FIP10601 – Text 4

Energy bands in two limits

After presenting a general view of the independent-electron approximation, highlighting the importance and consequences of Bloch's theorem, it is time to explore specific models of the periodic potential so that actual calculations can be performed.

We will do this in two extreme limits, which define further approximations to the problem of independent electrons. They are

- **the *tight-binding* approximation**, in the limit of electrons strongly connected to the atoms;
- **the *nearly-free-electron* approximation**, for very loose electrons, near the free-particle limit.

These approximations are applicable respectively to **narrow** and **wide** energy bands. Intermediate cases can also be addressed, but involve a large variety of theoretical and computational methods, which need a more specialized course to be presented in detail.

Energy bands in the *tight-binding* limit

We have already observed that a flat band corresponds to localized electron states. Thus, in a narrow band the electrons are almost localized, which means that they are tightly bound to the atoms. In this case, it is convenient to express Bloch states in terms of fairly localized wavefunctions, centered on lattice points, which are known as *Wannier functions*. The starting point to define these functions is the periodicity property of Bloch functions **in k-space**,

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \psi_{n,\mathbf{k}+\mathbf{K}}(\mathbf{r}) \ , \tag{1}$$

due to the equivalence between $\mathbf{k}+\mathbf{K}$ and $\mathbf{k}$ for any reciprocal-lattice vector $\mathbf{K}$. We saw earlier that a function of real-space coordinates that is periodic in the Bravais lattice has a discrete Fourier representation with non-zero coefficients only for reciprocal-lattice vectors. Similarly, a Bloch function, viewed as a function of $\mathbf{k}$ that is periodic in the reciprocal lattice, can be expressed as a Fourier representation with non-zero coefficients only for Bravais-lattice vectors, i.e.,

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_i w_n(\mathbf{R}_i, \mathbf{r}) \mathrm{e}^{\mathrm{i}\mathbf{k}.\mathbf{R}_i} \ . \tag{2}$$

This last equation may be viewed as a counterpart to Eq. (28) of Text 3, exchanging the roles of $\mathbf{r}$ and $\mathbf{k}$ as variable and parameter, consequently changing the sum over reciprocal-lattice vectors ($\mathbf{K}$) to a sum over real-lattice vectors ($\mathbf{R}_i$). Reversing the Fourier representation (2), we have

$$w_n(\mathbf{R}_i, \mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} \mathrm{e}^{-\mathrm{i}\mathbf{k}.\mathbf{R}_i} \psi_{n\mathbf{k}}(\mathbf{r}) \ . \tag{3}$$

Now, remembering the Eq. (12) of Text 3,

$$\psi_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r})\, \mathrm{e}^{\mathrm{i}\mathbf{k}.\mathbf{r}} \ , \tag{4}$$

we can write

$$w_n(\mathbf{R}_i, \mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}.(\mathbf{r}-\mathbf{R}_i)} u_{n\mathbf{k}}(\mathbf{r}) \ . \tag{5}$$

Then, taking into account the real-lattice periodicity of $u_{n\mathbf{k}}(\mathbf{r})$, Eq. (13) of Text 3, we see that Wannier functions depend only on the difference $\mathbf{r} - \mathbf{R}_i$, that is,

$$w_n(\mathbf{R}_i, \mathbf{r}) \equiv w_n(\mathbf{r} - \mathbf{R}_i) \ . \tag{6}$$

An important property is that Wannier functions corresponding to different bands and/or centered on different lattice sites are orthogonal. The demonstration is simple:

$$\begin{aligned} \int \mathrm{d}^3 r\, w_n^*(\mathbf{r} - \mathbf{R}_i) w_m(\mathbf{r} - \mathbf{R}_j) &= \frac{1}{N} \sum_{\mathbf{k}\mathbf{k}'} \mathrm{e}^{\mathrm{i}(\mathbf{k}.\mathbf{R}_i - \mathbf{k}'.\mathbf{R}_j)} \int \mathrm{d}^3 r\, \psi_{n\mathbf{k}}^*(\mathbf{r}) \psi_{m\mathbf{k}'}(\mathbf{r}) \\ &= \frac{1}{N} \sum_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}.(\mathbf{R}_i - \mathbf{R}_j)} \delta_{nm} = \delta_{nm}\delta_{ij} \ . \end{aligned} \tag{7}$$

Bloch and Wannier representations

In Text 3 we observed that Bloch states (including the spin index σ) could be represented by vectors $|n\mathbf{k}\sigma\rangle$ in the abstract one-electron Hilbert space. Equivalently, Wannier functions are the coordinate representation of a set of vectors $|ni\sigma\rangle$, where i refers to the lattice vector $\mathbf{R}_i$. These two sets of vectors are connected by transformations like Eqs. (2) and (3), i.e,

$$|n\mathbf{k}\sigma\rangle = \frac{1}{\sqrt{N}} \sum_i \mathrm{e}^{\mathrm{i}\mathbf{k}.\mathbf{R}_i} |ni\sigma\rangle \ , \qquad |ni\sigma\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} \mathrm{e}^{-\mathrm{i}\mathbf{k}.\mathbf{R}_i} |n\mathbf{k}\sigma\rangle \ . \tag{8}$$

Both sets are orthonormal,

$$\langle n\mathbf{k}\sigma | m\mathbf{k}'\sigma'\rangle = \delta_{nm}\delta_{\mathbf{k}\mathbf{k}'}\delta_{\sigma\sigma'} \ , \qquad \langle ni\sigma | mj\sigma'\rangle = \delta_{nm}\delta_{ij}\delta_{\sigma\sigma'} \ , \tag{9}$$

and complete,

$$\sum_{n\mathbf{k}\sigma} |n\mathbf{k}\sigma\rangle\langle n\mathbf{k}\sigma| = 1 \ , \qquad \sum_{ni\sigma} |ni\sigma\rangle\langle ni\sigma| = 1 \ . \tag{10}$$

So, both can be used as basis vectors for representations in the Hilbert space.

Since Bloch states are energy eigenstates, the **Bloch representation** of the Hamiltonian is simply

$$\mathcal{H} = \sum_{nm} \sum_{\mathbf{k}\mathbf{k}'} \sum_{\sigma\sigma'} |n\mathbf{k}\sigma\rangle\langle n\mathbf{k}\sigma|\mathcal{H}|m\mathbf{k}'\sigma'\rangle\langle m\mathbf{k}'\sigma'| = \sum_{n\mathbf{k}\sigma} |n\mathbf{k}\sigma\rangle \varepsilon_{n\mathbf{k}} \langle n\mathbf{k}\sigma| \ . \tag{11}$$

The **Wannier representation** is

$$\mathcal{H} = \sum_{nm} \sum_{ij} \sum_{\sigma\sigma'} |ni\sigma\rangle\langle ni\sigma|\mathcal{H}|mj\sigma'\rangle\langle mj\sigma'| \equiv \sum_{nm} \sum_{ij} \sum_{\sigma} |ni\sigma\rangle \mathcal{H}_{ni,mj} \langle mj\sigma| \ , \tag{12}$$

in terms of the matrix elements

$$\mathcal{H}_{ni,mj} \equiv \langle ni|\mathcal{H}|mj\rangle \ , \tag{13}$$

which do not depend on spin for the presently assumed form of the Hamiltonian.

