SPECIFIC HEAT OF SOLIDS

As far back as 1819 it was discovered that for many solids the heat capacity per atom is

$$C = 3k_B, \quad k_B = 1.38 \cdot 10^{-23}\,\frac{\mathrm{J}}{\mathrm{K}}$$

This is true for many monoatomic solids, however, there are a few exceptions, such as diamond. This was later explained quite accurately, by Stefan Boltzmann (in 1896), using a classical model, treating each atom as a harmonic oscillator. However, this law could not sufficiently describe the low temperature behaviour of solids.

Einstein Model of solids

Einstein's idea was similar to that of Boltzmann, however, he treated the harmonic oscillators quantum mechanically. The n-th eigenenergy of the harmonic oscillator is

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right)$$

where ω is the eigenfrequency of the harmonic oscillator, sometimes referred to as the Einstein frequency. Using this we can define the partition function:

$$Z = \sum_n e^{-\beta\hbar\omega\left(n+\frac{1}{2}\right)} = \frac{e^{-\beta\hbar\omega/2}}{1 - e^{-\beta\hbar\omega}} = \frac{1}{2\sinh\left(\beta\hbar\omega/2\right)}$$

From here we can define most thermodynamic quantities, such as, for example the average energy

$$\langle E\rangle = \hbar\omega\left(n_B(\beta\hbar\omega) + \frac{1}{2}\right)$$

where the Bose-Einstein distribution is given by

$$n_B(x) = \frac{1}{e^x - 1}$$

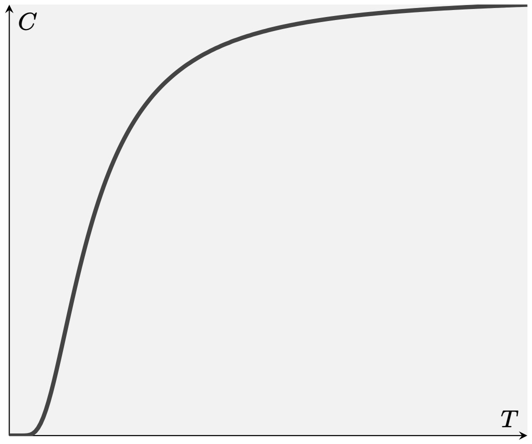
This is easy to generalise to three dimensions ($n + \frac{1}{2} \rightsquigarrow (n_x + \frac{1}{2}) + (n_y + \frac{1}{2}) + (n_z + \frac{1}{2})$), using the rules for the exponential function:

$$Z_{3\mathrm{D}} = \left(Z_{1\mathrm{D}}\right)^3$$

The derivative of the average energy with respect to temperature gives us the heat capacity:

$$C = 3k_B\left(\beta\hbar\omega\right)^2 \frac{e^{\beta\hbar\omega}}{\left(e^{\beta\hbar\omega} - 1\right)^2}$$

This means the heat capacity is effectively constant for high temperatures, compared to the Einstein temperature $T_{\mathrm{Einstein}} \equiv \frac{\hbar\omega}{k_B}$, and decays exponentially for low temperatures:



Figur 1: *Heat capacity as a function of T.*

Note that we recover $C = 3k_B$ for high temperatures, and that the vibrational degree of freedom "freezes out"at low temperatures. For most materials the Einstein temperature is low compared to room temperature, however, for diamond $T_{\mathrm{Einstein}} \approx 1320\mathrm{K}$, which is why $C < 3k_B$ per atom for Diamond. This is due to the strong atomic bonds between carbon atoms.

Debye Model of solids

Einstein's result did quite well, however, only qualitatively in the low temperature region. It was shown experimentally that there would be a T^3 dependence for low temperatures, which is not at all what Einstein's calculations predicted.

Debye realised that the vibrations inside solids are waves, and that waves should be quantised similarly to how Planck quantised light in 1900. However, there are a few differences between light and sound. Naturally the speed of sound is far lower than the speed of light, but additionally light only has two polarisations for any given $\boldsymbol{k}$, whereas sound has three vibrational modes (two transversal and one longitudinal).

More often than not we will look at waves using *periodic boundary conditions.* For a sample length of L this restricts the possible values of k to

$$k = \frac{2\pi n}{L}, \quad n \in \mathbb{Z}$$

If we are required to sum over all k values, we can instead integrate as follows:

$$\sum_k \rightsquigarrow \frac{L}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k$$

Generally though, in three dimensions this looks quite similar:

$$\boldsymbol{k} = \frac{2\pi}{L}\begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}, \quad n_1, n_2, n_3 \in \mathbb{Z}$$

and the summation can also be replaced by a volumetric integral in k-space:

$$\sum_{\boldsymbol{k}} \rightsquigarrow \left(\frac{L}{2\pi}\right)^3 \int \mathrm{d}\boldsymbol{k}$$

It's often beneficial to use spherical coordinates to carry out this integral.

Debye *assumed*[*] a linear dispersion relation:

$$\omega(\boldsymbol{k}) = v\|\boldsymbol{k}\|$$

Now we would like to sum over all values of $\boldsymbol{k}$:

$$\begin{aligned} \langle E\rangle &= 3\sum_{\boldsymbol{k}} \hbar\omega(\boldsymbol{k})\left(n_B(\beta\hbar\omega(\boldsymbol{k})) + \frac{1}{2}\right) \\ &= \frac{3L^3}{(2\pi)^3}\int \mathrm{d}\boldsymbol{k}\ \hbar\omega(\boldsymbol{k})\left(n_B(\beta\hbar\omega(\boldsymbol{k})) + \frac{1}{2}\right) \end{aligned}$$

Using spherical symmetry we convert this into a spherical integral, which we additionally can simplify, by using $k = \frac{\omega}{v}$, to

$$\langle E\rangle = \int_0^\infty \mathrm{d}\omega\ g(\omega)(\hbar\omega)\left(n_B(\beta\hbar\omega) + \frac{1}{2}\right)$$

where we have taken everything from the integral that is not $\varepsilon(\omega)$ and combined it into the **density of states**:

$$g(\omega) = \frac{9N\omega^2}{\omega_{\mathrm{D}}^3}, \quad \omega_{\mathrm{D}} = \left(6\pi^2 n\right)^{1/3} v$$

where ω_{D} is known as the Debye frequency. The term with $\frac{1}{2}$ integrates up to a temperature independent constant[†], therefore it does not contribute to the heat capacity. We get

$$\langle E\rangle = \frac{9N\hbar}{\omega_{\mathrm{D}}^3(\beta\hbar)^4}\int_0^\infty \mathrm{d}x\ \frac{x^3}{e^x - 1} + C$$

where C is the temperature independent constant. The integral is just a constant, so we get

$$\langle E\rangle = \frac{9N\hbar}{\omega_D^3(\beta\hbar)^4}\frac{\pi^4}{15} + C$$

This gives us an expression for the heat capacity:

$$C = \frac{\partial\langle E\rangle}{\partial T} = Nk_B\frac{T^3}{T_{\mathrm{D}}^3}\frac{12\pi^4}{5}, \quad T_{\mathrm{D}} \equiv \frac{\hbar\omega_D}{k_B}$$

This gives us the T^3 dependence, as expected. Unfortunately this expression is proportional to T^3 for all temperatures, however, this is because it is not reasonable to assume that all sound modes are activated (it is certainly not the case that we have an infinite number of sound modes, all with increasing energy). Debye fixed this problem by introducing a cutoff frequency:

$$3N = \int_0^{\omega_{\text{cutoff}}} \mathrm{d}\omega g(\omega)$$

Using this expression for the cutoff frequency, we can set an upper limit for the frequency in our integral for the average energy. At high temperatures the Bose-Einstein distribution becomes $\approx \frac{k_B T}{\hbar\omega}$, so we get

$$\langle E\rangle = k_B T\int_0^{\omega_{\text{cutoff}}} \mathrm{d}\omega g(\omega) = 3Nk_B T$$

which is exactly Dulong-Petit. Evaluating the cutoff frequency:

$$3N = \int_0^{\omega_{\text{cutoff}}} \mathrm{d}\omega g(\omega) = 9N\int_0^{\omega_{\text{cutoff}}} \frac{\omega^2}{\omega_{\mathrm{D}}^3} = 3N\frac{\omega_{\text{cutoff}}^3}{\omega_{\mathrm{D}}^3}$$

hence the cutoff frequency is exactly the Debye frequency. Interestingly enough $k_{\mathrm{D}} = \frac{\omega_D}{v} = \left(6\pi^2 n\right)^{\frac{1}{3}}$ is of the order of inverse interatomic distance.

Debye's theory has a few shortcomings:

1. The cutoff frequency is more of a hack than good physics.
2. The linear dispersion is not accurate, especially for large value of k.
3. Debye theory is not accurate at intermediate temperatures.
4. This model cannot explain the linear temperature dependent term that appears in metals in the low temperature limit.

VIBRATIONS OF A ONE-DIMENSIONAL MONATOMIC CHAIN

Consider a harmonic monatomic chain with spring constant κ and lattice constant a. Denote the position of

---

[*] This assumption turns out to be accurate for long wavelengths, however not for shorter wavelengths, as we will see.

