Bloch's Theorem: Some Notes

MJ Rutter

Michaelmas 2005

1 Bloch's Theorem

$$\left[\nabla^2 + V(\mathbf{r})\right]\psi(\mathbf{r}) = E\psi(\mathbf{r})$$

If V has translational symmetry, it does not follow that $\psi(\mathbf{r})$ has translation symmetry. At first glance we need to solve for ψ throughout an infinite space.

However, Bloch's Theorem proves that if V has translational symmetry, the solutions can be written

$$\psi_{\mathbf{k}} = \exp(i\mathbf{k}.\mathbf{r})u_{\mathbf{k}}(\mathbf{r})$$

where $u(\mathbf{r})$ has the same periodicity as V, and $\mathbf{k}$ lies within the reciprocal space unit cell. As $u(\mathbf{r})$ is periodic, one can take its discrete Fourier transform:

$$\psi_{\mathbf{k}} = \sum_{\mathbf{G}} c_{\mathbf{k}+\mathbf{G}}\exp(i(\mathbf{k}+\mathbf{G}).\mathbf{r})$$

where $\mathbf{G}.\mathbf{l}$ is an integer multiple of 2π for all lattice vectors $\mathbf{l}$.

How does this help, for $\mathbf{G}$ still has infinite range, and $\mathbf{k}$ needs to be sampled with infinite density?

As the kinetic energy term of the Hamiltonian is proportional to the square of the wavevector, states of high $|\mathbf{G}+\mathbf{k}|$ have high kinetic energy. Above some limit which will depend on the atoms present, this value will be so high that the weight in these basis functions will be negligible.

It is then assumed that ψ is continuous in $\mathbf{k}$, so that one $\mathbf{k}$ value can be taken as representative of a region of k-space, and the integral over all $\mathbf{k}$'s in the reciprocal space unit cell can be replaced by discrete sums.

Note that though $\psi(\mathbf{r})$ does not have the same periodicity as $V(\mathbf{r})$, $\rho(\mathbf{r}) = \psi(\mathbf{r})\psi^*(\mathbf{r})$ does, so if V is a function both of some external potential (e.g. a collection of pseudopotentials) and a potential arising from ρ (e.g. Coulomb and LDA exchange-correlation terms), its periodicity will be unaffected.

2 Properties of k-space

2.1 Periodicity and Special Points

$$
\begin{aligned}
\psi_{\mathbf{k}+\mathbf{G}} &= \exp(i(\mathbf{k}+\mathbf{G}).\mathbf{r})u_{\mathbf{k}+\mathbf{G}}(\mathbf{r}) \\
&= \exp(i\mathbf{k}.\mathbf{r})\exp(i\mathbf{G}.\mathbf{r})u_{\mathbf{k}+\mathbf{G}}(\mathbf{r}) \\
&= \exp(i\mathbf{k}.\mathbf{r})u'_{\mathbf{k}}(\mathbf{r})
\end{aligned}
$$

where $u'_{\mathbf{k}}(\mathbf{r})$ is still periodic in the unit cell, as $\exp(i\mathbf{G}.\mathbf{r})$ must be periodic in the unit cell. We can always choose to reduce any $\mathbf{k}$ in this fashion, and do always choose to work with $|k| \leq G/2$.

This also implies that the points $(\pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2})$ are all identical, for they are related via the addition/subtraction of a reciprocal lattice vector. This point is called the L point, and the reciprocal space point $(0,0,0)$ is called the Γ point. A small zoo of other, high-symmetry, points in k-space has also been named.

2.2 Inversion Symmetry

$$\left[\nabla^2 + V(\mathbf{r})\right] e^{i\mathbf{k}\cdot\mathbf{r}} u_\mathbf{k}(\mathbf{r}) = E_\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{r}} u_\mathbf{k}(\mathbf{r})$$

Assuming that V and E are real, one can take the complex conjugate of the above to give:

$$\left[\nabla^2 + V(\mathbf{r})\right] e^{-i\mathbf{k}\cdot\mathbf{r}} u_\mathbf{k}^*(\mathbf{r}) = E_\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{r}} u_\mathbf{k}^*(\mathbf{r})$$

and then one concludes that

$$u_{-\mathbf{k}}(\mathbf{r}) = u_\mathbf{k}^*(\mathbf{r})$$

and hence

$$\psi_{-\mathbf{k}}(\mathbf{r}) = \psi_\mathbf{k}^*(\mathbf{r})$$

and that these have the same energy, and will give rise to the same charge density.

2.3 Matrix Notation

Given that the wavefunctions etc. will be considered on a discrete grid, one can rewrite the above equations in matrix notation. This is most obvious for the 1D case where the grid is just a linear set of points. Using the convention that matrices are given bold capital letters, one has:

$$\left[\nabla^2 + \mathbf{V}\right] \psi = E\psi$$

It is clear that $\mathbf{V}$ is simply a diagonal matrix. But what is ∇^2? It is not simply the matrix for finding second derivatives by finite differences: the grid is far too coarse for that to give reasonable answers.

However, in reciprocal space ∇^2 can be represented as a diagonal matrix, $\mathbf{Q}$, whose diagonal elements are g^2, and a Fourier transform can also be represented as a matrix.

$$\left[\mathbf{F}^{-1}\mathbf{Q}\mathbf{F} + \mathbf{V}\right] \psi = E\psi$$

This form is useful as the collection of matrices can clearly be collapsed into a single matrix $\mathbf{H}$ for mathematical convenience, but for numerical convenience one can note that whereas dense matrix-vector multiplication is order N^2, diagonal matrix-vector multiplication is order N, and the application of a Fourier transform is order $N \ln N$. (One can write $\mathbf{F}$ as the product of $\ln N$ sparse matrices, each of which can multiply a vector in order N operations, but that is another subject.)

