The tight-binding model

The electrons in carbon based materials (like graphene) or in semiconductors, are not free to move like in metals. Instead, they spend most of their lives sitting on atomic sites. But every once in a while, due to a small overlap in atomic orbitals, the electrons may tunnel from one site to another.

(this drawing is terrible!)

This is how an electron in your phone moves: it *hops* from one site to another. *Tight-binding* is the generic name we give to models of electrons hopping through a lattice. These models are widely used to understand the electronic properties of organic molecules, graphene and semiconductors.

Most scientific papers use 2nd quantization to describe the tight-binding model. This, as we will learn, has some nice advantages. But it is not a necessity. Here we will describe the model using the usual rules of QM. Later we will revisit it in 2nd quantized notation. I think that comparing the two approaches will help make the physics clearer.

The Eq numbers in the next page start on (8).
Sorry about that

Tight-binding in one dimension

Imagine that you have a 1D lattice with N sites and lattice spacing a

We label each site by $m = 1, \dots, N$ and we choose a reference frame such that the position of each site is

$$x_m = a m \tag{8}$$

Now imagine you have an electron hopping from one site to another. It spends most of its time sitting in some lattice site but, every once in a while, it has a small probability amplitude for jumping to a neighboring site. We assume only nearest neighbor hopping since second nearest neighbor should be much less likely.

This problem can be described quantum mechanically as follows. Let $|m\rangle$ denote the state representing "the electron at site m." then the only non-zero matrix elements of the Hamiltonian will be

$$\langle m|H|m+1\rangle = \langle m+1|H|m\rangle = -g \tag{9}$$

where g is the hopping amplitude and the minus sign is placed only for aesthetic reasons.

we may now introduce 2 completeness relations and write

$$H = \sum_{n,m} |n\rangle\langle n| H |m\rangle\langle m|$$

$$= \sum_{n} \left\{ |n\rangle\langle n|H|n+1\rangle\langle n+1| + |n+1\rangle\langle n+1|H|n\rangle\langle n| \right\}$$

thus H may be written as

$$H = -g \sum_{n} \left[|n+1\rangle\langle n| + |n\rangle\langle n+1| \right] \qquad (10)$$

If you want, you could also write H as a big matrix. For instance, if N = 5 we would get

$$H = \begin{pmatrix} 0 & -g & 0 & 0 & -g \\ -g & 0 & -g & 0 & 0 \\ 0 & -g & 0 & -g & 0 \\ 0 & 0 & -g & 0 & -g \\ -g & 0 & 0 & -g & 0 \end{pmatrix} \qquad (11)$$

Note the two lonely terms in the corners. they represent the periodic boundary conditions, meaning an electron at 1 can hop either to 2 or to N. this term is usually left implicit in Eq (10) by assuming that $|N+1\rangle = |N\rangle$.

writting H as a matrix, however, is not recommended. We are interested in arbitrarily large N and, very soon, we will generalize this to an arbitrary lattice in arbitrary dimension. the outer product notation (10), albeit more abstract, is much more robust.

Before we diagonalize Eq (10), it is interesting to note that we can readily write down more general Hamiltonians describing different hopping dynamics. For instance, we could write a Hamiltonian like

$$H = -\sum_{n,m} g(n,m)\,|n\rangle\langle m| \qquad (12)$$

for some function $g(n,m)$, which describe the tunneling amplitude to go from n to m.

We could use this, for instance, to model tight-binding in some weird looking graph

Label each node by n and put a non-zero $g(n,m)$ when hopping from $n \to m$ is allowed. this idea is extensively used to model organic molecules.