[†] Which unfortunately is infinite, but never mind that

the n-th atom by x_n and its equilibrium position as x_n^{eq}. Using this we can define the n-th atom's displacement

$$\delta x_n = x_n - x_n^{\text{eq}}$$

The total potential is a sum over the individual harmonic oscillators:

$$V_{\text{tot}} = \sum_i V(x_{i+1} - x_i) = \sum_i \frac{\kappa}{2} \left(\delta x_{i+1} - \delta x_i\right)^2$$

There are only two terms that depend on δx_n, so luckily the force $F = -\nabla V = m\delta\ddot{x}_n$ is evaluated quite easily:

$$m\delta\ddot{x}_n = \kappa \left(\delta x_{n+1} + \delta x_{n-1} - 2\delta x_n\right)$$

We will now use the Ansatz, that the oscillations will be of the form:

$$\delta x_n = Ae^{i(\omega t - kx_n^{\text{eq}})} = Ae^{i(\omega t - kna)}$$

where I have used that the spacing between atoms is a and assumed that $x_0^{\text{eq}} = 0$. This gives us

$$-m\omega^2 Ae^{i(\omega t - kna)} = \kappa Ae^{i(\omega t - ikna)} \left(e^{-ika} + e^{ika} - 2\right)$$

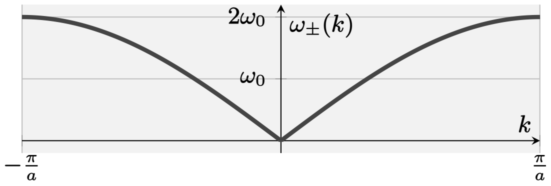
hence

$$\omega^2 = 2\omega_0^2 \left(1 - \cos(ka)\right) = 4\omega_0^2 \sin^2\left(\frac{ka}{2}\right), \quad \omega_0^2 = \frac{\kappa}{m}$$

Thus we obtain

$$\omega = 2\omega_0 \left|\sin\left(\frac{ka}{2}\right)\right|$$

which looks like this



Figur 2: *The first Brillouin zone for a monatomic harmonic chain.*

Notice that $\omega(k)$ is periodic in k, so there is no difference between $\omega(k)$ and $\omega(k + 2\pi/a)$. This is actually far more general than it appears: there is no physical difference between modes with k and $k + 2\pi/a$, hence we can look at the Brillouin zone alone, which is defined as the values of $|k|$ that are not equivalent to a smaller value of $|k|$. This is related to residue classes, as

$$k \equiv k + G \mod G, \quad G = \frac{2\pi n}{a}, \quad n \in \mathbb{Z}$$

hence we just need to look at all different $[k]$ – the Brillouin zone.

This leads us to a very general and important result:

> If a lattice is periodic in real space with periodicity a, then the reciprocal lattice is periodic, with periodicity $2\pi/a$.

For example this implies that if the unit cell has width a then the Brillouin zone with have width $2\pi/a$.

Given a direct lattice with lattice points denoted by x_n we can find the reciprocal lattice points, G_m by

$$e^{iG_m x_n} = 1$$

Properties of the one-dimensional chain

The velocity of wave packets if defined by the *group velocity*, which follows from Hamiltonian mechanics:

$$v_{\text{group}} \equiv \frac{\mathrm{d}\omega}{\mathrm{d}k}$$

additionally we can define the *phase velocity* as

$$v_{\text{phase}} \equiv \frac{\omega}{k}$$

which is the velocity at which individual maxima and minima. Using our dispersion relation we see that long wavelength waves (small values of k) have an approximately constant group velocity (these are sound waves). At higher values of k the group velocity decreases, and it reaches zero at the Brillouin zone boundary[‡].

The total number of normal modes is the size of the Brillouin zone $(2\pi/a)$ divided by the spacing between neighbouring k values $(2\pi/(Na))$, hence there are N normal modes, which is precisely what Debye *assumed* previously.

Quantum Modes: Phonons

Let us introduce the quantum correspondence principle:

[‡] this holds generally

> If a classical harmonic system has a normal mode with frequency ω then the corresponding quantum system will have eigenstates with energy
>
> $$E_n = \hbar\omega\left(n + \frac{1}{2}\right)$$

This fits very well with what found in QM with the quantum harmonic oscillator, however, this also holds for coupled harmonic oscillators. This leads us to define the **phonon**: a quantum of vibration. There is no constraint to how many phonons can occupy the same energy simultaneously, hence phonons are bosons. Therefore the occupancy of any mode is described with the Bose-Einstein distribution:

$$n_B(\beta\hbar\omega) = \frac{1}{e^{\beta\hbar\omega} - 1}$$

and hence the expectation value of the energy is

$$\langle E_k \rangle = \hbar\omega(k)\left(n_B(\beta\hbar\omega) + \frac{1}{2}\right)$$

The total energy is given by the sum of these individual terms, which can be evaluated by an integral instead, as discussed previously. We get

$$U = \frac{Na}{2\pi}\int_{-\pi/a}^{\pi/a} \mathrm{d}k\hbar\omega(k)\left(n_B(\beta\hbar\omega) + \frac{1}{2}\right)$$

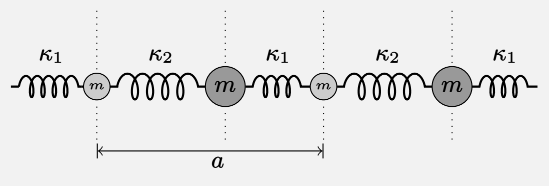
using a change in variables this integral is written as an integral with respect to ω, where we need to include the density of states over frequency, as we did previously. However now we can see that

$$g(\omega) = \frac{Na}{\pi}\left|\frac{\mathrm{d}k}{\mathrm{d}\omega}\right|$$

where the extra factor of two is because each value of ω corresponds to two values of k (one positive and one negative).

VIBRATIONS OF A ONE-DIMENSIONAL DIATOMIC CHAIN

Let us now instead consider a one-dimensional diatomic chain. They will have equal masses, say m, and two different spring constants, say κ_1 and κ_2. As in the following figure



We could from here define a *basis*, which is where each atom is with respect to the unit cell. For example the basis for the diatomic chain in the figure above would be

> light gray atom at the position of the lattice point
>
> dark gray atom $\frac{17}{30}a$ to the right of the lattice point.

Normal modes of the diatomic chain

The calculation for a diatomic chain is very similar to that of the monoatomic chain, however, now we have *two* displacements we need to keep track of. Let us denote the displacement of the light gray atoms by δx_n and the dark gray atoms by δy_n. We would then get

$$\begin{aligned} m\delta\ddot{x}_n &= \kappa_2\left(\delta y_n - \delta x_n\right) + \kappa_1\left(\delta y_{n-1} - \delta x_n\right) \\ m\delta\ddot{y}_n &= \kappa_1\left(\delta x_{n+1} - \delta y_n\right) + \kappa_2\left(\delta x_n - \delta y_n\right) \end{aligned}$$

Once again assuming planar waves we get that

$$\begin{aligned} \delta x_n &= \delta x_{n+1} e^{ika} \\ \delta y_n &= \delta y_{n-1} e^{-ika} \end{aligned}$$

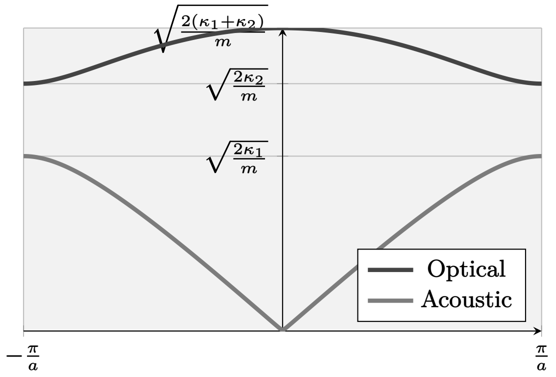
and also that $\delta\ddot{x}_n = \omega^2 \delta x_n$, thus

$$m\omega^2 \begin{pmatrix} \delta x_n \\ \delta y_n \end{pmatrix} = \begin{pmatrix} \kappa_1 + \kappa_2 & -\kappa_2 - \kappa_1 e^{ika} \\ -\kappa_2 - \kappa_1 e^{-ika} & \kappa_1 + \kappa_2 \end{pmatrix} \begin{pmatrix} \delta x_n \\ \delta y_n \end{pmatrix}$$

The eigenvalue equation gives us

$$\omega_\pm = \sqrt{\frac{\kappa_1 + \kappa_2}{m} \pm \frac{1}{m}\sqrt{\kappa_1^2 + \kappa_2^2 + 2\kappa_1\kappa_2\cos(ka)}}$$

Which looks like



Figur 3: *Dispersion relation for the diatomic chain in the reduced scheme.*

For $k \to 0$ on the acoustic mode we have the sound waves, for which there is a linear dispersion. We can define the speed of sound in this limit:

$$v_{\text{sound}} = \frac{\mathrm{d}\omega_-}{\mathrm{d}k} = \sqrt{\frac{a^2\kappa_1\kappa_2}{2m(\kappa_1+\kappa_2)}}$$

ELECTRONS IN METALS

Drude theory

The first model of electrons discussed in the course is the Drude theory, which makes the following assumptions:

1. Electrons have a scattering time τ, which corresponds to the average time between collisions. The probability of scattering in the time interval $\mathrm{d}t$ is $\mathrm{d}t/\tau$.
2. Once a scattering event occurs we assume the momentum returns to $\boldsymbol{p} = \boldsymbol{0}$. This is not true for individual electrons, but is the case for the expectation value of $\boldsymbol{p}$.
3. In between scattering events electrons respond to external electric and magnetic fields as per the Lorentz force.

