces it will only be appreciable near the interfaces and one would usually neglect it (but see comments in the next paragraph). The other two terms corresponding to $n = c, n' = v$ and $n = v, n' = c$ are best taken together and give

$$(\hbar/mE_g)[\Delta \tilde{V}_{cv}(k - k')p_{vc}k' + kp_{cv}\Delta \tilde{V}_{vc}(k - k')] \quad (9.21)$$

which gives a contribution

$$\begin{aligned} (\hbar/mE_g)[\Delta V_{cv}(z)p_{vc} + p_{cv}\Delta V_{vc}(z)]dF_c^{WS}/dz \\ + (\hbar/mE_g)p_{cv}(d\Delta V_{vc}(z)/dz)F_c^{WS} \end{aligned} \quad (9.22)$$

to the real-space equation. The term involving the derivative of the envelope function leads to a linear k term in the band structure in the bulk, so, at best, it is only non-zero near interfaces. In fact it vanishes identically: the $U_n(z)$ are real within overall phase factors and elementary manipulations show that the expression within the square brackets vanishes. The other term, involving $d\Delta V_{vc}(z)/dz$, is only non-zero near interfaces and we neglect it (but, again, see comments in the next paragraph); it corresponds to the term involving the first derivative of L , the composition profile in Young's equation (26). Finally, one is now in a position to write down the real-space version of (9.11) working to second order in the derivatives and first order in the ΔV . Ignoring the interface terms one gets

$$\begin{aligned} -(\hbar^2/2m_c^{(0)})d^2F_c^{WS}/dz^2 + (\hbar|p_{vc}|/mE_g)^2(d/dz)(\Delta E_g(z)dF_c^{WS}/dz) \\ + E_c(z)F_c^{WS} = EF_c^{WS} \end{aligned} \quad (9.23)$$

where $m_c^{(0)}$ is the conduction band mass of the reference crystal, $\Delta E_g(z) = \Delta V_{cc}(z) - \Delta V_{vv}(z)$ is the first-order change in the band gap, and $E_c(z) = E_c^{(0)}(0) + \Delta V_{cc}(z)$ is the conduction band-edge energy to first order. Both $\Delta E_g(z)$ and $E_c(z)$ will display Gibbs oscillations in the region of abrupt interfaces. To first order in $\Delta E_g(z)/E_g^{(0)}$, where $E_g^{(0)}$ is the band gap of the reference crystal, the equation may be rewritten as an effective-mass equation,

$$-(\hbar^2/2)(d/dz)\{[1/m_c(z)]dF_c^{WS}/dz\} + E_c(z)F_c^{WS} = EF_c^{WS} \quad (9.24)$$

where $m_c(z)$ is the conduction band mass at z defined by

$$m/m_c(z) = 1 + (2|p_{vc}|^2/mE_g(z)) \quad (9.25)$$

where $E_g(z) = E_g^{(0)} + \Delta E_g(z)$ is the band gap at z . We could, using the arguments of section 5, replace the functions $m_c(z)$ and $E_c(z)$ with their piecewise-continuous counterparts to arrive at the conventional equation.

It should be noted that in the above derivation terms involving $d\Delta V_{vc}(z)/dz$ and $d^2\Delta V_{cc}(z)/dz^2$ have been neglected. It is not obvious that this is justified. It is true that these terms only appear near interfaces; their influence on the eigenvalue is another matter. These terms are different in character to the interface terms referred

to in section 8 on numerical results. The terms referred to there had plane-wave expansion coefficients that had large wavevector denominators. The terms here do not have that property (except $d\Delta V_{vc}(z)/dz$ for the case in which $U_c(z)$ and $U_v(z)$ have opposite symmetry) and whether they can be discarded or not requires further investigation.

One should not be surprised at the effective-mass equation for the WS envelope functions differing from that for the LK envelope functions. To derive the equation for the WS case we have had to work to third order in small quantities (ka and $\Delta V/E_g^{(0)}$). One should also be able to derive the WS equation by starting from the LK equation and transforming the basis functions, the transformation differing from the identity by quantities of first order. That one gets a different equation to third order is only to be expected.

If, as in the derivation of the effective-mass equation for the LK case (section 6), one regards the $k \cdot p$ energy and the H_{nn} , as the same order of magnitude, then the derivation of the effective-mass equation is incomplete, for while we have gone to third order to get the position-dependent effective mass, we have not considered the $\sum_{k'} \langle ck|\Delta V|vk'\rangle \tilde{F}_v^{WS}(k')$ term, which is non-negligible. From the exact envelope-function equation for $n = v$ one has approximately

$$\tilde{F}_v^{WS}(k) = \frac{1}{E - E_v(k)} \sum_{k'} \langle vk|\Delta V|ck'\rangle \tilde{F}_c^{WS}(k'). \tag{9.26}$$

