# Fundamentals of envelope function theory for electronic states and photonic modes in nanostructures

**INVITED PAPER**

# Fundamentals of envelope function theory for electronic states and photonic modes in nanostructures

M G Burt

BT Laboratories, Martlesham Heath, Ipswich IP5 3RE, UK

**Abstract.**   The increasing sophistication used in the fabrication of semiconductor nanostructures and in the experiments performed on them requires more sophisticated theoretical techniques than previously employed. The philosophy behind the author's exact envelope function representation method is clarified and contrasted with that of the conventional method. The significance of globally slowly varying envelope functions is explained. The difference between the envelope functions that appear in the author's envelope function representation and conventional envelope functions is highlighted and some erroneous statements made in the literature on the scope of envelope function methods are corrected. A perceived conflict between the standard effective mass Hamiltonian and the uncertainty principle is resolved demonstrating the limited usefulness of this principle in determining effective Hamiltonians. A simple example showing how to obtain correct operator ordering in electronic valence band Hamiltonians is worked out in detailed tutorial style. It is shown how the use of out of zone solutions to the author's approximate envelope function equations plays an essential role in their mathematically rigorous solution. In particular, a demonstration is given of the calculation of an approximate wavefunction for an electronic state in a one dimensional nanostructure with abrupt interfaces and disparate crystals using out of zone solutions alone. The author's work on the interband dipole matrix element for slowly varying envelope functions is extended to envelope functions without restriction. Exact envelope function equations are derived for multicomponent fields to emphasize that the author's method is a general one for converting a microscopic description to a mesoscopic one, applicable to linear partial differential equations with piecewise or approximately piecewise periodic coefficients. As an example, the method is applied to the derivation of approximate envelope function equations from the Maxwell equations for photonic nanostructures.

## 1. Introduction

The physics of nanostructures is a prominent part of modern semiconductor physics and the computation of electronic states, both excited as well as ground states, in such structures plays an important role in the interpretation of experiment. In an ideal world, all experiments would be interpreted using the results of *ab initio* solutions of the many electron Schrödinger equation. But, at present, approximate calculations of even the one electron Green function using the GW approximation for the self-energy (see e.g. Hedin and Lundqvist 1969, Stolz 1974) are computationally intensive and are focused on bulk crystals, surfaces and clusters (see e.g. Rohlfing *et al* 1998, Rohlfing and Louie 1998a, Aryasetiawan and Gunnarson 1998). The computational effort needed to calculate the two particle Green function via the solution of the Bethe–Salpeter equation is even greater, though its solution for GaAs and LiF bulk crystals has just been accomplished (Rohlfing and Louie 1998b). Clearly, any approach, such as the use of envelope function methods, to reduce the computational burden of similar calculations for nanostructures is highly desirable. With the advent, two to three decades

ago, of superlattices and quantum wells (Esaki and Tsu 1969, Dingle *et al* 1974), even semi-empirical calculations (pseudopotential, tight-binding etc see e.g. Smith and Mailhiot 1990) were hardly practical and then only became so for the specialist theoretician. In the meantime, workers in the field had to resort to 'particle in a box' effective mass type methods as a rough and ready alternative for simple systems such as GaAs/GaAlAs quantum wells (type I), but a more sophisticated approach was needed in such systems as InAs/GaSb superlattices (type II). To meet this need, Bastard (1981) introduced a prescriptive envelope function method, in which the real space equations satisfied by the envelope functions were equivalent to the $\boldsymbol{k} \cdot \boldsymbol{p}$ bandstructure eigenvalue problem in bulk crystals (see e.g. Kane 1975), but the band edges were allowed to be functions of position; it was implicitly assumed that the equations would be valid at any atomically abrupt interfaces. The justification for this approach was reasonable agreement with the results of semi-empirical tight-binding calculations, though problems were encountered subsequently with the comparison not being as direct as one would like in some cases (see e.g. Chang *et al* 1985). The method has certainly been popular because, for the most part, it gives reasonable agreement with experiment for the systems of common interest, any lack of accuracy being balanced by the ease of use. It also has the advantage over microscopic methods of giving results that are easy to interpret: the output gives the envelope functions, the slowly varying mesoscopic parts of the wavefunction, rather than the complete wavefunction which contains the rapidly varying Bloch functions. One can also readily understand the physical significance of the input parameters.

Having obtained reasonable agreement with experiment one might suppose that all is well with the Bastard approach (Bastard 1988, Bastard *et al* 1991). However, there are two major objections to this view. The first is the practical one: if the justification for the method is agreement with microscopic calculations, then, in reality, the method is essentially a sophisticated interpolation tool, very useful though it is. One cannot use it with confidence outside the regime for which it has already been tested. This means that it has limited predictive power. Indeed, when interface effects (Aleiner and Ivchenko 1992, Ivchenko *et al* 1993, Krebs and Voisin 1996, Krebs *et al* 1998) were seen experimentally, the equations had to be modified in an *ad hoc* way to obtain agreement with experiment, because the earlier equations had implicitly ignored the physics involved in contrast to the fundamental approach introduced by the present author (Burt 1988a), which can readily encompass such phenomena (Foreman 1998a). And with an empirical theory such as Bastard introduced, what does one do if one cannot obtain agreement with experiment? Is it the method that is wrong or is one's characterization of the nanostructure inadequate or in error? The second major objection is a philosophical one: the scientific method involves a reductionist approach striving to understand all experimental phenomena using the smallest number of postulates. The envelope function method as introduced by Bastard implicitly ignores this approach; it postulates the envelope function equations rather than deriving them from the more fundamental microscopic Schrödinger equation.

At about the same time and in the years following Bastard's original paper a number of attempts were made to justify the application of this envelope function method and variations thereof to quantum wells. Prominent and most cogent among these early attempts were those of White and Sham (1981), Altarelli (1983a–c) and Pötz *et al* (1985). While the arguments in these papers made the validity of the envelope function equations proposed very plausible, they did not offer ways of estimating the errors involved nor a systematic method of deriving more accurate equations. These issues are discussed in Burt (1988a, 1992) and, in particular, the weakness of the prevalent assumption that the zone centre Bloch functions in the well and barrier crystals can be assumed to be the same was highlighted in Burt (1992). To overcome these problems the author introduced (Burt 1988a) a method for deriving exact envelope

function equations starting from the Schrödinger equation. Using reasonable approximations this led to a systematic derivation of the effective mass equation including the multiband case (Burt 1988a, b, 1992) for nanostructures incorporating atomically abrupt heterojunctions.

Despite these advances and a numerical demonstration (Burt 1994a) in which the author's effective mass theory is shown to be capable of generating the correct wavefunction in the presence of abrupt heterojunctions, there seems to be a widespread belief that this envelope function method is bedevilled with limitations and problems such as not being valid for abrupt heterojunctions (Wood and Zunger 1996); being incapable of tackling interface mediated intervalley couplings (Edwards and Inkson 1994, Wang *et al* 1997); containing ambiguities arising from the presence of 'out of zone' solutions (Godfrey and Malik 1996) and supposedly using the restrictive assumption that the Bloch functions are the same in the component crystals (von Allmen 1992). One of the purposes of this paper is to correct these and other misconceptions. Indeed, an example will be given of a calculation of the wavefunction by an envelope function method for a superlattice with abrupt heterojunctions, using out of zone solutions alone; this superlattice has disparate potentials—form factors differing by almost a factor of 2 with respect to normal semiconductor heterostructures—and significantly different zone centre Bloch functions for the well and barrier crystals.

The basis of the author's approach to electron states in nanostructures (Burt 1988a, 1992) has been the use of an envelope function *representation* rather than using envelope functions in the form of trial solutions. The envelope function expansion used, for the most part, is indeed just that introduced by Luttinger and Kohn (1955), but the emphasis in its application is very different. Firstly, exact rather than approximate equations for the envelope functions are derived. Secondly, rather than focus on 'gentle' perturbations to periodic structures, the focus for systematic approximations is on states which have envelope functions that have plane wave expansions dominated by small wavevector (crystal momentum) components. As demonstrated by the author (Burt 1994a, b), and contrary to popular belief, this does not mean that the envelope functions have to be smooth everywhere: occasional kinks and discontinuities are permissible. Foreman (1996) has coined the phrase 'globally slowly varying' to describe such envelope functions.

Why should focusing on 'globally slowly varying' envelope functions be appropriate? It may seem that such a restriction is just a consequence of wishful thinking. However, there is good reason for concentrating on such envelope functions. If one views a nanostructure as a perturbation on some bulk reference crystal, then the largest matrix elements occur between unperturbed states separated by small wavevectors (crystal momenta), a characteristic of piecewise periodic perturbations. Because small wavevector separation usually means small energy separation, the effect of these dominant matrix elements is all the more pronounced. So there is usually a tendency for the plane wave expansion coefficients of the envelope functions plotted against wavevector to be concentrated in the region around the wavevector(s) of the unperturbed state of the bulk reference crystal and states of similar energy in other parts of the Brillouin zone. (For bound states of a quantum well, for instance, the unperturbed state is not a single Bloch state of the unperturbed bulk crystal, but a combination of such Bloch states). This is particularly true of states near a band edge, because the energy separation for a given wavevector difference is usually smaller there than in the middle of a band; a good demonstration of this is seen in the numerical work of Williamson *et al* (1998). So one can understand why focusing on 'globally slowly varying' envelope functions is not nearly as restrictive as one might suppose.

The main theme of this paper is to show how the exact envelope function method and approximations based on 'globally slowly varying' envelope functions can resolve problems that arise in applying envelope function techniques and greatly expand the scope of such

techniques not only to electronic states, but to photonic nanostructures as well. Foreman (1995) has also successfully applied similar techniques to phonons (see also Ridley 1997).

In the next section the main features of the author's use of envelope function representations and its relation to more conventional envelope function methods is discussed, and some erroneous and misleading statements in the literature corrected. In the following section (section 3) we show how we can resolve an apparent conflict between the results of envelope function theory and the uncertainty principle demonstrating the limited usefulness of this principle in determining effective Hamiltonians. In section 4, we turn to the problem of the correct ordering of the differential operators in valence band Hamiltonians. Since this has been done formally in an earlier paper (Burt 1992), the approach here is distinctly didactic, to help anyone interested in doing their own calculations understand the issues involved and to provide simple to follow rules to construct correct Hamiltonians. We move on in section 5 to discuss the 'out of zone' solution conundrum and show how such solutions do have a role to play in a complete theory and give a practical demonstration of the construction of the *wavefunction* from out of zone solutions of properly derived approximate envelope functions equations. The correct evaluation of the interband dipole matrix element is a difficult issue in envelope function theory with erroneous expressions given in both textbooks and research papers implying that the dipole matrix element must always be of the order of atomic dimensions. This issue was addressed by the author (Burt 1993, 1995a, b) for globally slowly varying envelope functions, but in section 6 the basic results are extended to arbitrary envelope functions. The following two sections (7 and 8) before the summary, in section 9, are devoted to discussing, in general terms, the extension of the author's envelope function method to photonic nanostructures. Firstly, we build on previous work (Burt 1989) on electron states with spin–orbit interaction and strained structures, which highlighted the applicability of the author's method to nonscalar wave equations generally, to derive exact envelope function equations for multicomponent fields. This is followed by the application of these equations to the source free Maxwell equations for photonic nanostructures and bringing out the analogy with the traditional derivation of macroscopic Maxwell equations from their microscopic counterparts.

## 2. Representation and misrepresentation

The envelope function method developed by the author is based on a plane wave expansion of the wavefunction (spin is omitted in this discussion to avoid unnecessarily complex notation, but this is treated in Burt (1989))

$$\psi(r) = \sum_{kG} \tilde{\psi}(k+G) \exp(i(k+G)r) = \sum_{kG} \tilde{\psi}_G(k) \exp(i(k+G)r) \qquad (2.1)$$

where the $G$ are reciprocal lattice vectors of an underlying Bravais lattice and the $k$ are wavevectors confined to a primitive cell of the reciprocal lattice usually taken as the first Brillouin zone. $\tilde{\psi}(k+G)$ is the Fourier transform of $\psi$ and $\tilde{\psi}_G(k)$ is an alternative notation that emphasizes the decomposition of the wavevector, $k+G$, into a reciprocal lattice vector and a wavevector in the above mentioned primitive cell. A complete set of functions, $U_n(r)$, periodic in the Bravais lattice, is introduced

$$U_n(r) = \sum_G U_{nG} \exp(iG \cdot r) \qquad (2.2)$$

usually chosen to be orthonormal so that expressing the plane waves in terms of them is straightforward

$$\exp(iG \cdot r) = \sum_n U_{nG}{}^* U_n(r). \qquad (2.3)$$

Substitution for $\exp(\mathrm{i}\boldsymbol{G} \cdot \boldsymbol{r})$ from (3) into (1) gives

$$\psi(\boldsymbol{r}) = \sum_n F_n(\boldsymbol{r}) U_n(\boldsymbol{r}) \tag{2.4}$$

with the envelope functions, $F_n(\boldsymbol{r})$, given by

$$F_n(\boldsymbol{r}) = \sum_{kG} U_{nG}^* \tilde{\psi}(\boldsymbol{k} + \boldsymbol{G}) \exp(\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r}) = \sum_{kG} U_{nG}^* \tilde{\psi}_G(\boldsymbol{k}) \exp(\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r}). \tag{2.5}$$

As mentioned in Burt (1989), it is possible to extend this derivation of the envelope function expansion to include cases in which the basis functions $U_n(\boldsymbol{r})$ all change by the same phase factor, $\exp(\mathrm{i}\boldsymbol{K} \cdot \boldsymbol{a})$, other than unity, on translation by a Bravais lattice vector, $\boldsymbol{a}$. All this requires is for an overall factor $\exp(\mathrm{i}\boldsymbol{K} \cdot \boldsymbol{r})$ to be inserted into the definition of $U_n(\boldsymbol{r})$ which is equivalent to displacing all reciprocal lattice vectors by $\boldsymbol{K}$ and the summation over $\boldsymbol{k}$ remaining unchanged. Such definitions would be useful in problems in which, for instance, zone edge rather than zone centre states dominate.

Now, the envelope function expansion (2.4) is as complete as the original plane wave expansion by construction. It can therefore deal with any problem which can be solved by a plane wave expansion. So contrary to the claims made by Wang *et al* (1997) the method can include such effects as $\Gamma$X coupling in GaAs/AlAs structures. It is true that the expansion (2.4) does not make the $\Gamma$X coupling manifest, but the way to make it manifest has been outlined in Burt (1992) and worked out in detail by Foreman (1998a).

The expansion (2.4) was used by Luttinger and Kohn (1955) to derive effective mass equations for periodic structures in the presence of a 'gentle' perturbing potential. It has consequently become known as the effective mass representation (see e.g. Callaway 1991) and there is a widespread impression that this precludes its use in structures with abrupt heterojunctions (see Wood and Zunger (1996) and Yi and Razeghi (1997) for some recent examples) despite demonstrations to the contrary (see Burt 1994a, b) and perhaps heightened by an unnecessary and unrealistic focus by some authors on graded structures (Geller and Kohn 1993, Geller 1997). The author prefers to call this representation the envelope function, or perhaps in the light of the discussion below, an envelope function representation to try to redress the misleading impression of the term effective mass representation.