To restore this form from the Bloch form, premultiply both sides by $\exp(-i\mathbf{k}\cdot\mathbf{r})$, remembering that this will not commute with ∇^2, but will with E. If V is a simple real-space potential, it will with V too.

2.4 Chemistry

The k vector describes how the phase of ψ changes between adjacent unit cells. Some of this is familiar from chemistry: in a chain of atoms, one might have all of the atomic orbitals in phase, and overlapping to produce a bonding molecular orbital. This would be associated with $k = 0$, also called the Γ point. One might also have adjacent atomic orbitals in anti-phase, giving an antibonding orbital, and corresponding to $k = \frac{1}{2}$.

A continuum exists between these extremes (if one has a long chain, rather than a diatomic molecule), and, of course, the different orbitals have different energies and produce different charge densities (with nodes in antibonding orbitals, for instance).
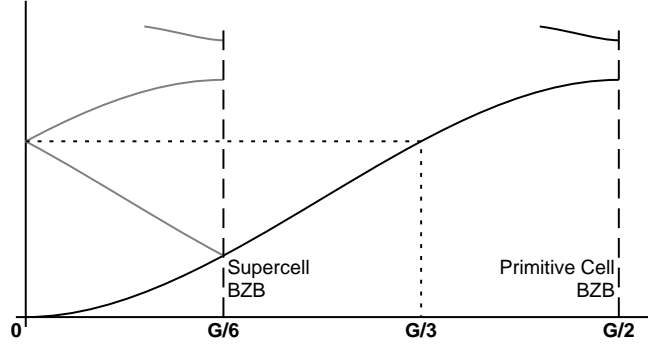
Figure 1: $E(k)$ for a unit cell, and in a supercell three times as long

2.5 Supercells, Equivalence, and the form of $E(\mathbf{k})$

A calculation at $k = 0$ in the supercell formed by three primitive cells should be identical to that at $k = 0$, $\frac{1}{3}$ and $\frac{2}{3}$ in the primitive cell. Furthermore, $k = \frac{2}{3}$ should be equivalent to $k = -\frac{1}{3}$, and by inversion symmetry $k = -\frac{1}{3}$ is equivalent to $k = \frac{1}{3}$.

It has been shown that $E(\mathbf{k})$ is periodic, and obeys $E(\mathbf{k}) = E(-\mathbf{k})$. Perturbation theory produces a discontinuity in $E(\mathbf{k})$ at points where $\mathbf{k}$ corresponds to a periodicity in the potential, and forces the derivative to zero at these points.

Considering figure 1 in detail, one sees a single band extending across the whole of the reciprocal space unit cell. The start of the next band is also shown. As $E(\mathbf{k})$ is even, there is no point in plotting negative values of $\mathbf{k}$. When one moves to a supercell, here of three times the length, the Brillouin Zone is a third of the size, and the k-points which now lie outside this Brillouin Zone can be reduced into it by subtracting multiples of the new, shorter, reciprocal lattice vectors.

The new supercell will contain three times as many electrons as the old unit cell, so it needs to have three times as many bands to put them in. The "folding" of the original band produces just this. Note that at the new BZ boundaries there are no band-gaps, and the derivative of $E(\mathbf{k})$ is not zero. Only at those points which correspond to a periodicity in the potential does this behaviour occur.

2.6 The Accuracy of Sums

When integrating functions numerically in one dimension, one often refers to a method as being exact for polynomials up to some order n, and assumes that the higher n, the better the integrator.

A similar argument holds for sampling k-space within the reciprocal space unit cell. As functions are periodic in k-space, the obvious basis set in which to expand an unknown function is a Fourier series in k.

As an example, consider a function in a 1-D reciprocal space unit cell which is real and symmetric (so can be expressed as a cosine series). Note that $E(k)$ is such a function.

If E(k) is a constant, then any point will integrate it. Otherwise, one needs to extract the coefficient of the first term in the cosine expansion, for all the other terms will integrate to zero.

If $E(k) = b_0 + b_1 \cos(2\pi k/G)$, then consider sampling at $k = G/4$.

$$E(G/4) = b_0 + b_1 \cos\left(\frac{\pi}{2}\right) = b_0 = \int E(k)dk$$

A perfect result still, but note that sampling at, for instance, $k = 0$ would not have given this result.

3

Consider one higher-order expansion:

$$E(k) = b_0 + b_1 \cos(2\pi k/G) + b_2 \cos(4\pi k/G)$$

and consider sampling at $k = G/8$ and $k = 3G/8$.

$$
\begin{aligned}
E(G/8) &= b_0 + b_1 \cos\left(\frac{\pi}{4}\right) + b_2 \cos\left(\frac{\pi}{2}\right) \\
E(3G/8) &= b_0 + b_1 \cos\left(\frac{3\pi}{4}\right) + b_2 \cos\left(\frac{3\pi}{2}\right)
\end{aligned}
$$

And

$$\frac{1}{2}E(G/8) + \frac{1}{2}E(3G/8) = b_0$$

One can define a first failure point for a set of k-points. As one is working in reciprocal space, the corresponding Fourier components are in real space, and one can find a smallest real-space vector for which the sum fails to reproduce the integral. The $k = G/4$ point fails to integrate a function with a term like $\cos(4\pi k/G)$, or $\cos(kx)$ where $x = 4\pi/G = 2L$. So one says that the first failure point occurs at $2L$.

In 3D similar arguments apply, and one is normally interested in the first set of points in the Fourier expansion of $E(\mathbf{k})$ which are not correctly integrated, for several points will have the same modulus and fail together. This set is referred to as the 'first failure star.'

