Graphene: an introduction

Geert Brocks

December 18, 2015

Computational Materials Science, Faculty of Science and Technology and MESA+
Institute for Nanotechnology, University of Twente, P. O. Box 217, 7500 AE Enschede, the
Netherlands.

Abstract

These notes form part of a master course on Nanophysics at the University of
Twente. Material based on part of these notes has also been presented at the Han-
sur-Lesse 2015 winter school "Theory and Spectroscopy". The audience often consists
of a mixture of physicists and chemists, so what sounds complicated to one, is com-
pletely trivial to another. Hopefully there is something of interest in these notes for
everybody.

18 December 2015.

Basic electronic structure of solids

Electronic structure calculations on crystalline solids are not so different from those performed on molecules. The main difference lies in the consistent use of translation symmetry. From the perspective of atomistic scale modeling, crystals are inifinitely large, but their internal structure consists of regularly spaced copies of a finite building block: the unit cell. This translation symmetry allows for mapping the infinite size crystal problem onto a finite size problem for the unit cell. It does not mean that a calculation for a crystal unit cell is identical to a calculation on a (finite) molecule; there are some differences.

Unfortunately these differences very often seem larger than they should be because the fields of molecular physics and solid states physics have a partially different history. Therefore, the two fields have developed a different folklore and sometimes use different labels for the same thing. In this section I will illustrate the similarities and the differences by a standard chemical bonding textbook model: the Hückel model for the π-electrons in conjugated molecules. Here is our first trivial difference: in solid state physics a Hückel model is called a *tight-binding* model.

Intermezzo: C atoms in conjugated molecules (chemists can skip this subsection)

An isolated carbon atom has the configuration $1s^2 2s^2 2p^2$. The $1s$ functions have a very low energy, (< -250 eV with respect to the vacuum level), are confined close to the C nucleus, and do not play a role in binding the C atom.[1] The $2s$ and $2p$ states are called "valence states". They have a higher energy and are perturbed considerably if the C atom is bonded to other atoms. If all C atoms are in one plane, one can form three equivalent, but independent, linear combinations of the $2s$ and the $2p_x$ and $2p_y$ orbitals on each C atom, which are called sp^2 hybrid orbitals. They make an angle of 120º with one another. Making σ-bonds between the C atoms then leads to a planar molecule like benzene or the hexagonal graphene lattice. On each C atom there is still a p_z orbital, which is oriented perpendicular to the graphene plane, and allows for making π-bonds. The energies of π-states are usually higher than those of σ-states, so the π-states are usually the ones that are involved in the low energy electronic excitations. Since σ- and π-states have different symmetries, one can treat them separately. More formally, the Hamiltonian matrix is block diagonal (with σ-states in one block and π-states in another block).
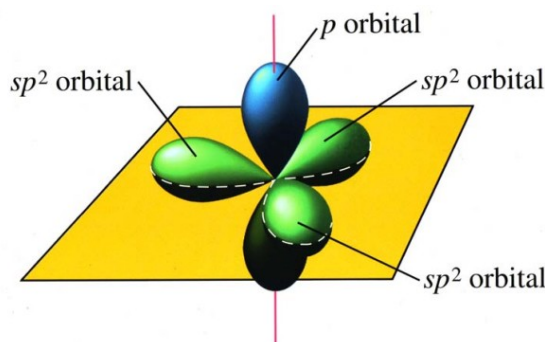

Figure 9.10
An *sp* hybridized C atom
Steven S. Zumdahl, *Chemistry*, Third Edition, © 1993 by D. C. Heath and Company

---

[1] The $1s$ state is very atomic like and does not depend too much on the state (molecule or crystal) the C atom is in. It is called a "core state".

The benzene molecule

The six carbon atoms of a benzene ring form a perfect planar hexagon, see the figure below. Each carbon atom has a p_z-orbital perpendicular to the plane, figure (a), which give rise to the π-states. The π-state with the lowest energy is sketched in figure (b). The remaining valence orbitals of carbon form sp^2-hybrids, which give the σ-bonds between the carbon atoms and between the carbon and hydrogen atoms, shown in figure (c). The (bonding) σ-states are deep down in energy, and the (anti-bonding) σ^*-states are high up in energy. We are not concerned with those.

(a)   (b)   (c)
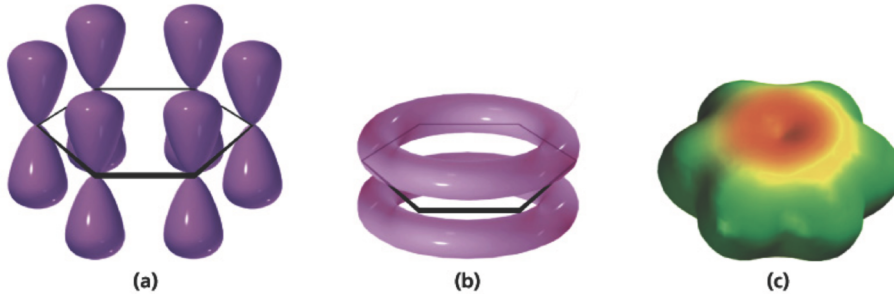
The Hückel model is a simple pen-and-paper model for the π-states of the molecule.[2] I am not interested in accurary right now, so I will use this model in its simplest form in order to minimize my computational effort.

I label the carbon atoms counter-clockwise from 0 to 5, as in the figure below, and call the corresponding p_z-orbitals $\phi_n(\mathbf{r})$; $n = 0, 1, 2, 3, 4, 5$. The (six) π-states are expressed as linear combination of these atomic orbitals: $\Psi_m(\mathbf{r}) = \sum_{n=0}^{5} c_{mn} \phi_n(\mathbf{r})$.
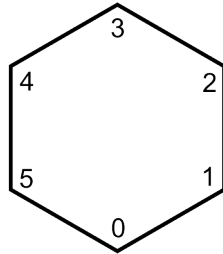
They should of course follow from the Schrödinger equation, which expressed in the atomic orbital basis becomes a matrix-vector eigenvalue equation

$$
\begin{bmatrix}
\epsilon & t & 0 & 0 & 0 & t \\
t & \epsilon & t & 0 & 0 & 0 \\
0 & t & \epsilon & t & 0 & 0 \\
0 & 0 & t & \epsilon & t & 0 \\
0 & 0 & 0 & t & \epsilon & t \\
t & 0 & 0 & 0 & t & \epsilon
\end{bmatrix}
\begin{bmatrix}
c_{m0} \\
c_{m1} \\
c_{m2} \\
c_{m3} \\
c_{m4} \\
c_{m5}
\end{bmatrix}
= E_m
\begin{bmatrix}
c_{m0} \\
c_{m1} \\
c_{m2} \\
c_{m3} \\
c_{m4} \\
c_{m5}
\end{bmatrix}.
$$

The matrix elements follow from the so-called *nearest neighbor* approximation: $\left\langle \phi_n \left| \hat{H} \right| \phi_n \right\rangle = \epsilon$; $\left\langle \phi_n \left| \hat{H} \right| \phi_{n\pm1} \right\rangle = t$ and $\left\langle \phi_n \left| \hat{H} \right| \phi_p \right\rangle = 0$; $p \neq n, n \pm 1$, where we assume that the indices

---

[2]"Paper" is an ultrathin "write once, read many" display invented in ancient times; "pen" is a wireless contact device with which you put information in the memory of the display; "pencil-and-paper" is an eraseable version of this technology. Many of such displays can be bundeled in storage devices called "books".

are cyclic: $n + 6 = n$. Note that both ϵ and t are in general negative numbers. I have also assumed that the atomic orbitals form an orthogonal set: $\langle \phi_n | \phi_p \rangle = \delta_{np}$.[3]

The eigenvalues follow from diagonalizing the matrix

$$E_0 = \epsilon + 2t;\ E_1, E_2 = \epsilon + t;\ E_3, E_4 = \epsilon - t;\ E_5 = \epsilon - 2t.$$

Translation symmetry

You may wonder how I have diagonalized a 6×6 matrix so fast. The answer is "by using symmetry". The key trick is to use a transformation

$$\psi_k(\mathbf{r}) = \frac{1}{\sqrt{6}} \sum_{n=0}^{5} e^{ikR_n} \phi_n(\mathbf{r});\ \ R_n = na;\ \ k = m\frac{2\pi}{6a}; m = 0, ..., 5.$$

Here R measures the distance around the perimeter of the hexagon, where the carbon atoms are positioned at integer multiples of a, see the figure below.
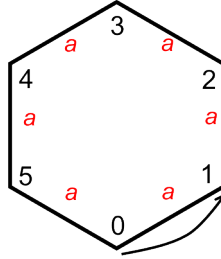
The transformation

$$
\begin{bmatrix}
\psi_{k_0} \\
\psi_{k_1} \\
\psi_{k_2} \\
\psi_{k_3} \\
\psi_{k_4} \\
\psi_{k_5}
\end{bmatrix}
=
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 \\
1 & e^{i\frac{\pi}{3}} & e^{i\frac{2\pi}{3}} & -1 & -e^{i\frac{\pi}{3}} & -e^{i\frac{2\pi}{3}} \\
1 & e^{i\frac{2\pi}{3}} & -e^{i\frac{\pi}{3}} & 1 & e^{i\frac{2\pi}{3}} & -e^{i\frac{\pi}{3}} \\
1 & -1 & 1 & -1 & 1 & -1 \\
1 & -e^{i\frac{\pi}{3}} & e^{i\frac{2\pi}{3}} & 1 & -e^{i\frac{\pi}{3}} & e^{i\frac{2\pi}{3}} \\
1 & -e^{i\frac{2\pi}{3}} & -e^{i\frac{\pi}{3}} & -1 & e^{i\frac{2\pi}{3}} & e^{i\frac{\pi}{3}}
\end{bmatrix}
\begin{bmatrix}
\phi_0 \\
\phi_1 \\
\phi_2 \\
\phi_3 \\
\phi_4 \\
\phi_5
\end{bmatrix},
$$

is a unitary transformation,[4] so

$$\langle \psi_k | \psi_{k'} \rangle = \delta_{kk'}.$$

The states $\psi_k(\mathbf{r})$ are called *Bloch states*. The usefulness of the transformation to Bloch states is that the Hamiltonian matrix becomes diagonal

$$\langle \psi_k | \hat{H} | \psi_{k'} \rangle = \delta_{kk'} \left(\epsilon + 2t \cos(ka) \right).$$

The eigenvalues of a diagonal matrix are then of course

$$E_k = \epsilon + 2t \cos(ka);\ \ k = m\frac{2\pi}{6a}; m = 0, ..., 5,$$

which are exactly the eigenvalues E_0 to E_5 given in the previous subsection.

---

[3]Introducing an overlap matrix would be more accurate of course, as would relaxing the nearest neighbor constraint for the integrals, but it would also be more tedious and obscure my message.

[4]If you are into group theory, you might identify the states ψ_k as the (projections onto the) irreducible representations of the 6-fold rotation group, where e^{ikR_n} are the characters. To make a connection to the stuff I am discussing below, I call it the "translation" group here. If you are not into group theory, just consider the transformation as clever inspiration.

My first band structure: a metal

A Bloch transformation can be use to solve the Hückel problem for any regularly shaped ring.

$$\psi_k(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} e^{ikR_n} \phi_n(\mathbf{r}); \ \ R_n = na; \ k = m\frac{2\pi}{Na}; m = 0, ..., N-1.$$
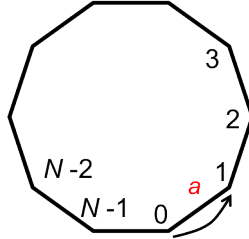
The Bloch states form an orthogonal basis $\langle \psi_k | \psi_{k'} \rangle = \delta_{kk'}$, and the Hamiltonian matrix is diagonal, giving the eigenvalues

$$E_k = \epsilon + 2t\cos(ka); \ \ k = m\frac{2\pi}{Na}; m = 0, ..., N-1.$$

Actually the shape of the ring does not matter too much. As long as the on-site energy ϵ stays the same for all atoms, and the coupling t of an atom with its neighbors stays the same, we can deform the ring any way we want without changing the Hückel model. For instance we could flatten the ring out into a linear chain as in the figure below. Just remember that the left end of the chain is still connected to the right end (not drawn in the figure below) with coupling strength t, i.e., topologically the chain is still a ring.
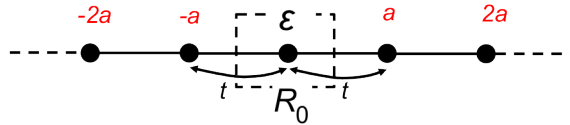
So we still apply cyclic indices

$$R_{n+N} = R_n.$$

In solid state physics folklore this is called the *Born-von Karman boundary condition.*[5] Choosing the origin in a ring chain is arbitrary. Just pick a site and call its position $R_0 = 0$. It is then custom to label the N unique atomic positions as

$$R_n = na; \ n = -\frac{N}{2} + 1, -\frac{N}{2} + 2, ..., 0, 1, ..., \frac{N}{2} - 1, \frac{N}{2}.$$

If we let $N \to \infty$, then the chain is a model for a simple one-dimensional (infinite) crystal. The atomic positions R_n form a regular (one-dimensional) lattice called the *crystal lattice*. It can be generated by choosing a *unit cell* that contains all the unique information (in this

---

[5]You may be wondering why we use this rather artificial construct, and why we don't take an "open" chain, i.e. one that ends at the left and right ends. One can of course. An open chain is however more difficult to handle technically. If the chain is sufficiently long, an open chain has the same energy spectrum as a "ring" chain, apart from some states that are localized at the edges of the chain. Unless one is specifically interested in these edge states, it is simpler to consider the ring chain.

case the one atom at the origin R_0), and create copies of that cell at positions na; $n = \pm 1, \pm 2, ..., \pm \infty$, where a is called the *lattice parameter*, or the *cell parameter*. The eigenvales of the chain are as before

$$E_k = \epsilon + 2t\cos(ka); \ k = m\Delta k; \ \Delta k = \frac{2\pi}{Na}; \ m = -\frac{N}{2} + 1, -\frac{N}{2} + 2, ..., 0, 1, ..., \frac{N}{2} - 1, \frac{N}{2}.$$

If $N \to \infty$, then $\Delta k \to 0$, and the energy spectrum becomes continuous. The spectrum does have an upper and a lower bound however, $E_{\max} = \epsilon - 2t$; $E_{\min} = \epsilon + 2t$. A continuous spectrum with an upper and a lower bound is called an *energy band*. Analytically, continuous spectra are tricky to work with, so one usually takes this limit somewhere near the end of the calculations. On the computer one has to use some form of discretization anyway.