Using this we can derive a differential equation that relates the momentum to the applied electric and magnetic fields – the Drude equation:

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = -e\left(\boldsymbol{E} + \boldsymbol{v}\times\boldsymbol{B}\right) - \frac{\boldsymbol{p}}{\tau}$$

At equilibrium we get that

$$E = \left(\frac{\boldsymbol{j}\times\boldsymbol{B}}{ne} + \frac{m}{ne^2\tau}\boldsymbol{j}\right)$$

where n is the electron density and we have used that $\boldsymbol{j} = -ne\boldsymbol{v}$. Using this we can define the **resistivity tensor**, ρ_{ij}, whose "diagonal"entries are the Drude resistivity:

$$\rho_{ii} = \frac{m}{ne^2\tau}$$

For $\boldsymbol{B} \parallel \hat{\boldsymbol{z}}$ we get

$$\rho_{xy} = -\rho_{yx} = \frac{B}{ne}$$

the Hall resistivity. The Hall coefficient R_H is defined as

$$R_H = \frac{\rho_{yx}}{|\boldsymbol{B}|} = -\frac{1}{ne}$$

where the final equality is the result from Drude theory. We can define the **conductivity tensor**, σ_{ij} as the inverse of ρ_{ij}, hence

$$j_\alpha = \sigma_{\alpha\beta}E_\beta, \quad \text{RISO}$$

RISO means repeated indices are summed over.

Thermal transport

The Drude theory of electrons, together with the kinetic theory of gas can help explain the Wiedemann-Franz law[§]. κ is the thermal conductivity, which is used in the diffusion equation. The ratio $\frac{\kappa}{T\sigma} \equiv L$ is known as the Lorenz number, and the two theories together predict that

$$L = \frac{\kappa}{T\sigma} = \frac{3}{2}\left(\frac{k_B}{e}\right)^2 \approx 1.11\cdot 10^{-8}\mathrm{W}\Omega\,\mathrm{K}^{-2}$$

which is very close (up to about a factor two) to the experimentally observed values for various metals.

However, this theory is wrong, but we get approximately the right results due to two mistakes that luckily cancel. We have vastly underestimated the velocity of electrons in metals, and vastly overestimated the specific heat of electrons in metals (both of which were used to calculate the heat conductance). In conclusion: we cannot use classical gas theory to describe electrons in a metal.

This can be seen when looking at the Seebeck and Peltier effects. The Peltier effect states that electric currents also cause heat currents:

$$\boldsymbol{j}^q = \Pi\boldsymbol{j}$$

where Π is the Peltier coefficient. In kinetic theory of gasses the thermal current is

$$\boldsymbol{j}^q = \frac{1}{3}c_V T n\boldsymbol{v}$$

where c_V is the specific heat. This leads to a Seebeck coefficient of

$$S = \frac{\Pi}{T} = \frac{-k_B}{2e} = -0.43\cdot 10^{-4}\mathrm{V}\ \mathrm{K}^{-1}$$

[§] Wiedemann and Franz found that the ratio $\frac{\kappa}{T\sigma}$ is approximately constant for most metals.

which is off by about a factor of 100 for most metals, and in fact some metals have a Seebeck coefficient that has the opposite sign.

Sommerfeld theory

Sommerfeld theory incorporates Fermi-Dirac statistics, into the description of electrons in metals. Given a system of non-interacting electrons with chemical potential, μ, the probability of an eigenstate with energy E being occupied is given by the Fermi-Dirac distribution[¶]:

$$n_F(\beta(E-\mu)) = \frac{1}{e^{\beta(E-\mu)}+1}$$

Now consider a box with volume $V = L^3$ with periodic boundary conditions[**]. The eigenstates will be planar waves, with quantised wave vectors:

$$\varepsilon(\boldsymbol{k}) = \frac{\hbar^2\|\boldsymbol{k}\|^2}{2m}, \quad \boldsymbol{k} = \frac{2\pi}{L}\begin{pmatrix} n \\ m \\ \ell \end{pmatrix}, \quad n, m, \ell \in \mathbb{Z}$$

At $T = 0$ the Fermi-Dirac distribution is a reversed Heaviside function, which flips at the Fermi energy:

$$n_F(\beta(E-\mu)) \overset{T\to 0}{\longrightarrow} \Theta(E_F - \varepsilon(\boldsymbol{k}))$$

Where the Fermi energy is defined as $E_F = \mu(T=0)$. The Heaviside function leads to a natural definition of the Fermi wavevector:

$$E_F - \varepsilon(\boldsymbol{k}) = 0 \rightsquigarrow E_F = \frac{\hbar^2 k_F^2}{2m}$$

At $T = 0$ we can integrate n_F with respect to $\boldsymbol{k}$ to get the total number of electrons, and use the spherical symmetry to get

$$N = \frac{V}{4\pi^3}\left(\frac{4}{3}\pi k_F^3\right) \rightsquigarrow E_F = \frac{\hbar^2\left(3\pi^2 n\right)^{\frac{2}{3}}}{2m}$$

Using the relation between ε and k we can change any integral over k into an integral over ε, for example

$$E_{\text{total}} = V\int_0^\infty \mathrm{d}\varepsilon \;\; \varepsilon g(\varepsilon) n_F(\beta(\varepsilon-\mu))$$

where $g(\varepsilon)$ is the density of states per unit volume, over energy. Unfortunately it is very difficult to calculate this integral for non-zero temperatures, due to the Fermi-Dirac distribution. The **Sommerfeld expansion** uses the fact that the integral is easily calculated at $T = 0$ and Taylor expands the integral from there, giving

$$E(T) = E(T=0) + \frac{\tilde{\gamma}}{2} V g(E_F)(k_B T)^2 + \cdots$$

where $\tilde{\gamma} = \frac{\pi^2}{3}$. Which gives us a linear heat capacity

$$C = \tilde{\gamma}\left(\frac{3Nk_B}{2}\right)\left(\frac{T}{T_F}\right)$$

which is the missing linear term for the heat capacity of metals – the linear term is hence because of the electrons in the metal, that do not behave classically.

Unfortunately this model too has its shortcomings, which is why we shall venture even further. Here are some things that are not explained by Sommerfeld theory:

1. Why aren't electrons scattered by the positively charged nuclei?
2. Why do core electrons not contribute to the Fermi energy and velocity, and why don't insulators have any free electrons?
3. We still do not understand why the Hall coefficient sometimes has the wrong sign.
4. Many optical features, such as high harmonic absorption are not explained by the Sommerfeld theory.
5. Sometimes measurements of the specific heat of electrons is still off by a factor 10. Additionally we sometimes measure the wrong electron mass
6. We cannot explain magnetism
7. Electron interaction – we ignored electron-electrons interactions, and there is no reason this should be accurate.

CHEMICAL BONDS

Covalent Bond

Suppose we have two hydrogen atoms that are close enough to each other for them to interact. We will use the *Born-Oppenheimer approximation*, which states that we can view the nuclei as stationary, and solve

---

[¶] Note the similarity to the Bose-Einstein distribution. The difference being the opposite sign in the denominator.

[**] It doesn't really matter whether we use periodic boundary conditions or fixed boundary conditions, as the main contributor to the physics of the system is the bulk, of which there is much more than there are borders.

for the electronic wave functions separately. Once this is done one would use the electronic energy surfaces as a function of the internuclear distancing to solve the nuclear wave functions, however, that is not what we are doing now.

Let us consider a single electron between the nuclei. This will have the following Hamiltonian

$$H = T + V_1 + V_2, \quad V_i = -\frac{e^2}{4\pi\varepsilon_0 \|\boldsymbol{r} - \boldsymbol{R}_i\|}$$

Using the variational principle we will take a trial wavefunction:

$$|\psi\rangle = \phi_1 |1\rangle + \phi_2 |2\rangle$$

where

$$\begin{aligned}(T + V_1)|1\rangle &= \varepsilon_0 |1\rangle \\ (T + V_2)|2\rangle &= \varepsilon_0 |2\rangle\end{aligned}$$

assume $\langle i|j\rangle = \delta_{ij}$ and define

$$\langle 1|V_2|1\rangle = \langle 2|V_1|2\rangle = V_{\text{cross}}$$

where the first equality is due to the indistinguishability of the two nuclei. Additionally let us define the *hopping* term:

$$\langle 1|V_1|2\rangle = t$$

let us assume $t > 0$. Now this gives us a Schrödinger equation of the form:

$$\begin{pmatrix} \varepsilon_0 + V_{\text{cross}} & t \\ t & \varepsilon_0 + V_{\text{cross}} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = E \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$$

which has the following eigenvalues:

$$E_\pm = \varepsilon_0 + V_{\text{cross}} \pm t$$

with eigenvectors

$$\begin{aligned}|\psi_{\text{bonding}}\rangle &= \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle) \\ |\psi_{\text{antibonding}}\rangle &= \frac{1}{\sqrt{2}}(|1\rangle + |2\rangle)\end{aligned}$$

This says that there exists a state $|\phi_{\text{bonding}}\rangle$ where it is more convenient for the hydrogen nuclei to be bonded together, hence forming H_2^+, as this has lower energy than the energy of the two individual hydrogen nuclei with the electron on one of them.

This is the basic principle behind covalent bonding.

Van der Waals, Fluctuating Dipole Forces, or Molecular Bonding

Atoms have a dipole moment, whose expectation is zero. However, the dipole moment can fluctuate momentarily, which will induce a dipole-force between atoms. This is what is known as the Van der Waals force. It is weak compared to ionic and covalent bonds, however the force can act over a further distance, because electrons don't need to hop.

Metallic Bonding

It is sometimes difficult to distinguish between covalent and metallic bonds – both involve the sharing of electrons through delocalisation. However for the metallic bonds the delocalisation is on the scale of the entire crystal, whereas covalent delocalisation only happens between two atoms (or a few). The delocalised electrons act as free particles, which conduct electricity. The delocalisation does not favour any direction particularly, therefore slightly moving atoms will not affect the bonds – metals are often ductile and malleable.

Hydrogen Bonding

The electronegativity and small size of hydrogen means that hydrogen can take on a positive partial charge. This special case is why we treat hydrogen bonding separately to other atomic bonds. The partial charges of hydrogen and the donor/acceptor cause a dipole-moment, which attracts neighbouring molecules. In the example of water the oxygen atoms take on a negative partial charge, which attracts neighbouring positively (and partially) charged hydrogen atoms.