3 Grids

The cut-off energy defines a highest frequency Fourier component which may be present in $u(\mathbf{r})$. In turn, this defines a maximum required sampling density in real space. One can therefore represent $u(\mathbf{r})$ on a discrete real-space grid, without any loss of information, and a Fourier transform moves one to a corresponding reciprocal space grid which includes components at this cut-off frequency. The spacing of the reciprocal space grid points are $\mathbf{G}$, the reciprocal space lattice vectors. The cut-off g_{cut} is more usually expressed as an energy, where one is referring to the kinetic energy of a plane wave with this g-vector.

The density, $\rho = uu^*$, will have components at frequencies up to twice the above cut-off frequency, leading to a grid stretching twice as far in reciprocal space, and of twice the density in real space, if one wishes to store ρ without loss of information.

The potential is an awkward function of ρ, and in the general case must be assumed to have components at all frequencies.

The number of grid points will scale as the cube of modulus of the cut-off g-vector, or the $\frac{3}{2}$ power of the cut-off energy.

CASTEP stores just the plane wave components which lie within the cut-off sphere. When an FFT needs to be done, they are copied onto an FFT grid, which is mostly empty, for the grid will be a parallelpiped which will enclose a sphere of (probably) $1.75\times$ the diameter of the cut-off sphere. The FFT grid thus takes approximately ten times as much memory as storing a single band. However, if there are more than ten bands, the bands are likely to dominate. (And yes, CASTEP needs more than a single copy of the wavefunction and a single copy of the FFT grid, so this order-of-magnitude calculation is just that.)

Note that for a cubic system, a reciprocal space grid containing all components up to g_{cut} will contain components up to $\sqrt{3}g_{\mathrm{cut}}$ along the (111) direction in reciprocal space. For non-orthorhombic systems, this waste is worse.
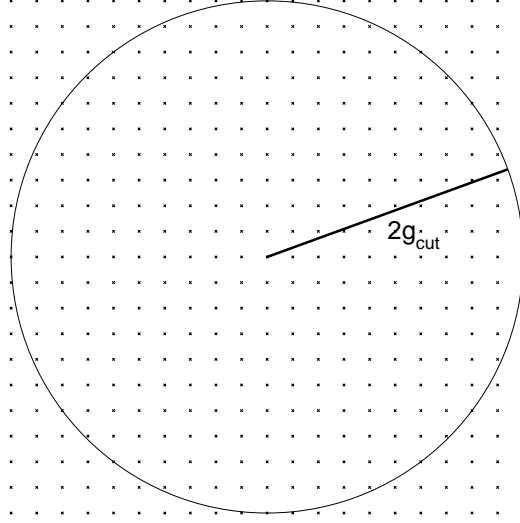
Figure 2: The cut-off sphere in a square reciprocal lattice. This is for the Γ point, for the grid includes the origin.

3.1 Orthogonality

If we desire that ψ_1 and ψ_2 be orthogonal, i.e.

$$\sum_{\mathbf{r}} \psi_1(\mathbf{r})\psi_2^*(\mathbf{r}) = 0$$

then, from the transformed ψ's.

$$\sum_{\mathbf{r}} \left(\sum_{\mathbf{g}} \psi_1(\mathbf{g}) \exp(i\mathbf{g}.\mathbf{r}) \right) \left(\sum_{\mathbf{g}'} \psi_2^*(\mathbf{g}') \exp(i\mathbf{g}'.\mathbf{r}) \right) = 0$$

One can then admire the orthogonal exponential functions eliminating the sums, and can conclude that

$$\sum_{\mathbf{g}} \psi_1(\mathbf{g})\psi_2^*(\mathbf{g}) = 0$$

Similar arguments hold for u, for the first equation can be multiplied by one as follows:

$$\sum_{\mathbf{r}} e^{-i\mathbf{k}.\mathbf{r}} \psi_1(\mathbf{r}) e^{i\mathbf{k}.\mathbf{r}} \psi_2^*(\mathbf{r}) = 0$$

which is then

$$\sum_{\mathbf{r}} u_1(\mathbf{r})u_2^*(\mathbf{r}) = 0$$

3.2 Scattering

The Schrödinger equation is usually written in real space. If one considers the effect of the Hamiltonian in k-space, one sees immediately that the ∇^2 term is diagonal: it operates on each component in k-space independently.

The $V(\mathbf{r})$ term does not have this property. However, if one considers $V(\mathbf{r})$ to have a single Fourier component

$$\left[\nabla^2 + V \exp(\mathbf{g}_v.\mathbf{r})\right] \psi(\mathbf{r})$$

can be Fourier transformed to give

$$-g^2 \psi(\mathbf{g}) + V\psi(\mathbf{g} - \mathbf{g}_v)$$

as the Fourier transform of $V(\mathbf{r})$ is now a delta function, and multiplication in real space is equivalent to convolution in reciprocal space. So this component of $V(\mathbf{r})$ has mixed states of $\mathbf{g}$ and $\mathbf{g} - \mathbf{g}_v$. As the Schrödinger equation is linear in the potential we can generalise this: the $\mathbf{g}$ and $\mathbf{g}'$ components of ψ will be independent unless $V(\mathbf{r})$ has a component with a Fourier co-efficient of $\mathbf{g} - \mathbf{g}'$

Fourier components of $V(\mathbf{r})$ which are of higher frequency than $2g_{\text{cut}}$ are uninteresting: these components will scatter a component of ψ within the g-space cut-off to something outside it, which is outside the basis set and thus disregarded. The highest frequency which can possibly be of interest is $2g_{\text{cut}}$ which will move across the diameter of the cut-off sphere.