Because E is close to the conduction band edge of the reference Hamiltonian and we need only small- k components, we can replace $E - E_v(k)$ by $E_g^{(0)}$, the band gap of the reference Hamiltonian. So the term

$$\sum_{k'} \langle ck|\Delta V|vk'\rangle \tilde{F}_v^{WS}(k') \tag{9.27}$$

in the equation for $\tilde{F}_c^{WS}(k')$ can be approximated by

$$\frac{1}{E_g^{(0)}} \sum_{k'} \sum_{k''} \langle ck|\Delta V|vk''\rangle \langle vk''|\Delta V|ck'\rangle \tilde{F}_c^{WS}(k'). \tag{9.28}$$

We evaluate this term using the plane-wave expansion of $\Delta V(z)$,

$$\Delta V(z) = \sum_{k,G} \Delta \tilde{V}_G(k) \exp[i(k + G)z] \tag{9.29}$$

and obtain

$$\frac{1}{E_g^{(0)}} \sum_{k'} \sum_{k''} \Delta \tilde{V}_{cv}(k - k'') \Delta \tilde{V}_{vc}(k'' - k') \tilde{F}_c^{WS}(k') \tag{9.30}$$

where the k dependence of the Bloch functions has been neglected. The $\Delta \tilde{V}_{nn'}(k)$ are just the plane-wave expansion coefficients of $\Delta V_{nn'}(z)$, the local part of $\Delta V_{nn'}(z, z')$. Going into real space our term becomes

$$(1/E_g^{(0)}) \Delta V_{cv}(z) \Delta V_{vc}(z) \tilde{F}_c^{WS}(z) \tag{9.31}$$

which acts like an effective band-edge term corresponding to a second-order correction to the conduction band edge at z due to the potential ΔV .

Complicated though these manipulations are, we have still not included the third-order terms needed for consistency! To do so is laborious and will not be pursued further here. Suffice it to say that the Luttinger-Kohn type approach gives an easier route for deriving an effective-mass equation for inhomogeneous material systems.

10. Non-parabolicity

A widespread application of effective-mass theory is to the motion of conduction band electrons in direct-gap semiconductors with the zincblende structure. In such crystals the conduction band can only be considered parabolic for energies small compared with the band gap. The question naturally arises as to how the effective-mass approximation is to be extended to the non-parabolic regime. There would appear at present broadly two ways of approaching this problem. If one starts with the Wannier–Slater approach, then expansion of the band structure of the reference crystal to include wavevectors to the fourth power immediately builds in some non-parabolicity effects, but to account for the different degrees in non-parabolicity for the constituent materials of a microstructure will involve even more algebra than was expended in deriving the ordinary effective-mass equation. Alternatively one can use the energy-dependent effective-mass approach that arises naturally in the derivation of the effective-mass equation using Luttinger–Kohn type envelope functions, an approach that will now be examined.

The energy-dependent effective-mass approach has been criticized on the grounds that it must fail in principle (Persson and Cohen 1988). These authors quote the argument given, for example, in the book by Schiff (1968) concerning the properties that any proposed time-dependent wave equation must have to allow a superposition principle involving states with different energies: the coefficients in the wave equation cannot depend on the energy. While this argument is valid for time-dependent wave equations, it is not applicable to energy eigenvalue equations; all one requires in this case is a superposition principle for states of the same energy. Examples of equations for energy eigenvalues with energy-dependent parameters are not difficult to find in solid-state theory: the orthogonalized plane-wave (OPW) pseudopotential equation (see e.g. Heine 1970) and the equation for the single-particle Green function (see e.g. Hedin and Lundqvist 1969, Stolz 1974) come readily to mind. Quite generally, an eigenvalue equation can be ‘folded down’ to produce an exact equation for a restricted set of the eigenvalues in which the coefficients depend on the eigenvalue (see e.g. Löwdin 1952, Heine and Cohen 1970), and the energy-dependent effective-mass equation derived in section 6 is just an example of this carried out approximately (for an early example see Shockley (1950)).

Having satisfied oneself that there is nothing to stop one using an energy-dependent effective mass, at least in principle, it is appropriate to ask under what circumstances it is a useful approach. Consider first the question of the band structure using the $k \cdot p$ method. If one band is coupled by the $k \cdot p$ interaction to the other bands but these other bands have no such coupling between them, owing to symmetry, then the energy-dependent effective-mass approach will be accurate (to order k^4) provided the free-electron kinetic energy terms can be neglected compared to the difference in the energy eigenvalue and the zone-centre energies of the other bands. This can be seen formally in the ‘folding down’ procedure mentioned above because in this case the matrix to be inverted in the ‘folding down’ procedure is diagonal. As an example, take the conduction band along the (001) direction of a direct-narrow-gap zincblende semiconductor. The $k \cdot p$ Hamiltonian matrix involving the conduction, light-hole and spin split-off bands has the above-mentioned structure (see e.g. Taylor and Burt 1987, equation (5)) and the energy-dependent mass follows naturally from the eigenvalue equation for the conduction band envelope function (Taylor and Burt 1987, equation (9)); note that the free-electron kinetic energy term

associated with the conduction band has been omitted in the derivation but one can easily see that this omission is not necessary. One notes, of course, that this procedure will not work for the light-hole and spin split-off bands, and more sophisticated approximations involving only a selective application of perturbation theory, such as in equation (6.3), are required. Such a scheme has been carried out by Eppenga *et al* (1987). This scheme has been tested (Burt and Tatham 1990) for the light-hole band structure for InP in the (001) direction against a 15-band $k \cdot p$ calculation (Pollak *et al* 1966, Pollak and Cardona 1966), including spin-orbit interaction, and found to be very accurate.