Tight-binding approximation

Let us analyze the tight-binding limit, starting with a simple case: a single atom per unit cell, and a single valence orbital per atom (non-degenerate atomic level), so that we may drop the band index. In this case, separating local and non-local terms in Eq. (12), we can write it as

$$\mathcal{H} = \varepsilon_0 \sum_{i\sigma} |i\sigma\rangle\langle i\sigma| - \sum_{ij\sigma} |i\sigma\rangle t_{ij} \langle j\sigma| \ , \tag{14}$$

where $\varepsilon_0 \equiv \mathcal{H}_{i,i}$, which should be close to the atomic eigenvalue, plays the role of a reference energy for the band, while $t_{ij} \equiv -\mathcal{H}_{i,j}$ for $i \neq j$ is called the *hopping integral* (or just *hopping*), and can be viewed as a measure of the kinetic energy due to tunneling across a potential barrier between two lattice sites. Note that the translation symmetry (equivalence of lattice sites) implies that t_{ij} must depend only on the **relative** position of the sites i and j. Within the tight-binding approach, it is assumed that hopping integrals are nonzero only between sites at short distances in the lattice, since the matrix elements $\mathcal{H}_{i,j}$ would be null for non-overlapping Wannier functions. If all nearest neighbors are equivalent (as, for example, in the cubic system), we are left with a single hopping parameter t, and the second sum in Eq. (14) is restricted to nearest neighbors.

Transforming the Hamiltonian (14) back to Bloch representation, one can easily verify (we leave the derivation as an **exercise**) that the electron energies depend on ε_0 and t_{ij} through the relation

$$\varepsilon(\mathbf{k}) = \varepsilon_0 - \sum_j t_{ij} \, \mathrm{e}^{\mathrm{i}\mathbf{k}.(\mathbf{R}_j - \mathbf{R}_i)} \ . \tag{15}$$

Note that the result of the sum in the above equation is independent of i.

Restriction to nearest-neighbor hopping yields the simpler form

$$\varepsilon(\mathbf{k}) = \varepsilon_0 - t \sum_{\boldsymbol{\delta}} \mathrm{e}^{\mathrm{i}\mathbf{k}.\boldsymbol{\delta}} \,, \tag{16}$$

where $\boldsymbol{\delta}$ is a vector connecting a given lattice site to one of its nearest neighbors. For example, applying this to a simple-cubic lattice (with lattice parameter a) yields

$$\varepsilon(\mathbf{k}) = \varepsilon_0 - 2t[\cos(k_x a) + \cos(k_y a) + \cos(k_z a)] \,. \tag{17}$$

If a single band is relevant, it is usual to choose $\varepsilon_0 = 0$.

Equation (17) explicitly reveals the periodicity of energies in $\mathbf{k}$-space. Here it appears as a period $k_\mu a = 2\pi$, for any direction $\mu = x, y, z$, consistent with the fact that the reciprocal-lattice vectors are combinations of primitive vectors of magnitude $2\pi/a$ along the coordinate axes. The same is seen (more formally) for the general case in Eq. (15) by adding a reciprocal-lattice vector to $\mathbf{k}$.

The choice of a negative sign for the hopping term in Eq. (14) implies that for $t > 0$ we have an energy minimum (lower band edge) at the origin of $\mathbf{k}$-space (which is the BZ center). It is important to note that the band width (difference between maximum and minimum energies in the band) is proportional to t. Since the hopping, by hypothesis, is small in the tight-binding limit, we are modeling a **narrow** band, as assumed at the beginning.

Equation (17) also allows us to verify that surfaces of constant energy are invariant under all symmetry operations of the cubic point group. For simplicity, we drop the k_z term, turning to a two-dimensional version of the model in the square lattice. Some constant-energy **lines** (not **surfaces** in $d = 2$) are shown in Fig. 1. The operations that take a

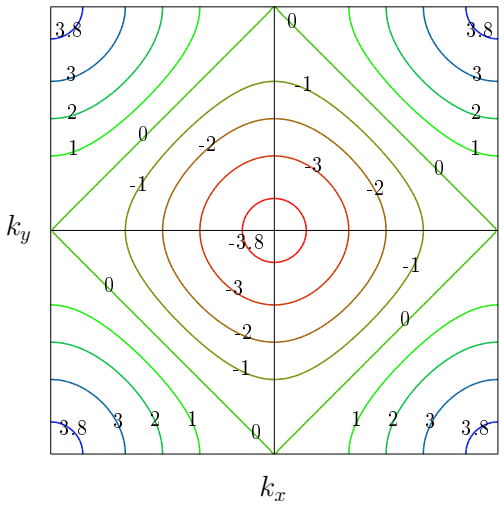

Figure 1: Constant energy lines for a square-lattice tight-binding band within the 1st BZ. The numbers on contour lines indicate the corresponding energy value (in units of t).

square into itself (apart from the identity) are rotations of 90° (and multiples) around an axis perpendicular to the plane of the square passing through its center, reflections on the coordinate axes (chosen as shown in Fig. 1), and reflections on the square's diagonals. We will discuss this group in more detail later.

At this point, it is important to look critically at what we have done so far. In Text 1 we mentioned that a general model is usually turned into a completely different form, depending on the specific problem being addressed, in order to get rid of irrelevant degrees of freedom, or unknown (but judged unnecessary) details. Here, by choosing a convenient representation of the Hamiltonian, and with reasonable assumptions concerning a specific class of systems, we have bypassed the need to know in detail the full periodic potential $V(\mathbf{r})$, ending up with a single (yet undetermined) parameter, the hopping integral t (or a few t_{ij} values in a less symmetric lattice). To determine this parameter for a specific system, one can follow two different lines. The **first-principles** approach involves actual (numerical) evaluation of matrix elements of the Hamiltonian, with Wannier functions built up from atomic orbitals appropriate to the system. The **phenomenological** approach consists in treating t as an adjustable parameter, to be determined from experimental values of physical quantities which can be obtained as functions of t in the model. Of course the latter is the simplest approach, and often employed.

Effective mass – electrons and holes

Before moving on to the large-band limit, let us look at the tight-binding solution in the low-energy sector, near the bottom of the band. Now, it is more convenient to change our choice of local energy ε_0 to have $\varepsilon(\mathbf{k}) = 0$ at the bottom, and only positive energies. Using $\varepsilon_0 = 6\,t$ in the simple-cubic example, if we expand the cosine functions in Eq. (17) around $\mathbf{k} = \mathbf{0}$ we easily arrive at the approximate expression

$$\varepsilon(\mathbf{k}) \simeq t\,a^2\,k^2\ , \tag{18}$$

with corrections of 4th order in $k \equiv |\mathbf{k}|$. This quadratic form in the wavevector magnitude reminds us of the purely kinetic energy of a free particle, $\hbar^2 k^2/2m$. We can then rewrite Eq.(18) as

$$\varepsilon(\mathbf{k}) \simeq \frac{\hbar^2 k^2}{2m^*}\ , \qquad m^* \equiv \frac{\hbar^2}{2ta^2}\,, \tag{19}$$

where m^* can be viewed as an *effective mass*. Therefore, the effect of the lattice potential has been incorporated into an effective mass of the electron, that now behaves as a free particle. We must keep in mind that this equivalence is only valid for electrons occupying states near the bottom of the band. So, we cannot in general replace a tight-binding model by a free-electron one, but we can do it if we are focusing on the ground-state (or very close to it), and the number of electrons in the system is such that the Fermi energy falls near the band minimum.

