Solid State Physics Notes

Shaojun Zan

12 April 2025

Contents

1   Crystal Structure Basics

1.1   Lattice & Bravais Lattice

A **lattice** is a recurring complex of atoms, ions, or molecules in a crystalline solid. A Bravais lattice, as the subset of lattices, is an abstract mathematical definition:

- Discrete Vector Space: A lattice is an infinite set of points defined by **integer sums** of a set of linearly independent vectors.

$$\mathbf{R}\left(n_1, n_2, n_3\right) = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \qquad n_1, n_2, n_3 \in \mathbb{Z} \tag{1}$$

  In the above equation, $\mathbf{R}$ is referred to as a lattice vector, and the linearly independent $\mathbf{a}_i$'s are primitive lattice vectors. In two and three dimensions, the choice of primitive lattice vectors is not unique.

- Closure: A lattice is an infinite set of vectors where addition of any two vectors in the set gives a third vector in the set.

  In this sense, 3D lattice vectors with vector addition form an Abelian $\mathbb{Z}^3$ group with three (non-unique) generators:

$$\mathbb{Z}^3 = \langle\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\rangle. \tag{2}$$

- Identical Ambience: A lattice is a set of points where the environment of any given point is equivalent to the environment of any other given point.

One should pay attention that, under this definition, a honeycomb is not a Bravais lattice, as it directly violates the identical ambience definition.
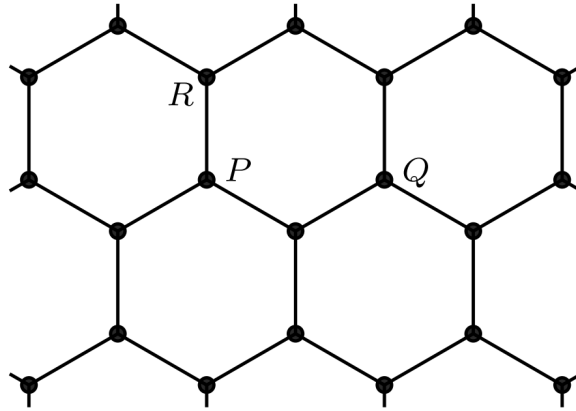
Figure 1: The lattice points P and Q obviously have different ambiences.

The Bravais lattice, as a mathematical construct, cannot by itself describe a real crystal. To reconstruct the actual crystal structure, we need to associate a specific group of atoms or molecules with each lattice point in an identical way — this is where the concept of unit cells and basis comes in.

1.2 Unit Cells & Basis

- A **unit cell** is a region of space such that when many idential units are stacked together it tiles (completely fills) all of space and reconstructs the full structure.

- A unit cell is the repeated motif which is the elementary building block of the periodic structure.

- A **primitive unit cell** for a periodic crystal is a unit cell containing exactly one lattice point.

  When a lattice point is on the boundary of the unit cell, it should only be counted fractionally depending on what fraction of the point is actually in the cell.

- Sometimes it is useful to define a unit cell which is not primitive in order to make it simpler to work with. This is known as a **conventional unit cell**. Almost always these conventional unit cells are chosen so as to have orthogonal axes.

- Given a lattice point, the set of all points in space which are closer to that given lattice point than to any other lattice point constitute the **Wigner–Seitz cell** of the given lattice point.
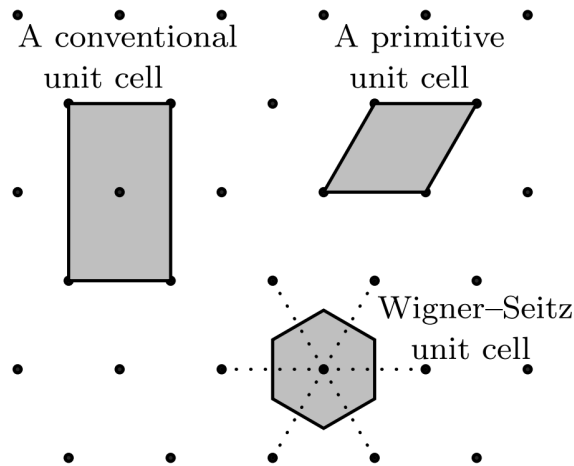
Figure 2: The perpendicular bisectors bound the Wigner-Seitz cell. This construction for a lattice always gives a primitive unit cell.

The description of objects in the unit cell with respect to the reference lattice point in the unit cell is known as a **basis**.

We think of reconstructing the entire crystal by associating with each lattice point a basis of atoms. The positions of the atoms in the crystal are "the lattice plus the basis".