In principle the $g(n,m)$ only need to satisfy one constraint:

$$g(n,m) = g(m,n)^* \qquad (13)$$

This is the requirement for having $H^\dagger = H$. But if our graph is actually a lattice, it will have the additional symmetry of translation invariance. this means that $g(n,m)$ can only depend on the distance $x_n - x_m$ between two sites

Thus,

$$\boxed{\text{Systems with trans. inv.} \;:\quad g(n,m) = g(x_n - x_m)} \tag{14}$$

From now on we will assume this is true. Then the Hamiltonian may be written as

$$\boxed{H = -\sum_{n,m} g(x_n - x_m)\,|n\rangle\langle m|} \tag{15}$$

Eq (10) is then a particular case where

$$g(+a) = g(-a) = g \tag{16}$$

Whenever you have translation invariance, the first thing to do is to move to Fourier space. Define a new set of kets $|k\rangle$ as

$$\boxed{|n\rangle = \frac{1}{\sqrt{N}} \sum_k e^{ikx_n} |k\rangle} \tag{17}$$

PBCs imply k is quantized as usual and the fact that $x_n = na$ are discrete will restrict k to the first BZ

$$\boxed{k = \frac{2\pi \ell}{Na}, \quad \ell = -\frac{N}{2}, \dots, \frac{N}{2} \quad \Rightarrow \quad k \in \left[-\frac{\pi}{a}, \frac{\pi}{a}\right]} \tag{18}$$

We now plug (17) into (15):

$$H = -\sum_{n,m} g(x_n - x_m)\left[\frac{1}{\sqrt{N}}\sum_k e^{ikx_n}|k\rangle\right]\left[\frac{1}{\sqrt{N}}\sum_q e^{-iqx_m}\langle q|\right]$$

$$= -\sum_{k,q} |k\rangle\langle q| \left\{ \frac{1}{N}\sum_{n,m} g(x_n - x_m)\, e^{i(kx_n - qx_m)} \right\}$$

We first work out the sums over n and m. Keep the sum over m, but change variables in the other to $x_l = x_n - x_m$. We then get

$$\frac{1}{N}\sum_{n,m} g(x_n - x_m)\, e^{i(kx_n - qx_m)} = \frac{1}{N}\sum_{l,m} g(x_l)\, e^{ik(x_l + x_m)}\, e^{-iqx_m}$$

$$= \left[\sum_l g(x_l)\, e^{ikx_l}\right]\underbrace{\left[\frac{1}{N}\sum_m e^{i(k-q)x_m}\right]}_{\delta_{kq}}$$

If we define

$$\boxed{E_k = -\sum_n g(x_n)\, e^{ikx_n}} \qquad (19)$$

we then finally get

$$\boxed{H = \sum_k E_k |k\rangle\langle k|} \qquad (20)$$

which is diagonal. Thus, by moving to Fourier space we have diagonalized the Hamiltonian (15). Its eigenvectors are $|k\rangle$ and the energy eigenvalues are E_k. It is quite remarkable that we have diagonalized a matrix of arbitrary size N. This is the magic of translation invariance.

→ "nearest neighbor"

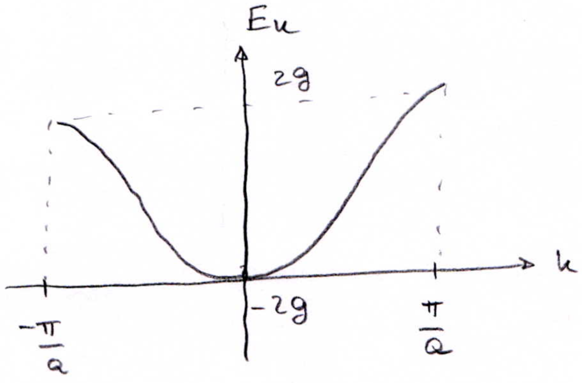
Let us understand the energies (19). For n. n. hopping the non-zero amplitudes are given by (16) so

$$E_k = -g(a)\,e^{ika} - g(-a)\,e^{-ika}$$

or

$$\boxed{E_k = -2g\cos k} \qquad (21)$$

This is the dispersion relation for an electron in a 1D lattice. It looks like



We only plot for k in the 1st BZ.