The effective mass that we obtained above is a single parameter due to the symmetries of the cubic lattice, for which the expansion contains squares of each wavevector component with equal coefficients. In a general case, the quadratic term in the expansion of $\varepsilon_n(\mathbf{k})$ involves different coefficients for distinct directions, or even products of different components. The concept of effective mass can be generalized to any independent-electron system by defining the inverse effective-mass **tensor**

$$\left[\mathbf{M}_n^{-1}(\mathbf{k})\right]_{\mu\nu} = \frac{1}{\hbar^2}\,\frac{\partial^2\varepsilon_n(\mathbf{k})}{\partial k_\mu \partial k_\nu}\,. \tag{20}$$

If we apply this expression to the simple-cubic tight-binding energies, Eq.(17), and take the limit $k \to 0$, we obtain a diagonal tensor with equal elements $1/m^*$.

The general effective-mass tensor of Eq. (20) depends on the band index, and is defined for any $\mathbf{k}$ in the 1st BZ. Of course, practical use of this definition only makes sense in an expansion of the energy eigenvalues around a reference point for which we have an energy minimum. In this case the effective mass tensor is evaluated **at** that point, and the connection with an effective free-electron model can be made.

From the above discussion, a somewhat puzzling conclusion is that electrons occupying states near the **top** of a band have **negative** effective mass. This is inconvenient since, for instance, in the classical limit we would have a particle accelerated contrarily to the force acting on it. This problem is circumvented in a clever way. The total energy of the electrons in a given band is equal to the total energy of a completely filled band **minus** the energy of the empty states. So, if we take the full band as reference, removing electrons from its top is equivalent to **adding holes** with energies that are the **negative** of those of the electron states being emptied. We then view this band as a *hole system*, with the holes having a positive effective mass!

Interpreting bands as electron-like or hole-like is completely arbitrary (provided we do not mix the two in the same band). But the appropriate choice is to have electrons when the Fermi level is near the band bottom, and holes when it is close to the top. In both cases we might be able to use an effective free-particle model if the "occupied" states are close enough to the energy extremum.

Later on, we will see how the effective mass enters into the description of dynamical processes, including the effect of applied external fields. Since electric and magnetic fields couple to the particle's electric charge, it is important to know what is the charge of a hole. It is probably intuitive that this charge is $+e$, since the suppression of an electron changes the charge associated to its state from $-e$ to zero, which is equivalent to adding a positive charge of the same magnitude.

Energy bands in the *nearly-free-electron* limit

We now turn to the opposite limit with respect to tight binding, which we could call "very loose binding". In this case, the electrons can easily move among the atoms.

The small hopping amplitude t of tightly bound electrons, which yields narrow energy bands, can be traced back to a weak overlap between Wannier functions, corresponding to somewhat large distances between neighboring atoms. With such quite localized Wannier functions the position probability amplitude of the electron varies strongly in space. On the other hand, if the interatomic distance is shorter, the overlap increases, resulting in a larger hopping amplitude, hence a larger bandwidth, but also a more smoothly varying position probability. The extreme limit of smooth variation is a constant probability, which is achieved in the free-electron limit with plane-wave eigenfunctions.

Therefore, when dealing with wide energy bands we are closer to the free-electron limit, which means that the lattice potential must be weak, and may be treated as a perturbation. The zeroth-order of perturbation theory corresponds to completely neglecting the lattice potential. This yields a pure free-electron model, for which the energy eigenfunctions are plane waves with wavevector $\mathbf{k}$, and the energy eigenvalues are $\varepsilon(\mathbf{k}) = \hbar^2 k^2/2m$. This is a single "band", which is unlimited (except for the fact that all energies are positive).

But we want to make contact with the generic prediction of a band structure for Bloch electrons in a lattice. This is done by taking the lattice into account to define a 1st BZ, to which the *reduction* of the free-electron energies is **imposed** through the general periodicity condition, Eq. (15) of Text 3. Then, the "folded" branches from higher energies define new bands inside the 1st BZ. This is called the *empty-lattice approximation*. Since the original energies were continuously growing (i.e., without gaps), the new bands will touch or cross at some points of the 1st BZ. With the aid of group theory, we can find out which of these crossing points are allowed by symmetry, and which just reflect accidental degeneracies. The latter are the points where perturbative corrections are mostly relevant.

We will first illustrate this procedure with the simplest possible model: a **one-dimensional crystal**. Let us consider a hypothetical crystal in one dimension (1D), consisting of a chain of identical atoms, equally spaced by a distance a. The Bravais lattice is a line of points with the same spacing (the only Bravais lattice in 1D). The reciprocal lattice is a line of points in reciprocal space with spacing $2\pi/a$. The 1st BZ is a segment of the k axis between $-\pi/a$ and π/a. Free-electron energies can be reduced to the 1st BZ by shifting the corresponding k's by appropriate multiples of $\pm 2\pi/a$.

The three lowest bands are shown on Fig. 2 (left side). We observe degeneracies at the center and edge of the 1st BZ, respectively denoted as Γ and L. These two points have the same wavevector group,

$$G_\Gamma = G_L = \{E, m\} \ , \tag{21}$$

where E is the identity operation and m (mirror) is the reflection (or inversion) through the origin. Note that the Γ point remains fixed under these operations. The two L points are interchanged by the operation m, but they are **equivalent** since they differ by a reciprocal lattice vector ($\pm 2\pi/a$).

Based on the generic discussion presented in Text 2, we can make the following observations:

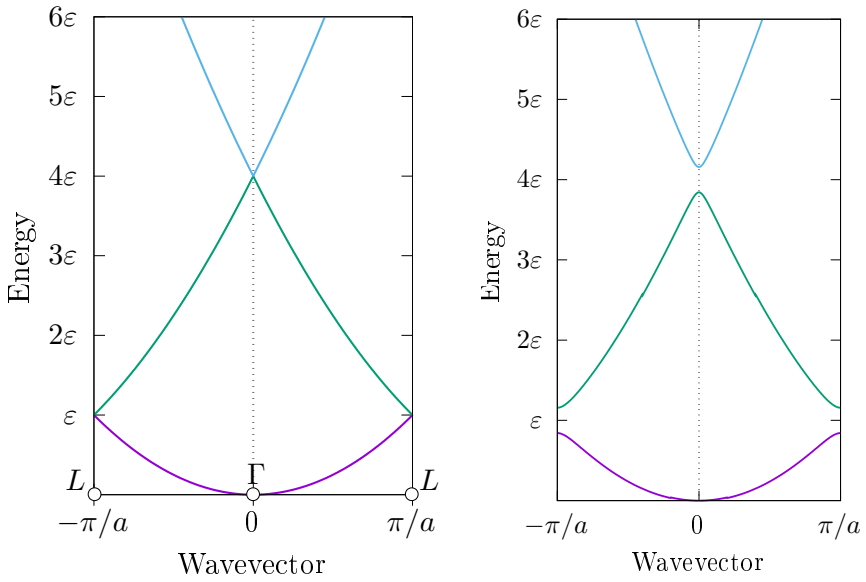

Figure 2: Some energy bands for a simple lattice in one dimension in the empty-lattice (left) and nearly-free-electron (right) approximations. The energy scale is given by $\varepsilon \equiv \hbar^2\pi^2/(2ma^2)$.