The range of k-values needs some explanation. First of all note that

$$E_{k+K} = E_k,$$

where $K = N\Delta k = \frac{2\pi}{a}$. The energy is a periodic function of k, where the period is K. This symmetry is a consequence of the translation symmetry of the lattice, and it does not depend on the details of the model used, Hückel or otherwise. With N atoms in the chain one should get N energy levels, which are given by the N eigenvalues within one period. The information outside this period is obsolete; it just duplicates what we already have. Whereas the size of the period is fixed, its origin is free to choose. It is custom to set the range of k-values to

$$-\frac{\pi}{a} < k \leq \frac{\pi}{a}.$$

This range is called the *Brillouin zone*. This setting has the advantage that it straight-forwardly expresses a symmetry in the spectrum

$$E_{-k} = E_k.$$

This symmetry is derived from time-reversal symmetry, which is fundamental and does not depend on any of the details of the model we have used. The figure below gives a plot of E_k.
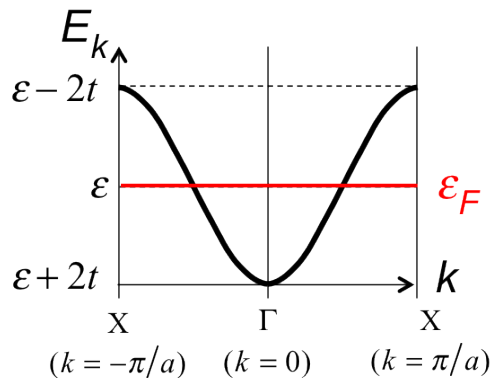
Such a plot is called a *band structure*. In solid state physics folklore special k-points have often special names: the point $k = 0$ is called Γ, and the point $k = \frac{\pi}{a}$ is called X.[6] There

---

[6]I guess one of the reasons for inventing these names is that spectra often have special features at these special points. In this particular case the points mark the extrema of the band, which give peaks in the density of states, and often peaks in the spectra (see the next subsection).

are N levels in this band. If each atom brings along one electron (as it would do in our Hückel model for the π-electrons of a ring or chain of carbon atoms), then in the ground state the lower $\frac{1}{2}N$ levels are filled (each level with a spin-up and a spin-down electron), and the upper $\frac{1}{2}N$ levels are empty. The energy of the highest filled state (in the ground state) is called the *Fermi energy* ϵ_F. In this particular case it is easy to determine. As the spectrum is symmetric around $E = \epsilon$, it means that $\epsilon_F = \epsilon$.

Excited states are created in this system by moving an electron from an energy level $E_i < \epsilon_F$ to an energy level $E_f > \epsilon_F$. If the Fermi level is inside an energy band, like here, the excitation energy $\Delta E = E_f - E_i$ can be made vanishingly small, as there are always levels available very close in energy to the Fermi level. Applying a low frequency electromagnetic field, for instance, is sufficient to excite electrons.[7] A system that strongly responds to such fields is a *metal*.

Density of states

Although the energy spectrum is continuous, it is not without structure. A quantity that characterizes this structure is the *density of states* $D(E)$: the number of energy levels per unit energy interval. For an infinite system (in the thermodynamic limit) it is proportional to the volume, so one defines the density of states per unit volume (or in our 1D example, per unit length). Simply count the number of energy levels with energies $E \leq E_k \leq E + dE$

$$D(E) = \frac{1}{a}\frac{1}{dE}\sum_k \left[\Theta(E_k - E) - \Theta(E_k - E - dE)\right],$$

with $\Theta(x) = 0; x < 0$ and $\Theta(x) = 1; x \geq 0$ the (Heaviside) step function. One can convert the sum into an integral using a standard trick

$$\sum_k f = \frac{1}{\Delta k}\sum_k f\Delta k = \frac{a}{2\pi}\sum_k f\Delta k = \frac{a}{2\pi}\int f dk,$$

which is allowed as we are letting $\Delta k \to 0$ anyway. Letting the energy interval $dE \to 0$, and noting that then $\left[\Theta(E_k - E) - \Theta(E_k - E - dE)\right]/dE \to \delta(E_k - E)$, the density of states becomes

$$D(E) = \frac{1}{2\pi}\int_{-\frac{\pi}{a}}^{\frac{\pi}{a}}\delta(E_k - E)dk.$$

Taking into account the spin degree of freedom of the electrons, this is the density of states per spin state. If we have spin degeneracy (spin-up has the same energy as spin-down), one has to multiply $D(E)$ by a factor of two.

For our model we can find a closed analytical expression for the density of states. As $E_k = \epsilon + 2t\cos(ka)$ and, $\delta(f(x)) = \sum_{a;f(x_0)=0;f'(x_0)\neq 0}\delta(x - x_0)/|f'(x_0)|$, this gives

$$D(E) = \frac{1}{2\pi}\frac{1}{a\,|t\sin(k_0a)|},$$

where k_0 is the solution of $E = E_k$, so $k_0 = \arccos\left(\frac{E-\epsilon}{2t}\right)$. Obviously this has only a solution if $-1 \leq \left|\frac{E-\epsilon}{2t}\right| \leq 1$, i.e. for energies within the band. For energies outside the band $D(E) = 0$ (as there are no states there). This density of states looks like

---

[7]Actually, even a zero-frequency field, i.e., a static field, suffices. In general, the excitations create an electric current.
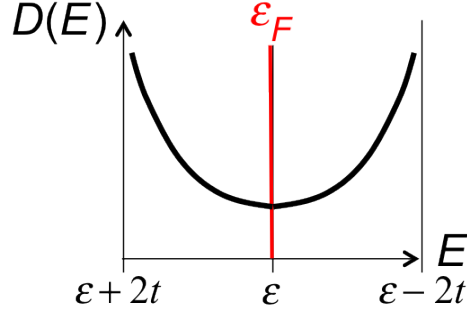
Note that $D(E) \to \infty$ if $E \to \epsilon \pm 2t$, i.e., the density of states becomes infinite if the energy approaches the band edges, where the extrema of the bands are. Such singularities are typical for one-dimensional systems.[8]

In two or three dimensions there are still singularities in the density of states, but they are of a milder type: in 2D the density of states shows discontinuities (but doesn't go off to infinity) and in 3D the derivative of the density of states has discontinuities. They are called *van Hove singularities*. The δ-function expression for the density of states is quite general; in two and three dimensions one gets

$$D(E) = \frac{1}{(2\pi)^n} \int \delta(E_{\mathbf{k}} - E) d^n k,$$

with $n = 2$ or $n = 3$. However, finding a closed analytical expression for $D(E)$ is not so simple anymore in 2D or 3D.

The density of states is a very useful object in solid state physics. One can use it to calculate the Fermi level for instance

$$\int_{-\infty}^{\epsilon_F} D(E) dE = n_e.$$

The right-hand side is the number of electrons per unit volume (which you should know, as you know on which material you do your calculation). Your band structure calculation gives you $D(E)$, and the integral just tells you to fill your levels until all electrons are used up; the energy of the highest level is then your Fermi energy.

One can also calculate the total energy, which in an independent electron model is simply the sum of all energies of the occupied states

$$E_{tot} = \int_{-\infty}^{\epsilon_F} E D(E) dE.$$

Bloch waves

The eigenstates of the simple Hückel Hamiltonian are given by

$$\psi_k(\mathbf{r}) = \sum_R e^{ikR} \phi(\mathbf{r} - \mathbf{R}); \ \mathbf{R} = \hat{\mathbf{x}}R; \ R = na; \ n = 0, \pm 1, \pm 2, ...,$$

in a slightly different notation as before, assuming that the chain lies along the $\hat{\mathbf{x}}$-direction. If the chain is infinite, the states is infinitely stretched out in space, so normalization is

---

[8]They can be observed in experimental spectra. Real 1D systems of course don't exist, but very anisotropic 2D or 3D systems can be quasi-1D. They show large peaks instead of the singularities.

ill-defined. The normalization used here has

$$\int\int\left[\int_0^a |\psi_k(\mathbf{r})|^2\,dx\right]dydz = 1,$$

i.e., the wave function is normalized per unit cell.

The state $\psi_k(\mathbf{r})$ obviously has a wave-like character. It is called a *Bloch wave*. A pictorial representation is
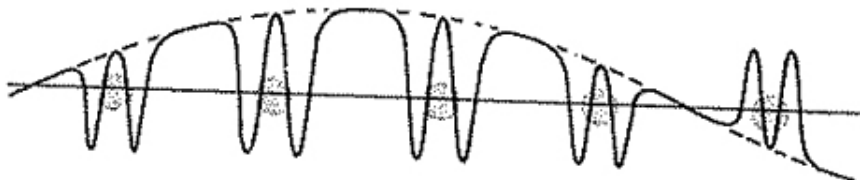
where the dotted line indicates the (real part of the) e^{ikR} factor, and the solid line the atomic orbital $\phi(\mathbf{r}-\mathbf{R})$.[9] You can do with this wave whatever you are used to do with plane waves in vacuum. The wave number is $k = \frac{2\pi}{\lambda}$ (with λ the wave length). According to de Broglie and Planck

$$p = \hbar k;\ \ \epsilon_k = \hbar\omega_k,$$

are the relations between the momentum and energy of the quantum particle (an electron in the chain), and the wave number and frequency of the quantum wave (the Bloch wave). The relation between frequency and wave number is called the *dispersion relation*. In this case it is

$$\omega_k = \hbar^{-1}(\epsilon + 2t\cos(ka)).$$

To represent a more or less localized particle one can make a wave packet. According to standard wave theory, a wave packet moves with a speed

$$v = \frac{d\omega_k}{dk} = \frac{d\epsilon_k}{dp},$$

called the *group velocity*. Going from the quantum mechanics of wave packets to classical mechanics of particles in the standard way, one obtains the classical equations of motions

$$\dot{r} = \frac{d\epsilon_k}{dp};\ \ \dot{p} = F,$$

where F is the force operating on the electron. The equations of motion have an obvious generalization in higher dimensions

$$\dot{\mathbf{r}} = \nabla_{\mathbf{p}}\epsilon_{\mathbf{k}};\ \ \dot{\mathbf{p}} = \mathbf{F},$$

In practice the dominant force operating on an electron will almost always be the Lorentz force caused by an external electromagnetic field

$$\mathbf{F} = -e\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right).$$

---

[9]The drawing is actually wrong. The wave is in the $\hat{\mathbf{x}}$-direction, so the atomic orbital apparently represents a p_x-orbital, and not a p_z-orbital. Sorry about that; I couldn't find a better picture in a hurry.

One can use these equations of motion to derive a semi-classical theory of electron transport, called *Drude-Sommerfeld*, to derive Ohm's law, for instance, and more interesting things. It is called semi-classical, because there is some quantum mechanics included: the band structure $\epsilon_{\mathbf{k}}$ is derived from quantum mechanics.[10]

In case the devices become sufficiently small, such that the wave length of the Bloch waves becomes a relevant length scale, one can get wave interference effects. In that case one has to switch to a completely quantum mechanical theory of electron transport, see my notes of the 2005 Han-sur-Lesse workshop.