In summary:

1. **Ionic bonds:** Electron is transferred from one atom to another. Formed by atoms with very different electronegativity. Hard, brittle, high melting temperature, insulator, hydrophilic.
2. **Covalent bonds:** Electron is shared between neighbouring atoms. Usually atoms in the groups *III* to *V*. A great example is silicon and diamonds. Very hard, high melting temperature, insulator or semiconductor.
3. **Metallic bonds:** Electrons are delocalised throughout the lattice. Left and middle of the periodic table. Ductile, malleable, hardened by impurities, lower melting temperature, good electrical and thermal conductors.
4. **Van der Waals bonding:** No transfer of electrons, only due to the oscillating dipole moment of

atoms. Noble gas solids, non-polar solids. Soft, weak, low melting temperature, insulator.

5. **Hydrogen bonding:** Involves partially charged hydrogen atoms. Important in organic materials. Holds together ice. Weak bond (stronger than Van der Waals).

TIGHT BINDING CHAIN

1D chain

Let us denote the wavefunction of electron at the n-th electron as $|n\rangle$. As we saw previously the interactions between these atoms will cause the electron orbits to hybridise. Hence the hybridised wave function will be some linear combination of atomic orbitals:

$$|\Psi\rangle = \sum_n \phi_n |n\rangle, \quad \langle n|m\rangle = \delta_{nm}$$

where the orthonormality is an *assumption*, the orbitals at each site needn't be orthogonal, and in fact aren't orthogonal for small interatomic distances. The time-independent Schrödinger equation in this Hilbert space is

$$\sum_m H_{nm}\phi_m = E\phi_n, \quad \langle\phi_n|H|\phi_m\rangle$$

Let us say that the electrons have an on-site energy of ε_0, and a hopping energy of t:

$$H_{nm} = \varepsilon_0\delta_{nm} + t\left(\delta_{n-1,m} + \delta_{n+1,m}\right)$$

Let us now assume that our wave function is of the form:

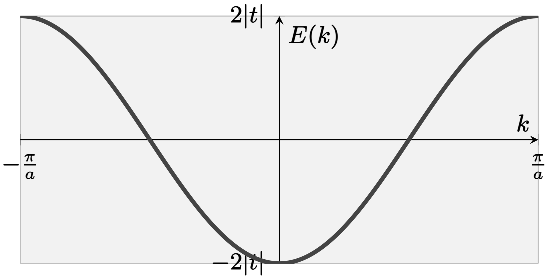
$$\phi_n = \frac{e^{-ikna}}{\sqrt{N}}$$

plugging this into our time-independent Schrödinger equation gives us

$$E = \varepsilon + 2t\cos(ka)$$

note that t is normally negative, hence it is convenient for the electrons to dissociate. This means its convenient to write

$$E = \varepsilon - 2|t|\cos(ka)$$



Figur 4: *The first Brillouin zone for a monatomic tight-binding chain.*

Note that in the long wavelength limit this is parabolic:

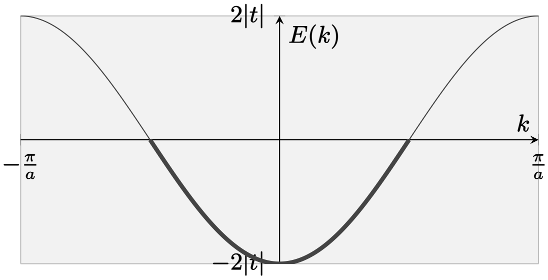
$$E \approx E_0 + |t|a^2k^2$$

where $E_0 = \varepsilon - 2|t|$. This means that our linear combination of atomic orbitals behave as a quasiparticle, whose mass is

$$\frac{\hbar^2k^2}{2m^*} = |t|a^2k^2 \quad \rightsquigarrow \quad m^* = \frac{\hbar^2}{2|t|a^2}$$

Filling bands

In the ground state $T = 0$ we fill the states from the bottom upwards. However, Pauli's exclusion principle says that we can fill each state with two electron (one with up and one with down spin). Thus for monovalent atoms we would only fill half the band:

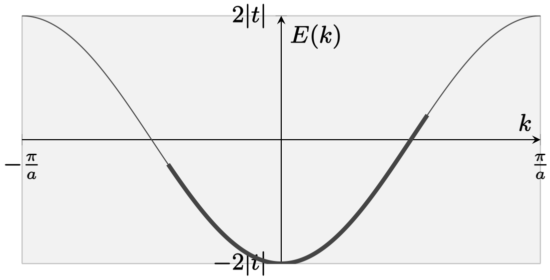


Figur 5: *The first Brillouin zone for a monatomic, monovalent tight-binding chain.*

Applying a constant electric field would apply a force $F = -eE$:

$$\hbar\frac{\mathrm{d}k}{\mathrm{d}t} = -eE \quad \rightsquigarrow \quad k(t) = -\frac{eE}{\hbar}t$$

which will cause the states to move slightly to the side – but not indefinitely, presumable because of the impurities? There do exist cases where you can drive the band on and on towards the right. Once it disappears on the right side of the Brillouin zone boundary, it will reappear on the left side – these are known as Bloch Oscillations.



Figur 6: *The first Brillouin zone for a monatomic, monovalent tight-binding chain. The electric field causes a displacement in* k*.*

The group velocity $\frac{1}{\hbar}\frac{\mathrm{d}E}{\mathrm{d}k} = v_\mathrm{g}$ is asymmetric now – there are more (quasi-) electrons with a positive group velocity than a negative group velocity– we have induced a current.

This leads to the fact that filled bands cannot carry currents – the band is periodic and filled, so a displacement will leave the band unchanged.

CRYSTAL STRUCTURE

A **lattice** is defined as an infinite set of points (lattice points) defined by integer linear combinations of a set of linearly independent primitive lattice vectors. In three dimensions we can write a lattice vector as

$$\boldsymbol{R}_{[n_1,n_2,n_3]} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3 = [n_1, n_2, n_3], \quad n_1, n_2, n_3 \in \mathbb{Z}$$

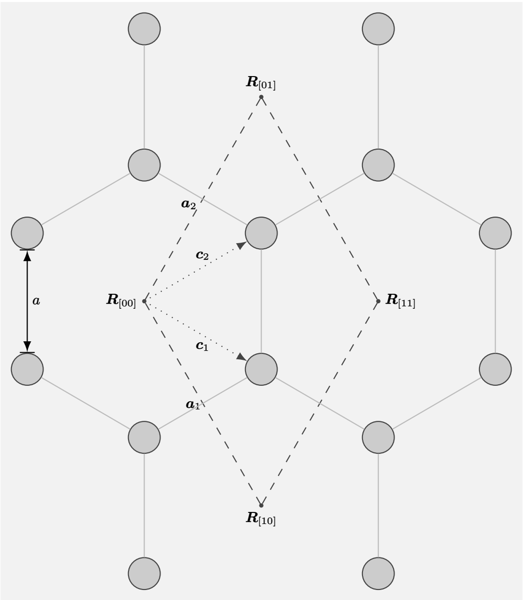
Within the lattice is the **unit cell**: a region of space such that when many identical units are stacked together it tiles (completely fills) all of space and reconstructs the full structure. If a unit cell contains only one lattice point it is a **primitive unit cell**.

There is an algorithm that can be used to always create a primitive unit cell from a lattice point. Given a lattice point, all points in space that are closer to the given lattice point than to any other lattice point constitute the **Wigner-Seitz cell**. Hence by finding perpendicular bisectors between the given and all of the closest lattice points you will find the Wigner-Seitz cell.

The unit cell describes the *lattice points*, which do not necessarily correspond to atoms. Hence given a unit cell, we need to define where the atoms lie within the cell – this is done using the **basis**. For instance graphene's basis[††] is

$$\boldsymbol{R}^{\mathrm{Carbon}_1}_{[n_1\ n_2]} = [n_1, n_2] + \left[\frac{2}{3}, \frac{1}{3}\right]$$
$$\boldsymbol{R}^{\mathrm{Carbon}_2}_{[n_1\ n_2]} = [n_1, n_2] + \left[\frac{1}{3}, \frac{2}{3}\right]$$

This basis is for the following unit cell:



Figur 7: *Graphene Lattice. The dashed lines denote the unit cell.*

In three dimensions there are only a finite number of lattices, I will name two of these presently, as well as their defining characteristics:

1. **Body centred cubic (bcc) lattice:** A simple cubic lattice where we have added an additional lattice point in the middle of the cube:

$$\boldsymbol{R}_{\mathrm{corner}} = [n_1, n_2, n_3]$$
$$\boldsymbol{R}_{\mathrm{centre}} = [n_1, n_2, n_3] + \left[\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right]$$

[††] Note that the basis is dependent on which unit cell is chosen – hence not unique.

with the following primitive unit vectors

$$\boldsymbol{a}_1 = [1,0,0]$$
$$\boldsymbol{a}_2 = [0,1,0]$$
$$\boldsymbol{a}_3 = \left[\frac{1}{2},\frac{1}{2},\frac{1}{2}\right]$$