An apparent problem with the way that the energy-dependent effective-mass approximation treats non-parabolicity in heterostructures has been raised by Nag and Mukhopadhyay (1991). They point out that the effective mass needed to describe the energy-wavevector relation in the non-parabolic region, the energy effective mass, is different from that needed, the velocity effective mass, to describe the group-velocity-wavevector relation. Further, they note that the energy effective mass does not appear to conserve probability current at an interface when it is used in the derivative boundary condition. They claim that the velocity effective mass is the correct mass to include in the derivative boundary condition to ensure conservation of probability current. (Essentially the same dilemma has been raised previously by Ekenberg (1989) but no firm recommendations were made there.) While this appears to be a reasonable proposal from a phenomenological point of view it appears less so from the author's fundamental approach: for slowly varying envelope functions the correct description of the band structures of the constituent materials and the conservation of the probability current are intimately connected via the matrix elements of momentum, as was shown in the section on boundary conditions (section 7). It is difficult to see how one can have a set of envelope-function equations that give a good description of the band structure and yet give a poor account of the current conservation. Indeed, we will show explicitly for the two-band model used by Nag and Mukhopadhyay that the energy-dependent effective-mass approximation does conserve the probability current in the limit of small band gaps.

The energy-wavevector relation quoted by Nag and Mukhopadhyay corresponds to a narrow-gap two-band model in which the electron effective mass is small compared to the free-electron mass, i.e. free-electron kinetic energies can be neglected compared to the electron energy relative to the band edge. In this approximation the band structure for the conduction band, with the energy zero taken at the band edge, is

$$E(E + E_g) = (\hbar k P / m)^2 \quad (10.1)$$

where $P = |P_{cv}|$ is the modulus of the interband momentum matrix element, P_{cv} , and E_g^{-1} is the parameter α_i in Nag and Mukhopadhyay (1991). The energy-dependent effective mass $m_c(E)$ is given by

$$m_c(E) = m_c(0)(1 + E/E_g) \quad (10.2)$$

with

$$m/m_c(0) = 2P^2/mE_g. \quad (10.3)$$

The group velocity v is given by

$$v = (1/\hbar)dE/dk = \hbar k/[m_c(0)(1 + 2E/E_g)] \quad (10.4)$$

where the velocity effective mass, $m_c(0)(1 + 2E/E_g)$, has twice as large a correction for non-parabolicity as the energy effective mass in (10.2). For a plane wave normalized to a length L , the current will be v/L . On the other hand, the energy-dependent effective-mass expression for the conserved current is (Burt 1988b)

$$[1/m_c(E)]\text{Re}[F_c^*(-i\hbar dF_c/dz)]. \quad (10.5)$$

For a plane wave $(A_c/L^{1/2})\exp(ikz)$ this gives a current

$$J = (|A_c|^2/L)\hbar k/[m_c(0)(1 + E/E_g)]. \quad (10.6)$$

It is now tempting to argue that $|A_c|^2 = 1$ for a normalized plane wave and that J given by (10.6) is incorrect because it does not agree with v/L deduced from (10.4). The error in the argument lies in assuming that $|A_c|^2 = 1$ gives the correct normalization. This is good enough in the parabolic regime, but not here. The correct normalization is $|A_c|^2 = 1/(1 + E/E_g)$ and is derived as follows. The wavefunction is given by

$$\Psi = F_c U_c + F_v U_v \quad (10.7)$$

where the U 's are normalized so that $|U_n|^2$ has mean value unity over a unit cell. Provided the components in the plane-wave expansion are in the inner half of the Brillouin zone, then the normalization integral (over interval L) is

$$\int |\Psi|^2 dz = \int |F_c|^2 dz + \int |F_v|^2 dz. \quad (10.8)$$

Now for the plane wave

$$F_c = (A_c/L^{1/2})\exp(ikz) \quad (10.9)$$

one has

$$F_v = (A_v/L^{1/2})\exp(ikz) \quad (10.10)$$

with

$$A_v = (\hbar k P_{vc}/m E_g) A_c. \quad (10.11)$$

For normalization one needs

$$|A_c|^2 = 1/[1 + (\hbar k P/m E_g)^2] = 1/(1 + E/E_g). \quad (10.12)$$

With (10.6) this gives a current that agrees with the group velocity expression to first order in E/E_g in keeping with the accuracy to which we have been working. One can readily extend this argument to include the case of a combination of an incident and a reflected wave.

So, for situations in which the energy-dependent effective mass gives a reasonable description of the band structure, the boundary condition $[1/m_c(E)] dF_c/dz$ continuous will conserve current.