3.3 Aliasing

If the FFT grid contains components of up to ag_{cut}, then any higher frequency components generated by the action of the potential on the wavefunction will be aliased back into this range by adding / subtracting multiples of $2ag_{\text{cut}}$. If one assumes that $a = 2$ and that the highest frequency component of the potential is also $2g_{\text{cut}}$, then this component will scatter a component of ψ at $-g_{\text{cut}}$ to $+g_{\text{cut}}$, and one of $+g_{\text{cut}}$ to $+3g_{\text{cut}}$ which will be aliased back to $-g_{\text{cut}}$. Increase a infinitesimally, and the aliased points lie outside the sphere of radius g_{cut} and are thus irrelevant.

If one chooses $a = 1.75$, then a component in the potential at $1.75g_{\text{cut}}$ will scatter a component in ψ from g_{cut} to $2.75g_{\text{cut}}$, which will be aliased back to $-0.75g_{\text{cut}}$. This now lies within the sphere of radius g_{cut}, so it is part of the basis set for the plane wave.

The smaller FFT grid has caused the $1.75g_{\text{cut}}$ component of the potential to scatter between the g_{cut} and $-0.75g_{\text{cut}}$ components of ψ, when no such scattering should occur.

3.4 Grids and Supercells

One can successfully solve the Schrödinger equation by forcing ψ to have the same periodicity as V. This is equivalent to using $\mathbf{k} = 0$ only.

One could then consider constructing a supercell, a multiple of the unit cell which is the periodic unit for V, and constraining ψ to be periodic in that supercell. Clearly as the supercell gets larger, this approximation becomes more accurate.

If one considers a supercell of twice the length of the original unit cell (and unchanged in the other dimensions), the real-space grid for ψ etc. needs twice as many points in the extended direction, for it needs to extend twice as far at the same spacing. Conversely, the reciprocal space grid now has its spacing halved, but needs to extend as far as before, so needs twice as many points. In the reciprocal space grid, half the points were in the original grid for the unit cell, but one extra point appears between each pair.
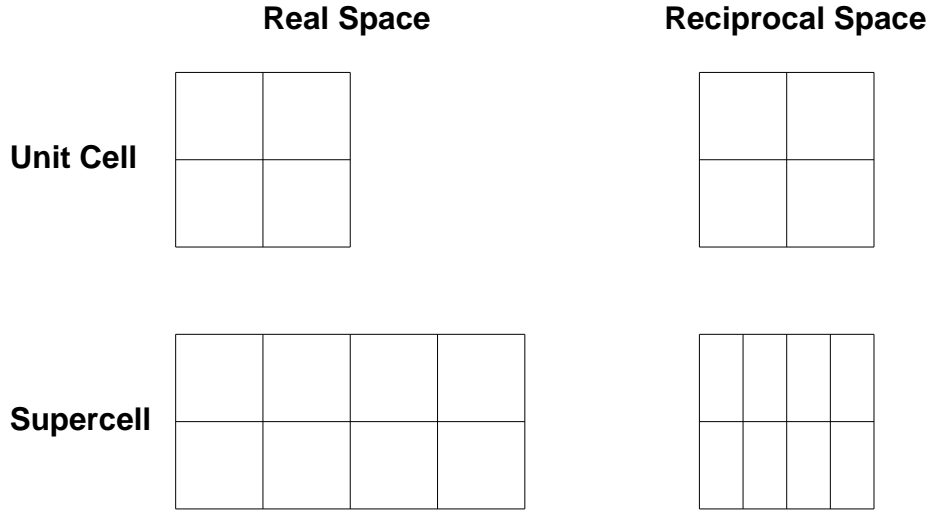
Real Space   **Reciprocal Space**

Unit Cell

Supercell

Figure 3: Real and reciprocal space grids in a cell and a 2×1 supercell

One could therefore consider the reciprocal space grid in the supercell to be made of the original grid, plus the original shifted by half a grid spacing.

But the potential is periodic in the original unit cell. Therefore the Fourier transform of the potential must be zero on all these extra grid points in reciprocal space. Although ψ is not necessarily periodic in the original unit cell, and therefore is not zero on the extra grid points in reciprocal space, the 'original' and 'extra' points will behave independently: there is no term in the Hamiltonian which can take a component of ψ on the original points and create one on the 'extra' points. So one can solve for these points as two distinct sets.

But this merely repeats what we know from Bloch's Theorem: the orginal set we equate with the $\mathbf{k} = 0$ k-point, and the new points the $\mathbf{k} = (\frac{1}{2}, 0, 0)$ k-point. The solutions at these two points are independent, and guaranteed to be orthogonal, for in k-space they are clearly orthogonal as each is zero when the other is non-zero.

4   Symmetry

If $V(\mathbf{r})$ is invariant under a symmetry operator, it does *not* follow that $\psi(\mathbf{r})$ will also be invariant.

Consider a rotation or reflection $\hat{S}$

$$\left[\nabla^2 + V(\mathbf{r})\right]\psi(\mathbf{r}) = E\psi(\mathbf{r})$$

$$\left[\nabla^2 + V(\hat{S}\mathbf{r})\right]\psi(\hat{S}\mathbf{r}) = E\psi(\hat{S}\mathbf{r})$$

as ∇^2 is invariant under such an operation, and as $V(\mathbf{r})$ is also

$$\left[\nabla^2 + V(\mathbf{r})\right]\psi(\hat{S}\mathbf{r}) = E\psi(\hat{S}\mathbf{r})$$

So $\psi(\hat{S}\mathbf{r})$ and $\psi(\mathbf{r})$ are either identical or degenerate. If $\psi(\mathbf{r})$ is known, then $\psi(\hat{S}\mathbf{r})$ is also a solution with the same eigenvalue.