- E and m are two different classes, due to the absence of other elements in the group to relate them by a unitary transformation.
- So, there are two irreducible representations. This is also the number of elements, which means that *all irreducible representations are one-dimensional*.
- Therefore, **there are no essential degeneracies**.

From this analysis, we conclude that the degeneracies observed in Fig. 2 are **accidental**. They are due to the choice of a null potential.

A weak potential may be taken into account to first order in Perturbation Theory, which characterizes the *nearly-free-electron approximation*. Since this approximation is usually studied in introductory courses of Solid State Physics, we will just present its main features, without any detailed development.

It is convenient to express Bloch functions as combinations of plane waves. The appropriate form is

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{K}} c_{\mathbf{k}-\mathbf{K}}\, \mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{K}).\mathbf{r}}\ , \tag{22}$$

which is the product of a plane wave by the Fourier representation of a function with the lattice periodicity in space. Using this form in the time-independent Schrödinger's equation, and taking into account that a periodic potential has nonzero Fourier components

only for reciprocal-lattice vectors, we obtain

$$\left[\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{K}}\right] c_{\mathbf{k}-\mathbf{K}} - \sum_{\mathbf{K}'} V_{\mathbf{K}'-\mathbf{K}}\, c_{\mathbf{k}-\mathbf{K}'} = 0 \; . \tag{23}$$

It is usual to choose the uniform component $V_{\mathbf{K}=0}$ to be null, so that the sum above does not include $\mathbf{K}' = \mathbf{K}$.

Up to this point there is no additional approximation. The above equation is completely equivalent to the time-independent Schrödinger's equation. We also see that the zeroth-order solutions are the empty-lattice energies. First, let us suppose that a specific eigenvalue $\varepsilon^0_{\mathbf{k}-\mathbf{K}}$ is non-degenerate. Then, the coefficient $c_{\mathbf{k}-\mathbf{K}} \sim 1$, while all the others depend at least linearly on the potential, since they are null if V is neglected. This means that the second term of Eq. (23) is quadratic in V, and therefore negligible at this level of approximation. On the other hand, if a zeroth-order eigenvalue is L times degenerate, i.e., if the energies $\varepsilon^0_{\mathbf{k}-\mathbf{K}_\ell}$ ($\ell = 1, 2, \ldots, L$) are equal, the coefficients $c_{\mathbf{k}-\mathbf{K}_\ell}$ are of the same order of magnitude, and contribute terms of first order in V to the sum in Eq. (23). This yields a set of L homogeneous equations with L unknowns. Its solution is completely equivalent to diagonalizing the Hamiltonian in the subspace defined by the L zeroth-order degenerate wavefunctions. The new eigenvalues are correct to first order in perturbation theory.

Applying the procedure described above to a point of two-fold degeneracy, as is the case in the example that we are analyzing, we find that $\varepsilon(\mathbf{k}) = \varepsilon_0(\mathbf{k}) \pm |V_{\mathbf{K}}|$, where we have used the fact that $V_{-\mathbf{K}} = V^*_{\mathbf{K}}$ because $V(\mathbf{r})$ is a real function. With this, we can qualitatively predict energy bands as shown on the right side of Fig. 2. Note that this simple structure, with a sequence of non-overlapping bands separated by gaps is unique to 1D. At higher dimensions, the bands may overlap or cross, as we will see in a second example below.

It is worth pointing out that, similarly to what we observed in the tight-binding approach, all we need to know about the lattice potential is reduced to a single parameter, $V_{\mathbf{K}}$, for each pair of bands.

Nearly free electrons – two-dimensional example

In this second example, our hypothetical crystal has a single-atom basis attached to each point of a square lattice in two dimensions (2D), with lattice parameter a. The reciprocal lattice is of the same type, with lattice parameter $2\pi/a$. This yields a 1st BZ as shown in Fig. 3. Its points of highest symmetry are Γ, M and X, connected by lines in which generic points are denoted by Δ, Z, and Σ. It is usual to plot the energies as functions of components of the wavevector along a closed line, which here we can choose as the triangle ΓXMΓ. Figure 4 shows some of the lowest energy bands in the empty-lattice limit, with numerical labels corresponding to the subscript n of the energies defined by

$$\begin{aligned} &\varepsilon_n(\mathbf{k}) = \hbar^2(\mathbf{k} + \mathbf{K}_n)^2/2m \; , \\ &\quad \mathbf{K}_1 = (0,0), \quad \mathbf{K}_2 = (-2\pi/a, 0), \quad \mathbf{K}_3 = (0, -2\pi/a), \quad \mathbf{K}_4 = (-2\pi/a, -2\pi/a). \end{aligned} \tag{24}$$

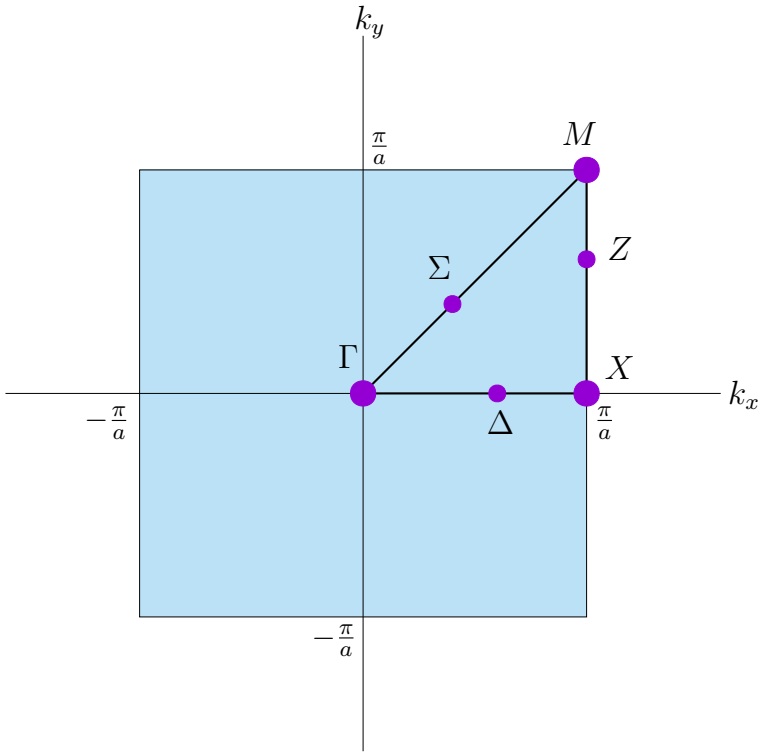

Figure 3: 1st BZ of the square lattice, showing high-symmetry points.