11. Summary

We have seen that the major problem with conventional justifications for applying the effective-mass approximation to microstructures is the assumption that the eigenfunctions at high-symmetry points in the Brillouin zone are approximately the same for the constituent crystals; the terms in the Hamiltonian that one throws away on making this assumption can easily be of the same order of magnitude as those one retains, i.e. the $k \cdot p$ type terms. Without this assumption, the almost ubiquitous assumption of continuity of the envelope function at an abrupt interface is a non-starter and undermines the usual discussions on the boundary condition for the derivative. It is also difficult to conceive of a well constructed mathematical theory that involves a change in periodic basis functions as one goes from one material to another.

These problems have been circumvented by applying the original Luttinger-Kohn envelope-function expansion to a microstructure in its entirety using the same periodic basis functions throughout even though they are not necessary eigenstates of the constituent crystals. In this way one has a well defined unique envelope-function expansion, the envelope functions and all their derivatives being continuous everywhere including at atomically abrupt interfaces. The boundary-condition problem is immediately sidelined thereby. Because this new envelope-function expansion is unique and invertible, exact equations for the envelope functions can be derived, equations that are completely equivalent to the original Schrödinger equation, the complexity of which is not restricted; spin-orbit interaction, including wavevector-dependent terms, non-local potentials and even strain can be catered for. Starting with these exact envelope-function equations it has been possible to derive effective-mass equations in a systematic way such that the errors in the approximations made can be estimated, if need be. The crucial point in these derivations is not to start from the limit of slowly varying composition changes, as have Karavaev and Tikhodeev (1991) for instance, but rather from the limit of slowly varying envelope functions. One asks the question, if the envelope function(s) are slowly varying, what approximate simplified equation(s) do they obey? The boundary conditions for these approximate equations have been examined and it has been found that kinks appear in the envelope functions when second derivatives are dropped and the coefficients in the envelope-function equations are approximated by piecewise-continuous functions representing abrupt composition changes.

By focusing on the equations satisfied by slowly varying envelope functions, it becomes clear why the effective-mass approximation works so well for quantum wells only a monolayer or so wide. The ground state is only weakly bound and the state necessarily extended and the envelope function slowly varying. We have managed to derive the effective-mass equation using only this slow variation and the usual assumption of the dominance of one envelope function; no assumption that the composition is slowly varying has been made. The slowly varying envelope function is only affected by the long-wavelength components of the compositional variation. Similarly, in short-period superlattices, where the electron tunnels easily through the barriers, the envelope function will have a very small derivative and the derivation of the effective-mass equation given here will still hold.

Our derivation of the new effective-mass equation includes the differences in the zone-centre eigenstates between the constituent crystals via the inclusion of the off-diagonal elements of the zone-centre Hamiltonian. The reason that one can still derive an effective-mass equation is that, in sufficiently high-symmetry crystals, there

are relatively few pairs of zone-centre eigenstates of the reference crystal that are coupled by both the Hamiltonian and the momentum, and those that are have large energy separations. The result is that, in the second-order perturbation theory, the perturbation being composed of a sum of $k \cdot p$ and Hamiltonian terms, the influence of cross terms, i.e. those linear in both $k \cdot p$ and the Hamiltonian, is minimal. Time-reversal symmetry also plays a role in reducing the overall effect of these terms since it is this symmetry that eliminates the linear k terms in the band structure in the absence of spin-orbit interaction (Kane 1966).

Much of the formal work has been extensively illustrated using a one-dimensional model superlattice, focusing on the ground state for one of the bands. The wavefunction shows a pronounced soft kink in its envelope and this is also displayed by the principal envelope function. It demonstrates what the author had already deduced algebraically, that the kink in the conventional envelope function due to the use of the boundary condition $(1/m^*) dF/dz$ continuous at a discontinuity of the effective mass is an artefact of too severe an approximation. The envelope function, as defined by the author, has a rapid change in its derivative that appears as a soft kink. The numerical work shows a similar behaviour in the wavefunction. For the model used, the new effective-mass equation proposed by the author, in which the coefficients are less severely approximated than is usual, gives an excellent approximation for the principal envelope function.

The use of the Wannier-Slater version of the new envelope functions as an alternative to the Luttinger-Kohn type has been examined. It has been shown that exact equations can be derived for these envelope functions as well. An effective-mass equation has been derived using these envelope functions for the case of small, but abrupt, compositional changes. The derivation is more cumbersome than that given for the Luttinger-Kohn type envelope functions. It becomes even more so when one tries to account for compositional changes to second order, as was managed with relative ease with the Luttinger-Kohn envelope functions. It appears, then, that, while the Wannier-Slater envelope functions are well suited to dealing with weak fields in bulk crystals, the Luttinger-Kohn envelope functions are better for deriving effective-mass-type approximations in microstructures. Of course, the position for numerical work may be different.

As an example of how the author's approach can help resolve controversy, we have investigated the recent proposal that for non-parabolic bands the velocity effective mass should be used in the derivative boundary condition, rather than the energy mass suggested by so-called $k \cdot p$ theory, in order to conserve probability current. We have shown that the approximate envelope-function equations derived in this paper do conserve current and that, in particular, for the two-band model in the limit of small band gap the energy effective mass is the correct mass to use in the derivative boundary condition.