In (2.3) we have expanded the Bravais lattice periodic plane waves, $\exp(\mathrm{i}\boldsymbol{G} \cdot \boldsymbol{r})$, in terms of just one set of orthonormal functions $U_n(\boldsymbol{r})$ independent of the wavevector $\boldsymbol{k}$, but there is no need to do so. For each $\boldsymbol{k}$ value one could introduce a different set of complete orthonormal functions, $U_n(\boldsymbol{r}, \boldsymbol{k})$, and end up with envelope functions

$$F_n(\boldsymbol{r}) = \sum_{kG} U_{nG}(\boldsymbol{k})^* \tilde{\psi}_G(\boldsymbol{k}) \exp(\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r}) \tag{2.6}$$

with the wavefunction being given by

$$\psi(\boldsymbol{r}) = \sum_n \sum_{kG} U_{nG}(\boldsymbol{k})^* \tilde{\psi}_G(\boldsymbol{k}) \exp(\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r}) U_n(\boldsymbol{r}, \boldsymbol{k}). \tag{2.7}$$

For the special case in which the orthonormal functions $U_n(\boldsymbol{r}, \boldsymbol{k})$ are just the Bloch functions of a single reference crystal, the envelope functions so defined are the Wannier–Slater envelope functions (Wannier 1937, Slater 1949) and an exact equation can be derived for them (Burt 1992). With a minor modification, the unperturbed energy operator $E_n^{(0)}(-\mathrm{i}\nabla)$ in (9.7) of Burt (1992) is no longer diagonal, the equation also covers the case in which a different reference crystal may be used to generate $U_n(\boldsymbol{r}, \boldsymbol{k})$ at each $\boldsymbol{k}$ value. Indeed, one can go one stage further and use any set of periodic functions $U_n(\boldsymbol{r}, \boldsymbol{k})$ at each $\boldsymbol{k}$ value, including functions that correspond to more than one reference crystal, provided the set is not over-complete. One then

arrives at the real space version of the linear combination of bulk bands method of Wang *et al* (1997).

Returning to the envelope function expansion (2.4), it is important to appreciate that the envelope functions, $F_n(r)$, so defined are a *bona fide* representation of the state, $|\psi\rangle$, on an equal footing, in the first instance, with any other representation such as the coordinate representation, $\psi(r)$, itself. As discussed elsewhere (Burt 1992), there is no need to restrict the application of (2.4) to stationary states. Contrast this with the conventional envelope function expansion (see e.g. Altarelli 1983a, b, c)

$$\psi(r) = \sum_n f_n(r)U_n(r) \tag{2.8}$$

in which the envelope functions, $f_n(r)$, are defined so that $\psi(r)$ is the solution of the stationary Schrödinger equation for a particular crystal perhaps in the presence of an external potential. (The $U_n(r)$ are usually defined to be the zone centre eigenfunctions for the crystal in question.) For flat band cases, the $f_n(r)$ will be a combination of propagating and evanescent plane waves and will obey a set of coupled local differential equations. The plane wave expansion of $f_n(r)$ will not be restricted, in general, to plane waves with wavevectors inside the first Brillouin zone in contrast to the $F_n(r)$ defined by (2.5). So $F_n(r)$ and $f_n(r)$ are not the same functions even if the same $U_n(r)$ are used in their definition and the $f_n(r)$ are constructed solely from 'in zone' solutions (see the example at the end of this section) though there are situations in which they differ little (Burt 1994b). The $F_n(r)$ obey nonlocal rather than local equations. Further discussion on the relation between $F_n(r)$ and $f_n(r)$ can be found in section 9 of Burt (1988a) and in Foreman (1998b).

It is perhaps worth looking at a simple extreme example of an envelope function representation to illustrate how envelope functions can differ greatly from the original functions they represent. Consider the case of a delta function, $\delta(x)$, at the origin. Let the reference Bravais lattice have period $a$ and apply periodic boundary conditions over a large length $L = Na$. The Fourier transform $\tilde{\delta}_G(k)$ is

$$\tilde{\delta}_G(k) = \int \frac{dx}{L} \delta(x) \exp(i(k + G)x) = \frac{1}{L}. \tag{2.9}$$

Using the plane waves $\exp(iGx)$ as the periodic basis functions, we find that the envelope functions, $F_G(x)$, are given by

$$F_G(x) = \sum_k \tilde{\delta}_G(k) \exp(ikx) = \frac{1}{L} \sum_k \exp(ikx) = \frac{\sin(\pi x/a)}{\pi x}. \tag{2.10}$$

It would now appear that we have a very poor representation of the delta function, displaying all the Gibbs oscillations effects one would expect for a truncated plane wave expansion. But the representation is accurate since there is not just one envelope function representing the delta function, but infinitely many: an $F_G(x)$ for each and every $G$. We recover the original delta function precisely on evaluating the envelope function expansion:

$$\sum_G F_G(x) \exp(iGx) = \sum_G \frac{\sin(\pi x/a)}{\pi x} \exp(iGx) = \frac{\sin(\pi x/a)}{\pi x} a \sum_n \delta(x - na) = \delta(x). \tag{2.11}$$

So we see that the representation is exact even though each envelope function appears inadequate.

In the discussion following equation (2.8) it was noted that the plane wave expansion of $f_n(r)$ will not be restricted, in general, to plane waves with wavevectors inside the first Brillouin zone. Even if the $f_n(r)$ are constructed solely from 'in zone' propagating solutions

this is still true and we can illustrate this using a simple textbook example—the wavefunction for an electron reflected/transmitted at a simple step. The electron is incident on the step situated at $x = 0$ from the left so the wavefunction is

$$\psi = \exp(ik^{(-)}x) + r\exp(-ik^{(-)}x) \qquad \text{for } x < 0 \qquad (2.12a)$$

$$\psi = t\exp(ik^{(+)}x) \qquad \text{for } x > 0 \qquad (2.12b)$$

where $r$ and $t$ are the amplitudes for reflection and transmission respectively and $k^{(\pm)}$, in an obvious notation, are the wavevectors either side of the step. The Fourier transform of the wavefunction is given by

$$\tilde{\psi}(q) = \int \psi(x)\exp(-iqx)\,dx = i\left[\frac{1}{q - k^{(-)}} + \frac{r}{q + k^{(-)}} - \frac{t}{q - k^{(+)}}\right]. \qquad (2.13)$$

Now consider the envelope function representation of this function using the 'empty lattice' with lattice period $a$ so, as in the case of the delta function treated above, our periodic basis functions are the plane waves $\exp(iGx)$ with $\exp(iGa) = 1$. Even if $k^{(\pm)}$ both lie in the first Brillouin zone i.e. $k^{(\pm)} < G/2$, $\tilde{\psi}(q)$ is nonzero outside the first Brillouin zone and, in principle, needs more than one term in an envelope function expansion to represent it. So, even if one is using 'in zone' propagating Bloch waves one should not confuse the exponentials associated with them with the envelope functions $F_n(r)$.

## 3. Use of the uncertainty principle

Hagston *et al* (1994) have claimed that the use of the 'kinetic energy' operator $T = -(\hbar^2/2)(d/dx)(1/m^*)(d/dx)$ is in conflict with the uncertainty principle. They point out that, if all other parameters are kept constant, the energy of the ground state tends to zero as the effective mass in the barrier tends to infinity. Zero energy would seem to imply zero uncertainty, $\Delta p$, in the momentum. Since the electron is confined to the well and there is no penetration of the barrier, we expect the uncertainty, $\Delta x$, in position to be less than, or in the order of, the well width and therefore finite. This leads to a zero $\Delta x\,\Delta p$ product in the limit of large barrier mass and hence a contradiction with the uncertainty principle. This suggests that the author's derivations (Burt 1988a, b, 1992, 1994a) of this Hamiltonian are in error. Indeed, the author was at one time (Burt 1989) misled by this argument into believing that the validity of his then current derivation of the effective mass equation had much more limited validity than subsequently turned out to be the case. We shall see, however, that the above mentioned argument is not a valid objection.

It is easy to provide what appears to be a resolution of this conflict by claiming that one should apply the uncertainty principle to the uncertainties evaluated using the wavefunction rather than to the envelope function. Then the zero energy state is seen to have a nonzero value, in general, for $\Delta p$ because of the confinement due to bonding, i.e. we expect $\Delta p \sim \hbar/a$ where $a$ is of atomic dimensions. This would lead to a resolution of the conflict except perhaps in some extreme cases, e.g. strong confinement in narrow quantum wells where the globally slowly varying envelope assumption, on which the simple effective mass Hamiltonian is based, would probably not hold for this particular and unusual envelope function or where interface effects may also be important.

However, the above resolution does not take into account that one can derive the uncertainty principle for almost any function of physical interest and this includes envelope functions as well as wavefunctions. All one needs to apply it to an envelope function, $f$, is the existence of the norm for $f$, $xf$ and $pf$. Since these all exist for the ground state of a single isolated quantum well, the $\Delta x\,\Delta p$ product must obey the uncertainty principle. To demonstrate this

for a simple case we will evaluate the uncertainties in the limit of large barrier mass, $m_b$. Take the origin at the centre of the well of width $2l$ and use the appropriate approximation to the envelope function:

$$f = \sqrt{\frac{1}{2l}} \cos kx \qquad\qquad |x| \leqslant l$$

$$f = \sqrt{\frac{1}{2l}} \exp(-q(x - l)) \qquad x \geqslant l$$

$$f = \sqrt{\frac{1}{2l}} \exp(+q(x + l)) \qquad x \leqslant -l$$

where we have used $\cos kl \approx 1$ and ignored the contribution from the barriers to the normalization integral because $q \propto \sqrt{m_b}$ is large. The energy eigenvalue is given by

$$\frac{k}{m_w} \tan kl = \frac{q}{m_b}$$

where $m_w$ is the well mass.

Because $q \propto \sqrt{m_b}$ the right-hand side of the transcendental equation tends to zero as $m_b \to \infty$ and hence $k$ and the ground state energy also tend to zero.

Since the expectation values of both $x$ and $p$ are zero, the uncertainties $\Delta x$ and $\Delta p$ are just given by the square root of the norm of $|xf\rangle$ and $|pf\rangle$ respectively. The contribution to the norm of $|xf\rangle$ from the barriers vanishes in the limit of large barrier mass, because there is no appreciable penetration into the barrier. One readily finds

$$\langle x^2 \rangle = \langle xf|xf\rangle = \frac{l^2}{3}.$$

On the other hand, the contribution to the norm of $|pf\rangle$ i.e. to $\langle pf|pf\rangle = \hbar^2 \int (\mathrm{d}f/\mathrm{d}x)^2 \, \mathrm{d}x$ from the well vanishes in the limit of large barrier mass because the envelope function is constant there. One might suppose, as in the evaluation of the norm of $|xf\rangle$, that there is no contribution from the barriers because there is vanishing penetration. However, the slope of the envelope is tending to infinity with the barrier mass and hence we cannot just assume that the contribution vanishes. Indeed, the contribution from the barriers is easily evaluated to give $\langle p^2 \rangle = \langle pf|pf\rangle = \hbar^2 q/2l$. Hence our uncertainty product is

$$\Delta x \Delta p = \hbar \sqrt{\frac{ql}{6}} \propto (m_b)^{1/4} \to \infty \text{ as } m_b \to \infty$$

which is not in conflict with the uncertainty principle.

Of course, in the limit $m_b \to \infty$, $f$ has a rectangular shape and is discontinuous at the well–barrier interface; the globally slowly varying envelope function approximation on which the derivation $T = (\hbar^2/2)(\mathrm{d}/\mathrm{d}x)(1/m^*)(\mathrm{d}/\mathrm{d}x)$ is based still works well, however, for wide wells as can be seen from the work of Hemmer and Wang (1993). While this robustness of the globally slowly varying envelope function approximation is satisfying, it is by no means essential to the above argument. All we are concerned with there is the size of the $\Delta x \Delta p$ product predicted by the given form of $T$ to show that no contradiction or conflict with the uncertainty principle arises.

What this whole exercise brings to the fore is the limited help the uncertainty principle can afford us in determining effective Hamiltonians. As mentioned above, the principle is based on minimal assumptions such as the existence of various norms and the Hermiticity of the operators. It is for this very reason that the principle is so widely applicable. By the same token, it is very difficult to find an effective Hamiltonian that yields eigenfunctions for which

the uncertainty principle is violated. If one thinks one has found such a Hamiltonian, then there is more likely to be a mistake in one's evaluation of the uncertainties than a genuine violation of the uncertainty principle. So while any proposed Hamiltonian must not lead to violations of the uncertainty principle, such a test is only a minimal requirement and unlikely to be very useful in practice for eliminating putative effective Hamiltonians.

## 4. Valence band Hamiltonians

The correct way to construct valence band Hamiltonians for inhomogeneous systems from Luttinger Hamiltonians for the bulk constituents is best illustrated using the simplest of examples. Consider the Luttinger (1956) Hamiltonian for a zincblende semiconductor in the absence of spin–orbit interaction where we have excluded the effect of all remote bands save the lowest conduction band i.e.

$$
\begin{bmatrix}
E_V + ak^2 - bk_xk_x & -bk_xk_y & -bk_xk_z \\
-bk_yk_x & E_V + ak^2 - bk_yk_y & -bk_yk_z \\
-bk_zk_x & -bk_zk_y & E_V + ak^2 - bk_zk_z
\end{bmatrix}
\tag{4.1}
$$

where $E_V$ is the energy of the three degenerate states at the top of the valence band, which we chose to be real and denote by $X$, $Y$ and $Z$. Following Kane (1957) we denote the lowest conduction band state by $iS$, where $S$ is real, as this makes the interband matrix element $p = \langle iS|p_x|X\rangle = \langle iS|p_y|Y\rangle = \langle iS|p_z|Z\rangle$ real. $a = \hbar^2/2m$ and $b = P^2/E_G$ where $P = \hbar p/m$. This Hamiltonian may be derived easily from the four band $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian by the Löwdin (1951) perturbation method (see also Heine and Cohen 1970) or put more simply by the approximate elimination of the small conduction band amplitude in favour of the valence band amplitudes. However, for the reader unfamiliar with these techniques, it is probably easier to start afresh with the treatment below which is applicable to both homogeneous and inhomogeneous systems. Equation (1) can then be verified as a by-product of the wider consideration of inhomogeneous systems. Anyway, given (1), the problem is how to generalize it to inhomogeneous structures in which the material parameters, just the coefficient $b$ and the energy $E_V$ in this simple case, are a function of position. Transforming the Hamiltonian (4.1) for the bulk, homogeneous system case is easy: one just makes the replacement $\mathbf{k} = -i\boldsymbol{\nabla}$. But in inhomogeneous systems $b$ depends on position and there is ambiguity as to how to order the derivatives and $b$. The most commonly adopted procedure is to make every element of (4.1) individually Hermitian by symmetrization (for an early example see Lin-Lui and Sham 1985, for evidence of wide acceptance Bastard 1988, Bastard *et al* 1991, its continued use Bockelmann and Bastard 1992, Chao and Chuang 1992, Ando *et al* 1993 Edwards *et al* 1994 and for a recent example Bobbert *et al* 1997), e.g. the element in row 1 column 2 in (4.1), $-bk_xk_y$, is replaced by

$$
-\tfrac{1}{2}(bk_xk_y + k_xk_yb)
\tag{4.2}
$$

and then the substitution $\mathbf{k} = -i\boldsymbol{\nabla}$ is made to obtain

$$
\frac{1}{2}\left(b\frac{\partial^2}{\partial x\partial y} + \frac{\partial^2}{\partial x\partial y}b\right).
\tag{4.3}
$$

The unphysical nature of solutions using Hamiltonians derived from this symmetrization procedure was highlighted by Foreman (1993). For a quantum well structure he showed that more physically plausible solutions were obtained using the Hamiltonian derived from first principles by Burt (1992). Meney *et al* (1994) demonstrated numerically that the Hamiltonian derived by Burt (1992) is consistent with the four band '$\mathbf{k} \cdot \mathbf{p}$' Hamiltonian while the symmetrized version was not. The calculations of both Foreman (1993) and

Meney *et al* (1994) were for (001) oriented quantum wells. Stavrinou and van Dalen (1997) have derived a specific Hamiltonian for (110) oriented quantum wells of zincblende crystals and performed calculations for GaInAs/InP. They show that there are similar differences between the results derived from the author's (Burt 1992) Hamiltonian and those derived from symmetrized Hamiltonians to those seen for the (001) orientation. van Dalen and Stavrinou (1998) have also derived Hamiltonians for general directions. Definitive evidence that the traditional symmetrized Hamiltonian is incorrect has been provided by comparison between semiempirical pseudopotential calculations and envelope function calculations for (001) quantum wells (Coles and Stavrinou 1998). The results show strong support for the author's Hamiltonian.