My second band structure: an insulator

The picture below shows a slightly more complicated chain. The unit cell with lattice constant a contains two different atoms. Instead of a chain of carbon atoms, think of a chain with alternating boron and nitrogen atoms, for instance. We stick to the same Hückel model for the p_z-orbitals, but now the on-site energies ϵ_1 and ϵ_2 are different, as the atoms are different. Keeping the interatomic distances equal, the nearest neighbor Hamiltonian matrix element t stays the same for all sites.
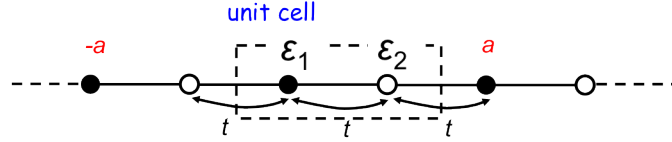
Inspired by the results of the previous sections we start with *Bloch orbitals*.

$$\chi_{1,k}(\mathbf{r}) = \sum_{n=0}^{N-1} e^{ikR_n}\phi_{1,n}(\mathbf{r}); \ \chi_{2,k}(\mathbf{r}) = \sum_{n=0}^{N-1} e^{ikR_n}\phi_{2,n}(\mathbf{r}),$$

where $\phi_{1,n}(\mathbf{r})$ is a p_z orbital on atom 1 in cell n, and $\phi_{2,n}(\mathbf{r})$ is a p_z orbital on atom 2 in cell n. As before, matrix elements of the Hamiltonian between states with different k are zero, so the Hamiltonian is diagonal in k. The Bloch state that is an eigenstate of the Hamiltonian must be a linear combination of the Bloch orbitals

$$\psi_k(\mathbf{r}) = c_{1,k}\chi_{1,k}(\mathbf{r}) + c_{2,k}\chi_{2,k}(\mathbf{r}).$$

A matrix representation of the Hamiltonian can be found by the following trick

$$\underline{H}_k = \begin{bmatrix} \langle\phi_{1,0}|\,\hat{H}\,|\chi_{1,k}\rangle & \langle\phi_{1,0}|\,\hat{H}\,|\chi_{2,k}\rangle \\ \langle\phi_{2,0}|\,\hat{H}\,|\chi_{1,k}\rangle & \langle\phi_{2,0}|\,\hat{H}\,|\chi_{2,k}\rangle \end{bmatrix} = \begin{bmatrix} \epsilon & t\left(1 + e^{-ika}\right) \\ t\left(1 + e^{ika}\right) & \epsilon \end{bmatrix}.$$

Because $c_{1,k} = \langle\phi_{1,0}|\psi_k\rangle$ and $c_{2,k} = \langle\phi_{2,0}|\psi_k\rangle$, a matrix-vector representation of the Schrödinger becomes

$$\hat{H}\psi_k(\mathbf{r}) = E_k\psi_k(\mathbf{r}) \rightarrow \underline{H}_k\underline{c}_k = E_k\underline{c}_k.$$

The 2×2 eigenvalue problem is easily solved to give

$$E_{1(2),k} = \frac{1}{2}\left(\epsilon_1 + \epsilon_2\right) \pm \sqrt{\frac{1}{4}\left(\epsilon_1 - \epsilon_2\right)^2 + |t_k|^2},$$

[10]If one looks at thermal transport properties, one also needs the Fermi-Dirac distribution, which also involves quantum mechanics.

with $t_k = t\left(1 + e^{-ika}\right)$. As $|t_k|^2 = t^2\left(2 + 2\cos(ka)\right) = 4t^2\cos^2(\frac{1}{2}ka)$, we get

$$E_{1(2),k} = \frac{1}{2}\left(\epsilon_1 + \epsilon_2\right) \pm \sqrt{\frac{1}{4}\left(\epsilon_1 - \epsilon_2\right)^2 + 4t^2\cos^2(\frac{1}{2}ka)}.$$

A picture is given below for the case $\epsilon_1 < \epsilon_2$.
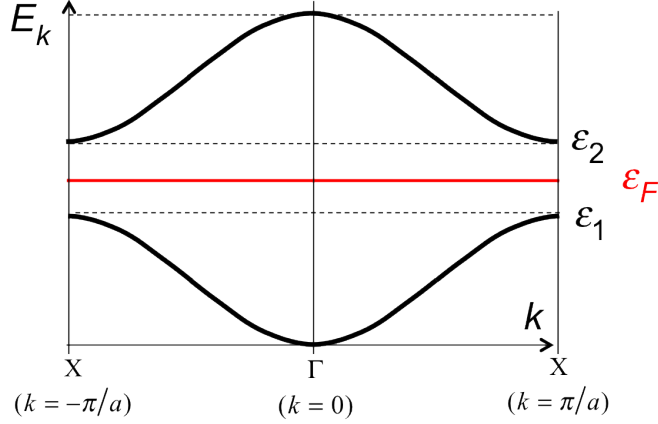
We have two bands now. Calculating the eigenvalues at the special points

$$E_{1,\Gamma} = \frac{1}{2}\left(\epsilon_1 + \epsilon_2\right) - \sqrt{\frac{1}{4}\left(\epsilon_1 - \epsilon_2\right)^2 + 4t^2}; \quad E_{2,\Gamma} = \frac{1}{2}\left(\epsilon_1 + \epsilon_2\right) + \sqrt{\frac{1}{4}\left(\epsilon_1 - \epsilon_2\right)^2 + 4t^2};$$

$$E_{1,X} = \epsilon_1; \quad E_{2,X} = \epsilon_2,$$

and knowing that $E_{1(2),k}$ are monotonic functions for k between Γ and X, shows that there is a gap in the middle of the energy spectrum

$$\epsilon_1 < E < \epsilon_2,$$

in which there are no energy levels. This is called a *band gap*. As again each atom brings one electron, in the ground state of the system the lowest half of the energy states are occupied. With the two bands being mirror images of one another, it means that the lower band with energies $E_k \leq \epsilon_1$ is occupied, whereas the upper band with energies $E_k \geq \epsilon_2$ is empty. The Fermi level is thus $\epsilon_F = \epsilon_1$. This is actually only true at zero temperature. At finite temperature the Fermi level is close to the middle of the gap, as drawn in the figure above.[11] If we try to excite the system by electromagnetic fields, a minimum frequency is needed to excite an electron across the gap

$$\hbar\omega \geq E_g = \epsilon_2 - \epsilon_1.$$

The system does not respond at all to radiation with a lower frequency, i.e., it is completely transparent to such radiation.[12] Such a system is called an *insulator*.[13]

---

[11] Actually one should speak of the chemical potential μ instead of the Fermi level ϵ_F. Solid state physicists are sloppy; they make no distinction between the two quantities.

[12] Vibrations of the atoms in the lattice can also be excited and thus absorb radiation. We are only considering electronic excitations here.

[13] The system also does not respond to fields with frequency zero, i.e., static fields. No electric currents are thus generated.

At a finite temperature there is always a finite probability that the system is in one of its excited states, the probability being determined by the Fermi-Dirac distribution. If that probability is sufficiently high at a certain temperature to give a significant number of electrons in the upper band, then those electrons can carry an electric current if an external field is applied. Such a material is called a *semiconductor*. There is no hard boundary between insulators and semiconductors; it all depends on the temperature you wish to work on, and what you call a significant number (of electrons in the upper band). Most commercially applied semiconductors work at room temperature, and have gaps between 1 and 2 eV, e.g., for Ge, Si, GaAs, $E_g = 0.7, 1.2, 1.5$ eV, respectively. Materials with larger gaps are sometimes euphemistically called "wide band gap semiconductors", but if $E_g \gtrsim 3$ eV this becomes just another word for insulators.

Tight-binding band structures in 2D and 3D

Tight-binding in 2D or 3D is conceptually not different from 1D. One starts with Bloch orbitals

$$\chi_{j,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \phi_j(\mathbf{r} - \mathbf{R}),$$

where $\phi_j(\mathbf{r} - \mathbf{R})$ is an atomic orbital in a unit cell at position $\mathbf{R}$. These positions form a regular lattice in real space; an example will be given in the next section. A Bloch orbital is a wave, and the wave vector $\mathbf{k}$ is now indeed a (2D or 3D) vector. It is a continuous variable, but its size is limited: $\mathbf{k}$ has to lie in the Brillouin zone; again an example is given in the next section. Again, Bloch orbitals with different $\mathbf{k}$ don't mix (the Hamiltonian is diagonal in $\mathbf{k}$), so the eigenstates have the form

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = \sum_i c_{jn,\mathbf{k}} \chi_{j,\mathbf{k}}(\mathbf{r}).$$

The Schrödinger equation becomes

$$\widehat{H}\psi_{n,\mathbf{k}}(\mathbf{r}) = E_{n,\mathbf{k}}\psi_{n,\mathbf{k}}(\mathbf{r}).$$

The collection of eigenvalues $E_{n,\mathbf{k}}$ is called the *band structure*, where $n = 1, 2, ...$ is the *band index*, which simply numbers the bands, and $\mathbf{k}$ denotes the wave vector (or $\hbar\mathbf{k}$ the momentum, if you prefer to speak in particle language). In principle n runs to ∞, as the spectrum is unbounded at the top. In practice the number of orbitals $\phi_j(\mathbf{r})$ used in the basis set is finite, say N. A band structure calculation then yields N bands.

In principle this representation of Bloch states is exact provided the atomic orbital basis set is complete. On this basis the Schrödinger equation becomes a matrix eigenvalue equation. The derivation is straightforward; considering the inner product of both sides of the Schrödinger equation with a basis function one gets

$$
\begin{aligned}
\langle\chi_i|\widehat{H}|\psi_{n,\mathbf{k}}\rangle &= E_{n,\mathbf{k}}\langle\chi_i|\psi_{n,\mathbf{k}}\rangle \Leftrightarrow \\
\sum_j H_{ij,\mathbf{k}}c_{jn,\mathbf{k}} &= E_{n,\mathbf{k}}\sum_j S_{ij,\mathbf{k}}c_{jn,\mathbf{k}}
\end{aligned}
$$

with the Hamiltonian and overlap matrices being defined by

$$
\begin{aligned}
H_{ij,\mathbf{k}} &= \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\langle\chi_i|\widehat{H}|\chi_{j\mathbf{R}}\rangle \\
S_{ij,\mathbf{k}} &= \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\langle\chi_i|\chi_{j\mathbf{R}}\rangle
\end{aligned}
$$

with $\chi_{j\mathbf{R}}(\mathbf{r}) \equiv \chi_j(\mathbf{r} - \mathbf{R})$. The matrix elements $\langle \chi_i | \widehat{H} | \chi_{j\mathbf{R}} \rangle$ and $\langle \chi_i | \chi_{j\mathbf{R}} \rangle$ get smaller the larger $|\mathbf{R}|$ (usually exponentially). For simple calculations the Hamiltonian matrix elements are set to zero, unless the atomic functions refer to the same site, or to neighboring sites. This is then called the nearest neighbor tight-binding approximation. A further approximation is then to assume that the atomic orbitals are orthogonal, i.e. $S_{ij,\mathbf{k}} = \delta_{ij}$.

Graphene

We will establish the band structure of the π-bands of graphene using the simple nearest neighbor tight-binding model (i.e., the Hückel model). The C atoms in graphene form a regular 2D hexagonal lattice. First we have to establish what this lattice looks like mathematically. We will also introduce an auxilary lattice called the reciprocal lattice, for reasons that will become clear below.

Graphene lattice

The 2D hexagonal lattice of graphene is sketched below. It can be constructed by choosing a unit cell with lattice vectors

$$\mathbf{a}_1 = a \left(\frac{\sqrt{3}}{2}, -\frac{1}{2} \right); \ \mathbf{a}_2 = a \left(0, 1 \right),$$

with a the lattice constant. The rhombic unit cell contains two C atoms at positions

$$\boldsymbol{\tau}_1 = a \left(\frac{1}{\sqrt{3}}, 0 \right); \ \boldsymbol{\tau}_2 = a \left(\frac{1}{2\sqrt{3}}, \frac{1}{2} \right),$$

respectively. A schematic representation of the lattice is given in the figure below. The positions of all the atoms in the lattice can be generated by

$$\boldsymbol{\tau}_1 + \mathbf{R}; \ \boldsymbol{\tau}_2 + \mathbf{R}; \ \mathbf{R} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2; \ \ n_1, n_2 = 0, \pm 1, \pm 2, \dots$$

It is clear that this gives two sublattices, which are indicated by the indices 1 and 2 in the figure below. This particular construction of the lattice is not unique; one can choose alternative unit cells with different orientations and origins. The lattice itself however is unique. The hexagonal symmetry of the lattice is not immediately obvious from a single unit cell; it becomes visible only after constructing the lattice. If one wants the unit cell to reflect the symmetry of the lattice, one can use the *Wigner-Seitz cell*.[14] In this case the Wigner-Seitz cell is the hexagon that is also the building block of the graphene structure.