The new envelope-function method described here is very powerful. It can deal with spin-orbit interaction, non-local potentials and strained structures (Burt 1989). In this paper it has been presented with the basis functions as periodic with respect to a Bravais lattice. However, it is not restricted to this case and perhaps it is appropriate to finish by giving a couple of examples. First, one could extend it to cases in which the basis functions all change by the same phase factor on translation by a lattice vector (Burt 1989). This would be suitable for electron states associated with zone-edge states such as those at X and L in zincblende structures. Second, in some structures, one might have states that are a mixture of zone-centre and zone-

edge states such as the Γ and X states in GaAs/AlAs (001) layered structures. In that case, one might try using a reference Bravais lattice twice the size of that of the constituent crystals. Both the zone-centre and zone-edge states of the constituent crystals would be periodic with respect to this reference Bravais lattice. The periodic basis functions for the envelope function expansion could be chosen so that each basis function had the translational symmetry of either the zone-centre or the zone-edge states of the constituent crystals. The new envelope-function equations would have the same form as derived here, but there would now be coupling between zone-centre and zone-edge-type basis functions via the Hamiltonian, but not the $k \cdot p$, terms near the interfaces.

Acknowledgments

The author would like to thank Professors V Heine and A M Stoneham for their support, Professor Sir Rudolf Peierls and Professor J B Pendry for correspondence and Dr G A Baraff for a valuable comment on an earlier version of the manuscript.

Appendix 1. A note on the choice of Brillouin zones

The envelope-function expansion used by the author involves a complete set of functions, $U_n(\mathbf{R})$, that are periodic with respect to a Bravais lattice. The associated envelope functions, $F_n(\mathbf{R})$, are defined to have a plane-wave expansion restricted to wavevectors within the Brillouin zone. If unnecessary complications are to be avoided, it is important to choose the Brillouin zone one uses with care.

A Brillouin zone is a range of wavevectors k such that the set of functions $\exp(i\mathbf{k} \cdot \mathbf{R})$ is a complete set of non-degenerate eigenfunctions for each of the three primitive translations, α_i , of the Bravais lattice, so that the eigenvalue $\exp(i\mathbf{k} \cdot \alpha_i)$ appears only once. A simple way of doing this is to take a parallelepiped representing the unit cell of the reciprocal lattice rather than the more complex conventional construction. Another way suitable for multilayer structures is to follow the method used for generating the band structure for complex wavevector. One takes a two-dimensional Bravais lattice with the same translational symmetry as that of the multilayer parallel to its interfaces. The two-dimensional Bravais lattice will have a two-dimensional Brillouin zone constructed in the usual way. By projecting this two-dimensional zone perpendicular to its plane one can construct a prism that contains sufficient k -values to provide the eigenfunctions for the third primitive translation of the reference Bravais lattice, i.e. the primitive translation that is not parallel, but also not necessarily perpendicular, to the interface planes. In this way the three-dimensional Brillouin zone has the same height for all points in the two-dimensional Brillouin zone. An example of this type of construction can be found in Burt (1980) in connection with the band structure for complex wavevector in the (111) direction for GaAs. The two-dimensional Brillouin zone in this case is a regular hexagon and the three-dimensional zone for GaAs is a prism with its axis along (111) and a hexagonal cross section. Such a construction would circumvent the problem encountered by Cuypers and van Haeringen (1991) of having more than one envelope function per band for a given component of k parallel to the interfaces. For multilayer structures with interfaces along low-index planes, the three-dimensional Bravais lattice recommended here would not usually correspond to the Bravais lattice of the constituent materials.

Appendix 2. Resolution of potential term into local and non-local parts

The effective potential for envelope functions is given by

$$V_{nn'}(\mathbf{R}, \mathbf{R}') = \frac{1}{\Omega} \sum_{\mathbf{k}, \mathbf{k}'} \exp(+i\mathbf{k} \cdot \mathbf{R}) \tilde{V}_{nn'}(\mathbf{k}, \mathbf{k}') \exp(-i\mathbf{k}' \cdot \mathbf{R}') \quad (\text{A2.1})$$

where

$$\tilde{V}_{nn'}(\mathbf{k}, \mathbf{k}') = \sum_{\mathbf{G}, \mathbf{G}'} U_{n\mathbf{G}}^* (\mathbf{k} + \mathbf{G} | V | \mathbf{k}' + \mathbf{G}') U_{n'\mathbf{G}'}. \quad (\text{A2.2})$$

When the original potential V is local, the effective potential $V_{nn'}(\mathbf{R}, \mathbf{R}')$ can be resolved into local and non-local parts, though, as noted previously (Burt 1988a), this resolution may not be unique. To demonstrate the resolution, write the local potential V as a plane-wave expansion

$$V(\mathbf{R}) = \sum_{\mathbf{k}, \mathbf{G}} \tilde{V}_{\mathbf{G}}(\mathbf{k}) \exp[i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{R}] \quad (\text{A2.3})$$

with

$$\langle \mathbf{k} + \mathbf{G} | V | \mathbf{k}' + \mathbf{G}' \rangle = \tilde{V}_{\mathbf{G}-\mathbf{G}'-\mathbf{G}_1}(k_1) \quad (\text{A2.4})$$