Consider benzene, and its six-fold rotation. There are six possible real solutions for a wave as one moves from atom to atom with these boundary conditions:

- All atoms in phase

- Adjacent atoms in anti-phase

- Three atoms in phase, the middle one having a greater amplitude, and the other three atoms in antiphase to the first three

- A diagonally opposite pair stationary, the pair between them in phase

- Three atoms in anti-phase, the middle one having a greater amplitude, and the other three atoms being identical to their diagonal partners

- A diagonally opposite pair stationary, the pair between them in anti-phase

The rotation operator maps the first solution to itself, and the second to a multiple (-1) of itself. The third it maps to a multiple (-1) of itself if applied thrice. If applied once or twice (thus changing the position of the nodal line), the result is a linear combination of the third and forth states, which are degenerate. Similarly for the fifth and sixth.

5   k-points and Symmetry

Assume some symmetry operator leaves $V(\mathbf{r})$ and hence $\hat{H}$ invariant. Although it need not leave $\psi(\mathbf{r})$ invariant, it may produce a collection of degenerate ψs whose total contribution to the charge density is invariant under the symmetry operator. (Degeneracy implies symmetry, symmetry does not imply degeneracy.)

This also implies that quantities such as forces must obey the symmetry operator, once the wavefunction is converged (i.e. is an eigenfunction of the equation). Thus in a cubic system the forces will zero no matter how poor the cut-off and pseudo-potential!

Once one adds a k-point to the equation:

$$\left[\nabla^2 + V(\mathbf{r})\right] e^{i\mathbf{k}.\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) = E_{\mathbf{k}} e^{i\mathbf{k}.\mathbf{r}} u_{\mathbf{k}}(\mathbf{r})$$

symmetry disappears (except for certain choices of $\mathbf{k}$). If one calculates a cubic system with the k-point (0.1,0.2,0.3) one will find all sorts of forces arise.

When one applies a symmetry operator to the co-ordinates, one must also apply it to the k-point, and one finds (assuming $\hat{H}$ invariant):

$$\left[\nabla^2 + V(\mathbf{r})\right] e^{i\hat{S}_{\mathbf{k}}\mathbf{k}.\hat{S}\mathbf{r}} u_{\mathbf{k}}(\hat{S}\mathbf{r}) = E_{\mathbf{k}} e^{i\hat{S}_{\mathbf{k}}\mathbf{k}.\hat{S}\mathbf{r}} u_{\mathbf{k}}(\hat{S}\mathbf{r})$$

In other words, if we know $\psi(\mathbf{r})$ at $\mathbf{k}$, then $\psi(\hat{S}\mathbf{r})$ is the solution at $\hat{S}_k\mathbf{k}$ and has the same eigen-value. Unless $\hat{S}_{\mathbf{k}}\mathbf{k} = \mathbf{k}$ these solutions must be orthogonal.

We state without proof that relections and rotations in real space are equivalent to the same reflection or rotation in k-space.

So solving for one k-point can readily yield solutions for all symmetry-related k-points. Indeed, one is probably not interested in the ψs themselves, only their eigenvalues (all the same), and their contribution to the charge density, which can be obtained by acting on the charge density from one ψ with the symmetry operators.

5.1 Real Space Inversion Symmetry

It may be the case that one has real space inversion symmetry: $V(\mathbf{r}) = V(-\mathbf{r})$. However, once one uses a k-point other than Γ (or L, or a few of its friends), this symmetry is removed from the new Hamiltonian. It is the case that at Γ the eigenfunctions are not simply real, but also either odd or even, if real space inversion symmetry exists.

Martin, in *Electronic Structure, Basic Theory and Practical Methods*, section 4.4, claims that $u(\mathbf{r})$ is real at all $\mathbf{k}$ if real space inversion symmetry exists. I have failed to follow his proof, beyond showing that $u(\mathbf{r}) = u^*(-\mathbf{r})$, which follows from the Hamiltonian being Hermitian in reciprocal space, and hence in reciprocal space the eigenvalues are real, so that $u(\mathbf{g})$ is real, but $u(\mathbf{g})$ is not necessarily $u(-\mathbf{g})$.

6 Monkhorst Pack

The k-point selection scheme favoured by Monkhorst and Pack[1] is for a regular grid of k-points, usually shifted by one-half of the grid spacing.

6.1 Monkhorst Pack and Symmetry

Suppose one wished to use the k-point grid of the eight k-points given by all combinations of ($\frac{1}{4}$ and $\frac{3}{4}$, $\frac{1}{4}$ and $\frac{3}{4}$, $\frac{1}{4}$ and $\frac{3}{4}$). This would be more normally written as ($\pm\frac{1}{4}, \pm\frac{1}{4}, \pm\frac{1}{4}$), and, in a cubic system, all those points are related by symmetry operators, so just one needs to be calculated, and the ψs (or ρs) for the others then determined by symmetry.

This grid would be called a $2 \times 2 \times 2$ MP grid, for obvious reasons.

Consider the $4 \times 4 \times 4$ MP grid in a cubic system. It clearly contains 64 k-points.

The k-point $(\frac{1}{8}, \frac{1}{8}, \frac{1}{8})$ maps to seven others (all possible sign combinations).

The k-point $(\frac{3}{8}, \frac{1}{8}, \frac{1}{8})$ maps to 23 others (all possible sign cominations, and interchanging x, y and z).

The k-point $(\frac{3}{8}, \frac{3}{8}, \frac{1}{8})$ maps to 23 others (all possible sign combinations, and interchanging x, y and z).