To see how the correct (Burt 1992) Hamiltonian is derived for the simple case mentioned above, start with the four band envelope function $\mathbf{k} \cdot \mathbf{p}$ equations,

$$\left(E_V - E - \frac{\hbar^2}{2m}\nabla^2\right) F_X - \frac{\mathrm{i}\hbar}{m}p\frac{\partial F_S}{\partial x} = 0 \tag{4.4a}$$

$$\left(E_V - E - \frac{\hbar^2}{2m}\nabla^2\right) F_Y - \frac{\mathrm{i}\hbar}{m}p\frac{\partial F_S}{\partial y} = 0 \tag{4.4b}$$

$$\left(E_V - E - \frac{\hbar^2}{2m}\nabla^2\right) F_Z - \frac{\mathrm{i}\hbar}{m}p\frac{\partial F_S}{\partial z} = 0 \tag{4.4c}$$

$$\left(E_S - E - \frac{\hbar^2}{2m}\nabla^2\right) F_S - \frac{\mathrm{i}\hbar}{m}p\left(\frac{\partial F_X}{\partial x} + \frac{\partial F_Y}{\partial x} + \frac{\partial F_Z}{\partial z}\right) = 0. \tag{4.4d}$$

The wavefunction is given by $\mathrm{i}F_S S + F_X X + F_Y Y + F_Z Z$. We will not dwell, in detail, on the general circumstances in which we are justified in writing down such equations, as this has been discussed elsewhere (Burt 1992, 1994a, Foreman 1996, Burt and Foreman 1998). If the reader would like to focus on a simple example, albeit artificial, in which these equations can undoubtedly be justified, then consider the case of a structure in which all the component crystals have band structures well described by the nearly free electron limit. In that limit, the zone centre Bloch functions in any component crystal are the same as in any other: they are just symmetrized sets of plane waves of equal kinetic energy; for the highest valence band and lowest conduction band considered here the Bloch functions are symmetrized sets of (111) plane waves. So, if these nearly free electron or 'empty lattice' Bloch functions are used as a basis, then there will be no off-diagonal terms in the bulk Hamiltonian matrix. Band gaps will be small, but there is no need for them to be similar and in relative terms they could vary dramatically so that correct operator ordering is still important. If the resultant envelope functions are sufficiently slowly varying, then interface terms (see both section 5 and appendix 4 of Burt 1992) that couple the conduction band envelope, $F_S$, to the valence band envelopes, $F_X$, $F_Y$ and $F_Z$, can be neglected. In principle, the interface terms in (4.4a), (4.4b) and (4.4c) connecting $F_X$, $F_Y$ and $F_Z$ could still be important (Foreman 1998a). But including them would detract from the clarity of the following and they will be omitted, a step which can be justified if there is sufficient smoothing of the microscopic potential in the interface regions. Also, to keep things as simple as possible, neglect the $\mathbf{k} \cdot \mathbf{p}$ coupling to other bands, though that coupling to the $\Gamma_{15}^c$ band will be treated in the appendix.

For valence band problems, the valence band envelope functions are dominant and one may eliminate $F_S$ from the equations for $F_X$ etc. From equation (4.4d) one obtains on putting $E_S - E \approx E_S - E_V = E_G$, the bandgap, and dropping the free electron kinetic energy term

$$F_S \approx \frac{\mathrm{i}\hbar p}{mE_G}\left(\frac{\partial F_X}{\partial x} + \frac{\partial F_Y}{\partial y} + \frac{\partial F_Z}{\partial z}\right). \tag{4.5}$$

We can use this expression for $F_S$ in any of the equations (4.4a), (4.4b) or (4.4c) to obtain

$$\left(E_V - E - \frac{\hbar^2}{2m}\nabla^2\right)F_X + \frac{\partial}{\partial x}\left(\frac{P^2}{E_G}\left(\frac{\partial F_X}{\partial x} + \frac{\partial F_Y}{\partial y} + \frac{\partial F_Z}{\partial z}\right)\right) = 0 \tag{4.6a}$$

and cyclic permutations (4.6b) and (4.6c). Equations (4.6) correspond to Burt's (1992) more general equations. $E_V$ and $E_G$ can depend on all three position coordinates.

Now suppose for simplicity that $E_G$ depends on $z$ alone, i.e. we are dealing with a (001) multilayer structure. Then the term

$$\frac{\partial}{\partial x}\left(\frac{P^2}{E_G}\left(\frac{\partial F_Z}{\partial z}\right)\right) \tag{4.7}$$

in (4.6a) is asymmetric, but, since it is derived, there is no doubt about its validity. It does not represent any departure from Hermiticity because there is a corresponding term

$$\frac{\partial}{\partial z}\left(\frac{P^2}{E_G}\left(\frac{\partial F_X}{\partial x}\right)\right) \tag{4.8}$$

in the equation (4.6c) for $F_Z$. So this Hamiltonian, which has been derived systematically, is Hermitian in its entirety as it must be, but the individual differential operators are not.

It is interesting to see the boundary conditions that the equations (4.6) provide. If we integrate (4.6a) wrt $z$ across an interface $z = $ const, then one obtains

$$\frac{\partial F_X}{\partial z} \text{ continuous} \tag{4.9a}$$

(note that no effective mass enters since we have neglected the higher bands that cause the heavy hole mass to differ from the free electron mass). There is no coupling of this heavy hole band (heavy, that is, in the $z$ direction) to other bands via the boundary. On the other hand, the corresponding equation for the light hole envelope function ($x \to z$ in (4.6a)) i.e.

$$\left(E_V - E - \frac{\hbar^2}{2m}\nabla^2\right)F_Z + \frac{\partial}{\partial z}\left(\frac{P^2}{E_G}\left(\frac{\partial F_X}{\partial x} + \frac{\partial F_Y}{\partial y} + \frac{\partial F_Z}{\partial z}\right)\right) = 0 \tag{4.6c}$$

gives the boundary condition as

$$-\frac{\hbar^2}{2m}\frac{\partial F_Z}{\partial z} + \frac{P^2}{E_G}\left(\frac{\partial F_X}{\partial z} + \frac{\partial F_Y}{\partial y} + \frac{\partial F_Z}{\partial z}\right) \tag{4.9c}$$

continuous.

It is, perhaps, not surprising that the boundary affects heavy and light holes differently. After all, the light holes have underlying p orbitals that are perpendicular to the boundary while those for heavy holes are parallel to it.

If, incorrectly, we had used the symmetrization procedure, then the boundary condition for the heavy hole would become

$$-\frac{\hbar^2}{2m}\frac{\partial F_X}{\partial z} + \frac{P^2}{2E_G}\frac{\partial F_Z}{\partial x} \text{ continuous} \tag{4.10a}$$

which introduces a kink into the envelope $F_X$ when (4.9a) demonstrates that no kink should exist for this case. The corresponding case for the light hole envelope, $F_Z$ gives

$$-\frac{\hbar^2}{2m}\frac{\partial F_Z}{\partial z} + \frac{P^2}{E_G}\left(\frac{1}{2}\left(\frac{\partial F_X}{\partial x} + \frac{\partial F_Y}{\partial y}\right) + \frac{\partial F_Z}{\partial z}\right) \text{ continuous} \tag{4.10c}$$

reducing by half the influence of the variation of the heavy hole band envelopes in the $x$ and $y$ directions.

Now, in this simple case it has been quite easy to determine the correct ordering of the differential operators in the Hamiltonian by working from scratch. But for most cases this is a lengthy procedure and a short cut would be welcome. If we rewrite the equations (4.6) using $k = -i\nabla$, then the corresponding Hamiltonian matrix becomes

$$\begin{bmatrix} E_V + ak^2 - k_x b k_x & -k_x b k_y & -k_x b k_z \\ -k_y b k_x & E_V + ak^2 - k_y b k_y & -k_y b k_z \\ -k_z b k_x & -k_z b k_y & E_V + ak^2 - k_z b k_z. \end{bmatrix} \tag{4.11}$$

One can see a simple rule emerging. For instance, if we did not make use of the symmetry condition

$$p = \langle iS|p_x|X\rangle = \langle iS|p_y|Y\rangle = \langle iS|p_z|Z\rangle$$

and that $p$ is real, then the $b = P^2/E_G$ in row $X$ column $Y$ would be $(\hbar^2/m^2 E_G)$ $\langle X|p_x|iS\rangle\langle iS|p_y|Y\rangle$. That the $p_x$ comes in the first matrix element means that the $k_x$ must come in front in inhomogeneous systems. That the $p_y$ comes in the second matrix element means that the $k_y$ must come behind. It is clear that the often used symmetrization procedure that would give

$$-\tfrac{1}{2}(k_x b k_y + k_y b k_x) \tag{4.12}$$

is wrong.

The case we have treated is simple, but it is representative of the main points that must be appreciated to construct valence band Hamiltonians correctly. However, the reader may be uneasy that the case treated here is too simple because the heavy hole bands 'bend the wrong way'. To assuage such uneasiness the work here is extended in the appendix to include the lowest triplet state in the conduction band which is sufficient to obtain the main features of the heavy hole band—the correct sign for the curvature and the presence of anisotropy. It will also be seen that Luttinger parameters have to be resolved into their component parts and that there is, in principle, a different, though simply understood, ordering for each part.

Finally it should be noted that the possibility of Hamiltonian matrices where the individual differential operators are not Hermitian has arisen long ago in semiconductor physics. In his paper on cyclotron resonance in bulk semiconductors Luttinger (1956) explained that magnetic field effects necessitated such Hamiltonians†.

## 5. The out of zone solution conundrum

The envelope function method is particularly suited to finding electron states in nanostructures in which there are no external applied or internal mesoscopic fields and the chemical composition is piecewise continuous i.e. in so called flat band cases. This is especially true when the interface terms, which arise even for atomically abrupt heterojunctions (Burt 1988a, 1992 section 5 and appendix 4, Foreman 1998a), can also be neglected because the envelope functions are sufficiently slowly varying and the band edge is nondegenerate (Burt 1994a). Whenever one has such flat bands, one can solve the envelope function equations by selecting an energy, constructing suitable general solutions and finding the energies at which one can satisfy the boundary conditions. Such methods generate out of zone solutions which at first appear to be unphysical and it has been argued that they should be rejected (Godfrey and Malik 1996). However, if the envelope function equations have been derived systematically and are consequently valid at interfaces, all the solutions, and this includes the out of zone solutions, are needed to satisfy all the boundary conditions. In this section we will see how the

† My thanks to Dr B A Foreman for drawing my attention to this.

out of zone solutions arise and show that they are legitimate trial solutions in each flat band region and need to be retained to obtain a consistent overall solution to the envelope function equations for the whole structure. Under certain circumstances, as will be shown, the solution will lead to useful approximate energies and wavefunctions. A numerical example to support the author's earlier assertion (Burt 1992) that all such out of zone solutions should be retained in a consistent solution to the envelope function equations has already been given including the successful approximate generation of the wavefunction (Burt 1998). At the end of this section we will demonstrate the calculation of an approximate wavefunction from out of zone solutions alone.

To see how the out of zone solutions arise and to appreciate their role in the overall solution of the envelope function equations, it is helpful to follow the author's procedure (Burt 1994a) for deriving approximate envelope function equations directly from the Schrödinger equation in the plane wave representation. The wavefunction is expanded in plane waves

$$\psi(\mathbf{r}) = \sum_{kG} \tilde{\psi}_G(\mathbf{k}) \, e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{5.1}$$

where the wavevector has been expressed as $\mathbf{k} + \mathbf{G}$, $\mathbf{G}$ being a reciprocal lattice vector of the underlying Bravais lattice and $\mathbf{k}$ is inside a Brillouin zone defined with respect to that lattice. (Cyclic boundary conditions are applied to a large macroscopic volume completely enclosing the nanostructure.) The Schrödinger equation in this representation is

$$\frac{\hbar^2}{2m}(\mathbf{k}+\mathbf{G})^2\tilde{\psi}_G(\mathbf{k}) + \sum_{k'G'}\langle \mathbf{k}+\mathbf{G}|V|\mathbf{k}'+\mathbf{G}'\rangle\tilde{\psi}_{G'}(\mathbf{k}') = E\tilde{\psi}_G(\mathbf{k}) \tag{5.2}$$

where $V$ is the microscopic potential which is assumed to be local. (This is not essential for all that follows, but is in keeping with the numerical example given at the end of the section.) By taking a Fourier transform of this equation wrt to both $\mathbf{k}$ and $\mathbf{k}'$ one would obtain the author's exact envelope function equations, including nonlocality, with the plane waves $\exp(i\mathbf{G}\cdot\mathbf{r})$ as the periodic basis functions. Now consider a modified form of (5.2)

$$\frac{\hbar^2}{2m}(\mathbf{q}+\mathbf{G})^2\tilde{\phi}_G(\mathbf{q}) + \sum_{q'G'}\langle \mathbf{q}+\mathbf{G}|V|\mathbf{q}'+\mathbf{G}'\rangle\tilde{\phi}_{G'}(\mathbf{q}') = E'\tilde{\phi}_G(\mathbf{q}) \tag{5.3}$$

in which the wavevector $\mathbf{q}$ ranges over all values both inside and outside the Brillouin zone. The matrix element $\langle \mathbf{q}+\mathbf{G}|V|\mathbf{q}'+\mathbf{G}'\rangle$ is defined for fixed $\mathbf{G}$ and $\mathbf{G}'$ by $\tilde{V}(\mathbf{q}-\mathbf{q}'+\mathbf{G}-\mathbf{G}')$ where $\tilde{V}$ is the Fourier transform of the potential $V(\mathbf{r})$.