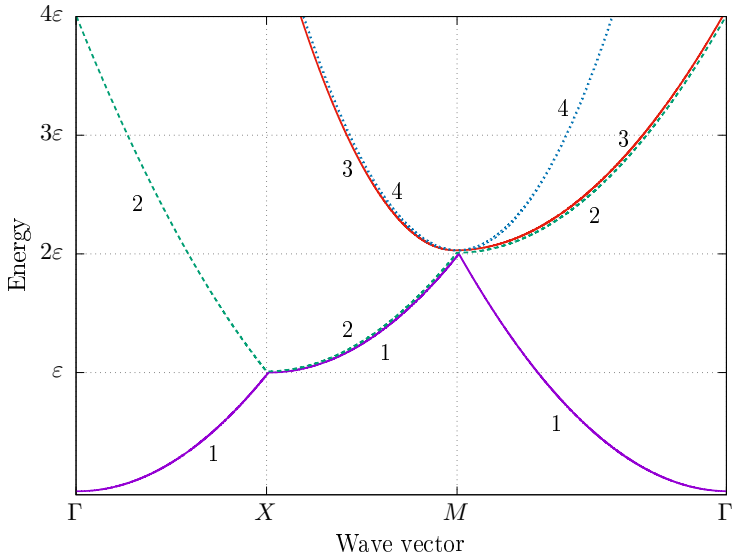

Figure 4: Lowest energy bands of the square lattice in the empty-lattice approximation. The wavevector varies along the line $\Gamma X M \Gamma$ shown in Figure 4.

It is clear that there are energy degeneracies in Fig. 4, not only at single points but also along entire lines. To find out whether they are all accidental (as was the case in 1D) or not, we have to study the **wavevector groups** of the points indicated in Fig. 3, which we will denote by G with a subscript specifying the particular point.

Groups G_Γ and G_M are identical and contain **all** symmetry operations of a square. These operations were already mentioned in connection to Fig. 1, but are listed again below, grouped in classes. We use the international notation, in which this group is called $4mm$. The symmetry operations are

- E — identity;
- 2_z — 180° rotations around the k_z axis (passing through the center of the square);
- $\{4_z, 4_z^3\}$ (class 4_z) — $\pm 90°$ rotations around the z axis;
- $\{m_x, m_y\}$ (class m_x) — reflections on the coordinate axes;
- $\{m_d, m_{d'}\}$ (class m_d) — reflections on the diagonals.

Note that point Γ remains **fixed** under any of these operations, while M is taken to another corner of the square, but **all** vertices of the 1st BZ are connected by reciprocal-lattice vectors, being therefore **equivalent**.

We have 8 elements in the group, divided into five classes. Hence, there are 5 irreducible representations, and their dimensions n_α must satisfy the sum rule

$$\sum_{\alpha=1}^{5} n_\alpha^2 = 8 \; . \tag{25}$$

Taking into account that there is always a one-dimensional identity representation (here denoted as A_1), we have

$$\sum_{\alpha=2}^{5} n_\alpha^2 = 7 \; . \tag{26}$$

For this equality to be verified, we necessarily have three one-dimensional representations (A_2, B_1, B_2) and a two-dimensional one (E). In total, the group has **4** one-dimensional and **1** two-dimensional irreducible representations. The irreducible representation E should not be confused with the identity operation E, despite the same notation.

Group G_X is composed of a subset of operations of $4mm$. Classes 4_z and m_d cease to be symmetries: they take point X to the middle point of a corner-sharing side, but these two points are not equivalent since they are **not** connected by a reciprocal-lattice vector. We then have

$$G_X = \{E, 2_z, m_x, m_y\} \; . \tag{27}$$

These operations now form independent classes, as the 90° rotations that would exchange m_x and m_y do not belong to this subgroup. Therefore, all irreducible representations of the group G_X are one-dimensional.

Groups G_Δ, G_Z and G_Σ contain only the identity and a reflection (m_y for Δ, m_x for Z, and m_d for Σ). Again, all irreducible representations are one-dimensional.

From this analysis, we conclude that

- the degeneracies along any lines as well as at point X are all **accidental**;
- there may exist two-fold degeneracies at points Γ and M.

Figure 5 shows a plot of the two lowest-energy bands, obtained from the empty-lattice bands 1 and 2 of Fig. 4 by the inclusion of **qualitative** corrections due to a weak crystal potential. Note that the bands do not cross, but they overlap in energy. Indicated in the plot is the Fermi energy evaluated for atoms of valence 2, i.e., two electrons per unit cell. It can be seen that the first band is mostly occupied, except for a hole pocket near the M point, while the second band has occupied states only around the X point. One can check how this situation is reflected in the shape of the Fermi surface (FS).

In two dimensions, the Fermi "surface" is actually a line. This line is a circle in the empty-lattice approximation, since the energies depend only on the magnitude of **k**. The FS radius, k_F, depends on the number of electrons, which in turn, because we are supposing a monoatomic lattice (a single atom per lattice point), depends on the atom's valence. Let us analyze two cases.

Valence 1: There is one electron per primitive cell, which implies that the number of electrons is **equal** to the number of Bloch states within the 1st BZ. Therefore, **half** of

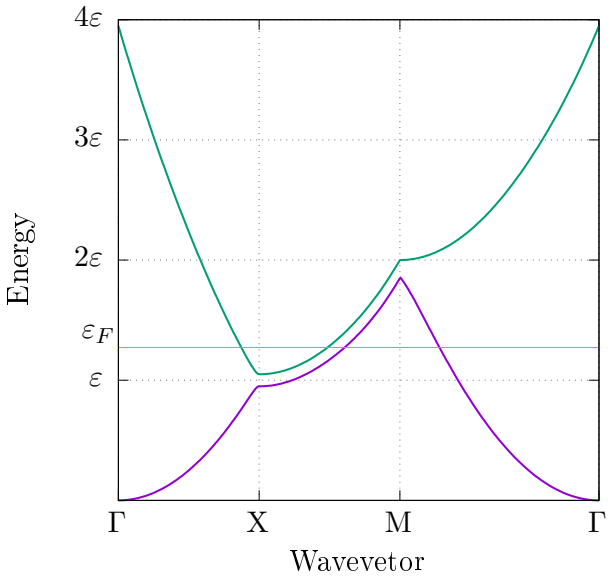

Figure 5: Lowest energy bands of a monoatomic square lattice, with qualitative nearly-free-electron corrections to the empty-lattice limit. The horizontal line indicates the Fermi energy for two electrons per unit cell, with the energy scale given by $\varepsilon \equiv \hbar^2\pi^2/(2ma^2)$.

these states are occupied (due to the spin degeneracy), and the FS area equals **half** the BZ area. Thus,

$$\pi k_F^2 = \frac{1}{2}\left(\frac{2\pi}{a}\right)^2 \quad \Rightarrow \quad k_F = \frac{\sqrt{2\pi}}{a} \, . \tag{28}$$