The k-point $(\frac{3}{8}, \frac{3}{8}, \frac{3}{8})$ maps to seven others (all possible sign combinations).

So the calculation should be done with just four k-points, but their weights when their charge densities are symmetrised and added should be in the ratio 8:24:24:8 (or 1:3:3:1).

6.2 Monkorst Pack and Hexagonal Symmetry

In a cubic (or similar) system one produces the MP set with the highest first failure star for a given number of points by shifting all the points by half their spacing so as to exclude the Γ point.

In a hexagonal (or similar) system, this is a bad strategy. If one shifts in this fashion, the symmetry operations will generate additional kpoints not in the original MP mesh, as shown in figure 4. This is probably not helpful, and is very confusing when attempting to compare symmetrised and unsymmetrised calculations.

This problem arises because on the cubic grid, each grid point has a four-fold axis through it, as does the point at the centre of each grid cell. In the hexagonal case, there is a six-fold axis through each grid point, but not through the centre of each grid cell. If one shifts the cubic grid by an amount other than $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ then again one is attempting to generate a grid which does not have the expected four-fold axis about the origin, and the symmetry operations will produce a mess.

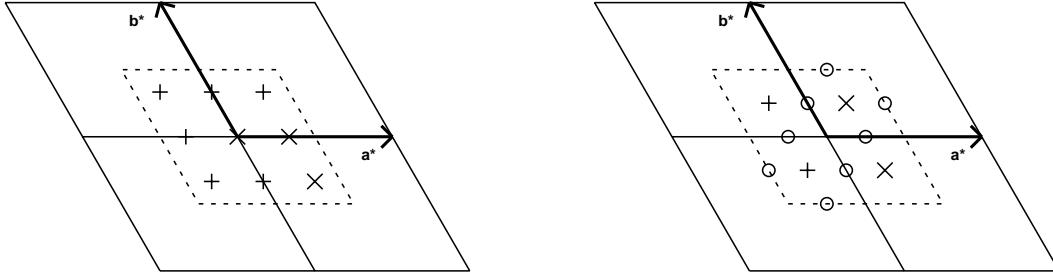[1]Monkhorst, H.J and Pack, J.D. *Phys Rev B* **13** 5188 (1976)

Figure 4: A 3×3 MP mesh in a hexagonal system, shifted to include the Γ point, and a 2×2 mesh shifted to exclude it. Required non-equivalent points are marked by 'X', required points generated by the symmetry operations on these by '+', and unwanted points generated are marked by '∘'.

6.3 Equivalent k-points

Suppose one wishes to compare the energetics of two systems, one of which reconstructs to form a unit cell of length two units in the z direction, the other of length three units. To ensure that one's calculated energy difference do not have terms arising from the use of different k-points, FFT grids, etc., one could perform both calculations in a cell of length six units. However, this would be expensive.

So, instead, one can perform calculations in the three-unit cell with k-points at $k_z = \frac{1}{8}$ and $\frac{3}{8}$, and in the two unit cell with k-points at $k_z = \frac{1}{12}$, $\frac{3}{12}$ and $\frac{5}{12}$. Assuming that one can also choose the FFT grid sizes to be in the ratio $3 : 2$, these calculations should be precisely equivalent to a calculation in the six-unit supercell with $k_z = \frac{1}{4}$.

7 Nasty Symmetry

Just because the crystal as a whole has a particular symmetry is no guarentee that the primitive unit cell, or, worse, a random unit cell, will have the same symmetry. This is obvious: a cubic system can be described by a cubic unit cell, which has all the right symmetry, a $1 \times 2 \times 3$ supercell of that cell, which has very little symmetry, or a strange '2-atom' cell which has always confused CASTEP somewhat.

In the case of hexagonal systems, life is worse. A unit cell is a parallelpiped described by three axes. A hexagonal prism, which is the obvious repeat unit with the symmetry of a hexagonal system, is not a parallelpiped.

This becomes painfully obvious when looking at the reciprocal space unit cell of a hexagonal system. The directions $\mathbf{a}^* + \mathbf{b}^*$ and $\mathbf{a}^* - \mathbf{b}^*$ are equivalent: they are related by a $\pi/3$ rotation about the 6-fold axis. However, the reciprocal space unit cell, and thus also the reciprocal space grid, extends twice as far in the $\mathbf{a}^* - \mathbf{b}^*$ direction as it does in the $\mathbf{a}^* + \mathbf{b}^*$ direction.

This results in the reciprocal space grid not mapping back onto itself under the operation of a symmetry operation of the Hamiltonian. By truncating the grid, we have reduced the symmetry of the system. The 'correct' thing to do is probably to set all the potential and wavefunction components which lie in this orphaned part of the reciprocal space grid to zero, which effectively produces a hexagonal cylinder as the FFT grid shape. Does your favourite code do this?

7.1 It's not just Hexes

Many people relax on reading the above section, resolving never to touch anything with non orthorhombic cells, believing this solves the problem. It doesn't. Consider a 2-D system with square symmetry, and a 4×4 grid, such as shown in figure 6.

Figure 5: The reciprocal space cell of a conventional hexagonal lattice

Figure 6: A 4×4 cubic grid does not have rotational symmetry about the origin.

The grid will run from -1 to 2 in each direction, which means that the four-fold axis will map things at $+2$ to non-existant points at -2. This problem does not arise if the grid is odd, and thus has equal numbers of points either side of the origin, but does for even grids.