We note that if $\tilde{\phi}_G(\mathbf{q})$ is a solution, then so is

$$\tilde{\phi}_G^{(n)}(\mathbf{q}) = \tilde{\phi}_{G+G_n}(\mathbf{q}-\mathbf{G}_n) \tag{5.4}$$

with envelope functions

$$\phi_G^{(n)}(\mathbf{r}) = \sum_q \tilde{\phi}_G^{(n)}(\mathbf{q})\exp(i\mathbf{q}\cdot\mathbf{r}). \tag{5.5}$$

Now, if the $\tilde{\phi}_G(\mathbf{q})$ are only appreciable for $\mathbf{q} \approx \mathbf{0}$ i.e. we have a globally slowly varying function, then $\tilde{\phi}_G^{(n)}$ will only be appreciable for $\mathbf{q} \approx \mathbf{G}_n$ and the envelope functions, $\phi_G^{(n)}(\mathbf{r})$, will be rapidly varying and be denoted as out of zone solutions even though the associated function in real space, $\phi^{(n)}(\mathbf{r})$, is, in fact, identical to $\phi(\mathbf{r}) = \sum_{qG} \tilde{\phi}_G(\mathbf{q})\exp(i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r})$ the analogue of $\psi(\mathbf{r})$. For,

$$\phi^{(n)}(\mathbf{r}) = \sum_{qG}\tilde{\phi}_G^{(n)}(\mathbf{q})\exp(i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}) = \sum_{qG}\tilde{\phi}_{G+G_n}(\mathbf{q}-\mathbf{G}_n)\exp(i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r})$$

$$= \sum_{qG}\tilde{\phi}_G(\mathbf{q})\exp(i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}) = \phi(\mathbf{r}). \tag{5.6}$$

So we see that these out of zone envelope function solutions, $\phi_G^{(n)}(r)$, just represent replicas of the original solution, $\phi(r)$, to the corresponding real space equation.

On the basis of the present discussion alone, there is no reason to believe that the functions $\phi(r)$ will be good approximations to the wavefunctions, $\psi(r)$, or that their eigenvalues $E$ and $E'$ will coincide because there is interzone coupling in equation (3) which is not present in (2). The matrix element $\langle q + G|V|q' + G'\rangle = \tilde{V}(q - q' + G - G')$ will generally be appreciable, that is we have interzone coupling, when $q$ and $q'$ are in different zones and will only be small when $q - q' + G - G'$ is sufficiently large.

Now there are situations for which the above mentioned interzone coupling may be small. Suppose one constructs a Brillouin zone in the form of a parallelepiped centred on the origin of the reciprocal lattice. If the primitive reciprocal lattice vectors are denoted by $G_1$, $G_2$ and $G_3$, then a wavevector $k$ can be expressed as $k_1G_1 + k_2G_2 + k_3G_3$ and will be in this Brillouin zone if $-\frac{1}{2} < k_i \leqslant +\frac{1}{2}$ for all three components. We will say it is in the 'inner region' of the Brillouin zone if $-\frac{1}{4} < k_i \leqslant +\frac{1}{4}$ for all three components and in the 'outer region' otherwise. Suppose we are only interested in wavefunctions for which $\tilde{\psi}_G(k)$ is zero for $k$ within the 'outer region' of the Brillouin zone. Then the expectation value for the RHS of (5.2) will only involve wavevectors $k$ and $k'$ such that $k - k'$ is within the Brillouin zone. Suppose, further, that we were to truncate the matrix element $\langle k + G|V|k' + G'\rangle$ for fixed $G$ and $G'$ so that it is zero when $k - k'$ is outside the Brillouin zone and consequently the matrix element $\langle q + G|V|q' + G'\rangle$ for fixed $G$ and $G'$ would only be nonzero if $q - q'$ is within the Brillouin zone. Then both $\tilde{\psi}_G(k)$ and the expectation value, $E$, of the energy given by (5.2) would be unchanged and $\tilde{\psi}_G(k)$ would also be a solution of (5.3) as would the replicas $\tilde{\psi}_{G+G_n}(k - G_n)$ and the eigenvalues $E$ and $E'$ would coincide. The truncation of the matrix element $\langle k + G|V|k' + G'\rangle$ looks severe at first since there is no reason to suppose that it will always be small when $k - k'$ is outside the Brillouin zone. However, when $\psi(r)$ is globally slowly varying, we have seen (Burt 1994a) that it is reasonable to neglect terms in $\langle k + G|V|k' + G'\rangle$ with large wavevector denominators as a first approximation and we will denote the matrix element so approximated as $\langle k + G|V|k' + G'\rangle_{\text{swd}}$, the suffix denoting that only terms with small wavevector denominators have been retained. Equation (5.2) in this approximation is

$$\frac{\hbar^2}{2m}(k + G)^2 \tilde{\psi}_G(k)_{\text{swd}} + \sum_{k'G'}\langle k + G|V|k' + G'\rangle_{\text{swd}}\tilde{\psi}_{G'}(k')_{\text{swd}} = E_{\text{swd}}\tilde{\psi}_G(k)_{\text{swd}} \qquad (5.7)$$

where $E_{\text{swd}}$ is the approximate energy and the $\tilde{\psi}_G(k)_{\text{swd}}$ are the plane wave expansion coefficients of the approximate wavefunction. This approximation is equivalent to ignoring all interface terms. $\langle k + G|V|k' + G'\rangle_{\text{swd}}$ becomes small for $k - k'$ outside the Brillouin zone, so we may anticipate that the solutions of the analogue of (5.3,) i.e.

$$\frac{\hbar^2}{2m}(q + G)^2 \tilde{\phi}_G(q)_{\text{swd}} + \sum_{q'G'}\langle q + G|V|q' + G'\rangle_{\text{swd}}\tilde{\phi}_{G'}(q')_{\text{swd}} = E'_{\text{swd}}\tilde{\phi}_G(q)_{\text{swd}} \qquad (5.8)$$

will be close to those of (5.7), which in turn are close to those of (5.2) and we can work with envelope functions regardless of whether or not they are slowly varying. In the above we have been careful to distinguish between the eigenvalues $E$, $E'$, $E_{\text{swd}}$ and $E'_{\text{swd}}$. It is the implicit assumption that $E$ and $E'_{\text{swd}}$ are identical that invalidates the claim of Karavaev and Krivorotov (1996) to have shown that the real space version of (5.8) is an exact equation for envelope functions. However, $E$ and $E'_{\text{swd}}$ may well be close in many cases.

As an example of the use of out of zone solutions in solving envelope function equations consider the one dimensional superlattice model used previously (Burt 1992, 1994a). The

matrix elements, $\langle k + G|V|k' + G'\rangle$, of the microscopic potential for $k \neq k'$ are given by

$$\langle k + G|V|k' + G'\rangle = 2\sin((k-k')L_B/2)\sum_{G''}\frac{V_{G''}^{(B)} - V_{G''}^{(W)}}{(k-k'+G-G'-G'')L} \tag{5.9}$$

and for small $k$ and $k'$ can be approximated by

$$\langle k + G|V|k' + G'\rangle_{\text{swd}} = \frac{2\sin((k-k')L_B/2)}{(k-k')L}(V_{G-G'}^{(B)} - V_{G-G'}^{(W)}). \tag{5.10}$$

One arrives at (5.10) by ignoring all terms with $G'' \neq G - G'$ in the sum on the RHS of (5.9) i.e. all the terms with large wavevector denominators. The matrix element $\langle k + G|V|k' + G'\rangle$ can be written as $\tilde{V}(k - k' + G - G')$ where the $\tilde{V}(q)$ are the plane wave expansion coefficients (or Fourier transform) of the microscopic potential. In figure 1, the matrix element $\langle k+G|V|k'+G'\rangle$ is given exactly and in the approximation (5.10) for $G = G'$. We see that the approximation is remarkably good and is even better for $G \neq G'$ (not shown). This is remarkable considering the large change, approaching a factor of 2, in form factors between well and barrier.
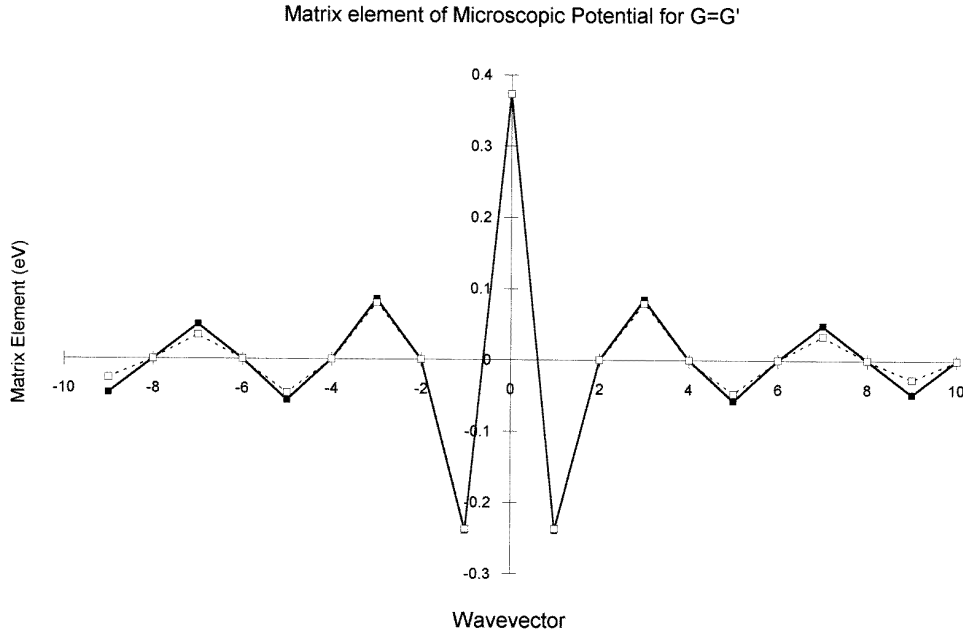
Matrix element of Microscopic Potential for G=G'



**Figure 1.** Plot of $\langle k + G|V|k' + G\rangle$ against $k - k'$ in units of $2\pi/L$ where $L$ is the superlattice period ($L = 20a$ where $a$ is the Bravais lattice period). The data points span the first Brillouin zone. The dark squares, which are joined by solid lines, give the values of the exact expression, (5.9), which for $G = G'$ is $2\sin((k-k')L_B/2)\sum_{G''}(V_{G''}^{(B)} - V_{G''}^{(W)})/((k-k'-G'')L)$. The light squares, which are joined by dotted lines, give the values of the approximate expression, (5.10), which for $G = G'$ is $[2\sin((k-k')L_B/2)/(k-k')L](V_0^{(B)} - V_0^{(W)})$, in which terms with large wavevector denominators are ignored.

When the terms with the large wavevector denominators in the matrix element $\langle k + G|V|k' + G'\rangle$ are ignored i.e. $\langle k + G|V|k' + G'\rangle$ is replaced by $\langle k + G|V|k' + G'\rangle_{\text{swd}}$ the approximate envelope function equations (Burt 1994a, 1998) derived from (5.8) may be written

$$-\frac{\hbar^2}{2m}\frac{d^2 F_n}{dx^2} - i\frac{\hbar}{m}\sum_{n'}p_{n,n'}\frac{dF_{n'}}{dx} + \sum_{n'}H_{n,n'}F_{n'} = E'_{\text{swd}}F_n \tag{5.11}$$

where the $H_{nn'}(x)$ are the matrix elements of the periodic crystal Hamiltonian of the material at $x$. $H_{nn'}(x)$ will be piecewise constant and depend only on the bulk properties of the crystals of which the superlattice is composed. One cannot solve this set of equations consistently (Burt 1992, 1998) unless one uses the out of zone solutions. However, it is clear that continuation of the approximation (5.10) for the matrix element of $V$ beyond the Brillouin zone will produce little interzone coupling. Hence we should be able to use the out of zone solutions of (5.11) in the well and barrier layers with confidence. As explained in Burt (1992), the boundary conditions demand continuity of both $F_n$ and $dF_n/dx$ for all bands and as a consequence the continuity of the wavefunction and its derivative are assured in this approximation. When these boundary conditions along with all the out of zone solutions are used for the state modelled in Burt (1992, 1994a) it has been found (Burt 1998) that the solution of (5.11) gives the wavefunction with remarkable accuracy. The out of zone solutions that do not give a good description of the band structure are present but with small amplitude; the mathematics has taken care of everything.