As $k_F < \pi/a$, the FS is totally contained within the 1st BZ, and sufficiently far from its boundaries not to show significant distortions relative to the circular shape in the nearly-free-electron approximation. The Fermi energy is

$$\varepsilon_F = \frac{\hbar^2 k_F^2}{2m} = \frac{2}{\pi}\frac{\hbar^2\pi^2}{2ma^2} = \frac{2}{\pi}\varepsilon \, , \tag{29}$$

where ε is the energy unit in Fig. 5

Valence 2: There are **two** electrons per primitive cell. Therefore, the FS area is **equal** to the 1st BZ area. In the free-electron limit, it is given by

$$\pi k_F^2 = \left(\frac{2\pi}{a}\right)^2 \quad \Rightarrow \quad k_F = \frac{2\sqrt{\pi}}{a} \, . \tag{30}$$

Now $k_F > \pi/a$, and the FS crosses the 1st-BZ boundaries. We then separate it in two branches, as sown in Fig. 6. The first branch is the area of occupied states inside the 1st BZ. The second is composed by the filled "slices" outside the 1st BZ, which are shown reduced to the inside by displacements through reciprocal lattice vectors. Now the free-electron Fermi energy is

$$\varepsilon_F = \frac{4}{\pi}\varepsilon \, . \tag{31}$$

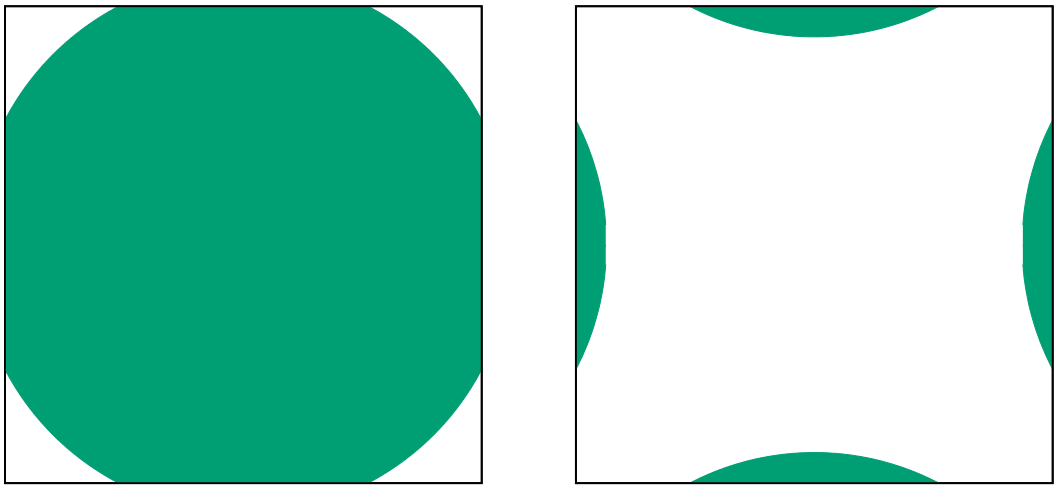

Figure 6: FS branches corresponding to the two bands shown in Fig. 5 (not including distortions due to the crystal potential). Coordinates of the 1st BZ are as in Fig. 3.

The FS pictures in Fig. 6 agree with the hole and electron pockets seen around points M and X in Fig. 5. Note also that both FS branches are **invariant under all symmetry operations of a square**.

Energy bands in semiconductors

As a more realistic example of energy bands, we show here some interesting characteristics of the relevant bands of semiconductors commonly employed in electronic devices. Figure 7 shows part of the band structure o Ge. It is obtained by heavily numerical calculation methods, taking into account, to a certain extent, the interaction between electrons, but still obtaining a set of independent-electron wave functions. This is done in the context of Density Functional Theory (DFT), which we will briefly discuss in Unit 2 of this course.

One of the interesting aspects to observe in Fig. 7 is the similarity of certain branches of the $\epsilon(\mathbf{k})$ curves with those obtained in our simple one-dimensional example of the nearly-free-electron approximation, as highlighted for the LΓ line. Another important characteristic is that the regions immediately above and below the gap show essentially parabolic forms, hole-like below the gap and electron-like above. This allows to employ a free-electron model for either electrons above the gap (conduction band) or holes below the gap (valence band), with effective masses obtained from the curvatures around the minimum and maximum.

Without going into much detail, the choice of which kind of charge carrier we will have is made through *doping*, i.e., controlling the chemical composition to have a certain amount

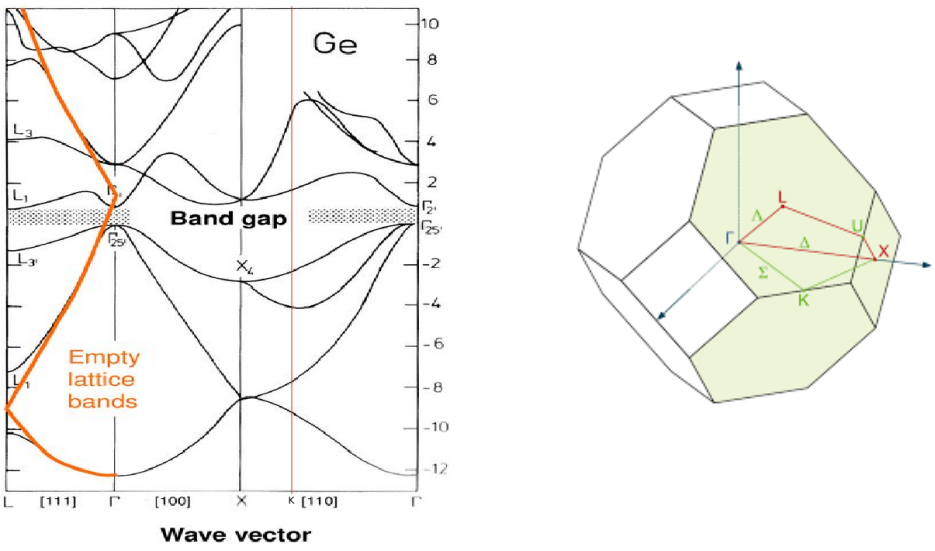

Figure 7: Calculated band structure of germanium (left) along lines of the 1st BZ (shown on the right). A comparison is made with empty-lattice bands for the line LΓ. [Figure extracted from https://slideplayer.com/slide/14335597/].