8 Unit cell to Supercell

A calculation at a k-point of $\frac{1}{3}$ should correspond to a calculation at the Γ point in a cell three times as long. To construct the full real-space wavefunction in the longer cell, one can write

$$
\begin{aligned}
\Psi(r) \quad &= e^{ikr}u(r) & 0 \le r \le L \\
&= e^{ikr}u(r-L) & L \le r \le 2L \\
&= e^{ikr}u(r-2L) & 2L \le r \le 3L
\end{aligned}
\tag{1}
$$

for $u(r)$ will have been calculated for $0 \le r \le L$ only. One can express the above in other ways too:

$$
\begin{aligned}
\Psi(r) \quad &= \psi(r) & 0 \le r \le L \\
&= e^{ikL}\psi(r-L) & L \le r \le 2L \\
&= e^{2ikL}u(r-2L) & 2L \le r \le 3L
\end{aligned}
$$

8.1 Continuity

It is clear physically that ψ must be continuous in real space: a discontinuity would be extremely expensive energetically. It is also clear from the form of equation 1 that ψ will be continuous as long as u is continuous with $u(L) = u(0)$. However, $\psi(L) = \psi(0)$ does not necessarily hold.

8.2 Real or Complex?

There appears to be an inconsistency here. At the Γ point, a wavefunction can be expressed as real, but the above functions are not obviously real.

However, the above wavefunctions do occur in degenerate pairs, for $\psi_{-\mathbf{k}}(\mathbf{r}) = \psi_{\mathbf{k}}^*(\mathbf{r})$. Clearly any linear combination of these degenerate eigenvectors must also be an eigenvector, and the two obvious combinations are

$$
\begin{aligned}
\psi_1(\mathbf{r}) &= \frac{1}{\sqrt{2}}\left(\psi_{\mathbf{k}}(\mathbf{r}) + \psi_{-\mathbf{k}}(\mathbf{r})\right) \\
\psi_2(\mathbf{r}) &= \frac{1}{\sqrt{2}i}\left(\psi_{\mathbf{k}}(\mathbf{r}) - \psi_{-\mathbf{k}}(\mathbf{r})\right)
\end{aligned}
$$

These must both be real by construction. Note that ψ_1 and ψ_2 constructed thus are real, degenerate, and orthogonal, but, in general, give rise to different charge densities, whereas $\psi_{\mathbf{k}}$ and $\psi_{-\mathbf{k}}$ are complex, degenerate, orthogonal, and give rise to identical charge densities. The charge density produced by ψ_1 and ψ_2 together is, of course, identical to that produced by $\psi_{\mathbf{k}}$ and $\psi_{-\mathbf{k}}$ togther.

Finally, given that ψ_1 and ψ_2 are real and degenerate, any linear combination of them (using real coefficents) will also be real and an eigenfunction. However, the $\psi_{\mathbf{k}}$ / $\psi_{-\mathbf{k}}$ pair are uniquely specified by the additional contraint that they are purely of the form $\exp(i\mathbf{k}.\mathbf{r})$ and not $\exp(-i\mathbf{k}.\mathbf{r})$ (and vice versa).

One can invert these equations:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{2}}\left(\psi_1(\mathbf{r}) + i\psi_2(\mathbf{r})\right)$$

$$\psi_{-\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{2}}\left(\psi_1(\mathbf{r}) - i\psi_2(\mathbf{r})\right)$$

and thus form two complex $\psi_{\mathbf{k}}$s which clearly produce the same densities from two degenerate real ψs which need not produce the same density.

8.3 Degeneracy and Symmetry

Finding a degeneracy is surprising: it implies the existance of a symmetry operator. In this case, there is a symmetry operation in the supercell: a displacement corresponding to the unit cell vector.

Furthermore, $\psi_{\mathbf{k}}$ is invariant under this operation, as it is simply multiplied by a phase factor.

8.4 Really Fast

If one uses the Γ point, then $\psi_{\mathbf{k}}$ and $u_{\mathbf{k}}$ can both be written as real functions. This halves memory requirements, and speeds up calculations. It is also true for the k-points whose components are zero and $\frac{1}{2}$ in any combination, for in these cases the phase difference between adjacent cells is a factor of ± 1.

9 Theory meets CASTEP

9.1 FFT grids and symmetry

In the eight-atom cubic cell, of, for example, silicon, the atoms at $(0,0,0)$ and $(0.5, 0.5, 0)$ are equivalent. (Indeed, the cubic cell contains just two types of atoms, distinguished by the orientation of their nearest neighbours.) A translation by $(0.5, 0.5, 0)$ is a symmetry operation of the Hamiltonian.

However, if one uses an odd FFT grid, one of these atoms is on a grid point in real space, and one is not. A translation by $(0.5, 0.5, 0)$ is no longer a symmetry operation of the Hamiltonian, whereas for an even grid it would be.

This can be seen in the resulting eigenvalues. Eigenvalues from a calculation using an eight-atom cell of silicon and two different FFT grids are shown in figure 7. States which are degenerate have acquired eigenvalues which are identical to ± 2 in the last digit, whereas the odd FFT grid has less degeneracy, splitting previously degenerate levels by ± 20 in the last digit.

Fortunately this is too small to matter to Physicists. It might annoy mathematicians though.

This is related to the rigid shift problem. If one displaces a CASTEP cell, one should get an identical answer. However, this is only strictly true if the displacement is an exact multiple of the real-space FFT grid cell vectors. Otherwise, energies and forces can change. So too will symmetries, as CASTEP cannot cope with rotation axes through arbitrary points. Again, these effects should be small enough to be ignorable in most circumstances: the thought that as one relaxes a system the atoms are bumping up and down over an artificial disturbance to the potential caused by the choice of location of the FFT grid points can be upsetting.