---

[14]It follows from a geometrical construction. Look on the web to see how it is done.
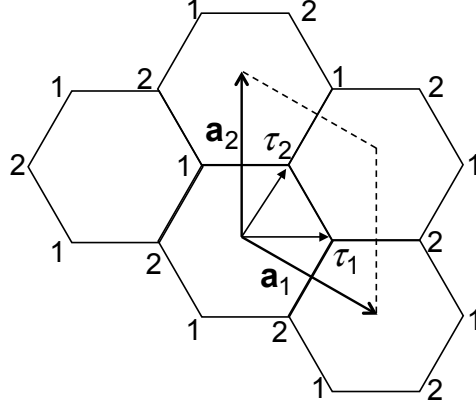
Note that the lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are not orthogonal. Dealing with Bloch orbitals one has to calculate lots of phase factors of the type $e^{i\mathbf{k}\cdot\mathbf{R}}$. Calculating inner products $\mathbf{k}\cdot\mathbf{R}$ in a non-orthogogal system is a bit of a nuisance. Suppose we write

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2; \quad n_1, n_2 = 0, 1, 2, ...$$
$$\mathbf{k} = k_1\mathbf{b}_1 + k_2\mathbf{b}_2,$$

with $\mathbf{R}$ real space lattice vectors and $\mathbf{k}$ expressed in some lattice vectors we choose to simplify calculations.[15]

$$\mathbf{k}\cdot\mathbf{R} = \sum_{i,j=1}^{2} k_i n_j \mathbf{b}_i \cdot \mathbf{a}_j.$$

This expression becomes simple if, and only if, the lattice vectors $\mathbf{b}_1, \mathbf{b}_2$ are chosen such that

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{i,j}.$$

The phase factors then get the simple form

$$e^{i\mathbf{k}\cdot\mathbf{R}} = e^{i2\pi(k_1 n_1 + k_2 n_2)}.$$

The lattice generated by the vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ is called the *reciprocal lattice*.[16]

$$\mathbf{K} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 \text{ with } m_1, m_2 \text{ integers},$$

for thr reciprocal lattice pertaining to graphene, see the figure below.
To construct a Bloch wave one can in principle take any vector $\mathbf{k}'$. However if $\mathbf{k}'$ is outside the unit cell at the origin, spanned by $\mathbf{b}_1$ and $\mathbf{b}_2$, one can always write

$$\mathbf{k}' = \mathbf{k} + \mathbf{K} \text{ with}$$

---

[15] One can of course choose $\mathbf{b}_1 = \mathbf{a}_1$ and $\mathbf{b}_2 = \mathbf{a}_2$, but this does not simplify things if the lattice vectors are non-orthogonal.

[16] For those of you who have seen similar things before: the representations in real and reciprocal lattice vectors are each other's "dual representations". One may call the real space representation "contravariant" and the reciprocal lattice representation "covariant", where $g_{ij} = \mathbf{a}_i \cdot \mathbf{a}_j$ is the metric.

$\mathbf{K}$ some reciprocal lattice vector and $\mathbf{k}$ a vector within the unit cell at the origin. As

$$e^{i\mathbf{K}\cdot\mathbf{R}} = e^{i2\pi(m_1 n_1 + m_2 n_2)} = 1,$$

this means

$$e^{i\mathbf{k}'\cdot\mathbf{R}} = e^{i\mathbf{k}\cdot\mathbf{R}}.$$

In other words, by taking only vectors $\mathbf{k}$ within the unit cell of the reciprocal lattice, we cover all possible wave functions and energies of the system.

Writing $\mathbf{b}_i = (b_{i1}, b_{i2})$, and demanding $\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{i,j}$, this leads to the equations

$$b_{11}a_{11} + b_{12}a_{12} = \frac{2\pi}{a} = b_{11}\frac{\sqrt{3}}{2} - b_{12}\frac{1}{2}$$

$$b_{11}a_{21} + b_{12}a_{22} = 0 = b_{12}$$

$$b_{21}a_{11} + b_{22}a_{12} = 0 = b_{21}\frac{\sqrt{3}}{2} - b_{22}\frac{1}{2}$$

$$b_{21}a_{21} + b_{22}a_{22} = \frac{2\pi}{a} = b_{22}$$

which are easily solved to give

$$\mathbf{b}_1 = \frac{2\pi}{a}(\frac{2}{\sqrt{3}}, 0); \quad \mathbf{b}_2 = \frac{2\pi}{a}(\frac{1}{\sqrt{3}}, 1)$$

The reciprocal lattice is sketched in the figure below. The special points $\Gamma = \frac{2\pi}{a}(0,0)$, $M = \frac{2\pi}{a}(\frac{1}{\sqrt{3}}, 0)$, and $K = \frac{2\pi}{a}(\frac{1}{\sqrt{3}}, \frac{1}{3})$ are indicated. In terms of the reciprocal lattice vectors we can write

$$\Gamma = 0\mathbf{b}_1 + 0\mathbf{b}_2; \quad M = \frac{1}{2}\mathbf{b}_1 + 0\mathbf{b}_2; \quad K = \frac{1}{3}\mathbf{b}_1 + \frac{1}{3}\mathbf{b}_2; \quad K' = \frac{2}{3}\mathbf{b}_1 + \frac{2}{3}\mathbf{b}_2.$$

As for the real lattice, the unit cell spanned by $\mathbf{b}_1$ and $\mathbf{b}_2$ does not immediately reflect the symmetry of the lattice. A Wigner-Seitz unit cell does. The Wigner-Seitz cell associated with the reciprocal lattice is called the Brillouin zone. In this case it is the hexagon around the Γ-point sketched below.
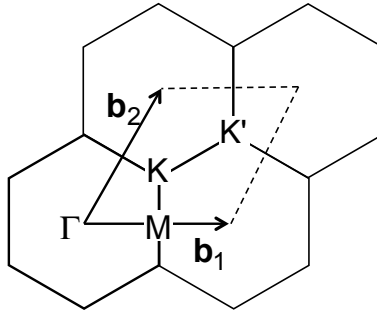
Intermezzo: Diffraction (if you are only interested in electronic structures, you may skip this)

The reciprocal lattice is a helpful mathematical tool for constructing Bloch waves and band structures. It is also directly observable. Consider the scattering of electromagnetic (EM) waves from a lattice of atoms. The incoming wave is a plane wave with electric field component $\mathbf{E}_I(\mathbf{r}, t) = \mathbf{E}_0 \exp\left[i\left(\mathbf{K}_I \cdot \mathbf{r} - \omega t\right)\right]$, where $|\mathbf{K}_I| = 2\pi/\lambda \equiv K'$ with λ the wave length, the unit vector $\mathbf{K}_I/K'$ giving the propagation direction of the wave, ω the (angular) frequency of the wave, and $\mathbf{E}_0$ the amplitude/polarization.[17] The charge distribution of an atom or molecule that is hit by an EM wave starts to oscillate with the same frequency as the incoming wave. An oscillating charge distribution emits EM radiation, which in this case has the same frequency and wave length as the incoming wave.[18] The radiated wave has the form of a spherical wave (at least for distances sufficiently far from the atom), with electric field component $\mathbf{E}_a(\mathbf{r}, t) = \mathbf{f} \exp\left[i\left(K'r - \omega t\right)\right]/r$, see the figure below.
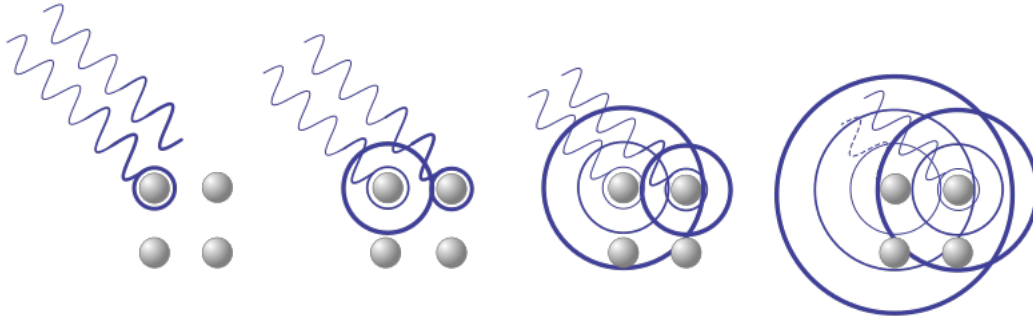
At the atomic positions $\mathbf{R}$, the incoming plane wave is $\mathbf{E}_I(\mathbf{R}, t) = \mathbf{E}_0 \exp\left[i\left(\mathbf{K}_I \cdot \mathbf{R} - \omega t\right)\right]$. As the atoms are driven by the incoming plane wave, there is a difference in phase between the spherical waves send out by the different atoms, simply because they are at different positions. The plane wave needs to travel different path lengths to the different atoms, so it arrives with a different phase. If $\mathbf{f}$ is the amplitude/phase of the spherical wave send out by an atom at the origin, then $\mathbf{f} \exp\left[-i\mathbf{K}_I \cdot \mathbf{R}\right]$ is the amplitude/phase of the spherical wave send out by an atom at position $\mathbf{R}$. For a lattice of atoms the sum of the phase differences coming from the incoming wave is $\sum_\mathbf{R} \exp\left[-i\mathbf{K}_I \cdot \mathbf{R}\right]$.

The spherical waves send out by the different atoms are summed, and the total wave is detected. We put our detector far from the lattice, and call $\mathbf{K}_O$ (with $|\mathbf{K}_O| = K'$) be the vector that gives the propagation direction from lattice to detector. The waves coming from the individual atoms have to travel different path lengths to the detector, hence they arrive with a different phase. Following the same reasoning as above, the sum of the phase differences for a lattice of atoms is $\sum_\mathbf{R} \exp\left[i\mathbf{K}_O \cdot \mathbf{R}\right]$. The electric field of the total wave is then given by

$$\mathbf{E}_O(\mathbf{r}, t) = S\left(\mathbf{K}_O - \mathbf{K}_I\right) \mathbf{f} \frac{\exp\left[i\left(K'r - \omega t\right)\right]}{r},$$

---

[17]There is a similar magnetic field component $\mathbf{B}_I(\mathbf{r}, t) = \mathbf{B}_0 \exp\left[i\left(\mathbf{K}_I \cdot \mathbf{r} - \omega t\right)\right]$, which is not terribly important in the present discussion.

[18]This is called *elastic scattering,* which is what the diffraction experiment is based upon. Inelastic scattering is also possible; part of the incoming EM energy is then converted into some other energy, electronic excitations, for instance. That is spectroscopy, however.

where the lattice sum

$$S(\mathbf{K}') = \sum_{\mathbf{R}} \exp\left[i\mathbf{K}' \cdot \mathbf{R}\right]$$

incorporates the interference caused by different path lenghts of the waves send out by the different atoms. For a regular lattice like the one discussed above, the interference will be destructive in most cases, i.e., $S(\mathbf{K}') \approx 0$.[19]

Only if $\mathbf{K}' \cdot \mathbf{R}$ is a multiple of 2π, then the interference is constructive, and $S(\mathbf{K}') = N$, where N is the number of atomic sites in the lattice (a macroscopic number usually). Following the discussion of the previous section, this is only the case if $\mathbf{K}' = \mathbf{K}$, a reciprocal lattice vector. The condition for constructive interference

$$\mathbf{K}' = \mathbf{K}_O - \mathbf{K}_I = \mathbf{K},$$

is called the *von Laue equations* (plural, as there is an equation for each component of the vector).[20]

The von Laue equations imply that the reciprocal lattice is directly visible in a diffraction experiment in terms of the directions in space along which constructive interference takes place. It is much more difficult to make the real space lattice visible (by scanning tunneling microscopy, for instance), So in a sense the reciprocal lattice is more real than the real space lattice. Below is an example of a diffraction pattern obtained for graphene.[21]

---
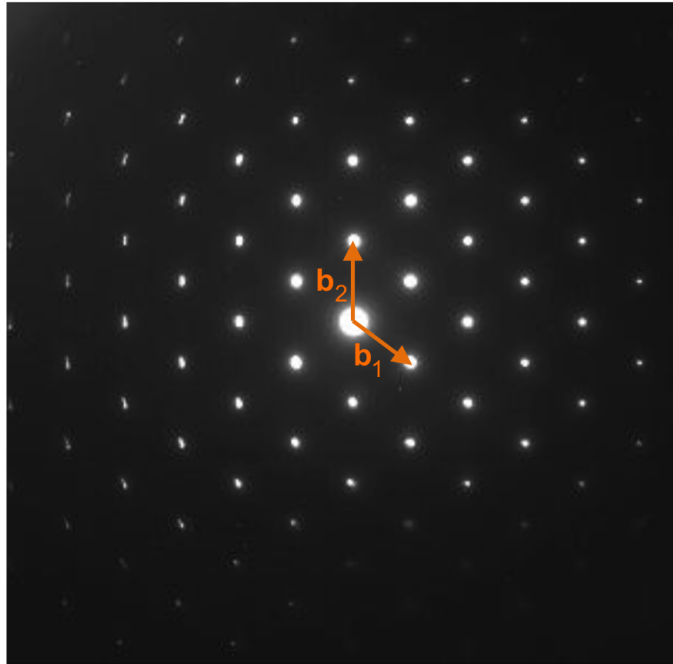
[19]Try to prove this yourself; start with a simple one-dimensional lattice.