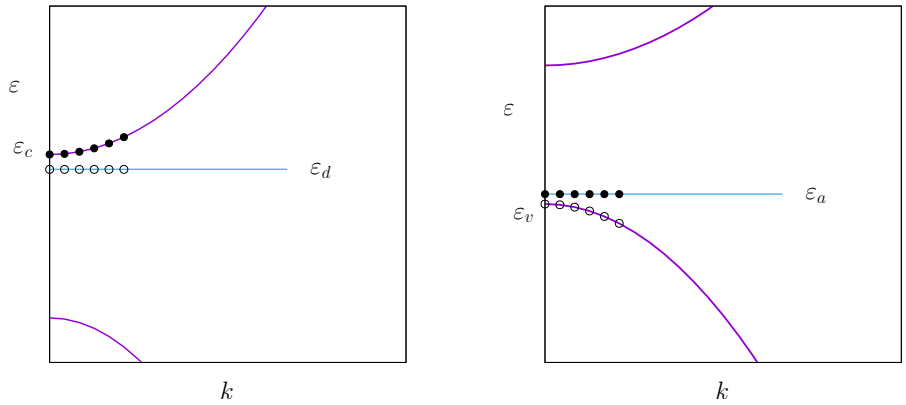

Figure 8: Schematic representation of conduction and valence bands of a semiconductor upon doping with electrons (left) or holes (right). The straight lines indicate the energies of localized donor (ε_d) and acceptor (ε_a) impurity levels.

of either *donor* or *acceptor* "impurity-atoms" in substitution to those of the pure system. As schematically depicted in Fig. 8, donnors have an extra electron with respect to the matrix atoms. But the corresponding extra proton in the nucleus generates a localized level closely below the conduction-band bottom. Acceptors have one electron less than the matrix atoms, but also one proton less, yielding an empty localized level closely above the valence-band top. At finite temperatures, electrons are thermally excited from the donor levels to the conduction band, or from the valence band to the acceptor levels. Thus, a non-zero carrier concentration will be present in one of the bands: either electrons in the conduction band or holes in the valence band. The carrier density is low in comparison to metallic conduction bands. So, the carriers will be in an energy region where the corresponding band is parabolic, thus behaving as free particles. However, the effective masses are in general significantly smaller than the electron mass.

Multiple bands in the tight-binding approximation

We have obtained multiple bands in the nearly-free electron approximation, since a sequence of bands with increasing energy appears as one folds the single free-electron dispersion relation to the 1st BZ. But tight-binding bands, as we saw at the beginning of this text, are automatically periodic in the reciprocal space. How can we generate multiple tight-binding bands?

The tight-binding approximation, in its simplest form, deals with non-degenerate Wannier states, yielding a single band. If this is not a good starting point, as in the case of degenerate atomic orbitals (e.g, p or d), one must use a more complex form of the tight-binding Hamiltonian, involving matrix elements between all degenerate Wannier states. Whereas the atomic orbitals provide a first approximation to Wannier states, it is important to take into account *crystal-field effects* on them. This begins by investigating what essential

degeneracies of atomic orbitals are predicted when the atom is part of a crystal structure.

For example, transition-metal atoms present partial filling of an atomic d level ($l = 2$). If the atom is isolated, spherical symmetry implies that the angular part of energy eigenfunctions is given by the spherical harmonics $Y_{lm}(\theta, \phi)$. For fixed l, they are basis functions of $(2l + 1)$-dimensional irreducible representations of the group of rotations in three-dimensional space (arbitrary rotation axis). Therefore, in the case of a d level we have a representation of dimension 5 (which is the degree of degeneracy of this atomic level), whose basis functions are the spherical harmonics $Y_{2m}(\theta, \phi)$ for $m = 2, 1, 0, -1, -2$.

Now suppose that the atom is in a crystal of cubic symmetry. It is straightforward to write appropriate linear combinations of the five spherical harmonics $Y_{2m}(\theta, \phi)$ that are basis functions of **two** irreducible representations of the cubic group, in general denoted e_g and t_{2g}, respectively two- and three-dimensional. The corresponding basis functions are usually written as

$$\{x^2 - y^2, \quad 3z^2 - r^2\} \quad \rightarrow \quad e_g\,, \tag{32a}$$

$$\{xy, \quad yz, \quad zx\} \quad \rightarrow \quad t_{2g}\,. \tag{32b}$$

If rewritten in spherical coordinates with fixed r, these functions are linear combinations of spherical harmonics $Y_{2m}(\theta, \phi)$.

In this example, the degeneracy observed in the isolate atom was reduced by a cubic-symmetry crystal field (which should yield different eigenvalues for e_g and t_{2g}), although not completely eliminated. Therefore, the problem of tight-binding d bands is not as simple as the case of non-degenerate orbitals, but the dimensions of the matrices involved are small.

A matrix also results in the case of a single orbital per atom but more than one atom per primitive cell. The Hamiltonian is written as in Eqs. (12) and (13), but the n index identifies atoms within the same primitive cell (same i) instead of different orbitals of a single atom. Keeping only hopping terms between nearest atoms and taking the Fourier transform, the Hamiltonian is given by a **k**-dependent matrix whose dimension is equal to the number of atoms per primitive cell, which thus gives the number of bands. We will see an example of this in the following.

Graphene – a tight-binding example with two bands

Graphene consists of a mono-layer of graphite, i.e., a two-dimensional array of carbon atoms, with sp^2 hybridization, forming the *honeycomb* structure shown in Fig. 9. The distance between two neighboring carbon atoms in graphene is $d_{CC} = 1.42$ Å. It is necessary to associate **two** carbon atoms (A and B in the figure) to each point of a 2D hexagonal lattice (also called *triangular*), generating two interpenetrating *sublattices* (shown with different colors).

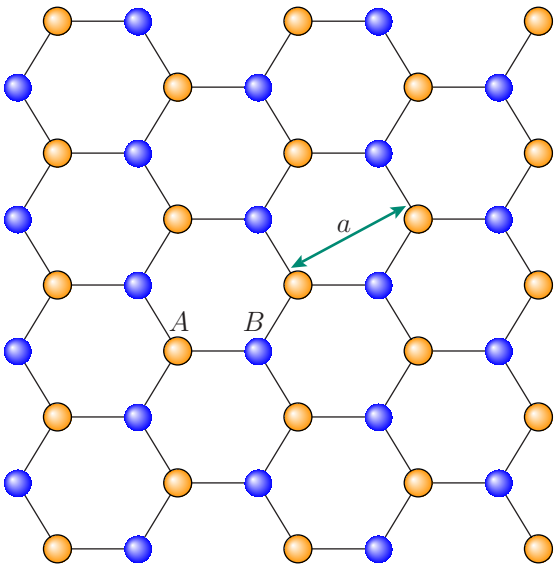

Figure 9: Partial representation of the honeycomb structure, revealing the existence of two hexagonal (or triangular) sublattices, A and B, both with lattice parameter a.

Carbon sp^2 orbitals form σ bonds that define the graphene structure. The p_z orbitals form out-of-plane π bonds which generate the relevant bands. Using the tight-binding approximation for π electrons, choosing the reference energy $\varepsilon_{p_z} = 0$, and taking into account hopping (t) only between nearest neighbors, which here means between atoms of different sublattices, we can write the Hamiltonian as

$$\mathcal{H} = -t{\sum_{ij}}' \left(|Ai\rangle\langle Bj| + |Bj\rangle\langle Ai| \right) , \tag{33}$$

where the primed sum indicates that it is restricted to sites directly connected. Note that $i = j$ is not discarded because there is hopping between atoms A and B belonging to the same primitive cell. It is also worth noticing that we are dealing with "pure" Wannier states, as we did with Bloch functions in Text 3, it being implicit that complete electronic states must include a spin part.