$24 \times 24 \times 24$		$25 \times 25 \times 25$	
Γ	$(0.5, 0.5, 0)$	Γ	$(0.5, 0.5, 0)$
-0.41712922	**-0.34294909**	**-0.41742090**	**-0.34301592**
-0.26646483	**-0.34294909**	-0.26675441	-0.34301587
-0.26646483	**-0.34294909**	-0.26675441	-0.34301587
-0.26646483	**-0.34294909**	**-0.26675435**	**-0.34301569**
-0.26646483	-0.18188123	**-0.26675435**	-0.18214242
-0.26646483	-0.18188123	**-0.26675435**	**-0.18214235**
-0.26646483	-0.18188123	-0.26675422	-0.18214228
-0.08665451	-0.18188123	**-0.08684657**	-0.18214228
-0.08665451	**-0.12035867**	**-0.08684657**	**-0.12050372**
-0.08665451	**-0.12035867**	**-0.08684657**	-0.12050361
-0.08665451	**-0.12035867**	-0.08684616	-0.12050360
-0.08665451	**-0.12035867**	-0.08684616	-0.12050360
-0.08665451	-0.03059775	-0.08684616	**-0.03065965**
0.01887506	-0.03059775	**0.01859783**	**-0.03065935**
0.01887507	-0.03059775	**0.01859784**	**-0.03065935**
0.01887509	-0.03059775	**0.01859786**	-0.03065922
0.04147361	**0.10551168**	0.04073573	**0.10463149**

Figure 7: Eigenvalues of an eight-atom cubic Si calculation with two different FFT grid sizes

9.2 ψ or u?

CASTEP outputs wavefunctions in reciprocal space. If what it outputs is FFTed using an unshifted grid, then $u(\mathbf{r})$ is produced. In order to obtain $\psi(r)$ one must either FFT using a shifted grid, for the g-vectors at which the wavefunction is stored are of the form $\mathbf{g} + \mathbf{k}$, or FFT and then multiply by $\exp(i\mathbf{k}.\mathbf{r})$, the latter being much easier numerically.

10 Aperiodic Systems

Sometimes the system one is attempting to model is not a 3D crystal. At this point one does not need to worry about sampling k-space in order to calculate accurately the unwanted interaction between the repeat units. For isolated molecules, one often uses the Γ point, preferably taking advantage of the real wavefunctions which result.

However, this approach maximises the unwanted interaction between the periodic images, as all the images are in phase, and thus forming bonding orbitals. If instead the $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ k-point were used, the interaction between the images would be reduced, and a smaller cell (with less vacuum) might be possible. However, this choice of k-point requires the use of a complex code, and favours the $(1, 1, 1)$ direction over other directions such as $(-1, 1, 1)$ which should perhaps be equivalent.

For 2D systems, the argument is between using a plane of k-points with $k_z = 0$, and noting that inversion symmetry in k-space will halve the number of points needed, although the interaction between the slab and its periodic image will be maximised, and using $k_z = \frac{1}{4}$, where inversion symmetry merely adds the $k_z = -\frac{1}{4}$ plane, so twice as many k-points will be needed, but the interaction between periodic images will be reduced, so less vacuum may be needed.

11 Plotting $\psi_{\mathbf{k}}$

Away from the Γ point, $\psi_{\mathbf{k}}$ encodes information about two distinct degenerate wavefunctions. These wavefunctions are periodic in the corresponding supercell, so merely tiling the fraction of them which
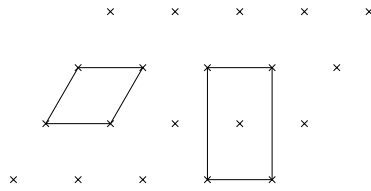
Figure 8: Two possible cells on a hexagonal lattice

appears in the unit cell may make no sense at all (and may introduce discontinuities).

Plotting a complex value throughout a 3D space also confuses humans.

Humans probably wish to see real wavefunctions, and a pair of these could be produced in the supercell in which the k-point considered maps to the Γ point.

Charge densities are easier, for $\psi_{\mathbf{k}}(\mathbf{r})\psi_{\mathbf{k}}^*(\mathbf{r})$ is periodic in the unit cell, and is identical to $\psi_{-\mathbf{k}}(\mathbf{r})\psi_{-\mathbf{k}}^*(\mathbf{r})$. However, it does not correspond to the charge density produced by the real wavefunction referred to above.

Quite simply, there exists a degeneracy, and therefore any linear combination of the degenerate eigenfunctions is an equally valid solution. But they are different.

12   Further Work

There follow some ideas for amusement, using your favourite plane-wave code.

- By how much does the energy change if one rigidly shifts all atoms by one half of an FFT grid cell in real space? How does this answer change if one uses an FFT grid extending to $2g_{\mathrm{cut}}$ instead of $1.75g_{\mathrm{cut}}$? If one uses norm conserving pseudopotentials, not ultrasoft? If one turns off the XC potential? Is it zero if one does all of these things?

- Do a symmetrised and unsymmetrised calculation give identical results? Or only for orthorhombic systems with odd FFT grids?

- If one has two different-shaped cells on the same lattice, how can one best set up calculations to give comparable energies? for instance, in figure 8, can one set up calculations such that the orthorhombic cell gives precisely twice the energy of the hexagonal cell?

- Figure 7 used norm conserving pseudopotentials. Would using grids of 15^3 and 16^3 with ultrasofts be reasonable? What is the artificial splitting in this case?

- Is Martin correct? (Section 5.1)

Acknowledgements

Credit is due to PDH for any concepts and equations which are correct, and to MJR for those which are not. The figures were prepared using emacs, and this document was typeset using LaTeX.