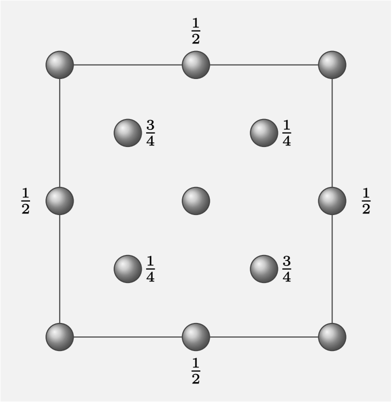
2. **Face centred cubic (fcc) lattice:** A simple cubic lattice where we have added additional lattice points at the centre of each face:

$$\boldsymbol{R}_{\text{corner}} = [n_1,n_2,n_3]$$
$$\boldsymbol{R}_{\text{face, }xy} = [n_1,n_2,n_3] + \left[\frac{1}{2},\frac{1}{2},0\right]$$
$$\boldsymbol{R}_{\text{face, }xz} = [n_1,n_2,n_3] + \left[\frac{1}{2},0,\frac{1}{2}\right]$$
$$\boldsymbol{R}_{\text{face, }yz} = [n_1,n_2,n_3] + \left[0,\frac{1}{2},\frac{1}{2}\right]$$

with primitive lattice vectors

$$\boldsymbol{a}_1 = \left[\frac{1}{2},\frac{1}{2},0\right]$$
$$\boldsymbol{a}_2 = \left[\frac{1}{2},0,\frac{1}{2}\right]$$
$$\boldsymbol{a}_3 = \left[0,\frac{1}{2},\frac{1}{2}\right]$$

It can be quite difficult to show 3D crystal structures, therefore one often uses the **plan view**, where you project the 3D image onto the plane and explicitly denote the lost-component for each atom:



Figur 8: *Plan view of the diamond crystal. Unmarked points are at $z = 0$ and $z = 1$.*

THE RECIPROCAL LATTICE

Given a lattice vector, $\boldsymbol{R}$, we can define the reciprocal lattice as all vectors, $\boldsymbol{G}$, such that

$$e^{i\boldsymbol{G}\cdot\boldsymbol{R}} = 1$$

We now make the following claims:

1. The reciprocal lattice is a lattice in reciprocal space.
2. The primitive reciprocal lattice vectors ($\boldsymbol{b}_i$) have the following property:

$$\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 2\pi\delta_{ij}$$

   where $\boldsymbol{a}_i$ are the primitive direct lattice vectors.

Using these properties, we can *construct* the vectors $\boldsymbol{a}_i$:

$$\boldsymbol{b}_i = \frac{\pi\varepsilon_{ijk}\boldsymbol{a}_j \times \boldsymbol{a}_k}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)}$$

for instance, this gives us

$$\boldsymbol{b}_1 = \frac{2\pi\boldsymbol{a}_2 \times \boldsymbol{a}_3}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)}$$

These vectors are linearly independent[‡‡], and therefore they span $\mathbb{R}^3$. Thus we can write all $\boldsymbol{G}$ as

$$\boldsymbol{G} = m_1\boldsymbol{b}_1 + m_2\boldsymbol{b}_2 + m_3\boldsymbol{b}_3$$

Additionally we have that

$$1 = e^{i\boldsymbol{G}\cdot\boldsymbol{R}} = e^{2\pi i(n_1m_1+n_2m_2+n_3m_3)}$$

this is true if and only if $m_1, m_2, m_3 \in \mathbb{Z}$.

Fourier transform of the direct lattice

Let us define the density of lattice points as

$$\rho(\boldsymbol{r}) = \sum_{n_1,n_2,n_3} \delta^3\left(\boldsymbol{r} - (n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3)\right)$$

The Fourier transform hereof would be

$$\mathcal{F}\left\{\rho(\boldsymbol{r})\right\} = \sum_{n_1,n_2,n_3} \int_{\mathbb{R}^3} \mathrm{d}\boldsymbol{r} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \delta^3\left(\boldsymbol{r} - (n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3)\right)$$

$$= \sum_{n_1,n_2,n_3} e^{i\boldsymbol{k}\cdot(n_1\boldsymbol{a}_1+n_2\boldsymbol{a}_2+n_3\boldsymbol{a}_3)} = \sum_{\boldsymbol{R}\in\text{ DL}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}}$$

$$= \frac{(2\pi)^3}{|\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)|} \sum_{\boldsymbol{G}} \delta^3\left(\boldsymbol{k} - \boldsymbol{G}\right)$$

[‡‡] because $\boldsymbol{a}_i$ are linearly independent

Where $|\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)| = V$ is the volume of the unit cell. However generally it is useful not to perform the integral over $\boldsymbol{r}$, and instead define the *structure factor*:

$$\mathcal{F}\{\rho(\boldsymbol{r})\} = \frac{(2\pi)^3}{V} \sum_{\boldsymbol{G}} \delta^3(\boldsymbol{k} - \boldsymbol{G}) S(\boldsymbol{k})$$

where

$$S(\boldsymbol{k}) = \int_{\text{unit cell}} \mathrm{d}\boldsymbol{x}\; e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \rho(\boldsymbol{x})$$

Reciprocal Lattice Points as Families of Lattice Planes

A *family of lattice planes* is an infinite set of equally separated parallel lattice planes, which taken together contain all points of the lattice.

Furthermore we claim that the family of planes are in one-to-one correspondence with the possible directions of reciprocal lattice vectors. The planes are normal to these reciprocal normal vectors. The spacing between the lattice planes is $d = 2\pi / \|\boldsymbol{G}_{\min}\|$, where $\boldsymbol{G}_{\min}$ is the reciprocal lattice vector that is normal to the set of planes and has the shortest length.

Lattice Planes and Miller Indices

Let us begin by defining the Miller indices:

$$(h, k, \ell) \equiv h\boldsymbol{b}_1 + k\boldsymbol{b}_2 + \ell\boldsymbol{b}_3 = \boldsymbol{G}_{(h,k,\ell)}$$

additionally it is customary to write, for example $(1, \bar{1}, 1)$ instead of $(1, -1, 1)$. The shortest distance between planes:

$$d_{(h,k,\ell)} = \frac{2\pi}{\|\boldsymbol{G}\|} = \frac{2\pi}{\sqrt{h^2\|\boldsymbol{b}_1\|^2 + k^2\|\boldsymbol{b}_2\|^2 + \ell^2\|\boldsymbol{b}_3\|^2}}$$

$$\rightsquigarrow \quad \frac{1}{|d_{(h,k,\ell)}|^2} = \left(\frac{h}{a_1}\right)^2 + \left(\frac{k}{a_2}\right)^2 + \left(\frac{\ell}{a_3}\right)^2$$

which is an ellipsoid in h, k and ℓ. Note that for cubic lattices ($a_1 = a_2 = a_3$) this simplifies to

$$d^{\text{cubic}}_{(h,k,\ell)} = \frac{a}{\sqrt{h^2 + k^2 + \ell^2}}$$

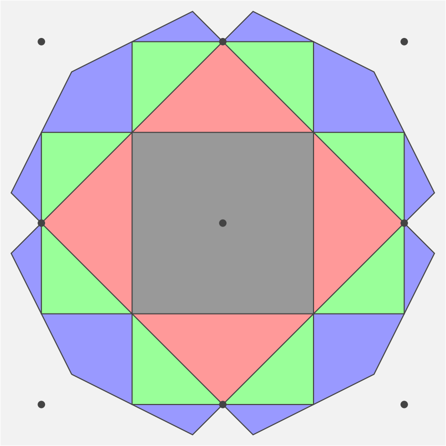
There is a point to make here:

1. For *unit cells* (h, k, ℓ) will describe a family of lattice planes, if and only if h, k and ℓ are (pairwise) coprime.
2. For non-primitive unit cells h, k and ℓ will generally not have to be pairwise coprime. For instance for a bcc crystal with the conventional cell $(0, 1, 0)$ does not represent a family of lattice planes, whereas $(0, 2, 0)$ does.

Brillouin Zones

The **Brillouin zone** is any primitive unit cell of the reciprocal lattice.

The first Brillouin zone is the set of $\boldsymbol{k}$ points that are closer to $\boldsymbol{0}$ than any other reciprocal lattice point. Similarly the second Brillouin zone is the set of $\boldsymbol{k}$ points where $\boldsymbol{0}$ is the next closest lattice point.



Figur 9: *First four Brillouin zones for a square lattice. Gray: first Brillouin zone, red: second Brillouin zone, green: third Brillouin zone, lavender: fourth Brillouin zone.*

WAVE SCATTERING BY CRYSTALS

According to Fermi's Golden Rule the scattering rate, $\Gamma(\boldsymbol{k}, \boldsymbol{k}')$, from $\boldsymbol{k} \rightsquigarrow \boldsymbol{k}'$ is given by:

$$\Gamma(\boldsymbol{k}, \boldsymbol{k}') = \frac{2\pi}{\hbar} |\langle \boldsymbol{k}'|V|\boldsymbol{k}\rangle|^2 \delta(E_{\boldsymbol{k}'} - E_{\boldsymbol{k}})$$

Where

$$\langle \boldsymbol{k}'|V|\boldsymbol{k}\rangle = \frac{1}{L^3} \int \mathrm{d}\boldsymbol{r}\; e^{-i(\boldsymbol{k}'-\boldsymbol{k})\cdot\boldsymbol{r}} V(\boldsymbol{r})$$

Now note that $V(\boldsymbol{r} + \boldsymbol{R}) = V(\boldsymbol{r})$, for any lattice vector $\boldsymbol{R}$. Hence we can rewrite this as

$$\langle \boldsymbol{k}'|V|\boldsymbol{k}\rangle = \frac{1}{L^3} \left(\sum_{\boldsymbol{R}} e^{-i(\boldsymbol{k}'-\boldsymbol{k})\cdot\boldsymbol{R}} \right) \left(\int_{\text{unit cell}} \mathrm{d}\boldsymbol{r}\; e^{-i(\boldsymbol{k}'-\boldsymbol{k})\cdot\boldsymbol{r}} V(\boldsymbol{r}) \right)$$