In the long wavelength limit, $ka \ll 1$, we may expand the cosine and write

$$E_k \simeq -2g + g a^2 k^2$$

The first term is just a constant, which is not important. We thus see that in the long wavelength limit

$$E_k \approx g a^2 k^2 \qquad (22)$$

This is a non-relativistic dispersion relation, $E_k = \frac{k^2}{2m}$, but with an effective mass

$$\frac{1}{2m_{eff}} = g a^2 \Rightarrow m_{eff} = \frac{1}{2 g a^2} \qquad (23)$$

An electron moving with low energies therefore behaves approximately like a free particle, but with an effective mass instead of the actual electron mass. The effective mass can be measured experimentally and it may differ substantially from the electron mass (it can be both larger or smaller).

The dispersion relation E_k corresponds to one energy band. We only got one band because we assumed each site only contains a single ket. We will learn how to generalize this in a second.

Filling an energy band

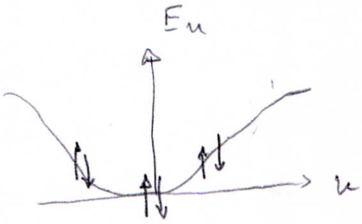
The tight-binding method neglects any electron-electron interaction and focuses on a single electron. To deal with the many-electron problem, we could at first continue to neglect their interactions. then all they will do is fill each energy level while respecting the Pauli principle.

To fill out the energy bands you need to remember 2 facts :

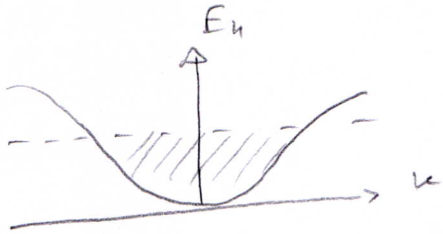
1) Each band has $2N$ available sites, where N is the number of lattice points and the factor of 2 is due to spin (each state can take one spin up and one down)

2) We need to know the valency Z of the atoms; ie, the number of valence electrons per atom.

For instance, if $Z=1$ then we have $2N$ sites available for N electrons. We will then get

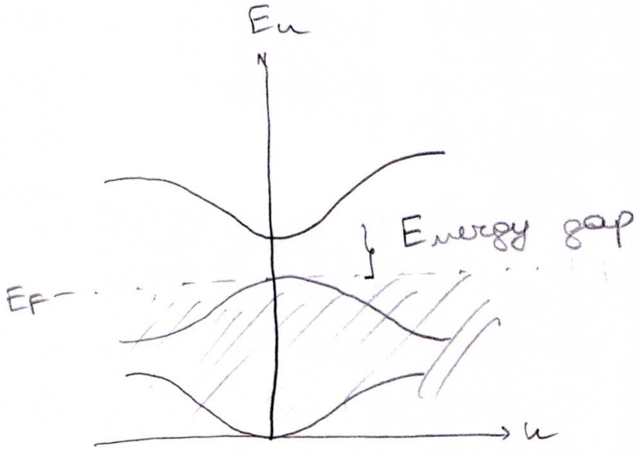


$\Rightarrow$

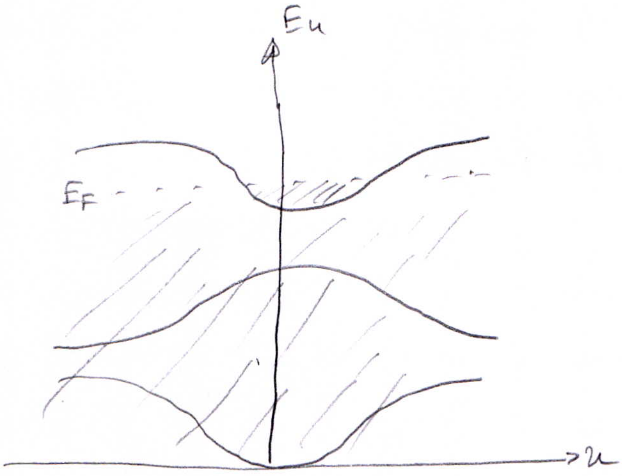


If $Z=2$ then we completely fill the band. If $Z=3$ then it will start filling some other band (we will discuss the multi-band case next).