However, the analysis given above would suggest that one should be able to retrieve the complete wavefunction from solutions restricted to only one Brillouin zone and this need not be the conventional Brillouin zone centred on $G = 0$. Indeed, we will show explicitly that this can be done for a state in the lowest conduction subband of the model mentioned above. We consider the state corresponding to the superlattice wavevector $0.25(2\pi/L)$, i.e. the state halfway between the zone centre and the zone edge. We solve the equations (5.11) and then project out those parts of the solution in each layer that have real part of their wavevectors in the zone centred about $G = +2\pi/a$ where $a$ is the period of the Bravais lattice (5.86 Å). After normalizing the wavefunction we obtain the plot in figure 2. On this scale the approximate
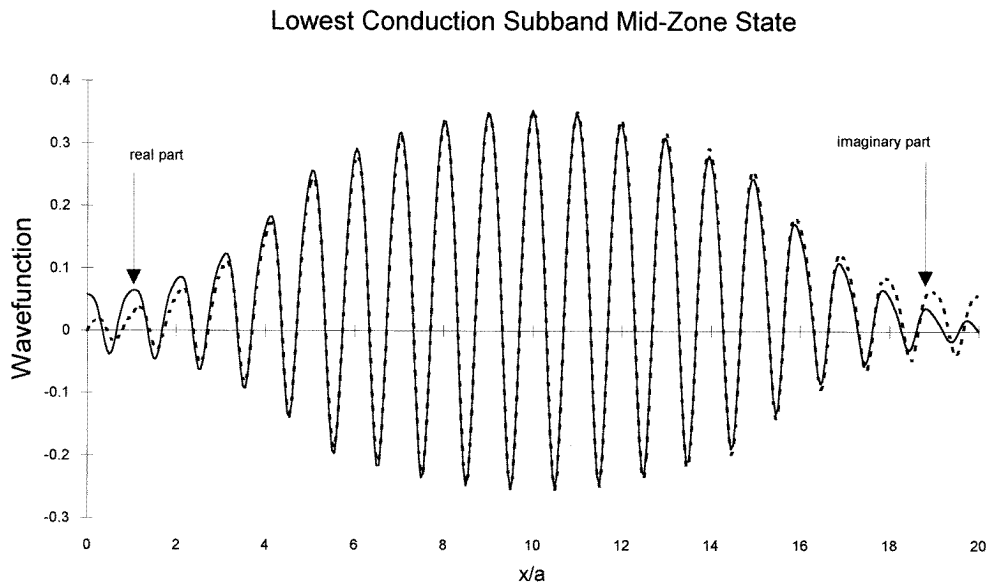


**Figure 2.** Approximate wavefunction for the $10 \times 10$ superlattice model obtained from the envelope function equations (5.11) and by projection out onto the subspace of solutions with real part of the wavevector in the zone centred on $G = +2\pi/a$. The real (imaginary) part is denoted by the solid (dashed) curve.

wavefunction is indistinguishable from that derived directly from diagonalizing the plane wave Hamiltonian. Therefore in figure 3 the plot of the absolute error in the approximate wavefunction depicted in figure 2 is given. Note that the scale in figure 3 is 10 times larger than in figure 2 so that one sees the approximation is very good.
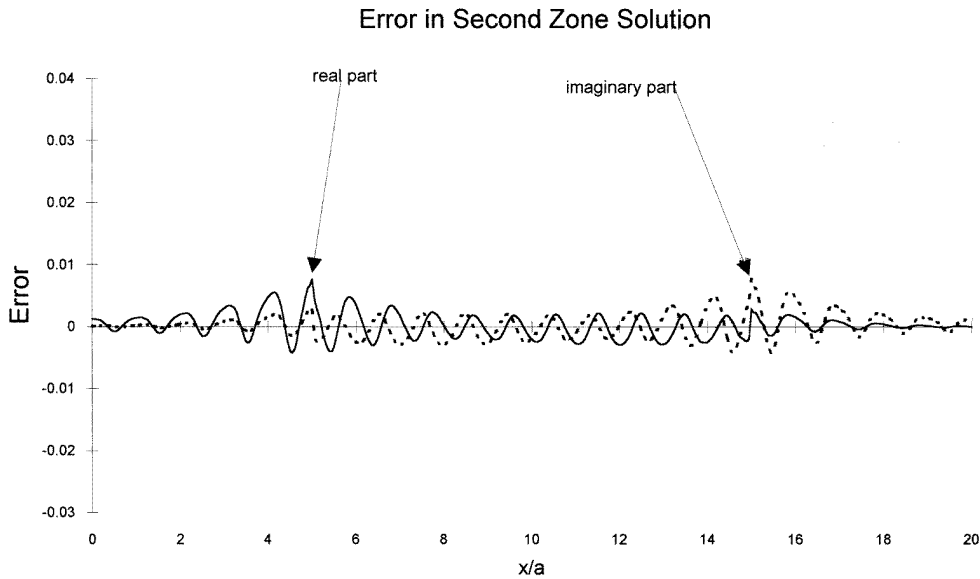
### Error in Second Zone Solution



**Figure 3.** Error in the approximate wavefunction plotted in figure 2. Note that the scale on the vertical axis is ten times larger than in figure 2.

In the matching problem all the out of zone solutions were used including oscillatory solutions with large wavevector in the barrier. The amplitude of such oscillatory solutions turned out to be small and consequently they did not cause a problem. Godfrey and Malik (1996) point out that such oscillatory trial solutions are unrenormalizable for an isolated quantum well. Imposing cyclic boundary conditions on the states of an isolated quantum well, i.e. regarding the quantum well as part of a large period superlattice, allows one to retain these otherwise unrenormalizable trial solutions in the barrier and gets round the problem. Of course, the whole procedure of using out of zone solutions, in general, depends on a justification such as small interzone coupling discussed earlier in the section and a sufficiently large envelope function expansion to give a good description of the wavefunction.

### 6. Interband dipole matrix element

The evaluation of interband dipole matrix elements needs great care especially so when using the envelope function representation (Burt 1993). It seems to be widely believed that the dipole matrix element for an allowed interband transition is of atomic dimensions and is given by the matrix element of the position operator essentially between the appropriate band edge Bloch functions (Weisbuch and Vinter 1991, Haug and Koch 1990, Rosencher *et al* 1996). In a series of papers (Burt 1993, 1995a, b, Coles *et al* 1998) the author and others have shown how to take the dipole matrix element correctly and that values much larger than atomic dimensions are possible for small bandgap semiconductors, though atomic size dipole matrix elements are obtained by the method in the tight-binding limit, as one would expect with large bandgaps

(Elliot 1957). The essential point is that the matrix element is not well approximated by using the dominant term in the envelope function expansion of each wavefunction and, indeed, in the limit of globally slowly varying envelope functions, this supposedly dominant contribution is irrelevant. However, it turns out that it is unnecessary to restrict the discussion to globally slowly varying envelope functions and one can extend the basic results with a zone edge related correction, usually small, to include all envelope functions. Since, as we have emphasized in earlier sections, the envelope functions can be viewed as forming a *bona fide* representation of the wavefunction even when they are not globally slowly varying, it is important to discuss these stronger results. As all the essential points can be explained for a one dimensional system we treat this case alone to make the discussion as transparent as possible.

When the wavefunction, $\psi$, in the interband dipole matrix element,

$$D^{(c,v)} = \int_{-\infty}^{+\infty} \psi^{(c)}(x)^* x \psi^{(v)}(x)\, dx \tag{6.1}$$

is approximated by just the dominant term in the envelope function expansion, the resulting integral

$$I_{cv} = \int_{-\infty}^{+\infty} (F_c^{(c)}(x) U_c(x))^* x F_v^{(v)}(x) U_v(x)\, dx \tag{6.2}$$

is in no way a good approximation to $D^{(c,v)}$ no matter how closely $FU$ approximates $\psi$. In the limit of slowly varying envelope functions $I_{cv}$ is zero. However, from Luttinger and Kohn (1955) and also e.g. Callaway (1991), it would appear that a much stronger result is possible: $I_{cv}$ is zero no matter how rapidly the envelope functions vary (provided their plane wave expansion is restricted to the Brillouin zone). The result would appear to follow naturally from Blount's (1962) treatment, just replacing the wavevector dependent Bloch functions with their zone centre wavevector independent counterparts. However, Blount (1962) was careful to include the whole of the wavefunction in his treatment and argued that there would be no discontinuities in the Fourier transform of the wavefunction under favourable circumstances. But, if one considers individual terms of the envelope function expansion, one cannot make this assumption and the more careful analysis below shows that $I_{cv}$ is not necessarily zero. It will become clear that while $I_{cv}$ is not strictly zero, it will often be small.

The more careful analysis proceeds as follows:

First expand the wavefunctions in plane waves

$$\psi(x) = \int_{-\infty}^{+\infty} dq\, \tilde{\psi}(q) \exp(iqx) \tag{6.3a}$$

$$\tilde{\psi}(q) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \psi(x) \exp(-iqx)\, dx \tag{6.3b}$$

and substitute into (6.1) for $D^{(c,v)}$.

$$
\begin{aligned}
D^{(c,v)} &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dq \int_{-\infty}^{+\infty} dq'\, \tilde{\psi}^{(c)}(q)^* \tilde{\psi}^{(v)}(q') x \exp(-i(q-q')x) \\
D^{(c,v)} &= i \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dq \int_{-\infty}^{+\infty} dq'\, \tilde{\psi}^{(c)}(q)^* \tilde{\psi}^{(v)}(q') \frac{\partial}{\partial q} \exp(-i(q-q')x) \\
D^{(c,v)} &= -2\pi i \int_{-\infty}^{+\infty} dq \int_{-\infty}^{+\infty} dq'\, \frac{d\tilde{\psi}^{(c)}(q)^*}{dq} \tilde{\psi}^{(v)}(q') \delta(q-q') \\
D^{(c,v)} &= -2\pi i \int_{-\infty}^{+\infty} dq\, \frac{d\tilde{\psi}^{(c)}(q)^*}{dq} \tilde{\psi}^{(v)}(q) = 2\pi i \int_{-\infty}^{+\infty} dq\, \tilde{\psi}^{(c)}(q)^* \frac{d\tilde{\psi}^{(v)}(q)}{dq}.
\end{aligned}
\tag{6.4}
$$

Now in terms of the $\tilde{F}(k)$, the plane wave expansion coefficients of the envelope functions, $F(x)$,

$$\tilde{\psi}(q) = \sum_n \tilde{F}_n(k) U_{n,G} \tag{6.5}$$

where $q = k + G$, $k$ lies in the first Brillouin zone and $G$ is a reciprocal lattice vector. Taking just the dominant term in (6.5) and using this in (6.4) gives

$$I_{\mathrm{cv}} = -2\pi\mathrm{i} \int_{-\infty}^{+\infty} \mathrm{d}q \frac{\mathrm{d}}{\mathrm{d}q} (\tilde{F}_{\mathrm{c}}^{(\mathrm{c})}(k) U_{\mathrm{c},G})^* \tilde{F}_{\mathrm{v}}^{(\mathrm{v})}(k) U_{\mathrm{v},G}. \tag{6.6}$$

Now, if we replace differentiation wrt $q$ with differentiation wrt $k$ and integration wrt $q$ with integration wrt $k$ and summation over the reciprocal lattice vectors, $G$, then we obtain

$$I_{\mathrm{cv}} = -2\pi\mathrm{i} \int_{-\pi/a}^{+\pi/a} \mathrm{d}k \frac{\mathrm{d}\tilde{F}_{\mathrm{c}}^{(\mathrm{c})}(k)^*}{\mathrm{d}k} \tilde{F}_{\mathrm{v}}^{(\mathrm{v})}(k) \sum_G U_{\mathrm{c},G}{}^* U_{\mathrm{v},G} \tag{6.7}$$

where $a$ is the lattice period. Of course, so far we have made no specific use that the indices c and v refer to specific bands so we can write for general bands $n$ and $n'$

$$I_{n,n'} = 2\pi\mathrm{i} \int_{-\pi/a}^{+\pi/a} \mathrm{d}k \, \tilde{F}_n^{(n)}(k)^* \frac{\mathrm{d}\tilde{F}_{n'}^{(n')}(k)}{\mathrm{d}k} \sum_G U_{n,G}{}^* U_{n',G}. \tag{6.8}$$

For matrix elements of $x$ for the case of plane wave envelope functions $(1/\sqrt{2\pi}) \exp(\mathrm{i}kx) U_n(x)$ and $(1/\sqrt{2\pi}) \exp(\mathrm{i}k'x) U_{n'}(x)$ we obtain

$$\mathrm{i} \frac{\mathrm{d}}{\mathrm{d}k'} \delta(k - k') \delta_{n,n'} \tag{6.9}$$

where we have used the orthonormal property of the $U$, i.e.

$$\sum_G U_{n,G}{}^* U_{n',G} = \delta_{n,n'}. \tag{6.10}$$

Equation (6.7) immediately leads to the result $I_{\mathrm{cv}} = 0$ no matter how rapidly the envelope functions vary and indeed (6.9) would have us believe that all matrix elements of $x$ between $FU$ products from different bands will vanish regardless of how quickly $F$ varies provided its plane wave expansion is restricted to the first Brillouin zone. The problem with this result is that it is does not take into account possible discontinuities in $\tilde{F}_{\mathrm{c}}^{(\mathrm{c})}(k) U_{\mathrm{c},G}$ or $\tilde{F}_{\mathrm{v}}^{(\mathrm{v})}(k) U_{\mathrm{v},G}$, regarded as a function of $q$, at the zone boundaries. If one applies the argument in going from (6.6) to (6.7) to the complete wavefunction, then discontinuities at the zone boundaries can generally be disregarded (Blount 1962), but not here. At each zone boundary $\tilde{F}_{\mathrm{c}}^{(\mathrm{c})}(k) U_{\mathrm{c},G}$ will change from $F_{\mathrm{c}}^{(\mathrm{c})}(\pi/a) U_{\mathrm{c},G}$ to $\tilde{F}_{\mathrm{c}}^{(\mathrm{c})}(-\pi/a) U_{\mathrm{c},G+2\pi/a}$. Consequently from (6.6)

$$I_{\mathrm{cv}} = \pi\mathrm{i} \sum_G \{ [\tilde{F}_{\mathrm{c}}^{(\mathrm{c})}(\pi/a) U_{\mathrm{c},G} - \tilde{F}_{\mathrm{c}}^{(\mathrm{c})}(-\pi/a) U_{\mathrm{c},G+2\pi/a}]^*$$
$$\times [\tilde{F}_{\mathrm{v}}^{(\mathrm{v})}(-\pi/a) U_{\mathrm{v},G+2\pi/a} + \tilde{F}_{\mathrm{v}}^{(\mathrm{v})}(\pi/a) U_{\mathrm{v},G}] \} \tag{6.11}$$

or

$$I_{\mathrm{cv}} = \pi\mathrm{i} \left\{ \tilde{F}_{\mathrm{c}}^{(\mathrm{c})}(\pi/a)^* \tilde{F}_{\mathrm{v}}^{(\mathrm{v})}(-\pi/a) \sum_G U_{\mathrm{c},G}{}^* U_{\mathrm{v},G+2\pi/a} \right.$$
$$\left. - \tilde{F}_{\mathrm{c}}^{(\mathrm{c})}(-\pi/a)^* \tilde{F}_{\mathrm{v}}^{(\mathrm{v})}(\pi/a) \sum_G U_{\mathrm{c},G+2\pi/a}{}^* U_{\mathrm{v},G} \right\}. \tag{6.12}$$

In general we would expect the RHS of (6.12) to be small because it only involves the $\tilde{F}$ evaluated at the zone edge, but the expression is nonetheless useful for checking and

understanding numerical results. In an earlier paper (Burt 1995a), $I_{cv}$ was not found to be zero as (6.9) would suggest it should, even taking into account computational precision, yet it was much smaller than would be expected if the globally slowly varying envelope function approximation used in Burt (1993) were really essential to establish $I_{cv} \approx 0$. Equation (6.12) provides a qualitative and potentially quantitative resolution of this puzzle.

Blount (1962) states that one can ignore discontinuities in the Fourier transform of a function $\varphi(x)$ when taking matrix elements of $x$, if, for instance, $x\varphi(x)$ is square integrable. This can be readily seen from the application of Parseval's theorem to the Fourier transform pairs $x\varphi(x)$ and $i\,d\tilde{\varphi}(q)/dq$ i.e.

$$\int \left| \frac{d\tilde{\varphi}}{dq} \right|^2 dq = \int \frac{dx}{2\pi} x^2 |\varphi|^2. \tag{6.13}$$

If there were discontinuities in $\tilde{\varphi}(q)$, then $d\tilde{\varphi}/dq$ would have delta function singularities, the square of any of which is not integrable. While many wavefunctions, when multiplied by $x$, will be square integrable, $FU$ products may not be.

The result (6.12) can be regarded as a generalization of the textbook result (Stoneham 1975) for taking matrix elements of slowly varying operators between two $FU$ products. While the author (Burt 1993) gave a reason for believing that the position operator could be regarded as slowly varying and demonstrated that this was the case for globally slowly varying envelope functions, the above shows that the position operator cannot be regarded as slowly varying in all circumstances and shows how to tackle such cases.