As before, the first term in parentheses must be zero unless

$$\boldsymbol{k}' - \boldsymbol{k} = \boldsymbol{G}$$

which is known as the **Laue condition**, which is equivalent to the conservation of crystal momentum, and hence $|\boldsymbol{k}'| = |\boldsymbol{k}|$. We use from before that the structure factor is given by

$$S(\boldsymbol{G}) = \int_{\text{unit cell}} \mathrm{d}\boldsymbol{r} e^{i\boldsymbol{G}\cdot\boldsymbol{r}} V(\boldsymbol{r})$$

The total potential is usually a linear combination of individual potentials:

$$V(\boldsymbol{r}) = \sum_{\text{atoms } j} V_j(\boldsymbol{r} - \boldsymbol{r}_j)$$

from where we get that

$$S(\boldsymbol{G}) = \sum_{\text{atoms } j \text{ in unit cell}} f_j(\boldsymbol{G}) e^{i\boldsymbol{G}\cdot\boldsymbol{r}_j}$$

where $f_j(\boldsymbol{G})$ is the Fourier transform of the scattering potential for atom j:

$$f_j(\boldsymbol{G}) = \int \mathrm{d}\boldsymbol{r}\ e^{i\boldsymbol{G}\cdot\boldsymbol{r}} V_j(\boldsymbol{r})$$

ELECTRONS IN A PERIODIC POTENTIAL

Nearly Free Electron Model

A free electron has the following Hamilton and energy levels:

$$H_0 = \frac{\|\boldsymbol{p}\|^2}{2m}, \quad \varepsilon_0(\boldsymbol{k}) = \frac{\hbar^2 \|\boldsymbol{k}\|^2}{2m}$$

given the three quantisation lengths[§§] L_x, L_y and L_z we can write $\boldsymbol{k}$ as

$$\boldsymbol{k} = \pi \left(\frac{n_x \hat{\boldsymbol{x}}}{L_x} + \frac{n_y \hat{\boldsymbol{y}}}{L_y} + \frac{n_z \hat{\boldsymbol{z}}}{L_z} \right), \quad n_x, n_y, n_z \in \mathbb{Z}$$

Suppose now we perturb the free electron with a periodic potential:

$$H = H_0 + V(\boldsymbol{r}), \quad V(\boldsymbol{r} + \boldsymbol{R}) = V(\boldsymbol{r})$$

First note the matrix element

$$\langle \boldsymbol{k}' | V | \boldsymbol{k} \rangle = \frac{1}{L^3} \int \mathrm{d}\boldsymbol{r} e^{i(\boldsymbol{k}' - \boldsymbol{k})\cdot\boldsymbol{r}} V(\boldsymbol{r}) \equiv V_{\boldsymbol{k}' - \boldsymbol{k}}$$

is zero unless $\boldsymbol{k}' - \boldsymbol{k} = \boldsymbol{G}$, a reciprocal lattice vector. This implies that the perturbation will only mix the state $|\boldsymbol{k}\rangle$ with $|\boldsymbol{k} + \boldsymbol{G}\rangle$. Thus, if we are far away from the Brillouin zone boundary we can use non-degenerate perturbation theory:

$$\varepsilon(\boldsymbol{k}) = \varepsilon_0(\boldsymbol{k}) + \langle \boldsymbol{k} | V | \boldsymbol{k} \rangle = \varepsilon_0(\boldsymbol{k}) + V_0$$

henceforth we will assume that $V_0 = 0$. Second-order perturbation theory will thus give us

$$\varepsilon(\boldsymbol{k}) = \varepsilon_0(\boldsymbol{k}) + \sum_{\boldsymbol{k}' \neq \boldsymbol{k}} \frac{|\langle \boldsymbol{k}' | V | \boldsymbol{k} \rangle|^2}{\varepsilon_0(\boldsymbol{k}) - \varepsilon_0(\boldsymbol{k}')}$$

where $\boldsymbol{k}'$ naturally is still restricted to be on the reciprocal lattice.

However, close to the Brillouin zone boundary we have that

$$\varepsilon_0(\boldsymbol{k}) \approx \varepsilon_0(\boldsymbol{k} + \boldsymbol{G})$$

therefore we need to perform degenerate perturbation theory. Each state will only be degenerate with one other state – however, a state can be approximately degenerate with numerous other states, which would require us to increase the size of our degenerate space. However we will stick to two degenerate state, $|\boldsymbol{k}\rangle$ and $|\boldsymbol{k}'\rangle$. The perturbation matrix is given by

$$W = \begin{pmatrix} \langle \boldsymbol{k} | V | \boldsymbol{k} \rangle & \langle \boldsymbol{k}' | V | \boldsymbol{k} \rangle \\ \langle \boldsymbol{k} | V | \boldsymbol{k}' \rangle & \langle \boldsymbol{k}' | V | \boldsymbol{k}' \rangle \end{pmatrix} = \begin{pmatrix} \varepsilon_0(\boldsymbol{k}) & V_G \\ V_G^* & \varepsilon_0(\boldsymbol{k} + \boldsymbol{G}) \end{pmatrix}$$

The eigenvalues of the perturbation matrix are

$$E_\pm = \varepsilon_0(\boldsymbol{k}) \pm |V_{\boldsymbol{G}}|$$

where we have assumed[*] that $\varepsilon_0(\boldsymbol{k}) = \varepsilon_0(\boldsymbol{k} + \boldsymbol{G})$.

Not quite at the boundary

This part will be done in one-dimension. If we are δ away from the Brillouin zone boundary, then it is the case that

$$\varepsilon_0(n\pi/a + \delta) \approx \varepsilon_0(-n\pi/a + \delta)$$

The unperturbed energies are

$$\varepsilon_0(\pm n\pi/a + \delta) = \frac{\hbar^2}{2m} \left(\left(\frac{n\pi}{a} \right)^2 \pm \frac{2n\pi\delta}{a} + \delta^2 \right)$$

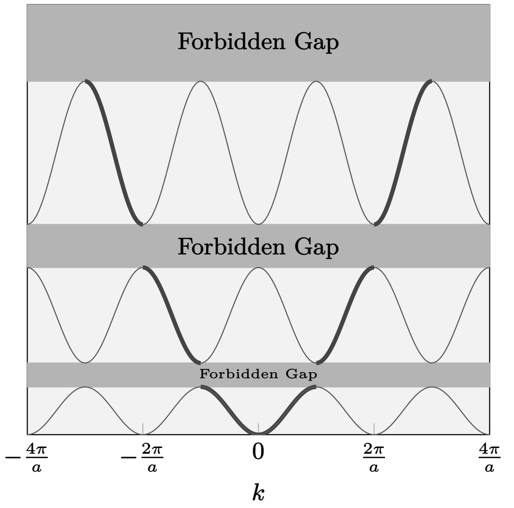
[§§] For periodic boundary conditions replace $\pi \rightsquigarrow 2\pi$

[*] that the energies are *equal*, not approximately equal.

and plugging these into the characteristic polynomial:

$$E_{\pm} = \frac{\hbar^2 n^2 \pi^2}{2ma^2} \pm |V_G| + \frac{\hbar^2 \delta^2}{2m} \left(1 \pm \frac{\hbar^2 n^2 \pi^2}{ma^2} \frac{1}{|V_G|} \right)$$

for small perturbations, compared to the eigenergies of the free electron, which is exactly what we are looking at, the second term in the parentheses is greater than unity.



Figur 10: *Dispersion of a nearly free electron model*

Note the fact that the dispersion is parabolic close to the Brillouin zone boundary:

$$E_{\pm}(G+\delta) = C_{\pm} \pm \frac{\hbar^2 \delta^2}{2m^*_{\pm}}, \quad m^*_{\pm} = \frac{m}{\left| 1 \pm \frac{\hbar^2 n^2 \pi^2}{2ma^2} \frac{1}{|V_G|} \right|}$$

Bloch's Theorem

The assumption that we can treat electrons as free particles, and hence describe their wave functions as planar waves seems rather bold. Especially because we previously spoke about scattering of electrons due to impurities. However **Bloch's theorem** states that the eigenfunctions of electrons in a periodic potential is of the form:

$$\Psi^{\alpha}_{\boldsymbol{k}}(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u^{\alpha}_{\boldsymbol{k}}(\boldsymbol{r})$$

where $u^{\alpha}_{\boldsymbol{k}}(\boldsymbol{r})$ is periodic in the unit cell and $\boldsymbol{k}$ can be chosen within the first Brillouin zone. Thus electrons behave as planar waves, with some periodic modulation of amplitude (and possibly phase, for $u^{\alpha}_{\boldsymbol{k}} \in \mathbb{C}$). In order for $u^{\alpha}_{\boldsymbol{k}}(\boldsymbol{r})$ to be periodic it must be possible to write it as a sum over $\boldsymbol{G}$ – every[†] periodic function has a Fourier transform:

$$u^{\alpha}_{\boldsymbol{k}} = \sum_{\boldsymbol{G}} \tilde{u}^{\alpha}_{\boldsymbol{G},\boldsymbol{k}} e^{i\boldsymbol{G}\cdot\boldsymbol{r}}$$

hence an alternative expression of Bloch's theorem would be

$$\Psi^{\alpha}_{\boldsymbol{k}} = \sum_{\boldsymbol{G}} \tilde{u}^{\alpha}_{\boldsymbol{G},\boldsymbol{k}} e^{i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{r}}$$

INSULATOR, SEMICONDUCTOR OR METAL

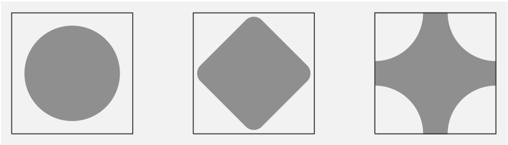
In short: insulators have filled bands, metals have partially filled bands. Semiconductors are a specific type of insulator, where the band gap is accessible through thermal fluctuations at room temperature[‡]. If the band gap is below 4eV we classify the insulator as a semiconductor.