[20]They are equivalent to Bragg's law for diffraction. In my humble opinion the von Laue approach is a much better way to explain diffraction than the Bragg approach with its reflecting planes. The atomic cross sections for X-rays of not-too-heavy atoms are quite small. To an X-ray a lattice largely looks like empty space with small atoms on the lattice sites. Point sources emitting spherical waves sounds more physical to me than reflecting planes. The mathematical results of both approaches are identical though.

[21]It is actually an electron diffraction pattern. Electrons scatter in much the same way as EM waves, but the atomic scattering cross sections are much larger. To an electron wave an atomic lattice is largely filled space. The electron wave hardly penetrates, which makes electron diffraction a surface sensitive technique, which is of course ideal for studying 2D materials. An X-ray is hardly scattered at all by a single atomic layer.

Nearest neighbor tight-binding model

Within the nearest neighbor approximation the diagonal matrix elements of the Hamiltonian matrix become

$$H_{11,\mathbf{k}} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \langle \chi_1 | \widehat{H} | \chi_{1\mathbf{R}} \rangle = \langle \chi_1 | \widehat{H} | \chi_1 \rangle = E_p$$

$$H_{22,\mathbf{k}} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \langle \chi_2 | \widehat{H} | \chi_{2\mathbf{R}} \rangle = \langle \chi_2 | \widehat{H} | \chi_2 \rangle = E_p$$

since only $\mathbf{R} = 0$ contributes, and the atoms of both sublattices 1 and 2 are equivalent. The off-diagonal matrix elements become

$$H_{12,\mathbf{k}} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \langle \chi_1 | \widehat{H} | \chi_{2\mathbf{R}} \rangle = \langle \chi_1 | \widehat{H} | \chi_2 \rangle \left(1 + e^{i\mathbf{k}\cdot\mathbf{a}_1} + e^{-i\mathbf{k}\cdot\mathbf{a}_2} \right)$$

$$H_{21,\mathbf{k}} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \langle \chi_2 | \widehat{H} | \chi_{1\mathbf{R}} \rangle = \langle \chi_2 | \widehat{H} | \chi_1 \rangle \left(1 + e^{-i\mathbf{k}\cdot\mathbf{a}_1} + e^{i\mathbf{k}\cdot\mathbf{a}_2} \right) = H_{12,\mathbf{k}}^*$$

We introduce the hopping parameter

$$t \equiv \langle \chi_1 | \widehat{H} | \chi_2 \rangle$$

and write

$$H_{12,\mathbf{k}} = t \left(1 + e^{i\mathbf{k}\cdot\mathbf{a}_1} + e^{-i\mathbf{k}\cdot\mathbf{a}_2} \right) = t \left(1 + e^{i2\pi k_1} + e^{-i2\pi k_2} \right) = t_{\mathbf{k}}$$

The overlap matrix is approximated by

$$S_{mn,\mathbf{k}} = \delta_{mn}$$

so the tight-binding problem becomes

$$\begin{pmatrix} E_p & t_{\mathbf{k}} \\ t_{\mathbf{k}}^* & E_p \end{pmatrix} \begin{pmatrix} c_{1,\mathbf{k}} \\ c_{2,\mathbf{k}} \end{pmatrix} = E_{\mathbf{k}} \begin{pmatrix} c_{1,\mathbf{k}} \\ c_{2,\mathbf{k}} \end{pmatrix}$$

The energy levels are then given by

$$E_{\mathbf{k}}^{\pm} = E_p \pm |t_{\mathbf{k}}| = E_p \pm \sqrt{t_{\mathbf{k}}^* t_{\mathbf{k}}} = E_p \pm t\sqrt{3 + 2\cos\left(2\pi k_1\right) + 2\cos\left(2\pi k_2\right) + 2\cos\left(2\pi(k_1 + k_2)\right)}$$

and the eigenvectors by

$$\boldsymbol{\psi}_{\mathbf{k}}^{+} = \begin{pmatrix} c_{1\mathbf{k}} \\ c_{2\mathbf{k}} \end{pmatrix} = \frac{1}{|t_{\mathbf{k}}|\sqrt{2}} \begin{pmatrix} t_{\mathbf{k}} \\ |t_{\mathbf{k}}| \end{pmatrix} \text{ and } \boldsymbol{\psi}_{\mathbf{k}}^{-} = \frac{1}{|t_{\mathbf{k}}|\sqrt{2}} \begin{pmatrix} -t_{\mathbf{k}} \\ |t_{\mathbf{k}}| \end{pmatrix}$$

Choosing the xy-plane to represent $\mathbf{k} = (k_x, k_y)$, and plotting $E_{\mathbf{k}}^{\pm}$ along the z-axis gives a picture like the one below. It clearly illustrates why the special points K and K' (at the corners of the hexagon) are called the conical points.
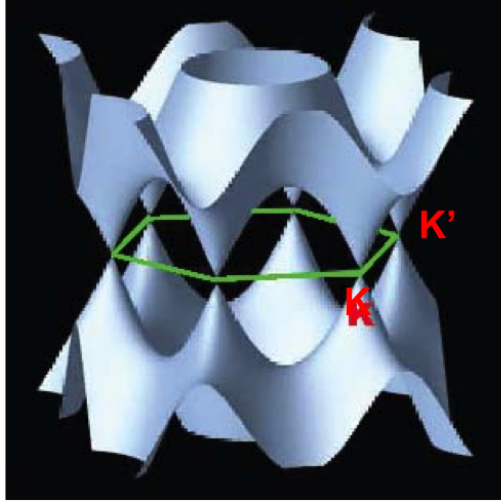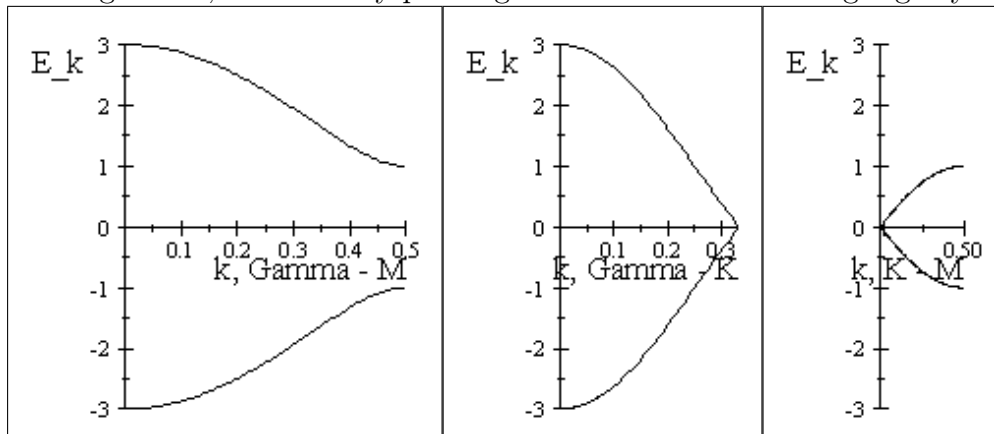
Special points

For the special **k**-points we find the energy levels

$$E_\Gamma^\pm = E_p \pm 3t; \quad E_M^\pm = E_p \pm t; \quad E_K^\pm = E_p; \quad E_{K'}^\pm = E_p$$

showing that we have a degeneracies at the K-point and at the K'-point. In fact, we have band crossings there, as shown by plotting the band structure along high symmetry directions.

The eigenvectors at the special **k**-points are given by

$$\Gamma : \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$M: \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$K, K': \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Long wave length limit

The band structure (the collection of all energy values $E_\mathbf{k}; \mathbf{k} \in 1^{st}$ Brillouin zone) is symmetric in energy around E_p. Each carbon atom has one electron in a p_z state. Assuming spin degeneracy, it means that exactly half of the bands are filled. Therefore the Fermi energy E_F lies exactly at the K and K' points (or at the green hexagon in the picture above). As

transport is determined by electrons with energies close to the Fermi level, it is worth while to look at states around the K,K' points in more detail. Write

$$\mathbf{k} = \mathbf{K} + \mathbf{q}; \mathbf{k}' = \mathbf{K}' + \mathbf{q}'$$

with $|\mathbf{q}| \ll |\mathbf{K}|$, or as $|\mathbf{K}| = K \sim 1/a$, with $qa \ll 1$. In a similar way $q'a \ll 1$. So we are interested in states with $\mathbf{k}$-vectors within circles of radius q centered around K and K', see the figure below. The name "long wave length limit" comes from writing $q = 2\pi/\lambda$, so $\lambda \gg a$.
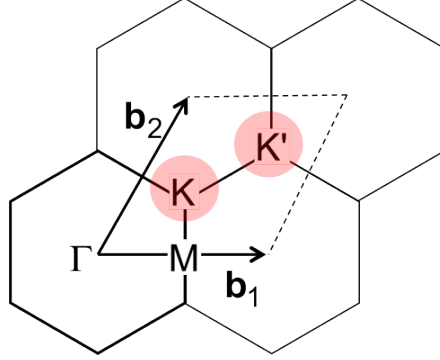
One can write

$$e^{i\mathbf{k}\cdot\mathbf{a}_1} = e^{i\mathbf{K}\cdot\boldsymbol{a}_1}e^{i\mathbf{q}\cdot\boldsymbol{a}_1} \approx e^{i\frac{2\pi}{3}}(1 + i\mathbf{q}\cdot\boldsymbol{a}_1) = e^{i\frac{2\pi}{3}}\left(1 + iq_x a + iq_y a\sqrt{3}\right).$$

The last substitution assumes a specific cartesian coordinate system to represent the vector $\mathbf{q}$, with its x-axis along the direction $\mathbf{a}_1 + \mathbf{a}_2$, see the figure of real space lattice. In the same way one finds

$$e^{-i\mathbf{k}\cdot\mathbf{a}_2} = e^{-i\mathbf{K}\cdot\boldsymbol{a}_2}e^{-i\mathbf{q}\cdot\boldsymbol{a}_2} \approx e^{-i\frac{2\pi}{3}}(1 - i\mathbf{q}\cdot\boldsymbol{a}_2) = e^{-i\frac{2\pi}{3}}\left(1 - iq_x a + iq_y a\sqrt{3}\right).$$

Using this in the hopping matrix element one gets

$$t_{\mathbf{k}} = t\left(1 + e^{i\mathbf{k}\cdot\mathbf{a}_1} + e^{-i\mathbf{k}\cdot\mathbf{a}_2}\right) \approx -\frac{3at}{2}(q_x - iq_y)$$

and in a similar way

$$t_{\mathbf{k}'} = t\left(1 + e^{i\mathbf{k}'\cdot\mathbf{a}_1} + e^{-i\mathbf{k}'\cdot\mathbf{a}_2}\right) \approx \frac{3at}{2}\left(q'_x + iq'_y\right)$$

Below I will do the q case, i.e. the Brillouin zone close to the K-point, the q' case around the K'-point follows from similar algebra.