## 7. Envelope function equations for multicomponent fields

Consider equations of the type

$$\int d^3r' O(r, r')F(r') = 0 \tag{7.1}$$

where the field, $F$, is a multicomponent quantity. In quantum mechanics $F$ may be a spinor. In electromagnetic theory it may represent various components of the electromagnetic fields. $O$ will them generally be a matrix, $2 \times 2$ in the case of electrons with spin–orbit effects included, $6 \times 6$ if (7.1) were to represent the source free Maxwell equations for the curls of the fields in a combined form. The nonlocality could arise from nonlocal pseudopotentials in solid state theory or the nonlocal nature of the constitution relations in electromagnetic theory on the microscopic scale. There is no need to restrict the form of (7.1) to time independent problems. In quantum mechanics the operator $O$ could be $H - i\hbar\partial/\partial t$. $O$ could even be nonlocal in time such as is required for Maxwell's equations with dispersion or the equation of motion for the one electron Green function. To develop envelope function equations we define a Bravais lattice and introduce plane waves with wavevectors $k + G$, with $G$ a reciprocal lattice vector and $k$ in the first Brillouin zone (which does not have to be the conventional one, see Burt (1992), but any unit cell in the reciprocal lattice as pointed out in section 2). It is convenient to write (7.1) using the bra–ket notation

$$\sum_{k'G'} \langle k + G|O|k' + G' \rangle \langle k' + G'|F \rangle = 0 \tag{7.2}$$

where for ease of notation $G$ now also denotes an index, $\alpha$, labelling rows or columns of $O(r, r')$ as well as the relevant reciprocal lattice vector; using the explicit index notation, for instance, (7.1) would be written

$$\sum_{\alpha'} \int d^3r' O_{\alpha\alpha'}(r, r')F_{\alpha'}(r') = 0. \tag{7.1'}$$

We now introduce a similarity transformation, $S$, to use suitable periodic functions as the basis rather than the plane waves, $\exp(\mathrm{i}\boldsymbol{G} \cdot \boldsymbol{r})$. In application to electrons the transformation is usually chosen to be unitary because of the simple form of $O$ and the simple orthogonality that arises from the Schrödinger equation. However, for applications to electromagnetic fields, such a simple orthogonality does not usually exist for the modes and a more general similarity transformation is desirable. The elements of this transformation are written $S_{Gn}$ where, as in the convention mentioned above, $\boldsymbol{G}$ also includes the index $\alpha$, i.e. $S_{Gn}$ is short-hand for $S_{G\alpha,n}$. The associated periodic basis 'functions', $S_n(\boldsymbol{r})$, are column vectors with as many rows as $O(\boldsymbol{r}, \boldsymbol{r}')$ has columns; for the electron problem they would be spinors, with just two rows. The individual elements $(S_n(\boldsymbol{r}))_\alpha$ of $S_n(\boldsymbol{r})$ are related to the plane wave basis by

$$(S_n(\boldsymbol{r}))_\alpha = \sum_G \exp(\mathrm{i}\boldsymbol{G} \cdot \boldsymbol{r}) S_{G\alpha,n} \tag{7.3}$$

where, just for this equation, $\boldsymbol{G}$ represents a reciprocal lattice vector alone.

Naturally we have to chose $S$ to be invertible i.e.

$$\sum_G (S^{-1})_{nG}(S)_{Gn'} = \delta_{nn'} \tag{7.4a}$$

$$\sum_n (S)_{Gn}(S^{-1})_{nG'} = \delta_{GG'}. \tag{7.4b}$$

We can now write (7.2) as

$$\sum_{\boldsymbol{k}'} \sum_{GG'G''} \sum_{n'} (S^{-1})_{nG}\langle \boldsymbol{k} + \boldsymbol{G}|O|\boldsymbol{k}' + \boldsymbol{G}'\rangle(S)_{G'n'}(S^{-1})_{n'G''}\langle \boldsymbol{k}' + \boldsymbol{G}''|F\rangle = 0. \tag{7.5}$$

Define a new set of basis vectors

$$|n\boldsymbol{k}\rangle = \sum_G |\boldsymbol{k} + \boldsymbol{G}\rangle(S)_{Gn} \tag{7.6a}$$

$$\langle n\boldsymbol{k}| = \sum_G (S^{-1})_{nG}\langle \boldsymbol{k} + \boldsymbol{G}|. \tag{7.6b}$$

Note that $|n\boldsymbol{k}\rangle$ and $\langle n\boldsymbol{k}|$ are not dual vectors as is usual in the notation of quantum mechanics. This should present no problem if one remembers it is a notational convenience and always refers to equations (7.6) when evaluating matrix elements.

We can now express (7.5) as

$$\sum_{n'\boldsymbol{k}'} \langle n\boldsymbol{k}|O|n'\boldsymbol{k}'\rangle\langle n'\boldsymbol{k}'|F\rangle = 0 \tag{7.7}$$

with

$$\langle n\boldsymbol{k}|O|n'\boldsymbol{k}'\rangle = \sum_{G,G'} (S^{-1})_{nG}\langle \boldsymbol{k} + \boldsymbol{G}|O|\boldsymbol{k}' + \boldsymbol{G}'\rangle(S)_{G'n'}. \tag{7.8}$$

Using the orthogonality of the plane waves $\exp(\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r})$

$$\int \frac{\mathrm{d}^3\boldsymbol{r}}{\Omega} \exp(\mathrm{i}(\boldsymbol{k} - \boldsymbol{k}') \cdot \boldsymbol{r}) = \delta_{\boldsymbol{k},\boldsymbol{k}'} \tag{7.9}$$

where $\Omega$ is the normalizing volume over which cyclic boundary conditions have been imposed, (7.7) can be written as

$$\sum_{n'} \int \mathrm{d}^3\boldsymbol{r}' O_{nn'}(\boldsymbol{r}, \boldsymbol{r}') F_{n'}(\boldsymbol{r}') = 0 \tag{7.10}$$

where

$$O_{nn'}(r, r') = \frac{1}{\Omega} \sum_{kk'} \exp(ik \cdot r) \langle nk|O|n'k' \rangle \exp(-ik' \cdot r') \tag{7.11}$$

and

$$F_n(r) = \frac{1}{\sqrt{\Omega}} \sum_{k} \exp(ik \cdot r) \langle nk|F \rangle = \frac{1}{\sqrt{\Omega}} \sum_{kG} (S^{-1})_{nG} \exp(ik \cdot r) \langle k + G|F \rangle. \tag{7.12}$$

Equation (7.10) is the envelope function form of equation (7.1). The nonlocality in (7.10), however, is not solely a consequence of the nonlocality in (7.1). Even if $O(r, r')$ were a local operator, $O_{nn'}(r, r')$ would be nonlocal, because the range of $k$ is restricted to the Brillouin zone. This restriction can be lifted when the plane wave expansion coefficients, $\langle nk|F \rangle$, are zero outside the 'inner region' (defined in section 5) of the Brillouin zone. Then the summation over $k$ and $k'$ in (7.11) can be extended outside the Brillouin zone to all values. If $O(r, r')$ were local, then $O_{nn'}(r, r')$ would be as well. If, in addition, the $\langle nk|F \rangle$ are only appreciably different from zero for small $k$, then one can approximate $\langle nk|O|n'k' \rangle$ even further such as in Burt (1994a). It turns out that in the lowest approximation $O_{nn'}(r, r')$ is local even if $O(r, r')$ is nonlocal, provided the nonlocality is completely contained in the unit cell such as can occur in OPW pseudopotential theory when the boundary of the unit cell does not cut through a core region of any atom.

In this treatment of the envelope function equations we have taken the basis 'functions', $S_n(r)$, to be nominally independent of $k$. There is no need to be so restrictive. One can easily allow the $S_n(r)$ to be different for each $k$. Usually the $S_n(r)$ will be periodic solutions for some reference crystal, but one could use a different reference crystal for each and every $k$. Or one could go even further and use some complete set of 'functions' not related to any particular reference crystal and again use a different set for each and every $k$. Also, in the same way as pointed out in section 2, one could also use sets of basis functions that all change with a phase factor, $\exp(iK \cdot a)$, different from unity, on translation by a Bravais lattice vector, $a$. It is just a matter of displacing all reciprocal lattice vectors, $G$, by $K$ in the formulas above.

## 8. Photonic nanostructures

The application of envelope function methods to photonic analogues of semiconductor nanostructures would appear to be particularly appropriate. The number of plane waves needed to represent the typical dielectric profile in a unit cell is much larger than that required for the pseudopotential in a semiconductor. Supercell plane wave calculations then involve the diagonalization of very large matrices. Envelope function methods offer the possibility of making such calculations much more manageable. The formal approach is also important. For the case of electronic nanostructures the requirements of epitaxy force the constituents to be very similar chemically. As a first approximation, one could neglect the difference between the zone centre Bloch functions, at least for very small compositional changes (e.g. GaAs–Ga$_{0.9}$Al$_{0.1}$As). However, for photonic nanostructures there is no requirement for the component photonic crystals to have similar dielectric profiles, in which case it is all the more important to have a systematic approach to derive approximate envelope function equations.

In the previous section we have already given the formal theory needed to derive exact envelope function equations for any set of linear equations. However, the envelope function method is only really useful if one can dispense with the nonlocality generated by the finite extent of the Brillouin zone of the underlying Bravais lattice. So we will assume in this section

that the envelope functions involved all have a plane wave expansion the coefficients of which are all zero, or at least sufficiently small to have negligible influence, in the outer region (see section 5 for the definition of this region) of the Brillouin zone. The specific simplified form of the Maxwell equations that is most convenient varies widely with the nature of the problem in hand, so we will treat the general case of the source free Maxwell equations to illustrate the principles. We will work in the frequency domain which implies that the media are lossless. Since the derivation of the envelope function equations focuses on the space variables, one could also carry out the derivation in the time domain so that one could, in principle, include dissipation. However, working in the time domain introduces extra complexity in the notation so we will use the simpler frequency domain treatment.

We write the source free Maxwell equations for time variation $\exp(-i\omega t)$ as

$$\nabla \times E = i\frac{\omega}{c}\mu H \tag{8.1}$$

and

$$\nabla \times H = -i\frac{\omega}{c}\varepsilon E \tag{8.2}$$

where $H$ has been used to denote $\sqrt{\mu_0/\varepsilon_0}$ times the magnetic field and the magnetic permeability $\mu$ and the dielectric constant $\varepsilon$ are $3 \times 3$ matrices. We can write the equations formally as $OF = 0$ where $F$ is a six component field with $H_x$, $H_y$ and $H_z$ as the first three components and $E_x$, $E_y$ and $E_z$ as the last three components. Then equations (8.1) and (8.2) can be combined in the form $OF = 0$ with

$$O = \begin{bmatrix} i\frac{\omega}{c}\mu & -\nabla\times \\ \nabla\times & i\frac{\omega}{c}\varepsilon \end{bmatrix}. \tag{8.3}$$

Using the methods of the previous section one arrives at a set of envelope function equations

$$\sum_{n'} K_{n,n'}(r)F_{n'}(r) + \sum_{n'} \Gamma_{n,n'} F_{n'}(r) = \frac{\omega}{c} \sum_{n'} \Lambda_{n,n'}(r)F_{n'}(r). \tag{8.4}$$

The matrix $\Lambda$ is given by

$$\Lambda_{n,n'}(r') = \sum_{GG'} (S^{-1})_{nG}\lambda_{GG'}(r)(S)_{G'n'} \tag{8.5}$$

where the $6 \times 6$ matrix $\lambda_{GG'}(r)$ is

$$\lambda_{GG'}(r) = \begin{bmatrix} \mu_{GG'}(r) & 0 \\ 0 & \varepsilon_{GG'}(r) \end{bmatrix} \tag{8.6}$$

and, for example,

$$\mu_{GG'}(r) = \sum_{k} \tilde{\mu}(k + G - G')\exp(ik \cdot r) \tag{8.7}$$

with $\tilde{\mu}(q)$ being the Fourier transform of $\mu(r)$. In (8.5) $S_{G'n'}$ for fixed $G'n'$ is to be regarded as a six component column matrix and $(S^{-1})_{nG}$ for fixed $Gn$ as a six component row matrix so that $\Lambda_{n,n'}(r)$ is just a function of $r$. Similarly the matrix $\Gamma$ is given by

$$\Gamma_{n,n'} = \sum_{G} (S^{-1})_{nG}\gamma_G(S)_{Gn'} \tag{8.8}$$

where $\gamma_G$ is the $6 \times 6$ matrix

$$\gamma_G = \begin{bmatrix} 0 & 0 & 0 & 0 & -G_z & G_y \\ 0 & 0 & 0 & G_z & 0 & -G_x \\ 0 & 0 & 0 & -G_y & G_x & 0 \\ 0 & G_z & -G_y & 0 & 0 & 0 \\ -G_z & 0 & G_x & 0 & 0 & 0 \\ G_y & -G_x & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{8.9}$$

And finally, in the same vein, the **K** matrix is given by

$$K_{n,n'}(\mathbf{r}) = \sum_G (S^{-1})_{nG} \kappa(\mathbf{r})(S)_{Gn'} \tag{8.10}$$

where the $6 \times 6$ matrix $\kappa(\mathbf{r})$ is

$$(-\mathrm{i}) \begin{bmatrix} 0 & 0 & 0 & 0 & -\frac{\partial}{\partial z} & \frac{\partial}{\partial y} \\ 0 & 0 & 0 & \frac{\partial}{\partial z} & 0 & -\frac{\partial}{\partial x} \\ 0 & 0 & 0 & -\frac{\partial}{\partial y} & \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial z} & -\frac{\partial}{\partial y} & 0 & 0 & 0 \\ -\frac{\partial}{\partial z} & 0 & \frac{\partial}{\partial x} & 0 & 0 & 0 \\ \frac{\partial}{\partial y} & -\frac{\partial}{\partial x} & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{8.11}$$

It should be borne in mind that equations (8.1) and (8.2) admit zero frequency solutions which, in general, have nonzero divergences for $D$ and $B$. The solutions of interest, of course, are those with nonzero frequency for which both $\nabla \cdot D$ and $\nabla \cdot B$ vanish i.e. $D$ and $B$ are transverse fields. If one is making a Luttinger–Kohn type envelope function expansion, $\sum_n F_n(\mathbf{r}) S_n(\mathbf{r})$, then it is not permissible, in principle at least, to restrict the summation to include 'functions', $S_n(\mathbf{r})$, that correspond to transverse $D$ and $B$ because the overall envelope function expansion will not then, in general, correspond to transverse $D$ and $B$. The $S_n(\mathbf{r})$ with zero frequency and which have $D$ and $B$ with nonzero longitudinal components are needed to give a complete representation of all possible solutions of (8.1) and (8.2) with transverse $D$ and $B$. This point is potentially relevant to the application of '$k \cdot p$' theory to photonic bandstructures (Johnston *et al* 1994), though its importance in practice generally remains to be seen.