Energy bands in two and three dimensions

If the unit cell is square, then the reciprocal unit cell is also square – hence the Brillouin zone is square. If we now fill up "empty space"with electrons from $E = 0$ to $E = E_F$ we will first fill low energy states, and then slowly build upwards once all the states filled, resulting in a circular Fermi-"surface"[§]. If we now add a strong periodic potential, we will reduce the band-energies close to the inside of the first Brillouin zone boundary, as before, and increase the energies close to the outside of the first Brillouin zone.

Monovalent atoms

For varying potential strengths in a monovalent lattice we get:



Figur 11: *Left: free electrons. Middle: intermediate potential. Right: strong potential*

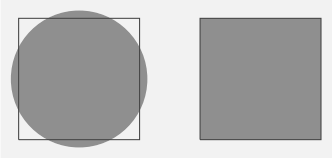
Divalent atoms

[†] Every periodic function with only a finite number of discontinuities per period, which has a bounded variation and must be absolutely integrable over a period.

[‡] The image that only states below the Fermi energy are occupied is only true at $T = 0$

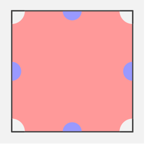
[§] in 2D the Fermi-surface is a line.

For varying potential strengths in a divalent lattice we get:



Figur 12: *Left: free electrons. Right: electrons in a very strong periodic potential*

For any potential in the spectrum between no potential and strong potential, there will be partial filling of the second Brillouin zone:



Figur 13: *Electrons in a square lattice with intermediate potential strength. The red shaded area is the occupied part of the first Brillouin zone, and the blue shaded area is the occupied part of the second Brillouin zone*

Here the same principle holds as in 1D: filled bands do not conduct electricity – for example a divalent square lattice will not conduct electricity. However, now it is possible for divalent square lattices to conduct electricity, providing the potential is weak enough.

In three dimensions the Fermi surface is a surface, which behaves similarly to what we discussed now. Weak potentials have spherical Fermi surfaces. Intermediate potentials will have 6 bulges, which touch the Brillouin zone boundary for strong potentials.

Failures of the Band-structure picture of metals

We have discussed how to determine the electric properties of solids, by first calculating the band structure, which depends on the geometry of the solid, and then filling the band up with however many electrons we have, following Pauli's exclusion principle. From there we had the simple principle: filled bands are inert.

However, this would cause us to assume that all monovalent lattices are metals, however, we have completely ignored the interaction between electrons, which does play a significant role in some systems. An example of this picture failing is in magnetism: the magnetic moments of electrons can interact – if they all align you have a ferromagnet, which shows complexity that simply cannot be described by our model.

Another example is the Mott insulator: if the electronic repulsion is so strong in a monovalent lattice that it simply costs too much to have two electrons at the same atom (irrespective of spin), then the material will be inert, even though it is monovalent.

Optical properties of metals and insulators

If a semiconductor has a *direct* band gap ΔE, then a photon with this energy can excite an electron from the valence to the conduction band. Conversely an electron in the conduction band can de-excite into the valence band, giving off a photon with the corresponding energy. Thus if a material has a band gap in the visible region of the electromagnetic spectrum, this material will absorb light.

If a material has a band gap greater than ~ 3.2eV then light will not interact: the material is transparent. For example this is the case with diamond and quartz.

Cadmium sulfide has a band gap of 2.6eV which corresponds to green-blue light. Thus CdS appears reddish (removing green-blue and upwards from white gives you red).

Metals conduct electricity, and therefore their interaction with light is slightly different. The electronic excitation is quickly undone, such that most absorbed photons are re-emitted quickly (hence reflected). If all photons are re-emitted the metal will look silver, like a mirror. However, some metals do absorb some frequencies, depending on the band structure and potential strength¶, giving them colours – normally yellow-ish or red-ish.

SEMICONDUCTOR PHYSICS

Holes

Relativistic quantum mechanics teaches us to think about negative energy states. And to all of our surprises one often comes across relativistic quantum mechanical equations in condensed matter physics. Hence it is natural to think about these negative energy states, in the context of condensed matter physics.

Imagine we have a filled valence band, and an empty conduction band. Now we excite a single electron,

---

¶ This is an oversimplified model of metals.

leaving a *hole* in its place. Now we can either do our calculations on the electron in the conduction band, or on the hole in the valence band. We will get the same result.

In fact, the negative energy states** I mentioned previously are the holes. It can be very convenient to perform calculations using holes instead of electrons, and in fact in some materials it is the holes and not the electrons that carry current, which is why we had Hall-coefficients with the wrong sign previously.

Effective Mass of Electrons and Holes

The dispersion relation for an electron near the bottom of the conduction band can be Taylor expanded:

$$E = E^{\mathrm{C}}_{\mathrm{min}} + \alpha \|\boldsymbol{k} - \boldsymbol{k}_{\mathrm{min}}\|^2 + \cdots$$

which, as we discussed previously leads us to define the effective mass:

$$m^*_{\mathrm{e}} = \frac{\hbar^2}{2\alpha}$$

and equivalently we can define the group velocity:

$$\boldsymbol{v} = \frac{\nabla_{\boldsymbol{k}} E}{\hbar} = \frac{2\alpha(\boldsymbol{k} - \boldsymbol{k}_{\mathrm{min}})}{\hbar} = \frac{\hbar(\boldsymbol{k} - \boldsymbol{k}_{\mathrm{min}})}{m^*_{\mathrm{e}}}$$

Similarly the energy of the hole can be Taylor expanded about the maximum of the valence band:

$$E = E^{\mathrm{V}}_{\mathrm{max}} - \alpha \|\boldsymbol{k} - \boldsymbol{k}_{\mathrm{max}}\|^2 + \cdots$$

From here it is conventional to define the hole mass to be positive, which has some implications later on, when we discuss hole momentum and velocity. Thus

$$\frac{\hbar^2}{m^*_{\mathrm{h}}} = -\frac{\partial^2 E}{\partial k^2} \quad \rightsquigarrow \quad m^*_{\mathrm{h}} = \frac{\hbar^2}{2\alpha}$$

It requires positive energy to push an electron *upwards* in the band. Viewing a hole as the lack of an electron will thus lead us to the conclusion that it requires positive energy to push a hole *downwards* in the valence band. This is rather counterintuitive, but is in fact the case. Hence holes prefer to maximise their energy, whereas electrons minimise theirs: The minimal energy state is an electron at the bottom of the conduction band and a hole at the top of the valence band.

Momentum of holes

If we remove an electron with momentum $\hbar\boldsymbol{k}$ from a filled valence band the valence band will have momentum $-\hbar\boldsymbol{k}$– remember filled bands carry no momentum. Thus it is convenient to define

$$\boldsymbol{k}_{\mathrm{hole}} = -\boldsymbol{k}_{\mathrm{electron}}$$

Thus

$$\boldsymbol{v}_{\mathrm{hole}} = \boldsymbol{v}_{\mathrm{missing\ electron}}$$

Where we have used that

$$E_{\mathrm{absence\ of\ electron}}(\boldsymbol{k}) = -E_{\mathrm{hole}}(\boldsymbol{k})$$

This is natural, because if we remove an electron then we decrease the energy by the energy that the electron had.

Adding electrons or holes: Doping

Imagine a lattice of Silicon atoms, with an occasional Phosphorous atom. If the concentration of Phosphorous is negligible compared to that of Silicon, then the impurities will not have a measurable effect on the dispersion relation. However, Phosphorous has an additional electron (and proton), which means that there is a free electron floating around – this is called n-doping, because we dope with negative charges.

Similarly, if we dope the Silicon with Aluminium we will have one fewer electron per Aluminium atom: we have "added"holes. This is p-doping, because we dope with positive charges.

Impurity states

The additional charges, both electrons and holes, only see the impurities – the lattice due to its periodicity is effectively invisible. However, the lattice does affect the dielectric constant ($\varepsilon_0 \rightsquigarrow \varepsilon$ where $\varepsilon = \varepsilon_0 \varepsilon_r$), which in turn affects the energy levels and characteristic lengths. Additionally it is not electrons that are bound by the impurities, but the quasi-particles made by the hybridised wave-functions: we need to take the effective mass into account.

For low concentrations of dopants (which is the limit we require for our model to work anyway) we can treat the impurities as hydrogenic atoms, with a relative permittivity. We get

$$\mathrm{Ry}^{\mathrm{eff}} = \mathrm{Ry}\left(\frac{m^*_{\mathrm{e}}}{m}\frac{1}{\varepsilon_r^2}\right), \quad \mathrm{Ry} \approx 13.6\mathrm{eV}$$

$$a_0^{\mathrm{eff}} = a_0\left(\varepsilon_r \frac{m}{m^*_{\mathrm{e}}}\right), \quad a_0 \approx 0.5\text{Å}$$

** The energy is not exactly negative in this context though – if you set the zero value of energy to be exactly the middle between the gaps then it is clear that the hole will have the negative of the electrons energy.

For typical materials $\mathrm{Ry}^{\mathrm{eff}}$ is below 0.1eV and a_0^{eff} is above 30Å.

These hydrogenic states are added to the band structure. If the material is n-doped the impurity band is just below[††] the bottom of the conduction band (donor impurity states), for p-doped materials the impurity band is just above the top of the valence band (acceptor impurity state).