For reasons that will become clear below, one defines a parameter v_F by

$$\hbar v_F = -\frac{3at}{2}$$

The parameter v_F is called the "Fermi velocity"; it has the dimension of velocity, and as $t < 0$ (in fact $t \sim -3\text{eV}$), v_F is positive; its numerical value is approximately $1 \times 10^6 \text{m/s}$. Choosing the atomic energy E_p as reference energy, i.e. $E_p = 0$, our Hamiltonian becomes

$$\left(\begin{array}{cc} E_p & t_{\mathbf{k}} \\ t_{\mathbf{k}}^* & E_p \end{array} \right) \approx \hbar v_F \left(\begin{array}{cc} 0 & q_x - iq_y \\ q_x + iq_y & 0 \end{array} \right) = \hbar v_F q_x \left(\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right) + \hbar v_F q_y \left(\begin{array}{cc} 0 & -i \\ i & 0 \end{array} \right)$$

The matrices in the final expression correspond mathematically to the Pauli spin matrices, so one can write formally

$$\hbar v_F q_x \hat{\sigma}_x + \hbar v_F q_y \hat{\sigma}_y = \hbar v_F \mathbf{q} \cdot \hat{\boldsymbol{\sigma}}$$

Defining a momentum $\mathbf{p} = \hbar \mathbf{q}$ then gives a very compact expression of the Hamiltonian

$$v_F \mathbf{p} \cdot \hat{\boldsymbol{\sigma}}$$

In this short-hand notation the eigenvalue equation becomes

$$\left(\begin{array}{cc} E_p & t_{\mathbf{k}} \\ t_{\mathbf{k}}^* & E_p \end{array} \right) \left(\begin{array}{c} c_{1\mathbf{k}} \\ c_{2\mathbf{k}} \end{array} \right) = E_{\mathbf{k}} \left(\begin{array}{c} c_{1\mathbf{k}} \\ c_{2\mathbf{k}} \end{array} \right) \Rightarrow (v_F \mathbf{p} \cdot \hat{\boldsymbol{\sigma}}) \boldsymbol{\psi}_{\mathbf{p}} = E \boldsymbol{\psi}_{\mathbf{p}}$$

with the "spinor"

$$\boldsymbol{\psi}_{\mathbf{p}} = \left(\begin{array}{c} c_{1\mathbf{p}} \\ c_{2\mathbf{p}} \end{array} \right)$$

The components of the spinor $\boldsymbol{\psi}_{\mathbf{p}}$ do not refer to amplitudes of spin-up and spin-down, of course, but to the amplitudes of the wave function on the carbon lattices 1 and 2. The equation

$$(v_F \mathbf{p} \cdot \hat{\boldsymbol{\sigma}}) \boldsymbol{\psi}_{\mathbf{p}} = E \boldsymbol{\psi}_{\mathbf{p}} \Rightarrow v_F \left(\begin{array}{cc} 0 & p_x - ip_y \\ p_x + ip_y & 0 \end{array} \right) \left(\begin{array}{c} c_{1\mathbf{p}} \\ c_{2\mathbf{p}} \end{array} \right) = E_{\mathbf{p}} \left(\begin{array}{c} c_{1\mathbf{p}} \\ c_{2\mathbf{p}} \end{array} \right)$$

is easily solved to give the eigenvalues

$$E_{\mathbf{p}}^{\pm} = \pm v_F \sqrt{p_x^2 + p_y^2} = \pm v_F p,$$

which shows that the dispersion, i.e. the relation between E and p, is linear. The energy is independent of the direction of the momentum, so in a plot of the dispersion relation one obtains perfect cones as in the figure below.
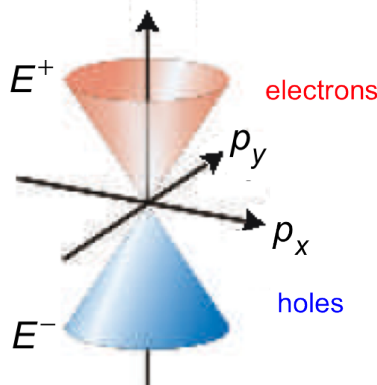
In undoped graphene the Fermi level is exactly at the K-point (and at the K'-point, of course), so only the $p = 0$ state is occupied. One can introduce electrons with $p > 0$ and energy $E^+ = v_F p$ by doping with electron donors, or by using a positive gate voltage.

Alternatively, one can introduce holes with $p > 0$ and energy $-E^- = v_F p$ by doping with electron acceptors, or by using a negative gate voltage.[22] Hole doping should lead to the same properties as electron doping, apart from a sign change of the charge. Such an "electron-hole symmetry" is not observed in conventional semiconductors, where for instance the value of the effective mass of an electron is usually quite different from that of a hole.

The eigenvectors are given by

$$\psi_{\mathbf{p}} = \begin{pmatrix} c_{1\mathbf{p}} \\ c_{2\mathbf{p}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \pm 1 \\ e^{i\varphi_{\mathbf{p}}} \end{pmatrix},$$

where $\varphi_{\mathbf{p}}$ is the phase angle in the complex plane, i.e. $p_x + ip_y = pe^{i\varphi_{\mathbf{p}}}$.

Obviously the long wave length approximation becomes less accurate if you move further away in $\mathbf{k}$-space from the K,K' points, which means that the Fermi level mover further away in energy from zero. The figure below shows the difference between the dispersion relations $E_{\mathbf{k}}$, calculated with the tight-binding model, and $E_{\mathbf{q}}$, the long wave length approximation. For energies within $E_F \pm 0.5$eV the two coincide.

Caption: band structure; $E_{\mathbf{k}} = \pm |t_{\mathbf{k}}|$; $E_{\mathbf{q}} = \pm \hbar v_F q$
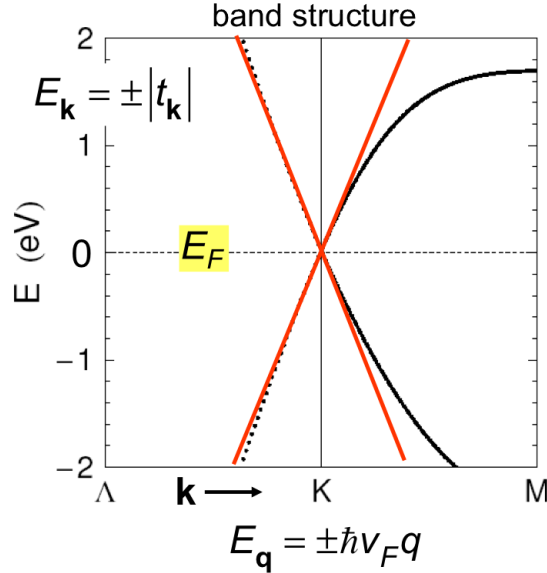
Magnetic field, Landau levels

The long wave length limit is more general that we have assumed up till now. Suppose the amplitudes of the wave function depend upon the position in (2-dimensional) space $\mathbf{r} = x, y$

$$\psi(\mathbf{r}) = \begin{pmatrix} c_1(\mathbf{r}) \\ c_2(\mathbf{r}) \end{pmatrix},$$

One can write down an equation

$$(v_F \hat{\mathbf{p}} \cdot \hat{\boldsymbol{\sigma}}) \psi(\mathbf{r}) = E\psi(\mathbf{r}) \Rightarrow v_F \begin{pmatrix} 0 & \hat{p}_x - i\hat{p}_y \\ \hat{p}_x + i\hat{p}_y & 0 \end{pmatrix} \begin{pmatrix} c_1(\mathbf{r}) \\ c_2(\mathbf{r}) \end{pmatrix} = E \begin{pmatrix} c_1(\mathbf{r}) \\ c_2(\mathbf{r}) \end{pmatrix}$$

assuming the usual quantum mechanical momentum operators $\hat{p}_x = \frac{\hbar}{i} \frac{\partial}{\partial x}$ and $\hat{p}_y = \frac{\hbar}{i} \frac{\partial}{\partial y}$. It is straightforward to show that the solution of this equation has the same eigenvalues $E_{\mathbf{p}}^{\pm}$ as

---

[22]Adding a hole should be the same as removing an electron, hence the "−" in front of E^-.

before. The eigenfunctions are given by a plane wave, typical of a free particle, multiplied with the spinor $\boldsymbol{\psi}_\mathbf{p}$ as before

$$\boldsymbol{\psi}(\mathbf{r}) = e^{\frac{i}{\hbar}\mathbf{p}\cdot\mathbf{r}}\boldsymbol{\psi}_\mathbf{p}$$

Mathematically the eigenvalue equation resembles the Dirac equation in two dimensions

$$\left(c\hat{\mathbf{p}}\cdot\hat{\boldsymbol{\sigma}} + mc^2\hat{\sigma}_z\right)\boldsymbol{\psi}(\mathbf{r}) = E\boldsymbol{\psi}(\mathbf{r}),$$

where c is the speed of light, m is the rest mass, and $\hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. The Dirac equation holds for spin-$\frac{1}{2}$ particles moving at relativistic speeds. In our graphene case we apparently have $m = 0$, as this term is missing. Such particles are called "ultra-relativistic", as according to Dirac they should move at a speed c.[23] One should be careful not to exaggerate the resemblance between graphene and Dirac particles. In graphene the particles move at a speed $v_F \sim 0.003c$. For Dirac particles the components of the spinor refer to amplitudes of spin-up and spin-down, of course, whereas for graphene they refer to amplitudes of the wave function on the carbon lattices 1 and 2. The electron-hole symmetry in graphene translates into Dirac language a a symmetry between matter and anti-matter, i.e. both have the same properties, apart from a sign change of the charge.

Our "Dirac equation" for graphene not only works for free particles, but also in the presence of external potentials, provided these vary on a length scale much larger than the size of the graphene unit cell.[24] In the presence of a magnetic field we replace $\hat{\mathbf{p}}$ by $\hat{\mathbf{p}} + e\mathbf{A}$ as usual

$$\left(v_F\left(\hat{\mathbf{p}} + e\mathbf{A}\right)\cdot\hat{\boldsymbol{\sigma}}\right)\boldsymbol{\psi}(\mathbf{r}) = E\boldsymbol{\psi}(\mathbf{r})$$

To obtain a constant magnetic field along the z-direction $\mathbf{B} = (0,0,B)$ we use $\mathbf{A} = (-yB, 0, 0)$, i.e. the Landau gauge [check $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A}$]. This gives

$$v_F\begin{pmatrix} 0 & \hat{p}_x - i\hat{p}_y - eBy \\ \hat{p}_x + i\hat{p}_y - eBy & 0 \end{pmatrix}\begin{pmatrix} c_1(\mathbf{r}) \\ c_2(\mathbf{r}) \end{pmatrix} = E\begin{pmatrix} c_1(\mathbf{r}) \\ c_2(\mathbf{r}) \end{pmatrix}$$

Writing $c_1(\mathbf{r}) = e^{ikx}\chi_1(y)$ and $c_2(\mathbf{r}) = e^{ikx}\chi_2(y)$ then results in two equations

$$\begin{cases} v_F\left(\hbar k - \hbar\frac{d}{dy} - eBy\right)\chi_2(y) = E\chi_1(y) \\ v_F\left(\hbar k + \hbar\frac{d}{dy} - eBy\right)\chi_1(y) = Eu\chi_2(y) \end{cases}$$

Two equations of type

$$\begin{cases} \hat{Q}\chi_2 = E\chi_1 \\ \hat{R}\chi_1 = E\chi_2 \end{cases}$$

can always be combined into one by

$$\hat{R}\hat{Q}\chi_2 = \hat{R}E\chi_1 = E\hat{R}\chi_1 = E^2\chi_2$$

---

[23]Photons have zero rest mass and move at speed c. However, photons are spin 1 particles. They do not obey the Dirac equation, but the Maxwell equations. Neutrino's have long been thought to be spin $\frac{1}{2}$ particles with zero rest mass.

[24]Proving this statement goes along the same lines as deriving the effective mass model for a normal electron gas.

In this case

$$\hat{R}\hat{Q} = v_F\left(\hbar k + \hbar\frac{d}{dy} - eBy\right)v_F\left(\hbar k - \hbar\frac{d}{dy} - eBy\right) = v_F^2\left(-\hbar^2\frac{d^2}{dy^2} + e^2B^2\left(y - \frac{\hbar k}{eB}\right)^2 - \hbar eB\right)$$

and the eigenvalue equation becomes

$$v_F^2\left(-\hbar^2\frac{d^2}{dy^2} + e^2B^2\left(y - \frac{\hbar k}{eB}\right)^2 - \hbar eB\right)\chi_2(y) = E^2\chi_2(y)$$

If we define parameters $\frac{1}{2m} \equiv v_F^2$, $\omega = 2eBv_F^2$, and $y' = y - y_k$, with $y_k = \frac{\hbar k}{eB}$ then the equation reads as
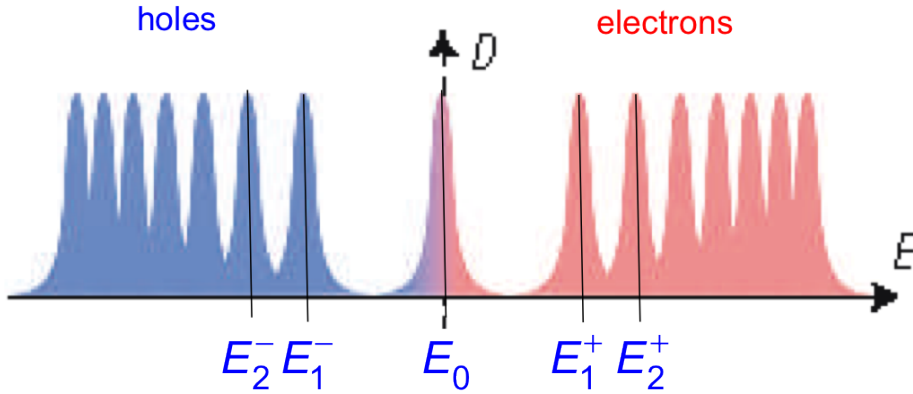
$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dy'^2} + \frac{1}{2}m\omega y'^2 - \frac{1}{2}\hbar\omega\right)u(y') = E^2(y')$$

which is a harmonic oscillator equation, except that the usual E is replaced by E^2 and a constant $C = -\frac{1}{2}\hbar\omega$ is added on the left handside. The solutions are of course

$$E^2 = \left(n + \frac{1}{2}\right)\hbar\omega + C = n\hbar\omega,$$

giving the Landau levels

$$E_n^{\pm} = \pm\sqrt{n\hbar\omega} = \pm v_F\sqrt{n\hbar eB}; \ n = 0, 1, 2, ...$$

Remarkably, the Landau levels depend on the quantum number as $\sqrt{n}$, and also on the magnetic field strength as $\sqrt{B}$, instead of a linear dependence found for a normal electron gas. Moreover, the lowest Landau level is $E_0 = 0$, instead of the usual $\frac{1}{2}\hbar\omega$ in a normal electron gas. Another difference is of course that the energy levels attain negative, as well as positive values. The positive ones are for electrons, and the negative ones for holes, or as Dirac would say, the positive ones are for particles and the negative ones for anti-particles. In undoped graphene the Fermi level $E_F = 0$, i.e., exactly at the lowest Landau level. As the spectrum concerning electrons and holes is symmetric, this must mean that the lowest Landau level in undoped graphene is exactly half-filled, see the figure above.