As an illustrative example of (8.4), consider the case in which both $\varepsilon$ and $\mu$ (taken as scalars) and the fields are independent of $z$ and the only nonvanishing electric field component is in the $z$ direction. Only the first and second components of (8.1) and the last component of (8.2) need be considered and if we use plane waves, $\exp(\mathrm{i}G \cdot r)$, as the periodic basis functions, then the envelope function equations are

$$-\mathrm{i}\frac{\partial E_{zG}}{\partial y} + G_y E_{zG} = \frac{\omega}{c} \sum_{G'} \mu_{GG'} H_{xG'} \tag{8.12a}$$

$$+\mathrm{i}\frac{\partial E_{zG}}{\partial x} - G_x E_{zG} = \frac{\omega}{c} \sum_{G'} \mu_{GG'} H_{yG'} \tag{8.12b}$$

$$-\mathrm{i}\left(\frac{\partial H_{xG}}{\partial y} - \frac{\partial H_{yG}}{\partial x}\right) + (G_y H_{xG} - G_x H_{yG}) = \frac{\omega}{c} \sum_{G'} \varepsilon_{GG'} E_{zG'}. \tag{8.12c}$$

These can be obtained directly from (8.4) by using the unit matrix for $S$ and eliminating the third, fourth and fifth rows and columns (those referring to variables $H_z$, $E_x$ and $E_y$).

By using the inverse, $(\mu^{-1})_{GG'}$, of $\mu_{GG'}$ at each point $r$, one can use the expressions for $H_{xG}$ and $H_{yG}$ obtained from (8.12a) and (8.12b) to substitute for these variables in (8.12c) to obtain

$$\left(\frac{\partial}{\partial x} + \mathrm{i}G_x\right) \sum_{G'} (\mu^{-1})_{GG'} \left(\frac{\partial}{\partial x} + \mathrm{i}G'_x\right) E_{zG'} + \left(\frac{\partial}{\partial y} + \mathrm{i}G_y\right) \sum_{G'} (\mu^{-1})_{GG'} \left(\frac{\partial}{\partial y} + \mathrm{i}G'_y\right) E_{zG'}$$

$$+ \left(\frac{\omega}{c}\right)^2 \sum_{G'} \varepsilon_{GG'} E_{zG'} = 0. \tag{8.13}$$

Now, (8.13) could have been derived by starting directly from the scalar wave equation

$$\frac{\partial}{\partial x}\left(\frac{1}{\mu}\frac{\partial E_z}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{\mu}\frac{\partial E_z}{\partial y}\right) + \left(\frac{\omega}{c}\right)^2 \varepsilon E_z = 0. \tag{8.14}$$

We would have regained (8.13) but with $(1/\mu)_{GG'}$ rather than the inverse of $\mu_{GG'}$. The former is found by taking the Fourier transform of the function $1/\mu$ and then forming $(1/\mu)_{GG'}$ while in the latter case one takes the Fourier transform of $\mu$, forms $\mu_{GG'}$ and then inverts this matrix. As one might expect, a closer analysis shows that these two quantities are equivalent when Fourier components in the outer region (see section 5 for the definition) of the Brillouin zone can be neglected in both $\mu$ and $1/\mu$.

To put (8.13) in nontrivial envelope function form one needs to apply a similarity transformation, $S$. Since we are now concerned with only one field component, $E_z$, the transformation is that between plane waves and periodic functions $S_n(r)$. So in (7.3), the index $\alpha$ can be dropped. For nonmagnetic media the envelope function equations become

$$-\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)E_n - 2\mathrm{i}\sum_{n'}\left(P_{nn'}^x\frac{\partial}{\partial x} + P_{nn'}^y\frac{\partial}{\partial y}\right)E_{n'} + \sum_{n'}T_{nn'}E_{n'} = \left(\frac{\omega}{c}\right)^2\sum_{n'}\varepsilon_{nn'}(r)E_{n'} \tag{8.15}$$

where

$$E_n(r) = \sum_G (S^{-1})_{nG}E_{zG}(r) \tag{8.16a}$$

$$P_{nn'}^i = \sum_{GG'}(S^{-1})_{nG}G_iS_{Gn'} \tag{8.16b}$$

$$T_{nn'} = \sum_G (S^{-1})_{nG}G^2S_{Gn'} \tag{8.16c}$$

and

$$\varepsilon_{nn'}(r) = \sum_{GG'}(S^{-1})_{nG}\varepsilon_{GG'}(r)S_{G'n'}. \tag{8.16d}$$

Now in the case of envelope function equations for electrons, we invariably use a unitary transformation for **S** and chose it to diagonalize the zone centre Hamiltonian in some important region. In a photonic crystal, rather than a photonic nanostructure, one could choose **S** to simultaneously diagonalize both **T** and $\varepsilon$, i.e. both $\tilde{\mathbf{S}}^*\mathbf{T}\mathbf{S}$ and $\tilde{\mathbf{S}}^*\varepsilon\mathbf{S}$ would be diagonal. But **S** would not be unitary, in general, and hence neither $T_{nn'}$ nor $\varepsilon_{nn'}$ as defined in (8.16c) and (8.16d) would be diagonal. They would, however, form diagonal matrices on premultiplication by $(\tilde{\mathbf{S}}^*)\mathbf{S}$. In photonic nanostructures it may turn out that premultiplying (8.15) by $\tilde{\mathbf{S}}^*\mathbf{S}$ is useful so that both **T** and $\varepsilon$ could be chosen to be diagonal somewhere with the drawback that the $-(\partial^2/\partial x^2 + \partial^2/\partial y^2)E_n$ term becomes more complex i.e. $-\sum_{n'}(\tilde{S}^*S)_{nn'}(\partial^2/\partial x^2 + \partial^2/\partial y^2)E_{n'}$.

In the above, we have been working with the macroscopic Maxwell equations and hence with local $\varepsilon$ and $\mu$. We could, however, imagine that we had been starting from microscopic Maxwell equations and deriving envelope function equations for them. All that is required is to generalize what we have done to include the case of nonlocal $\varepsilon$ and $\mu$. This, of course, implies that our fields are all to be considered as expectation values of the corresponding quantum mechanical operators and that a linear response has already been established between the electric displacement and electric field and also the magnetic induction and magnetic field. This generalization to nonlocal $\varepsilon$ and $\mu$ is not difficult, for instance using the method in Burt (1994a), if the nonlocality is always contained in a unit cell (though this is not always strictly necessary) in an analogous way to nonlocal OPW pseudopotentials (see e.g. Heine 1970) with sufficiently tightly bound core levels; then the $\varepsilon_{GG'}$ and $\mu_{GG'}$ turn out to be local functions and

our analysis is directly transferable. In this case, we expect only the field components with $G = 0$ to be appreciably different from zero because the periodicity of the Bravais lattice is less than or of the order of a nanometre and hence $cG$ corresponds to frequencies in the x-ray region while $\omega$ is in the optical region. To illustrate this consider (8.13) for the case of slowly varying envelopes so that one can neglect the derivatives. For nonmagnetic media one sees that for $G \neq 0$

$$E_{zG}(x, y) \approx -\frac{(\omega/c)^2}{G_x^2 + G_y^2}\varepsilon_{G0}(x, y)E_{z0}(x, y) \tag{8.17}$$

and so $E_{zG}$ is small for frequencies in the optical region. It is clear that (8.4) can take us from microscopic to the macroscopic regime if we use only plane waves, $\exp(iG \cdot r)$, as the periodic basis functions, consider only the $G = 0$ equations and neglect all field components for $G \neq 0$.

Of course, the derivation of macroscopic Maxwell equations from their microscopic counterparts goes back to the work of Lorentz (1909) and Rosenfeld (1951) where a coarse grained averaging was carried out on the microscopic variables, with further developments, including the article by Russakoff (1970), well summarized by Jackson (1975). An alternative way (Robinson 1973) of expressing this averaging procedure is to note that the microscopic equations

$$\nabla \cdot e = \frac{\rho}{\varepsilon_0}$$
$$\nabla \cdot b = 0$$
$$\nabla \times e = -\frac{\partial b}{\partial t}$$
$$\nabla \times b = \mu_0 \left(j + \varepsilon_0 \frac{\partial e}{\partial t}\right)$$

only involve translationally invariant operators, so that the long wavelength components of the variables, those that correspond to the Fourier transforms of the macroscopic variables, obey the same equations. Both of these former approaches then assume that the macroscopic polarization and magnetization are determined only by the macroscopic fields at the same point. These analyses, however, do not assume that the materials involved are crystalline. The envelope function approach, though restricted to systems made up of unit cells arranged on a Bravais lattice, does give us a more formal insight into how the macroscopic constitutive relations can be local even though the microscopic counterparts are nonlocal: $\varepsilon_{GG'}$ and $\mu_{GG'}$ and in particular $\varepsilon_{00}$ and $\mu_{00}$ turn out to be local operators in real space, at least, if unit cells can be defined so that the nonlocal aspect of the microscopic constitutive relations can be restricted to each unit cell, i.e. the polarization, for instance, in one cell is only dependent on the values of the electric field in the same cell and at the same time the dimensions of the unit cell are small compared to the wavelength. This dropping out of the nonlocality is not an issue for the derivation of the normal macroscopic Maxwell equations since there is really only one wavelength in the problem and that is large compared with atomic sizes. But when one is considering, say, the propagation of x-rays through a multilayer and near a Bragg condition in each of the constituent crystals, then elimination of nonlocal effects is not so straightforward and one would have to fall back on a more sophisticated argument such as that outlined above.

## 9. Summary

If there is one unifying message to be extracted from the preceding pages, it is the power of a systematic envelope function method, such as the author has developed, to resolve

controversies that arise when one has to rely solely on semiempirical methods. What does one do, for example, if one only has a semiempirical method and it is claimed that the kinetic energy operator, $T = -(\hbar^2/2)(\mathrm{d}/\mathrm{d}x)(1/m^*)(\mathrm{d}/\mathrm{d}x)$, one is using leads to solutions that violate the uncertainty principle and therefore must be rejected? If one has an exact envelope function method and one has derived this form of the kinetic energy operator systematically, one knows it cannot be generally wrong and one has the motivation and the tools to get to the root of the problem. Similarly, when constructing valence band Hamiltonians, with a semiempirical method one just has to guess at the correct form of the Hamiltonian out of many possible ones. With an exact envelope function theory one can systematically derive approximate Hamiltonians that would not have been obvious ones to try if one were just relying on guesswork. Similarly, with the out of zone solution conundrum and the problem of interband dipole matrix element, the presence of an exact envelope function method based on an envelope function representation gives one confidence that issues that look intractable from the semiempirical viewpoint are, indeed, resolvable. In connection with the out of zone solution problem, we have demonstrated the surprising result that the wavefunction can be entirely generated by out of zone solutions of the envelope function equations. The fallacy that just retaining the dominant term in the envelope function expansion for the wavefunction is sufficient when taking matrix elements of the position operator has been exposed and an intriguing corollary to this is that, in the limit of wide quantum wells, vanishingly small components of the wavefunction determine the size of the matrix element.

There are, of course, other messages. The author's method is a general one, at least for approximately piecewise periodic systems, for transforming from microscopic to mesoscopic descriptions and its application to photonic nanostructures has been outlined to emphasize this point. Indeed, the envelope function method may be particularly appropriate for such structures since the number of plane waves needed in the expansion of the fields to obtain an accurate bandstructure for a photonic crystal is much larger than for that required for the corresponding problem for electrons using pseudopotentials. Direct 'supercell' type methods for calculating modes of photonic nanostructures could be very computationally intensive.

Another important message is the elucidation of the subtle difference between the envelope functions that appear in the author's method as part of a representation and those that arise in heuristic or empirical methods. In the author's method each envelope function is defined to have a plane wave expansion that only contains wavevectors in a unit cell, usually the first Brillouin zone, of a reference reciprocal lattice. On the other hand, the envelope functions that are introduced in the method of Bastard are defined to be solutions of a given set of differential equations and have no restriction on their plane wave expansion. Typically, for flat band conditions, these latter envelope functions are exponentials, but even if for each layer of a multilayer structure the exponentials have real 'in zone' wavevector, these envelope functions would not be the same as those defined by the author's envelope function representation, though in many cases they are similar.

Perhaps the most important message of all concerns the scope of envelope function methods. Traditionally, they were justified by considering 'gentle' perturbations such those arising from shallow impurities, applied magnetic fields or slowly graded structures. The author, however, has taken a very different approach to deriving approximate envelope function equations. The main assumption is that the envelope functions are globally slowly varying and no explicit assumption about the perturbations is made, though, of course, the globally slowly varying envelope function assumption puts some implicit restrictions on the perturbations. This does not automatically rule out structures containing atomically abrupt junctions since the low lying states for such structures may well have globally slowly varying envelope functions. The great advantage of the author's approach is that after the approximate envelope function

equations have been solved one can check how well the solutions fit the globally slowly varying criterion and estimate corrections, providing one with a consistent approximation procedure.

Finally, a parting comment on the resolution of the basic boundary condition problem, one of the motivations for developing an exact envelope function method in the first place. When the relevant part of the bandstructure in each of the component crystals of a multilayer structure is a single parabolic band, does one impose continuity on the derivative of the envelope function or on the ratio of the derivative to the effective mass at an abrupt heterojunction? In the former case one has a smooth envelope function, which is consistent with the required absence of kinks in the wavefunction, but there is an apparent violation of current conservation. In the latter case one has current conservation, but a kink appears in the envelope function at the interface, which implies a corresponding kink in the wavefunction. The author's derivation of the effective mass equation shows that the effective mass should be included in the derivative boundary condition, but also shows that the kink in the approximate envelope function arising from a discontinuous change in effective mass at an abrupt heterojunction just models the smooth, but relatively rapid, change in derivative of the exact envelope function; there is a corresponding feature in the exact wavefunction related to the effective mass discontinuity. But in a delightful twist, smooth matching of the envelope function derivative is also a permissible procedure, when interface effects are not important, provided one is solving the complete multiband set of envelope function equations with the inclusion of the out of zone solutions instead of making the effective mass approximation; in the numerical example presented here, at least, the effective mass related feature is still present in the approximate wavefunction so derived.

## Acknowledgments

The author would like to thank Professor R A Abram, Professor B K Ridley, Professor A M Stoneham and Dr B A Foreman for many helpful and stimulating conversations, R A Coles and Dr P Stavrinou for communicating their results on valence band Hamiltonians prior to publication and Dr A J Williamson for a discussion on the linear combination of bulk bands method.

## Appendix. Inclusion of more remote bands

In section 4, only one remote band, the lowest conduction band, was included in the analysis and this led to unphysical behaviour i.e. the heavy hole bands bent upwards into the gap. A more realistic treatment including the lowest $\Gamma_{15}$ conduction bands will further illustrate the rule noted in section 4 on how to obtain correct ordering of the differential operators and provide a more realistic valence bandstructure.