For instance, we could have the following situations



or



Filled band

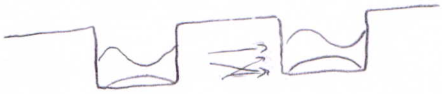
Partially filled band

The first Fig would occur for $Z = 4$ and the 2nd for $Z = 5$. The physics of each situation is completely different. Suppose you want to give some momentum to an electron. If the bands are partially filled (right image), that is very easy because there are a bunch of momentum states available. But for a filled band, any push we want to give requires us to supply a minimum energy, E_g, called the energy gap. Consequently, in a filled band it is very hard to get the electrons moving.

A filled band will therefore behave like an insulator, whereas a partially filled band will behave like a metal. A semiconductor is an insulator, but with a small band gap ($\lesssim 2$ eV). Then thermal aggitation facilitates the excitation of some of the electrons.

Tight-binding in an arbitrary lattice

It is very easy to generalize the previous example to an arbitrary lattice. We may also include the presence of a basis. In fact, what we can do is add additional degrees of freedom to each Bravais lattice site. This can be a basis of two or more atoms, or it can be the presence of additional orbitals (or spin!). For instance, we may consider a lattice where in each atom an electron may occupy one of 2 orbitals. We may then have transitions from an orbital in one site to an orbital in another site and all sorts of complicated combinations.



To generalize that we now denote our kets by $|n, a\rangle$, where a is an internal index. The general tight-binding Hamiltonian may then be written as

$$\boxed{H = -\sum_{\substack{n,m \\ a,b}} g_{ab}(\mathbf{R}_n - \mathbf{R}_m)\, |n, a\rangle\langle m, b|} \qquad (24)$$

All information about the hoppings is contained in g, and translation invariance is manifested in the fact that g depends only on $\mathbf{R}_n - \mathbf{R}_m$.

We assume there are N Bravais sites and d internal degrees. Then the matrix H will have size Nd.

Now we introduce the Fourier transform

$$\boxed{|n,a\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}_n} |\mathbf{k},a\rangle} \qquad (25)$$

Inserting this in (24) then gives

$$H = -\sum_{\substack{n,m\\a,b}} g_{ab}(\mathbf{R}_n - \mathbf{R}_m)\, \frac{1}{N} \sum_{\mathbf{k},\mathbf{q}} e^{i\mathbf{k}\cdot\mathbf{R}_n}\, e^{-i\mathbf{q}\cdot\mathbf{R}_m}\, |\mathbf{k},a\rangle\langle\mathbf{q},b|$$

$$= -\sum_{\substack{\mathbf{k},\mathbf{q}\\a,b}} |\mathbf{k},a\rangle\langle\mathbf{q},b|\, \frac{1}{N} \sum_{n,m} g_{ab}(\mathbf{R}_n - \mathbf{R}_m)\, e^{i(\mathbf{k}\cdot\mathbf{R}_n - \mathbf{q}\cdot\mathbf{R}_m)}$$

Next we change variables to $\mathbf{R}_l = \mathbf{R}_n - \mathbf{R}_m$. This will factor the sums in n, m as

$$\frac{1}{N} \sum_{n,m} g_{ab}(\mathbf{R}_n - \mathbf{R}_m)\, e^{i(\mathbf{k}\cdot\mathbf{R}_n - \mathbf{q}\cdot\mathbf{R}_m)} =$$

$$= \underbrace{\frac{1}{N} \sum_{m} e^{i(\mathbf{k}-\mathbf{q})\cdot\mathbf{R}_m}}_{\delta_{\mathbf{k},\mathbf{q}}} \sum_{l} g_{ab}(\mathbf{R}_l)\, e^{i\mathbf{k}\cdot\mathbf{R}_l}$$

As before, we define

$$\boxed{M_{ab}(\mathbf{k}) = -\sum_{n} g_{ab}(\mathbf{R}_n)\, e^{i\mathbf{k}\cdot\mathbf{R}_n}} \qquad (26)$$

The Hamiltonian then becomes

$$\boxed{H = \sum_{\mathbb{k}\, a,b} M_{ab}(\mathbb{k})\, |\mathbb{k},a\rangle\langle \mathbb{k},b|} \tag{27}$$

This Hamiltonian is not yet diagonal. To diagonalize it, we must diagonalize the matrix M_{ab}. But note how this diagonalization does not mix different values of $\mathbb{k}$. The $\mathbb{k}$ part is already taken care off. Now we only need to look for the internal degrees of freedom. If we have d degrees of freedom, then M_{ab} will have dimension $d \times d$. That is much easier to work with than the original Hamiltonian, which had dimension Nd, with $N \sim 10^{23}$.