The wave functions $\chi_{2,n}(y) = u_n(y - y_k)$ are the displaced harmonic oscillator wave functions, with the same displacements y_k as for the normal electron gas. Given $\chi_{2,n}(y)$ one can calculate $\chi_{1,n}(y) = \frac{1}{E}\hat{Q}\chi_{2,n}(y) = u_{n-1}(y - y_k)$, provided $n \geq 1$, and $\chi_{1,n=0}(y) = 0$. It gives

$$\boldsymbol{\psi}_{0,k}(\mathbf{r}) = e^{ikx}\begin{pmatrix} 0 \\ u_0(y - y_k) \end{pmatrix}, \ \boldsymbol{\psi}_{n,k}^{\pm}(\mathbf{r}) = \frac{e^{ikx}}{\sqrt{2}}\begin{pmatrix} \pm u_{n-1}(y - y_k) \\ u_n(y - y_k) \end{pmatrix}, \ n \geq 1$$

where $+$ is for electrons, and $-$ is for holes.

Quantum Hall effect

The quantum Hall effect can be explained in the same way as for a normal electron gas, i.e. by counting the number of Landau levels that give rise to edge modes in a channel, given the Fermi level E_F. Normally one would have in a balistic channel a conductance

$$G = \frac{e^2}{h} \left\{ 2 \left(n_{max} + 1 \right) \right\}$$

where the factor 2 accounts for spin degeneracy, assuming that spin splitting can be neglected, and n_{max} the maximal harmonic oscillator quantum number for which $\left(n_{max} + \frac{1}{2} \right) \hbar \omega < E_F$.
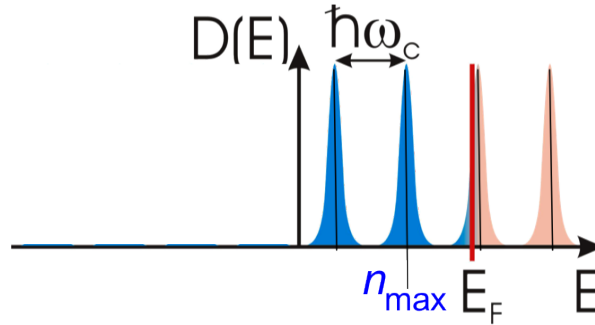
In graphene we must account for the fact that the $n = 0$ Landau level is already half-filled at $E_F = 0$. Only half the states are available for putting in extra electrons if the Fermi level is raised. The other half is available for holes if the Fermi level is lowered. So instead of $\left\{ 2 \left(n_{max} + 1 \right) \right\}$ one should use $\left\{ 2 \left(n_{max} + 1 \right) \right\} - 1$ to avoid double counting of the $n = 0$ level. Moreover, for graphene we have an extra degeneracy, as states with a momentum round K and K' are independent of one another, and both are both filled to the same amount at raising E_F. For graphene one expects to see a conductance

$$G = 2 \frac{e^2}{h} \left\{ 2 \left(n_{max} + 1 \right) - 1 \right\} = \frac{4e^2}{h} \left(n_{max} + \frac{1}{2} \right),$$

which is indeed observed. Of course for graphene the maximal harmonic oscillator quantum number is determined by $\sqrt{n_{max}} \hbar \omega < E_F$.
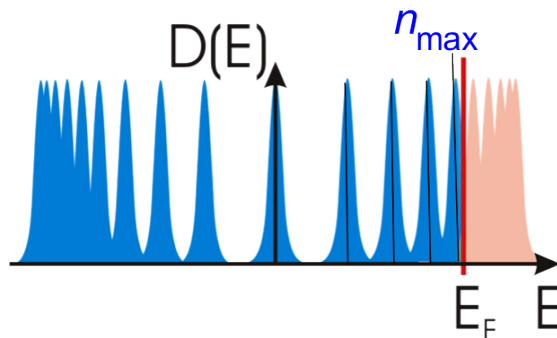

In the famous experiment by Novoselov et al., Nature **438**, 197 (2005), the Hall effect was measured at a fixed field strength ($B = 14$T) at low temperature ($T = 4$K). The density of charge carriers in graphene (effectively the position of E_F) could be varied continuously by sweeping the gate voltage, using graphene in a field-effect transistor (FET) geometry. As

each Landau level can hold the same amount of electrons, the number of occupied Landau levels is simply proportional to the density. The red curve below shows the Hall conductivity σ_{xy} (or G, as I called it above), as a function of the density of electrons in graphene (where the "−" sign stands for holes); a perfect example of the quantum Hall effect in graphene. The green curve shows the longitudinal resistivity ρ_{xx}, which shows beautiful Shubnikov-de Haas oscillations. The resistivity is (close to) zero if the Fermi level is in a gap between two Landau levels. Then the graphene strip used in the FET has a number of well-defined edge states at the Fermi level (equal to the number of occupied Landau levels), with left-going states at one edge, and right-going states at the other edge. The spatial separation of left- and right-going states strongly suppresses scattering between those states; therefore the resistance is very small. In contrast, if the Fermi level coincides with a Landau level, the "edge" states associated with that level, are close to the center of the strip (they aren't edge states anymore). There is no spatial separation between left- and right-going states, which enables scattering, and results in a high resistance.
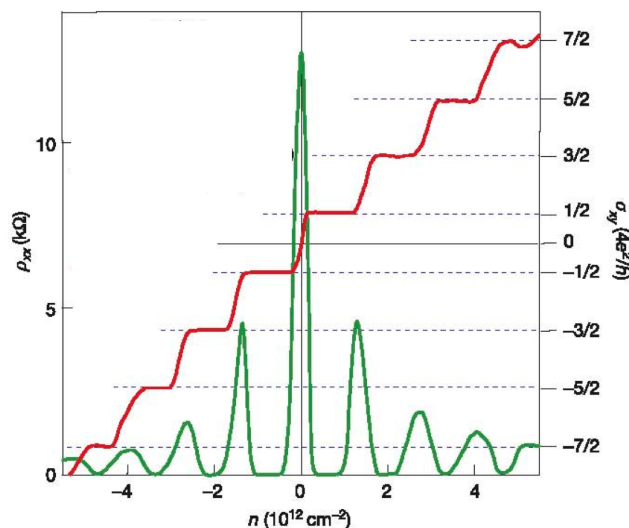
The difference between the quantum Hall effect in a normal electron gas and that in graphene is given schematically in the figure below (taken from Novoselov et al., Nature **2**, 167 (2006)). Figure (a) gives the Hall conductivity σ_{xy} versus the number of occupied Landau levels for a normal electron gas ($g = 2$ in this case), and figure (c) give s the same for graphene ($g = 4$ in this case, because of the degeneracy of the K and K' levels).

26
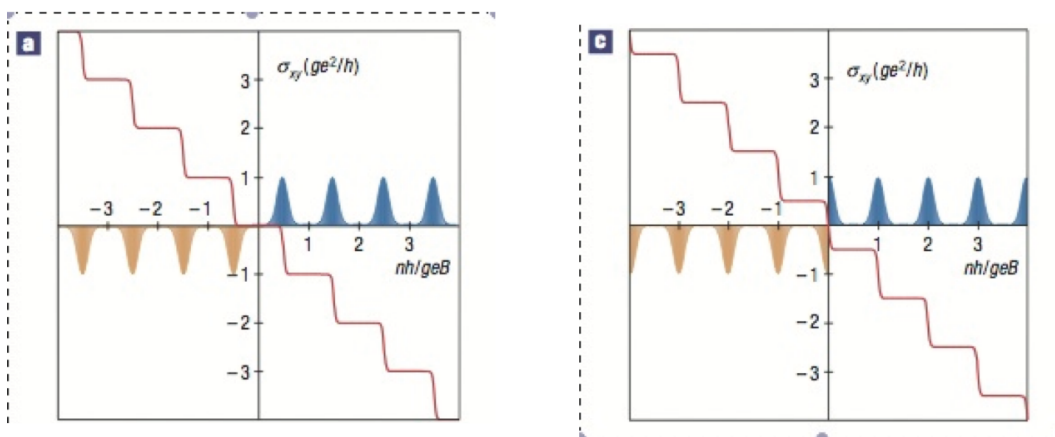
One of the beauties of graphene is that the quantum Hall effect can be observed at room temperature ($T = 300K$), albeit using a large magnetic field ($B = 29T$). The figure below is taken from Novoselov et al., Science **315**, 1379 (2007) (V_g stands for gate voltage). The fact that it is possible to see this quantization at a high temperature results from the high mobility of charge carriers in graphene, and from the peculiar energy level spectrum of graphene (see the exercizes).
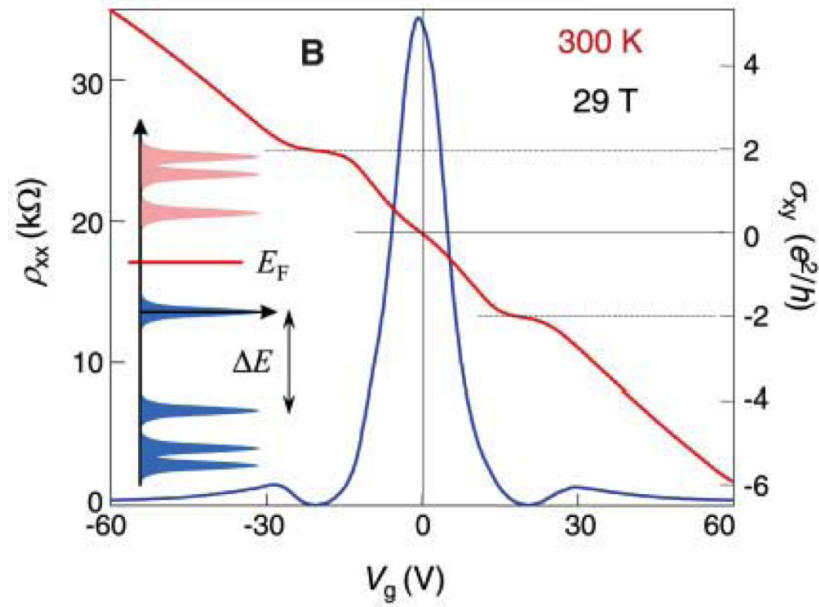
APPENDIX: YOUR OWN RELATIVISTIC QUANTUM HAMIL-TONIAN FOR ONLY $1.99 (batteries not included).

Non-relativistic quantum mechanics

For one particle in one dimension, the time-dependent Schrödinger equation is

$$\left[i\hbar\frac{\partial}{\partial t} - \widehat{H} \right] \psi(x,t) = 0. \tag{1}$$

For a free particle (no potential) we have

$$\widehat{H} = \frac{\widehat{p}^2}{2m} \text{ with } \widehat{p} = -i\hbar\frac{\partial}{\partial x}. \tag{2}$$

The plane wave

$$\psi(x,t) = Ce^{\frac{i}{\hbar}(px-Et)} = Ce^{i(kx-\omega t)}, \tag{3}$$

is a solution to the Schrödinger equation for a free particle, provided

$$E = \frac{p^2}{2m}, \tag{4}$$

as you can easily check for yourself by trying the plane wave expression in the Schrödinger equation. Eq. 4 looks like the relation between (kinetic) energy E and momentum p for a classical free particle. Using Planck $E = \hbar\omega$ and DeBroglie $p = \hbar k$, in wave mechanics one would call it a dispersion relation $\omega(k)$.