where $k_1 = k - k' + G_1$ is in the Brillouin zone and G_1 is a reciprocal-lattice vector. For fixed k' , k can be replaced in favour of k_1 to obtain

$$V_{nn'}(\mathbf{R}, \mathbf{R}') = \frac{1}{\Omega} \sum_{\mathbf{k}_1, \mathbf{k}'} \exp[+i(\mathbf{k}_1 + \mathbf{k}' - \mathbf{G}_1) \cdot \mathbf{R}] \left(\sum_{\mathbf{G}, \mathbf{G}'} U_{n\mathbf{G}}^* \tilde{V}_{\mathbf{G}-\mathbf{G}'-\mathbf{G}_1}(k_1) U_{n'\mathbf{G}'} \right) \times \exp(-i\mathbf{k}' \cdot \mathbf{R}'). \quad (\text{A2.5})$$

In the 'local' approximation G_1 is set to zero, noting that this only affects the summand when $k_1 + k'$ is outside the Brillouin zone, a region of little importance for slowly varying envelope functions. Denoting this local approximation to $V_{nn'}(\mathbf{R}, \mathbf{R}')$ by $V_{nn'}^{(\text{loc})}(\mathbf{R}, \mathbf{R}')$ we find that

$$V_{nn'}^{(\text{loc})}(\mathbf{R}, \mathbf{R}') = \sum_{\mathbf{k}_1} \left(\sum_{\mathbf{G}, \mathbf{G}'} U_{n\mathbf{G}}^* \tilde{V}_{\mathbf{G}-\mathbf{G}'}(k_1) U_{n'\mathbf{G}'} \right) \exp(+i\mathbf{k}_1 \cdot \mathbf{R}) \Delta(\mathbf{R} - \mathbf{R}') \\ = V_{nn'}(\mathbf{R}) \Delta(\mathbf{R} - \mathbf{R}'). \quad (\text{A2.6})$$

The non-local part is just $V_{nn'}(\mathbf{R}, \mathbf{R}') - V_{nn'}^{(\text{loc})}(\mathbf{R}, \mathbf{R}')$.

Appendix 3. Asymptotic form of the potential term far from an interface

Again we start with the effective potential for envelope functions

$$V_{nn'}(\mathbf{R}, \mathbf{R}') = \frac{1}{\Omega} \sum_{\mathbf{k}, \mathbf{k}'} \exp(+i\mathbf{k} \cdot \mathbf{R}) \tilde{V}_{nn'}(\mathbf{k}, \mathbf{k}') \exp(-i\mathbf{k}' \cdot \mathbf{R}') \quad (\text{A3.1})$$

where

$$\tilde{V}_{nn'}(\mathbf{k}, \mathbf{k}') = \sum_{\mathbf{G}, \mathbf{G}'} U_{n\mathbf{G}}^* \langle \mathbf{k} + \mathbf{G} | V | \mathbf{k}' + \mathbf{G}' \rangle U_{n'\mathbf{G}'}. \quad (\text{A3.2})$$

Using

$$\langle \mathbf{k} + \mathbf{G} | V | \mathbf{k}' + \mathbf{G}' \rangle = \frac{1}{\Omega} \int V(\mathbf{R}'') \exp[-i(\mathbf{k} - \mathbf{k}' + \mathbf{G} - \mathbf{G}') \cdot \mathbf{R}''] d^3\mathbf{R}'' \quad (\text{A3.3})$$

for a local potential, one can write

$$V_{nn'}(\mathbf{R}, \mathbf{R}') = \sum_{\mathbf{G}, \mathbf{G}'} U_{n\mathbf{G}}^* U_{n'\mathbf{G}'} \int \Delta(\mathbf{R} - \mathbf{R}'') V(\mathbf{R}'') \exp[-i(\mathbf{G}' - \mathbf{G}) \cdot \mathbf{R}''] \times \Delta(\mathbf{R}'' - \mathbf{R}') d^3\mathbf{R}'' \quad (\text{A3.4})$$

Now $\Delta(\mathbf{R})$ (defined in (3.10)) is only appreciable near $\mathbf{R} = 0$ and tends to zero at large R , so for \mathbf{R} and \mathbf{R}' far from any interface the integrand in (A3.4) will only be appreciable in the region around \mathbf{R} and \mathbf{R}' . Hence in this asymptotic region one can replace $V(\mathbf{R}'')$ in the integrand by $V^{\text{bulk}}(\mathbf{R}'')$, the potential of the appropriate bulk crystal. Expanding $V^{\text{bulk}}(\mathbf{R})$ in plane waves

$$V^{\text{bulk}}(\mathbf{R}) = \sum_{\mathbf{G}} V_{\mathbf{G}}^{\text{bulk}} \exp(i\mathbf{G} \cdot \mathbf{R}) \quad (\text{A3.5})$$

one has

$$\langle \mathbf{k} + \mathbf{G} | V^{\text{bulk}} | \mathbf{k}' + \mathbf{G}' \rangle = V_{\mathbf{G}-\mathbf{G}'}^{\text{bulk}} \delta_{\mathbf{k}\mathbf{k}'}. \quad (\text{A3.6})$$