The eigenstuff Eq for M may be written as

$$\boxed{M(\mathbb{k})\, \psi_\lambda(\mathbb{k}) = E_\lambda(\mathbb{k})\, \psi_\lambda(\mathbb{k})} \tag{28}$$

where ψ_λ is a d-dimensional vector with components $\psi_{\lambda a}(\mathbb{k})$. If we define

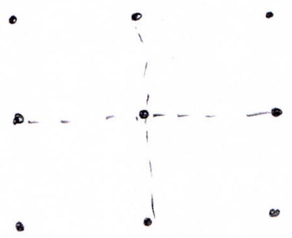
$$|\mathbb{k},a\rangle = \sum_\lambda \psi_{\lambda a}(\mathbb{k})\, |\mathbb{k},\lambda\rangle \tag{29}$$

then the Hamiltonian (27) is finally put in diagonal form, as

$$\boxed{H = \sum_{\mathbb{k},\lambda} E_\lambda(\mathbb{k})\, |\mathbb{k},\lambda\rangle\langle \mathbb{k},\lambda|} \tag{30}$$

Example: square lattice

Consider a square lattice with nearest-neighbor hopping



$a_1 = a(1,0)$

$a_2 = a(0,1)$

the only non-zero hopping amplitudes will be

$$g(\pm a_1) = g(\pm a_2) = g$$

so the dispersion relation will be

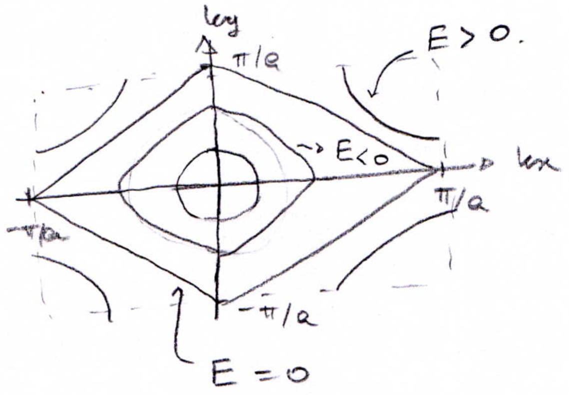
$$E_k = -\sum_m g(R_m)\, e^{ik\cdot R_m}$$

$$= -g\left[e^{ik\cdot a_1} + e^{-ik\cdot a_1} + e^{ik\cdot a_2} + e^{-ik\cdot a_2}\right]$$

or

$$\boxed{E_k = -2g\left[\cos(k_x a) + \cos(k_y a)\right]}$$

the constant energy surfaces look like this



At low k the const. energy surfaces are roughly circular (the 2D version of the Fermi sphere). But as k increases the circles bend and, when $E = 0$, they become a straight line

—//—

We can also add second nearest neighbors

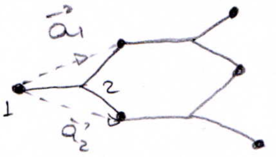
$$g(\pm(Q_1 + Q_2)) = g(\pm(Q_1 - Q_2)) = g'$$

Physically, we expect $g' < g$. This adds to the energy a contribution

$$\boxed{\begin{aligned} E_k = & -2g\left[\cos(k_x a) + \cos(k_y a)\right] \\ & -2g'\left[\cos((k_x + k_y)a) + \cos((k_x - k_y)a)\right] \end{aligned}}$$

Example : graphene

The most important example, in view of current research, is graphene:

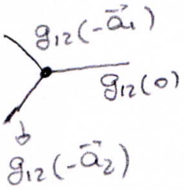


The Bravais lattice is the hexagonal/triangular with two atoms in the basis. The primitive vectors are

$$\vec{a}_{1,2} = \frac{a}{2}(\sqrt{3}, \pm 1) \qquad (31)$$

Since we have two atoms in the basis, the matrix gab will be 2×2. Moreover, an atom of type 1 only has 3 connections and all with atoms of type 2, so $g_{11} = g_{22} = 0$. The only connections are

$$g_{12}(0) = g_{12}(-\vec{a}_1) = g_{12}(-a_2) = g \qquad (32)$$



The entry M_{12} in Eq (26) then becomes

$$M_{12}(k) = -g(1 + e^{-ik\cdot a_1} + e^{-ik\cdot a_2}) \qquad (33)$$

Since H must be Hermitian, $M_{21} = M_{12}^*$. Let

$$f = 1 + e^{-ik\cdot a_1} + e^{-ik\cdot a_2} \qquad (34)$$

then the matrix M will be

$$M = -g\begin{pmatrix} 0 & f \\ f^* & 0 \end{pmatrix} \qquad (35)$$

The eigenvalues of M are

$$E_k = \pm g \, | f(k) |$$

To get an actual formula, we may massage $f(\mathbb{k})$ a bit. For instance, substitute

$$\mathbb{k} \cdot a_{1,2} = \frac{a}{2} \left(\sqrt{3}\, k_x \pm k_y \right)$$

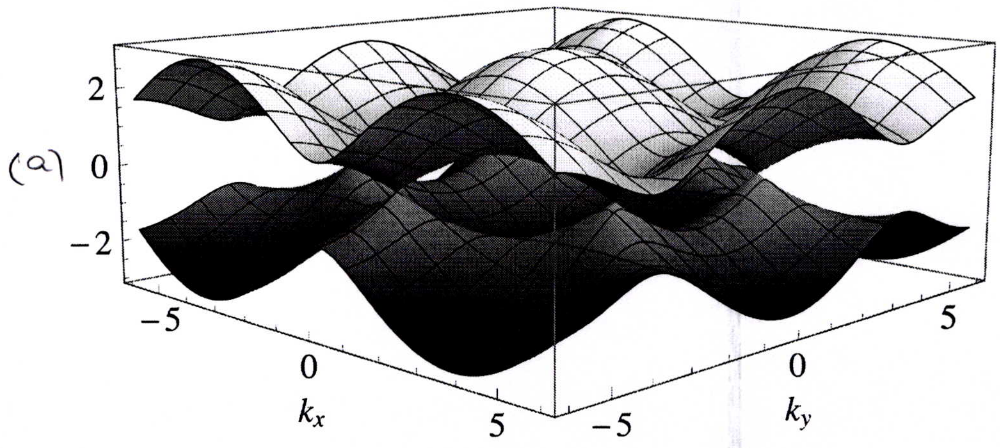
After boring simplifications we then get

$$\boxed{E_k = \pm g \sqrt{3 + 2\cos(k_x a) + 4 \cos\left(\frac{k_x a}{2}\right) \cos\left(\frac{\sqrt{3}\, k_y a}{2}\right)}} \qquad (36)$$

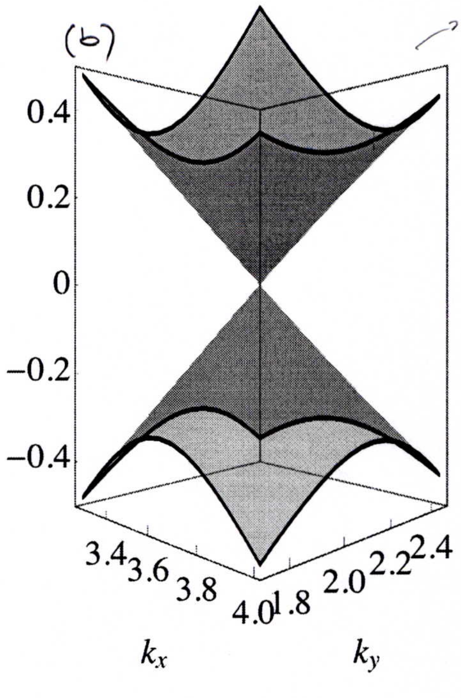
Since we have two internal degrees of freedom, we now get two energy bands. These bands are plotted as pretty figures in the next page.