Now imagine that you are Schrödinger and you are inventing quantum mechanics. Start from the classical relation eq. 4 and write it as

$$\left[E - \frac{p^2}{2m} \right] = 0. \tag{5}$$

Now comes a leap of imagination. Replace E and p by operators

$$E \rightarrow i\hbar\frac{\partial}{\partial t} \text{ and } p \rightarrow -i\hbar\frac{\partial}{\partial x}, \tag{6}$$

let them operate on a wave function, and rewrite eq. 5 as

$$\left[i\hbar\frac{\partial}{\partial t} + \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} \right] \psi(x,t) = 0, \tag{7}$$

which is of course the Schrödinger equation for a free particle, eq. 1, with the plane wave solution, eq. 3. Note that if (i) you like to have a plane wave describe a free particle (which seems logical, what else would you use to describe a wave in free space), (ii) you want to obey Planck and DeBroglie, and (iii) obey the dispersion relation eq. 4, the wave equation eq. 7 is almost inescapable. That is, the substitutions, eq. 6, seem almost predestined: the time and space derivatives result from (i), the $\hbar$'s come from (ii), the i's come from (iii). One could argue about the "$-$" sign for the momentum operator; that comes from the fact that the plane wave should also be an eigenstate of the momentum operator, i.e.

$$\widehat{p}\psi(x,t) = p\psi(x,t). \tag{8}$$

Again this seems logical: a particle in free space should have a well-defined momentum. Obviously the classical relation we started from, eq. 5, is a non-relativistic one. This means that the quantum mechanics defined by the Schrödinger equation, eq. 7, is non-relativistic.

Relativistic quantum mechanics; 1st attempt

In quantum mechanics, as in classical mechanics, particles with a large momentum become relativistic. Such particles are no longer described by the Schrödinger equation and we have to invent a new equation. One could start from the relativistic classical equivalent to eq. 4, i.e. $E = \sqrt{c^2 p^2 + m^2 c^4}$,[25] or

$$\left[E - \sqrt{c^2 p^2 + m^2 c^4} \right] = 0. \tag{9}$$

Applying the substitutions of eq. 6 then gives the following wave equation

$$\left[i\hbar \frac{\partial}{\partial t} - \sqrt{-c^2 \hbar^2 \frac{\partial^2}{\partial x^2} + m^2 c^4} \right] \psi(x, t) = 0. \tag{10}$$

This is a strange-looking equation because of the square root of an operator. I could try and give it a meaning using a Taylor expansion of the square root, $\sqrt{a + y} = \sqrt{a} + \frac{1}{2\sqrt{a}} y - \frac{1}{6\sqrt[3]{a}} y^2 + ...$, where $y = \partial^2/\partial x^2$, but this makes not less strange, as I have to use an infinite series of differential operators. Putting $\psi(x, t)$ under the square root is not allowed as such, as the wave function has a dimension.[26] Anyway, that would make it into a non-linear differential equation, and as far as we know from experiment, relativistic quantum mechanics is linear, just like non-relativistic quantum mechanics.[27] In conclusion, this attempt of creating relativistic wave mechanics is a failure.

Relativistic quantum mechanics; the Klein-Gordon equation

As obviously the square root in eq. 10 is the key problem, why not rewrite eq. 9 as

$$\left[E^2 - c^2 p^2 - m^2 c^4 \right] = 0. \tag{11}$$

Applying the substitutions of eq. 6 now leads to

$$\left[-\frac{1}{c^2} \frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2} - \frac{m^2 c^2}{\hbar^2} \right] \psi(x, t) = 0. \tag{12}$$

This looks like a decent wave equation; it is called the **Klein-Gordon equation**. For $m = 0$ it becomes the classical wave equation we know from optics (in one dimension, of course). If we try the plane wave, eq. 3, as solution we see that it works, provided E and p obey eq. 11.
So far, so good. There is a catch however. The relation eq. 11 allows for

$$E = \pm\sqrt{c^2 p^2 + m^2 c^4}, \tag{13}$$

i.e., it gives solutions for negative energy, as well as for positive energy, and the spectrum of allowed energies runs from $-\infty$ to ∞. Putting a particle in the ground state, i.e. the lowest energy level, would mean giving it energy $-\infty$, and having many particles with such energy becomes somewhat embarrassing. One could of course throw away the negative energy

---

[25]In popular folklore also called $E = mc^2$, which is not true, unless $p = 0$. At least it is closer to truth than most things in popular folklore.

[26]$|\psi|^2 \, dx$ is a probability, i.e. a dimensionless number, so ψ has the dimension of $\mathrm{m}^{-\frac{1}{2}}$.

[27]Linear means, if ϕ_1 and ϕ_2 are solutions of the differential equation, then also $a\phi_1 + b\phi_2$, for any set of constants a, b.

solutions; after all in physics we constantly throw away unwanted mathematical solutions as non-physical (functions that diverge, cannot be normalized, etc.), but throwing away half the spectrum only for the reason that these solutions are embarrassing, does not seem to be proper physics.

Because of the problem with the negative energy solutions, the Klein-Gordon was set aside for some time. By now we have found an interpretation for the negative energy levels: they belong to anti-particles, whereas the positive energy levels belong to particles. Think of a semi-conductor in solid state physics and put the zero of energy at the Fermi level inside the band gap. Electrons in the conduction band are in positive energy levels; they are the particles. Holes in the valence band are in negative energy levels; they are the anti-particles. This interpretation saves the Klein-Gordon equation.[28] However, it turns out to be an equation that is only valid for particles with spin zero.

Relativistic quantum mechanics; the Dirac equation

In the days of Klein and Gordon the negative energy problem was unsolved however. The problem is obviously in the quadratic relation of eq. 11, which gives two solutions, eq. 13. Dirac thought it may be possible to factorize eq. 11

$$\left[E^2 - c^2 p^2 - m^2 c^4\right] = \left[E + \alpha cp + \beta mc^2\right]\left[E - \alpha cp - \beta mc^2\right] = 0. \tag{14}$$

Then take only the last factor on the righthand side, and apply the substitutions of eq. 6

$$\left[i\hbar\frac{\partial}{\partial t} - \alpha c\left(-i\hbar\frac{\partial}{\partial x}\right) - \beta mc^2\right]\psi(x,t) = 0. \tag{15}$$

This equation is first order in time, like the Schrödinger equation. There is something odd about the factor α and β, however. For the factorization of eq. 14 to work, they have to obey

$$\alpha^2 = 1; \ \beta^2 = 1; \ \alpha\beta = -\beta\alpha. \tag{16}$$

It will be clear that the last relation cannot be obeyed if α and β are simple numbers. However, eq. 16 can be obeyed by matrices. The simplest solution is to take 2×2 matrices, for instance the Pauli matrices

$$\alpha = \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \ \beta = \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}; \tag{17}$$

The wave function in eq. 15 then becomes a vector with two elements, called a **spinor**.

$$\left[i\hbar\frac{\partial}{\partial t} - \sigma_x c\left(-i\hbar\frac{\partial}{\partial x}\right) - \sigma_z mc^2\right]\begin{pmatrix} \psi_1(x,t) \\ \psi_2(x,t) \end{pmatrix} = 0, \tag{18}$$

which gives the two coupled equations

$$\begin{cases} \left[i\hbar\frac{\partial}{\partial t} - mc^2\right]\psi_1(x,t) + ic\hbar\frac{\partial}{\partial x}\psi_2(x,t) = 0 \\ \left[i\hbar\frac{\partial}{\partial t} + mc^2\right]\psi_2(x,t) + ic\hbar\frac{\partial}{\partial x}\psi_1(x,t) = 0 \end{cases}. \tag{19}$$

---

[28]There is more to it than this, however. The Klein-Gordon equation does not describe a single particle, but many particles (and anti-particles), and $\psi(x,t)$ is not a wave function, but a field operator. Field operaters create or annihilate particles, just like the step-up/step-down operators in the quantum harmonic oscillator. Field operators form the basis of quantum field theory, where particles and anti-particles are described as excitations from the vacuum. Such is the fate of all relativistic quantum mechanics; the number of particles is not conserved. At these high energies there is always the possibilty of creating (and annihilating) particles.

Eq. 18 is called the **Dirac equation** (for a free particle in one dimension). By now you will have gathered that this equation is for spin-$\frac{1}{2}$ particles, such as electrons.[29] It is easy to check that the plane wave solutions to eq. 19 are

$$\begin{pmatrix} \psi_1(x,t) \\ \psi_2(x,t) \end{pmatrix} = C \begin{pmatrix} cp \\ E - mc^2 \end{pmatrix} e^{\frac{i}{\hbar}(px - Et)}, \tag{20}$$

provided $E^2 = c^2p^2 + m^2c^4$. It seems that the Dirac equation has negative energy solutions, just like the Klein-Gordon equation. Dirac was somewhat disappointed because of this, but luckily it did not deter him from publishing his idea, and from our particle-antiparticle interpretation (inspired by semiconductor physics), we know how to deal with the problem.[30]

Relativistic quantum mechanics; the Dirac equation in more than one dimension

So far we have found the Dirac equation for a free particle in one dimension, eq. 18

$$\left[i\hbar \frac{\partial}{\partial t} - c\sigma_x \widehat{p}_x - \sigma_z mc^2 \right] \boldsymbol{\psi}(x,t) = 0, \tag{21}$$

with $\boldsymbol{\psi}(x,t)$ a spinor. This has an obvious generalization for two dimensions

$$\left[i\hbar \frac{\partial}{\partial t} - c\left(\sigma_x \widehat{p}_x + \sigma_y \widehat{p}_y\right) - \sigma_z mc^2 \right] \boldsymbol{\psi}(x,y,t) = 0. \tag{22}$$

Now replace c by v_F, put $m = 0$, and try a solution of type $\boldsymbol{\psi}(x,y,t) = \boldsymbol{\phi}(x,y)e^{-\frac{i}{\hbar}Et}$. Then the equation becomes the one we derived for graphene in the long wave length approximation!! Because $m = 0$, we have $E = pc$, particles that move at light speed. Because of this mathematical similarity, the electrons and holes in graphene are called *ultra-relativistic* fermions, i.e. they obey an equation that is mathematically identical to the Dirac equation for massless particles.[31]

Note that in two dimensions we have ran out of Pauli matrices, i.e., we have used alle three $\sigma_x, \sigma_y, \sigma_z$. So in order to generalize the Dirac equation to three dimensions (arguably the most relevant case, as we live in three dimensions), one has to come up with something new. For the factorization of eq. 14 to work, one has to find four matrices $\alpha_x, \alpha_y, \alpha_z, \beta$ with properties that generalize eq. 16,

$$\alpha_i^2 = 1; \ \beta^2 = 1; \ \alpha_i\alpha_j = -\alpha_j\alpha_i; \ j \neq i; \ \alpha_i\beta = -\beta\alpha_i; \ i = x,y,z; \ j = x,y,z. \tag{23}$$

A bit of pondering shows that the simplest set of four matrices that have these properties are 4×4 matrices. I spare you their form here, as they are not relevant to graphene.[32]

---

[29]By now you will also have gathered that the relativistic wave equations for elementary particles depend on their spin; the Klein-Gordon equation for spin-0 particles, the Dirac equation for spin-$\frac{1}{2}$ particles. The one for spin-1 particles is called the Proca equation. The version for spin-1 particles with zero mass (i.e. $m = 0$) you know already. These are photons, so the relevant equations are the Maxwell equations. Interpreted as quantum field equations, the Maxwell equations remain valid, and, as you know well, they are relativistic.

[30]But, like in other relativistic wave equations, the spinors do not consist of simple wave functions, but of field operators that allow for the creation and annihilation of particles and anti-particles.

[31]As their speed is v_F instead of c, they are not really relativistic, of course.

[32]Pick up any advanced quantum mechanics book, and you will find them, but it will cost you more than $1.99. By the way, the "spinor" now also becomes a vector with 4 elements.

So far we have discussed free particles. But like free lunches, free particles are rare. To guess what the proper equation is for relativistic particles is in a potential, again we can learn from the (non-relativistic) Schrödinger equation. One modifies the latter from

$$\left[i\hbar \frac{\partial}{\partial t} - \frac{1}{2m} |\widehat{\mathbf{p}}|^2 \right] \psi(\mathbf{r}, t) = 0, \tag{24}$$

for a free electron to

$$\left[i\hbar \frac{\partial}{\partial t} - \frac{1}{2m} |\widehat{\mathbf{p}} + e\mathbf{A}(\mathbf{r}, t)|^2 + e\phi(\mathbf{r}, t) \right] \psi(\mathbf{r}, t) = 0, \tag{25}$$

for an electron in a scalar potential $\phi(\mathbf{r}, t)$ and a vector potential $\mathbf{A}(\mathbf{r}, t)$. Similarly, the Dirac equation of eq. 22 becomes

$$\left[i\hbar \frac{\partial}{\partial t} - c \left(\sigma_x \left(\widehat{p}_x + eA_x(x, y, t) \right) + \sigma_y \left(\widehat{p}_y + eA_y(x, y, t) \right) \right) - \sigma_z mc^2 + e\phi(x, y, t) \right] \boldsymbol{\psi}(x, y, t) = 0. \tag{26}$$

This concludes my little introduction to relativistic quantum mechanics.[33]

---

[33]As you would expect, I will end with a footnote. In quantum field theory, $\widehat{\boldsymbol{\psi}}(\mathbf{r}, t)$ become field operators that deal with the creation/annihilation of electrons, and $\widehat{\phi}(\mathbf{r}, t)$, $\widehat{\mathbf{A}}(\mathbf{r}, t)$ field operators that deal with the creation/annihilation of photons. The latter are described by Maxwell's equations.