Using (A3.6) in (A3.2) and then (A3.2) in (A3.1) gives asymptotically

$$V_{nn'}^{\text{bulk}}(\mathbf{R}, \mathbf{R}') = V_{nn'}^{\text{bulk}} \Delta(\mathbf{R} - \mathbf{R}') \quad (\text{A3.7})$$

where

$$V_{nn'}^{\text{bulk}} = \sum_{\mathbf{G}, \mathbf{G}'} U_{n\mathbf{G}}^* V_{\mathbf{G}-\mathbf{G}'}^{\text{bulk}} U_{n'\mathbf{G}'}. \quad (\text{A3.8})$$

We note that in this asymptotic region far from interfaces the effective potential $V_{nn'}(\mathbf{R}, \mathbf{R}')$ corresponds to the local part, i.e. the non-local part is restricted to the interface regions.

Appendix 4. $V_{nn'}(\mathbf{R})$ for a superlattice

Take Cartesian axes with the (x, y) plane parallel to the interfaces but without assuming that the crystal constituents of the superlattice have cubic symmetry. The superlattice potential has the form

$$V(\mathbf{R}) = \begin{cases} \sum_{\mathbf{G}} V_{\mathbf{G}}^{(1)} \exp(i\mathbf{G} \cdot \mathbf{R}) & \text{for } 0 < z < L_1 \\ \sum_{\mathbf{G}} V_{\mathbf{G}}^{(2)} \exp(i\mathbf{G} \cdot \mathbf{R}) & \text{for } L_1 < z < L_1 + L_2 \end{cases} \quad (\text{A4.1})$$

where the \mathbf{G} 's are the reciprocal-lattice vectors for a Bravais lattice that has the same translational symmetry as the superlattice in the (x, y) plane. For superlattices with interfaces parallel to high-index planes, the Bravais lattice will be that of the constituent bulk crystals. The plane-wave expansion coefficients $\tilde{V}_{\mathbf{G}}(\mathbf{k})$ will be zero except when the component of \mathbf{k} in the (x, y) plane is zero. So $V_{nn'}(\mathbf{R})$ will be a function of z only and from equation (A2.6) we see that (we use K for the z component of \mathbf{k})

$$V_{nn'}(z) = \sum_K \sum_{\mathbf{G}, \mathbf{G}'} U_{n\mathbf{G}}^* \tilde{V}_{\mathbf{G}-\mathbf{G}'}(K) U_{n'\mathbf{G}'} e^{iKz} \quad (\text{A4.2})$$

where

$$\tilde{V}_{\mathbf{G}}(K) = \frac{1}{\Omega} \int V(\mathbf{R}) \exp[-i(K\hat{z} + \mathbf{G}) \cdot \mathbf{R}] d^3\mathbf{R} \quad (\text{A4.3})$$

and \hat{z} is the unit vector in the z direction, i.e. perpendicular to the interfaces. Using manipulations similar to those used for the single interface in one dimension (Burt 1988a), one finds

$$V_{nn'}(z) = V_{nn'}^{(b)}(z) + V_{nn'}^{(i)}(z) \quad (\text{A4.4})$$

where $V_{nn'}^{(b)}(z)$ depends only on the bulk properties of the constituent crystals and the structure of the superlattice. $V_{nn'}^{(b)}(z)$ is periodic in z with period $L = L_1 + L_2$ as one would expect and in the region $0 < z < L$ is given by

$$V_{nn'}^{(b)}(z) = \theta(L_1 - z) V_{nn'}^{(1)} + \theta(z - L_1) V_{nn'}^{(2)} \quad (\text{A4.5})$$

where $V_{nn'}^{(1)}$ and $V_{nn'}^{(2)}$ are the bulk values of $V_{nn'}(z)$ for the constituent crystals and $\theta(z)$ is the usual step function *but with its plane-wave expansion limited to wavevectors within the Brillouin zone*, i.e. the abrupt step is softened by Gibbs oscillations. The interface term $V_{nn'}^{(i)}(z)$ is given by

$$V_{nn'}^{(i)}(z) = \frac{1}{L} \sum_{K \neq 0} \Delta \tilde{V}_{nn'}(K) [\exp(-iKL_1) - 1] e^{iKz} \quad (\text{A4.6})$$

with

$$\Delta \tilde{V}_{nn'}(K) = \sum_{\mathbf{G}, \mathbf{G}'} U_{n\mathbf{G}}^* \left(\sum_{\mathbf{G}'' \neq \mathbf{G}-\mathbf{G}'} \frac{\tilde{V}_{(\Delta\mathbf{G}, \mathbf{G}'')}^{(2)} - \tilde{V}_{(\Delta\mathbf{G}, \mathbf{G}'')}^{(1)}}{K + \mathbf{G} - \mathbf{G}' - \mathbf{G}''} \right) U_{n'\mathbf{G}'}. \quad (\text{A4.7})$$

Here $\Delta g = g - g'$ where g is the component of G in the (x, y) plane, i.e. (g, G) denotes G , G being the z component of G . On examining (A4.7) one sees that the wavevector denominators are always large and this makes the interface terms small. One can also readily verify from (A4.6) that, when the layer widths are sufficiently large, then $V_{nn}^{(i)}(z)$ will tend to zero except near the interfaces.