Statistical Physics of Electrons and Holes

If the chemical potential is "well"below the conduction band ($\beta(\varepsilon_c-\mu) \gg 1$) we can approximate Fermi-Dirac statistics by Boltzmann statistics, which gives us the standard expression for the total density of electrons in the conduction band:

$$n(T) = \frac{1}{4}\left(\frac{2m_{\mathrm{e}}^* k_B T}{\pi\hbar^2}\right)^{\frac{3}{2}} e^{-\beta(\varepsilon_{\mathrm{c}}-\mu)}$$

We used the density of states of the conduction band in the calculation ($n(T) = \int_{\varepsilon_c}^{\infty} \mathrm{d}\varepsilon g_c(\varepsilon) n_F(\beta(\varepsilon-\mu))$). When we calculate the total density of holes in the valence band we need to use the density of states of the valence band and replace the Fermi-Dirac distribution by $1-n_F$ – the states not occupied by electrons are occupied by holes. From here we get

$$p(T) = \frac{1}{4}\left(\frac{2m_{\mathrm{h}}^* k_B T}{\pi\hbar^2}\right)^{\frac{3}{2}} e^{-\beta(\mu-\varepsilon_{\mathrm{v}})}$$

the two different exponents are due to the limits of integration (the conduction band is integrated from ε_{c} to infinity, whereas the valence band is integrated from $-\infty$ to ε_{v}). An important quantity is

$$n(T)p(T) = \frac{\left(m_{\mathrm{h}}^* m_{\mathrm{e}}^*\right)^{\frac{3}{2}}}{16}\left(\frac{2k_B T}{\pi\hbar^2}\right)^3 e^{-\beta E_{\mathrm{gap}}}, \quad E_{\mathrm{gap}} = \varepsilon_{\mathrm{c}} - \varepsilon_{\mathrm{v}}$$

This is the *law of mass action.* In order to see that this is a law of mass action, consider an intrinsic semiconductor (the number of holes equals the number of electrons): Now think of the following reaction:

$$e^- + p^+ \rightsquigarrow 0$$

where p^+ is the hole and $[e^-] = n(T)$ and $[p^+] = p(T)$. The concentration $[e^-p^+]^0 = 1$, so

$$\frac{[e^-][p^+]}{[e^-p^+]^0} = n(T)p(T) = K(T)$$

Intrinsic semiconductors

As mentioned before $n = p$ for intrinsic semiconductors, hence by dividing our expressions for $n(T)$ and $p(T)$ and taking the logarithm we can solve for $\mu(T)$:

$$\mu(T) = \frac{1}{2}(\varepsilon_{\mathrm{c}} + \varepsilon_{\mathrm{v}}) + \frac{3}{4}k_B T \ln\left(\frac{m_{\mathrm{h}}^*}{m_{\mathrm{e}}^*}\right)$$

At $T = 0$ the chemical potential lies exactly between the extrema of the conduction and valence bands.

Interestingly enough for $m_{\mathrm{h}}^* = m_{\mathrm{e}}^*$, which means the shape of the bottom of the conduction is the same as the top of the valence band, there is no temperature dependence of the chemical potential.

Doped semiconductors

We can define the doping, D:

$$D = n - p$$

Negatively doping a material will begin to fill the conduction band from the bottom, which results in the chemical potential rising up towards the conduction band. Similarly positively doping a material will increase the number of holes in the valence band, which in turn causes the chemical potential to fall towards the top of the valence band.

SEMICONDUCTOR DEVICES

Designing Band Gaps

The band gaps of *GaAs* and *AlAs* are 1.4eV and 2.7eV respectively. Both have direct band gaps at $\boldsymbol{k} = \mathbf{0}$. If we now were to mix these two materials to $Al_xGa_{1-x}As$ one could hope that the band gap is given by the convex combination:

$$E_{\mathrm{gap}}(x) = (1-x)1.4\mathrm{eV} + x2.7\mathrm{eV}, \quad x \in (0, 0.4)$$

This turns out to be approximately true, for x between 0 and 0.4. Therefore by choosing the relative concentrations we can engineer a semiconductor with a band gap anywhere in the interval (1.4eV, 1.92eV).

Using this we can create a sandwich with $Al_xGa_{1-x}As$ on each side of *GaAs*. The energy gap will have an abrupt dip, in which one can fill with bound electron states. Effectively you can use this method to make a finite well potential. The importance is that electrons in the conduction band see the bottom of the band as

---

[††] The effective Rydberg constant is usually much smaller than the band gap.

a potential as a function of position, which they can be trapped in, or scattered by.

p-n junction

Putting an n-doped material into contact with a p-doped material has some very interesting effects. The n-doped material has:

1. A filled valence band
2. A mostly empty conduction band

whereas a p-doped material has:

1. A mostly filled valence band (mostly empty if you're thinking about holes)
2. An empty conduction band (full if you're thinking about holes)

When these are brought into contact the electrons in the n-doped material's conduction band will begin to annihilate the holes in the p-doped material's valence band. This results in a depletion zone – a zone where all the holes and electrons have been annihilated. In this depletion zone there will be an electric field, pointing from n-doped towards p-doped. This electric field obstructs further electrons from travelling through the depletion zone (and similarly prevents holes from doing so). The electric field is such that the electrochemical potential ($\mu - e\phi$) is equal on each side of the depletion zone, where ϕ is the electric potential.

Solar Cell

If one applies light to a p-n junction electron-hole pairs may appear (if the photon energy is greater than the band-gap). Once this pair is created, and if they have enough kinetic energy to avoid each other initially, they will be accelerated in opposite directions by the electric field in the depletion zone: electrons are sent towards the n-doped region whereas holes are sent towards the p-doped region. This results in a current going from n to p. This is, in principle, how photovoltaics work.

The Diode

By applying an additional, external, electric potential (voltage) to a p-n junction, and by applying it in the opposite direction of the electric field in the depletion zone one can get electrons (holes) to pass through the depletion zone and enter the p- (n-) doped material. Once they are there they will annihilate with holes (electrons). This results in a net current from p to n. The probability of this event occurring is proportional to $e^{-(E_{\text{gap}}-eV)/k_BT}$, where V is the applied voltage.

It is also possible for another effect to occur: an electron in the p-doped side is thermally excited into the conduction band. Once it is here it is free to move and the electric field in the depletion zone will accelerate it to the n-doped side. This however requires electrons (and holes) to thermally excite, which has a probability that is proportional to e^{-E_{gap}/k_BT}.

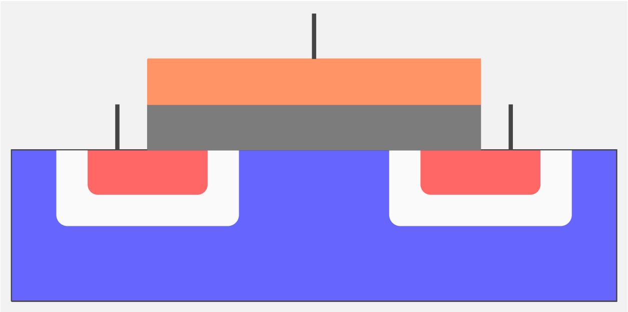
Combining the two effects we get

$$I_{\text{total}} = J_s(T)\left(e^{eV/k_BT} - 1\right)$$

where $J_s \propto e^{-E_{\text{gap}}/k_BT}$. This is known as the "Diode equation"and says that current flows far more easily from p to n, than the other way resulting in a very important electric component.

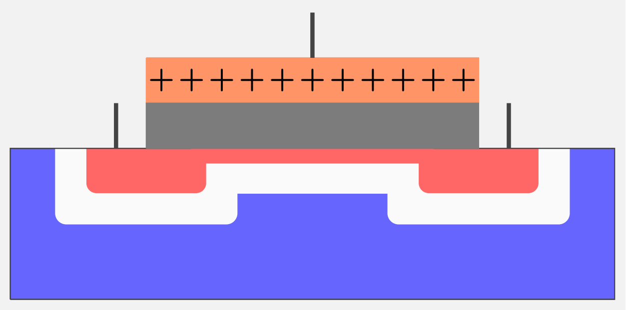
Transistor

A transistor is a component with three terminals (source, drain and gate). You can build a transistor by putting an n-doped semiconductor between two p-doped semiconductor, or vice versa. The former is called a p-MOSFET, and the latter a n-MOSFET. We will focus on the n-MOSFET:



Figur 14: *n-MOSFET. Red: n-doped semiconductor. Blue: p-doped semiconductor. White: depletion zone. Gray: Oxide insulator. Orange: Metal Gate. The two n-doped semiconductors are the source and drain terminals.*

Applying a positive voltage to the metal gate will push away positive charges in the top of the p-doped semiconductor. Once $V > V_{\text{threshold}}$ this will breach a gap through which current can flow:



Figur 15: *n-MOSFET with positive voltage at the metal gate. Red: n-doped semiconductor. Blue: p-doped semiconductor. White: depletion zone. Gray: Oxide insulator. Orange: Metal Gate. The two n-doped semiconductors are the source and drain terminals.*

And that is how a n-MOSFET works. Oh, and MOSFET stands for **M**etal on top of **O**xide Insulator on top of a **S**emiconductor- **F**ield **E**ffect **T**ransistor.

Remember:

$$N = V \int_0^\infty \mathrm{d}\varepsilon g(\varepsilon) n_F(\beta(\varepsilon - \mu))$$

$$N = \frac{2L^d}{(2\pi)^d} \int \mathrm{d}^d\boldsymbol{k}\, n_F(\beta(\varepsilon(\boldsymbol{k}) - \mu))$$

$$\langle E \rangle = V \int_0^\infty \mathrm{d}\varepsilon \varepsilon g(\varepsilon) n_F(\beta(\varepsilon - \mu))$$

$$\langle E \rangle = \frac{2L^d}{(2\pi)^d} \int \mathrm{d}^d\boldsymbol{k}\, \varepsilon(\boldsymbol{k}) n_F(\beta(\varepsilon(\boldsymbol{k}) - \mu))$$