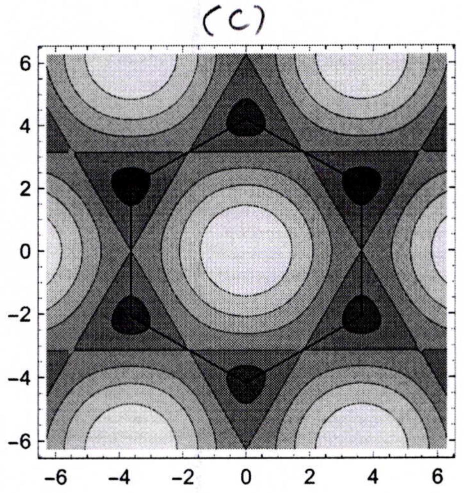
Each band can carry 2N electrons, where N is the number of Bravais points (giving a total of 2N atoms). Since the valency of carbon is $Z = 1$, there will be 2N electrons. Thus, the electrons will completely fill the lower band and leave the upper band empty. In other words, the Fermi level is

$$E_F = 0$$

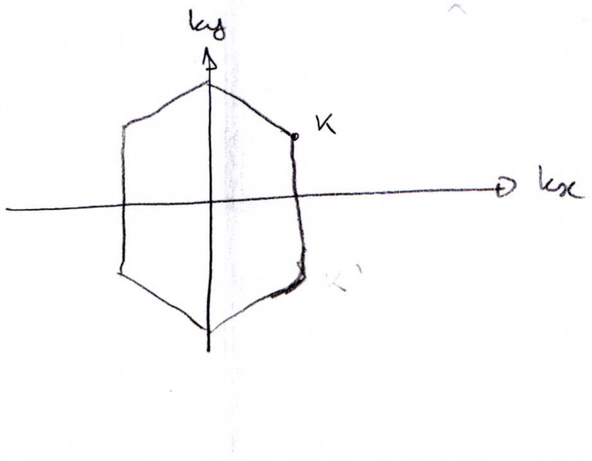


→ Dirac cones





$$\vec{K} = \frac{2\pi}{a}\left(\frac{1}{\sqrt{3}}, \frac{1}{3}\right)$$



The first BZ is drawn in page 16 and corresponds to a hexagon. The boundaries of the first BZ are special and, for reasons that will become clear in a second, they are called <u>Dirac points</u>. They correspond to the highest filled points, where the lower and upper bands touch.

Let us analyze one point, called the K-point,

$$\vec{K} = \frac{2\pi}{a}\left(\frac{1}{\sqrt{3}}, \frac{1}{3}\right) \tag{37}$$

Assume that we are in the vicinity of this point, by setting

$$\mathbb{k} = \mathbb{K} + \mathbb{q}, \qquad \mathbb{q} = \text{small}. \tag{38}$$

If we expand (36) around $\mathbb{K}$ we get

$$\boxed{E_{\mathbb{k}} \simeq \pm v|\mathbb{q}| \qquad v = \frac{\sqrt{3}\, t a}{2}} \tag{39}$$

This result is the reason why graphene is so important. Practically all of the remarkable properties of graphene can be traced back to this formula. It represents the dispersion relation of electrons close to the Fermi level. And we see that this dispersion relation is <u>linear</u>. It is just like the dispersion relation for photons, $\omega = c|\mathbb{k}|$. Electrons in graphene therefore move approximately like massless particles. You can see this in Fig (b) of page 16. Close to the Fermi levels we have two <u>Dirac cones</u> touching each other (we use the name Dirac because Dirac's Eq for a massless Fermion predicts a dispersion like (39)). We also say graphene is a <u>gapless</u> <u>semiconductor</u>, because the two bands just touch perfectly on one another.