References

- Altarelli M 1983a *Physica B* **117 + 118** 747
 — 1983b *Phys. Rev. B* **28** 842
 — 1983c *Springer Lecture Notes in Physics* vol 177, ed G Landwehr (Berlin: Springer) p 174
 — 1986 *Heterojunctions and Semiconductor Superlattices* ed G Allan, G Bastard, N Boccarda, M Lannoo and M Voos (Berlin: Springer) p 12
 Altarelli M and Bassani F 1982 *Handbook on Semiconductors vol 1 Band Theory and Transport Properties* vol 1, ed W Paul (Amsterdam: North-Holland) p 269
 Ando T, Wakahara S and Akera H 1989 *Phys. Rev. B* **40** 11 609
 Baraff G A and Gershoni D 1991 *Phys. Rev. B* **43** 4011
 Bastard G 1981 *Phys. Rev. B* **24** 5693
 Bastard G, Brum J A and Ferreira R 1991 *Solid State Physics* vol 44 (New York: Academic) p 229
 Burt M G 1980 *J. Phys. C: Solid State Phys.* **13** 1825
 — 1987 *Semicond. Sci. Technol.* **2** 460 (erratum **2** 701)
 — 1988a *Semicond. Sci. Technol.* **3** 739
 — 1988b *Semicond. Sci. Technol.* **3** 1224
 — 1988c unpublished calculations
 — 1989 *Bandstructure Engineering in Semiconductor Microstructures (Nato ASI Series B, vol 189)* ed R A Abram and M Jaros (New York: Plenum) p 99
 Burt M G and Tatham M C 1990 unpublished calculations
 Cunningham T, Barker R C and Chiu L C 1988 *J. Appl. Phys.* **63** 5393
 Cuypers J P and van Haeringen W 1991 *Physica B* **168** 58
 — 1992 *J. Phys.: Condens. Matter* **4** 2587
 Dingle R, Wiegmann W and Henry C H 1974 *Phys. Rev. Lett.* **33** 827
 Einevoll G T 1990 *Phys. Rev. B* **42** 3497
 Einevoll G T and Hemmer P C 1988 *J. Phys. C: Solid State Phys.* **21** L1193
 Einevoll G T, Hemmer P C and Thomsen J 1990 *Phys. Rev. B* **42** 3485
 Ekenberg U 1989 *Phys. Rev. B* **40** 7714
 Eppenga R, Schuurmans M F H and Colak S 1987 *Phys. Rev. B* **36** 1554
 Hedin L and Lundqvist S 1969 *Solid State Physics* vol 23 (New York: Academic) p 35
 Heine V 1970 *Solid State Physics* vol 24 (New York: Academic) p 12
 Heine V and Cohen M H 1970 *Solid State Physics* vol 24 (New York: Academic) p 78
 Kane E O 1966 *Semiconductors and Semimetals* vol 1 (New York: Academic) p 75
 Karavaev G F and Tikhodeev Yu S 1991 *Sov. Phys.-Semicond.* **25** 745
 Kohn W 1957 *Solid State Physics* vol 5 (New York: Academic) p 257
 Löwdin 1952 *J. Chem. Phys.* **19** 1396
 Luttinger J M and Kohn W 1955 *Phys. Rev.* **97** 869
 Morrow R A 1987a *Phys. Rev. B* **35** 8074
 — 1987b *Phys. Rev. B* **36** 4836
 Morrow R A and Brownstein K R 1984 *Phys. Rev. B* **30** 678
 Nag B R and Mukhopadhyay S 1991 *Appl. Phys. Lett.* **58** 1056
 Persson A and Cohen R M 1988 *Phys. Rev. B* **38** 5568
 Pollak F H and Cardona M 1966 *J. Phys. Chem.* **70** 423
 Pollak F H, Higginbotham C W and Cardona M 1966 *Proc. Int. Conf. Physics of Semiconductors (Kyoto); J. Phys. Soc. Japan* **21** (Suppl.) 20
 Pötz W, Porod W and Ferry D K 1985 *Phys. Rev. B* **32** 3868
 Schiff L I 1968 *Quantum Mechanics* 3rd edn (New York: McGraw-Hill) p 21
 Shockley W 1950 *Phys. Rev.* **78** 173
 Slater J C 1949 *Phys. Rev.* **76** 1592

- Smith D L and Mailhiet C 1986 *Phys. Rev. B* **33** 8345
- Stolz H 1974 *Einführung in die Vielelectronen-Theorie der Kristalle* (Berlin: Academic) p 108
- Stoneham A M 1975 *Theory of Defects in Solids* (Oxford: Oxford University Press) p 68
- Taylor R I and Burt M G 1987 *Semicond. Sci. Technol.* **2** 485
- Thomsen J, Einevoll G T and Hemmer P C 1989 *Phys. Rev. B* **39** 12783
- Tikhodeev S G 1991 *Solid State Commun.* **78** 339
- Wannier G H 1937 *Phys. Rev.* **52** 191
- Young K 1989 *Phys. Rev. B* **39** 13434