As each sub-lattice is a Bravais lattice, we can associate Bloch states to each, defining

$$|\alpha\mathbf{k}\rangle = \frac{1}{\sqrt{N}} \sum_i \mathrm{e}^{\mathrm{i}\mathbf{k}.\mathbf{R}_i^\alpha} |\alpha i\rangle \, , \tag{34}$$

where $\alpha = A, B$, and we are using a superscript on lattice vectors to label the sublattice.

Taking the inverse relation,

$$|\alpha i\rangle = \frac{1}{\sqrt{N}} \sum_i \mathrm{e}^{-\mathrm{i}\mathbf{k}.\mathbf{R}_i^\alpha} |\alpha\mathbf{k}\rangle \, , \tag{35}$$

and substituting into the Hamiltonian, we obtain

$$\mathcal{H} = -\frac{t}{N}\sum_{ij}{}' \sum_{\mathbf{k}\mathbf{k}'} \left[|A\mathbf{k}\rangle\langle B\mathbf{k}'| \, \mathrm{e}^{\mathrm{i}\,(\mathbf{k}.\mathbf{R}_i^A - \mathbf{k}'.\mathbf{R}_j^B)} + \mathrm{H.c.} \right] , \tag{36}$$

where H.c. indicates the Hermitian conjugate of the previous term. The exponential appearing on the right-hand side can be rewritten as

$$\mathrm{e}^{\mathrm{i}(\mathbf{k}.\mathbf{R}_i^A - \mathbf{k}'.\mathbf{R}_j^B)} = \mathrm{e}^{\mathrm{i}\,\mathbf{k}.(\mathbf{R}_i^A - \mathbf{R}_j^B)} \, \mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}').\mathbf{R}_j^B} . \tag{37}$$

The result of summing over i, which involves the first factor, is independent of $\mathbf{R}_j^B$ due to the lattice symmetry. Therefore, the second factor is the only one involved in the sum over j, which is unrestricted, and results in $\delta_{\mathbf{k}\mathbf{k}'}$ (orthogonality between characters of irreducible representations of the translations group). The restricted sum over i (for each j) can be transformed into a sum over vectors

$$\boldsymbol{\delta} \equiv \mathbf{R}_i^A - \mathbf{R}_j^B \tag{38}$$

satisfying the condition of connecting neighbors.

Thus, defining

$$t_{\mathbf{k}} \equiv t \sum_{\boldsymbol{\delta}} \mathrm{e}^{\mathrm{i}\,\mathbf{k}.\boldsymbol{\delta}} , \tag{39}$$

the Hamiltonian assumes the form

$$\mathcal{H} = -\sum_{\mathbf{k}} \left(t_{\mathbf{k}} \, |A\mathbf{k}\rangle\langle B\mathbf{k}| + t_{\mathbf{k}}^* \, |B\mathbf{k}\rangle\langle A\mathbf{k}| \right) . \tag{40}$$

This Hamiltonian can be rearranged as a sum of matrix products,

$$\mathcal{H} = -\sum_{\mathbf{k}} \begin{pmatrix} |A\mathbf{k}\rangle & |B\mathbf{k}\rangle \end{pmatrix} \begin{pmatrix} 0 & t_{\mathbf{k}} \\ t_{\mathbf{k}}^* & 0 \end{pmatrix} \begin{pmatrix} \langle A\mathbf{k}| \\ \langle B\mathbf{k}| \end{pmatrix} . \tag{41}$$

Diagonalizing the square matrix, we obtain the energy eigenvalues $\varepsilon_{\pm}(\mathbf{k}) = \pm|t_{\mathbf{k}}|$. With an appropriate choice of the vectors $\boldsymbol{\delta}$ in Eq. (39), these energies can be written in the form

$$\varepsilon_{\pm}(\mathbf{k}) = \pm t \left[1 + 4\cos(\sqrt{3}k_x a/2)\cos(k_y a/2) + 4\cos^2(k_y a/2) \right]^{\frac{1}{2}} , \tag{42}$$

where a is the lattice parameter, i.e., the shortest distance between atoms of the same sublattice (see Fig. 9).

A plot of the two energy bands of graphene in two-dimensional $\mathbf{k}$-space is shown in Fig. 10. A closer look shows that the **energy varies linearly** with the module of $\mathbf{k}$ near the BZ vertices, where the two bands touch.

EXERCISES: (1) Check that the band degeneracy at the contact points is allowed by symmetry. (2) Where is the Fermi level? (3) Can we evaluate the effective mass of an electron near the Fermi level?

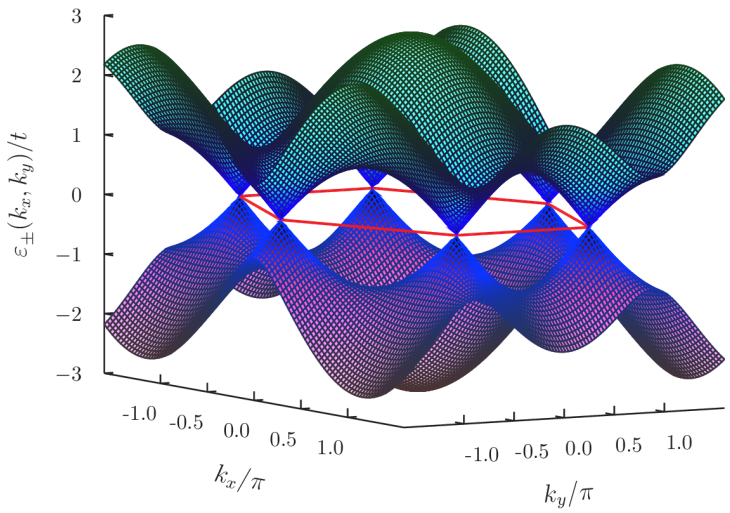

Figure 10: Graphene's energy bands. The hexagonal line delimits the 1st BZ.

Another view of one-electron energies in graphene is presented in Fig. 11, in the form of a gray-scale map of the top band projected on the $k_x k_y$ plane. It also contains contour lines, i.e., constant-energy lines, which clearly reflect the symmetries. Note that the symmetry is hexagonal with respect to the BZ center, but triangular in relation to the vertices.

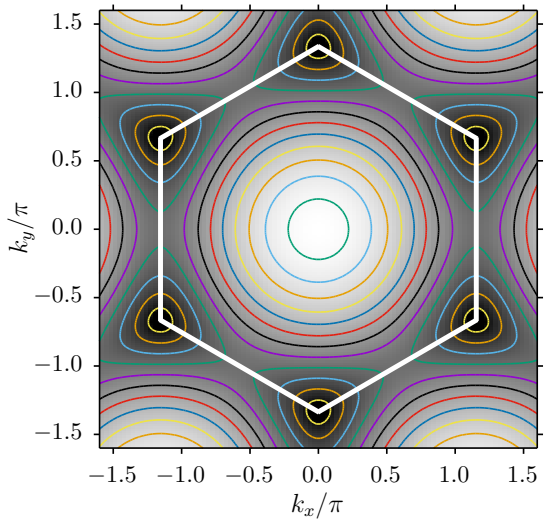

Figure 11: Gray-scale map of graphene's upper energy band, with lighter regions corresponding to higher energy. Color lines are constant-energy contours. White lines indicate the 1st BZ.