Denote the zone centre $\Gamma_{15}$ conduction band Bloch functions as $X'$, $Y'$ and $Z'$. The wave function is now written

$$iF_S S + F_X X + F_Y Y + F_Z Z + F_{X'} X' + F_{Y'} Y' + F_{Z'} Z'. \tag{A.1}$$

Equation (4.4a) for $F_X$ is now

$$\left( E_V - E - \frac{\hbar^2}{2m} \nabla^2 \right) F_X - \frac{i\hbar}{m} \langle X|p_x|iS \rangle \frac{\partial F_S}{\partial x} - \frac{i\hbar}{m} \langle X|p_y|Z' \rangle \frac{\partial F_{Z'}}{\partial y}$$
$$- \frac{i\hbar}{m} \langle X|p_z|Y' \rangle \frac{\partial F_{Y'}}{\partial z} = 0. \tag{A.2}$$

By symmetry there are no matrix elements of momentum between any pair of the states $X$, $Y$ and $Z$ nor between any pair of the states $X'$, $Y'$ and $Z'$. The only nonzero matrix elements between one of $X$, $Y$ and $Z$ and one of $X'$, $Y'$ and $Z'$ are $\langle X|p_y|Z'\rangle$ and $\langle X|p_z|Y'\rangle$ and their cyclic permutations.

We now substitute for $F_{Y'}$ and $F_{Z'}$ as well as $F_S$ in the equation (A.2) for $F_X$ using the approximate equation (4.5) and corresponding equations for $F_{Y'}$ and $F_{Z'}$; the equation for $F_{X'}$ is, by analogy with (A.2),

$$\left(E_{C'} - E - \frac{\hbar^2}{2m}\nabla^2\right)F_{X'} - \frac{i\hbar}{m}\langle X'|p_x|iS\rangle\frac{\partial F_S}{\partial x} - \frac{i\hbar}{m}\langle X'|p_y|Z\rangle\frac{\partial F_Z}{\partial y}$$
$$- \frac{i\hbar}{m}\langle X'|p_z|Y\rangle\frac{\partial F_Y}{\partial z} = 0 \tag{A.3}$$

from which the corresponding equations for $F_{Y'}$ and $F_{Z'}$ can be obtained by cyclic permutation. $E_{C'}$ is the energy of the states $X'$, $Y'$ and $Z'$.

We now use the fact that the envelope functions $F_X$, $F_Y$ and $F_Z$ are dominant, all the others are small, and that the envelopes are all slowly varying so that we can neglect the free electron kinetic energy term compared with energy gaps e.g. $E_{C'} - E \approx E_{C'} - E_V = E_{G'}$.

$$F_{X'} \approx \frac{i\hbar}{mE_{G'}}\left[\langle X'|p_y|Z\rangle\frac{\partial F_Z}{\partial y} + \langle X'|p_z|Y\rangle\frac{\partial F_Y}{\partial z}\right]. \tag{A.4}$$

Cyclic permutation will give the corresponding equations for $F_{Y'}$ and $F_{Z'}$. Substitution for $F_{Y'}$ and $F_{Z'}$ as well as $F_S$ in (A.2) gives

$$\left(E_V - E - \frac{\hbar^2}{2m}\nabla^2\right)F_X + \left(\frac{\hbar}{m}\right)^2\left\{\frac{\partial}{\partial x}\frac{\langle X|p_x|iS\rangle\langle iS|p_x|X\rangle}{E_G}\frac{\partial F_X}{\partial x}\right.$$
$$+ \frac{\partial}{\partial x}\frac{\langle X|p_x|iS\rangle\langle iS|p_y|Y\rangle}{E_G}\frac{\partial F_Y}{\partial y} + \frac{\partial}{\partial x}\frac{\langle X|p_x|iS\rangle\langle iS|p_z|Z\rangle}{E_G}\frac{\partial F_Z}{\partial z}\right\}$$
$$+ \left(\frac{\hbar}{m}\right)^2\left\{\frac{\partial}{\partial y}\frac{\langle X|p_y|Z'\rangle\langle Z'|p_x|Y\rangle}{E_{G'}}\frac{\partial F_Y}{\partial x} + \frac{\partial}{\partial y}\frac{\langle X|p_y|Z'\rangle\langle Z'|p_y|X\rangle}{E_{G'}}\frac{\partial F_X}{\partial y}\right.$$
$$+ \frac{\partial}{\partial z}\frac{\langle X|p_z|Y'\rangle\langle Y'|p_z|X\rangle}{E_{G'}}\frac{\partial F_X}{\partial z} + \frac{\partial}{\partial z}\frac{\langle X|p_z|Y'\rangle\langle Y'|p_x|Z\rangle}{E_{G'}}\frac{\partial F_Z}{\partial z}\right\} = 0. \tag{A.5}$$

The first pair of braces contain the terms that originate from coupling to the conduction band and are exactly the same as calculated previously. The second pair of braces contain the new terms, those that originate from coupling to the $\Gamma_{15}$ conduction band. Symmetry allows us to write

$$C_1 = \left(\frac{\hbar}{m}\right)^2\frac{\langle X|p_y|Z'\rangle\langle Z'|p_y|X\rangle}{E_{G'}} = \left(\frac{\hbar}{m}\right)^2\frac{\langle X|p_z|Y'\rangle\langle Y'|p_z|X\rangle}{E_{G'}} \tag{A.6a}$$

and

$$C_2 = \left(\frac{\hbar}{m}\right)^2\frac{\langle X|p_y|Z'\rangle\langle Z'|p_x|Y\rangle}{E_{G'}} = \left(\frac{\hbar}{m}\right)^2\frac{\langle X|p_z|Y'\rangle\langle Y'|p_x|Z\rangle}{E_{G'}}. \tag{A.6b}$$

So in terms of $k = -i\nabla$ our Hamiltonian with addition of the $\Gamma_{15}$ conduction band becomes

$$\begin{bmatrix} -k_xbk_x - k_yC_1k_y - k_zC_1k_z & -k_xbk_y - k_yC_2k_x & -k_xbk_z - k_zC_2k_x \\ -k_ybk_x - k_xC_2k_y & -k_xC_1k_x - k_ybk_y - k_zC_1k_z & -k_ybk_z - k_zC_2k_y \\ -k_zbk_x - k_xC_2k_z & -k_zbk_y - k_yC_2k_z & -k_xC_1k_x - k_yC_1k_y - k_zbk_z \end{bmatrix} \tag{A.7}$$

where the common $E_V + ak^2$ contribution to each diagonal element, for which operator ordering is not an issue, has been omitted. We see now that in the off-diagonal elements the different parts of the Luttinger parameter are treated differently in inhomogeneous systems. For instance, in the bulk, the element in row $X$ and column $Y$ of (A.7) would be written as

$$-(b + C_2)k_x k_y.$$

For the $-bk_x k_y$ term, since $p_x$ comes in the first matrix element of $b = (\hbar^2/m^2 E_G)\langle X|p_x|iS\rangle \langle iS|p_y|Y\rangle$ and $p_y$ comes in the second matrix element, $k_x$ must come in front of $b$ in inhomogeneous systems and $k_y$ must come behind. On the other hand, for the $-C_2 k_x k_y$ term, since $p_y$ comes in the first matrix element of $C_2 = (\hbar/m)^2 \langle X|p_y|Z'\rangle \langle Z'|p_x|Y\rangle/E_{G'}$ and $p_x$ comes in the second matrix element, $k_y$ must come in front of $C_2$ in inhomogeneous systems and $k_x$ must come behind.

## References

Aleiner I L and Ivchenko E L 1992 *JETP Lett.* **55** 693
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
Ando H, Nojima S and Kanbe H 1993 *J. Appl. Phys.* **74** 6383
Aryasetiawan F and Gunnarson O 1998 *Rep. Prog. Phys.* **61** 237
Bastard G 1981 *Phys. Rev.* B **24** 5693
——1988 *Wave Mechanics Applied to Semiconductor Heterostructures* (Paris: Editions de Physique) p 105
Bastard G, Brum J A and Ferreira R 1991 *Solid State Physics* vol 44, ed H Ehrenreich and D Turnbull (New York: Academic) p 229
Blount E I 1962 *Solid State Physics* vol 13, ed F Seitz and D Turnbull (New York: Academic) p 357
Bobbert P A, Wieldraaijer H, van der Weide R, Kemerink M, Koenraad P M and Wolter J H 1997 *Phys. Rev.* B **56** 3664
Bockelmann U and Bastard G 1992 *Phys. Rev.* B **45** 1688
Burt M G 1980 *J. Phys. C: Solid State Phys.* **13** 1825
——1988a *Semicond. Sci. Technol.* **3** 739
——1988b *Semicond. Sci. Technol.* **3** 1224
——1989 *Bandstructure Engineering in Semiconductor Microstructures (NATO ASI Series B vol 189)* ed R A Abram and M Jaros (New York: Plenum) p 99
——1992 *J. Phys.: Condens. Matter* **4** 6651
——1993 *J. Phys.: Condens. Matter* **5** 4091
——1994a *Phys. Rev.* B **50** 7518
——1994b *Appl. Phys. Lett.* **65** 717
——1995a *Superlatt. Microstruct.* **17** 335
——1995b *Semicond. Sci. Technol.* **10** 412
——1998 *Superlatt. Microstruct.* **23** 531
Burt M G and Foreman B A 1998 *Proc. 24th Int. Conf. on the Physics of Semiconductors* eds E Cohen and M Heiblum (Singapore: World Scientific) at press
Burt M G and Inkson J C 1976 *J. Phys. D: Appl. Phys.* **9** 43
Callaway J 1991 *Quantum Theory of the Solid State* (New York: Academic) p 518
Chang Y C, Schulman J N, Bastard G, Guldner Y and Voos M 1985 *Phys. Rev.* B **31** 2557
Chao C Y and Chuang S L 1992 *Phys. Rev.* B **46** 4110
Coles R A 1998 *PhD Thesis* University of Durham
Coles R A, Brand S, Abram R A and Burt M G 1998 *Phys. Rev.* B, submitted
Coles R A and Stavrinou P 1998 private communication and to be published
Dingle R, Weigmann W and Henry C H 1974 *Phys. Rev. Lett.* **33** 827
Edwards G and Inkson J C 1994 *Solid State Commun.* **89** 595
Edwards G, Valadares E C and Sheard F W 1994 *Phys. Rev.* **50** 8493
Elliot R J 1957 *Phys. Rev.* **108** 1384
Esaki L and Tsu R 1969 *IBM Research Note* RC-2418
Foreman B A 1993 *Phys. Rev.* B **48** 4964

Foreman B A 1994 *Phys. Rev.* B **49** 1757
——1995 *Phys. Rev.* B **52** 12 260
——1996 *Phys. Rev.* B **54** 1909
——1998a *Phys. Rev. Lett.* **81** 425
——1998b *Phys. Rev. Lett.* **80** 3823
Geller M R 1997 *Phys. Rev. Lett.* **78** 110
Geller M R and Kohn W 1993 *Phys. Rev. Lett.* **70** 3103
Godfrey M J and Malik A M 1996 *Phys. Rev.* B **53** 16 504
Hagston W E, Harrison P, Piroek T and Stirner T 1994 *Superlatt. Microstruct.* **15** 199
Haug H and Koch S W 1990 *Quantum Theory of the Optical and Electronic Properties of Semiconductors* (Singapore:
       World Scientific)
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 1
Heine V and Cohen M H 1970 *Solid State Physics* vol 24 (New York: Academic) p 78
Hemmer P C and Wang D T 1993 *Phys. Rev.* B **47** 6603
Ivchenko E L, Kaminskii A Yu and Aleiner I L 1993 *Sov. Phys.–JETP* **77** 609
Jackson J D 1975 *Classical Electrodynamics* (New York: Wiley)
Johnston N F, Hui P M and Luk K H 1994 *Solid State Commun.* **90** 229
Kane E O 1957 *J. Phys. Chem. Solids* **1** 249
Kane E O 1975 *Semiconductors and Semimetals* vol 1 (New York: Academic) p 75
Karavaev G F and Krivorotov I N 1996 *Fiz. Tekh. Poluprovodn.* **30** 177 (Engl. transl. *Semiconductors* **30** 102)
Krebs O, Rondi D, Gentner J L, Goldstein L and Voisin P 1998 *Phys. Rev. Lett.* **80** 5770
Krebs O and Voisin P 1996 *Phys. Rev. Lett.* **77** 1829
Lin-Liu Y R and Sham L J 1985 *Phys. Rev.* B **32** 5561
Lorentz H A 1909/1952 *Theory of Electrons* (Liepzig/Dover)
Löwdin P 1951 *J. Chem. Phys.* **19** 1396
Luttinger J M 1956 *Phys. Rev.* **102** 1030
Luttinger J M and Kohn W 1955 *Phys. Rev.* **97** 869
Meney A T, Gonul B and O'Reilly E P 1994 *Phys. Rev.* B **50** 10 893
Pötz W, Porod W and Ferry D K 1985 *Phys. Rev.* B **53** 3868
Ridley B K 1997 *Electrons and Phonons in Semiconductor Multilayers* (Cambridge: Cambridge University Press)
Robinson F N H 1973 *Macroscopic Electromagnetism* (Oxford: Pergamon)
Rohlfing M, Krüger P and Pollmann J 1998 *Phys. Rev.* B **57** 6485
Rohlfing M and Louie S G 1998a *Phys. Rev.* B **57** 9392
——1998b *Phys. Rev. Lett.* **81** 2312
Rosencher E, Fiore A, Vinter B, Berger V, Bois Ph and Nagle J 1996 *Science* **271** 168
Rosenfeld L 1951 *Theory of Electrons* (Amsterdam: North-Holland)
Russakoff G 1970 *Am. J. Physics* **38** 1188
Slater J C 1949 *Phys. Rev.* **76** 1592
Smith D L and Mailhiot C 1990 *Rev. Mod. Phys.* **62** 173
Stavrinou P N and van Dalen R 1997 *Phys. Rev.* B **55** 15 456
Stolz H 1974 *Einführung in die Vielelectronen-Theorie der Kristalle* (Berlin: Academie) p 108
Stoneham A M 1975 *Theory of Defects in Solids* (Oxford: Clarendon) p 75
Taylor R I, Burt M G and Abram R A 1987 *Superlatt. Microstruct.* **3** 63
von Allmen P 1992 *Phys. Rev.* B **46** 15 377
van Dalen R and Stavrinou P N 1998 *Semicond. Sci. Technol.* **13** 11
Wang L W, Franceschetti A and Zunger A 1997 *Phys. Rev. Lett.* **78** 2819
Wannier G H 1937 *Phys. Rev.* **52** 191
Weisbuch C and Vinter B 1991 *Quantum Semiconductor Structures* (Academic)
White S R and Sham L J 1981 *Phys. Rev. Lett.* **47** 879
Williamson A J, Zunger A and Canning A 1998 *Phys. Rev.* B **57** R4253
Wood D M and Zunger A 1996 *Phys. Rev.* B **53** 7949
Yi H and Razeghi M 1997 *Phys. Rev.* B **56** 3933