In this sense, we may say that the honeycomb is not a lattice, but matter with a honeycomb structure can be crystals.
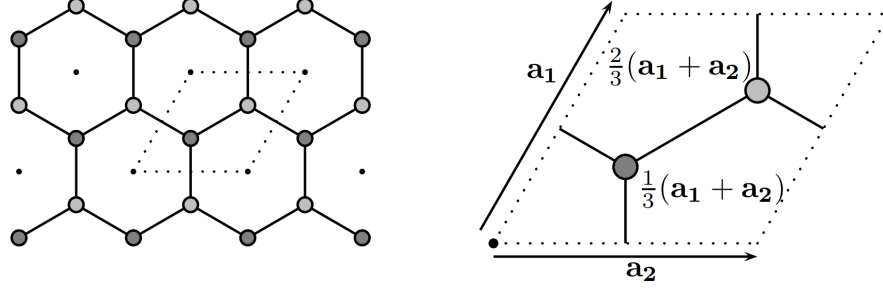
Figure 3: The honeycomb structure can be described as a triangular Bravais lattice with a two-atom basis.

There are 5 2D Bravais lattices and 14 3D Bravais lattices. We should pay attention to three of the 3D cases.

The simplest 3D lattice is of course the simple cubic lattice (SC). The body-centered cubic (BCC) lattice is a simple cubic lattice with an additional lattice point in the very center of the cube. Its conventional unit cell contains 2 lattice points and is thus not primitive. The face-centered cubic (FCC) lattice is a simple cubic lattice where there is an additional lattice point in the center of every face. Its conventional unit cell contains 4 lattice points.

	Body-centered Cubic (BCC)	Face-centered Cubic (FCC)
Wigner-Seitz Cell Shape	Truncated octahedron	Rhombic dodecahedron
Lattice Points in Conventional Cell	2	4
Interpretation	2 interpenetrating SC	4 interpenetrating SC

1.3   Reciprocal Lattice & Reciprocal Lattice Vectors

The reciprocal lattice is a (Bravais) lattice in reciprocal space. Given a lattice of points $\mathbf{R}$, a point $\mathbf{G}$ is a point in the reciprocal lattice iff

$$\exp(i\mathbf{G} \cdot \mathbf{R}) = 1 \tag{3}$$

for all points $\mathbf{R}$ of the direct lattice. $\mathbf{G}$ is called the reciprocal lattice vector.

The primitive lattice vectors of the reciprocal lattice $\mathbf{b}_i$ (**primitive reciprocal lattice vectors**) are defined by

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi \delta_{ij}. \tag{4}$$

By the cyclic quantity

$$V = \varepsilon_{ijk} \left(a_1\right)_i \left(a_2\right)_j \left(a_3\right)_k, \tag{5}$$

we find

$$\varepsilon_{ijk} \mathbf{b}_k = \frac{2\pi}{V} \mathbf{a}_i \times \mathbf{a}_j. \tag{6}$$

Therefore, an arbitrary point in reciprocal space may be represented as

$$\mathbf{G} = m_1 \mathbf{b}_1 + m_2 \mathbf{b}_2 + m_3 \mathbf{b}_3. \tag{7}$$

As the reciprocal space asks for $\exp\left(i\mathbf{G} \cdot \mathbf{R}\right) = 1$, we find that m_i must be integers.

2 Scattering Waves

2.1 Elastic Scattering

By my nonexistent memory for E&M, visible light cannot penetrate far into metals, and thus we use X-rays instead to probe into solids.

> By ignoring absorption and re-emission corresponding to atomic transitions, we assume that X-rays are scattered elastically.

Therefore, the photons before and after scattering will share the same energy and the same frequency, and, thus, the same magnitude of wavevector:

$$|\mathbf{k}_{in}| = |\mathbf{k}_{out}| \tag{8}$$

In general, the scattered direction is random, but when it comes to a periodic array (Bravais lattice), interference happens. Define a scattering wavevector $\mathbf{q}$ such that

$$\mathbf{q} \equiv \mathbf{k}_{out} - \mathbf{k}_{in}, \tag{9}$$

we may rewrite the elastic condition:

$$|\mathbf{k}_{out}|^2 = |\mathbf{k}_{in} + \mathbf{q}|^2 = |\mathbf{k}_{in}|^2$$
$$|\mathbf{k}_{in}|^2 + 2\mathbf{k}_{in} \cdot \mathbf{q} + |\mathbf{q}|^2 = |\mathbf{k}_{in}|^2$$
$$\Rightarrow 2\mathbf{k}_{in} \cdot \mathbf{q} + |\mathbf{q}|^2 = 0.$$

Therefore, we get the elastic condition:

$$2\mathbf{k}_{in} \cdot \mathbf{q} + |\mathbf{q}|^2 = 0. \tag{10}$$

2.2 Laue Condition

Let's examine when constructive interference takes place. One may assume, without loss of generality, that one unit cell is at the origin O and the other at $\mathbf{R}$ away.

Starting in the same phase, the plane waves may have the form

$$\exp(i\mathbf{k}_{in} \cdot \mathbf{r}). \tag{11}$$
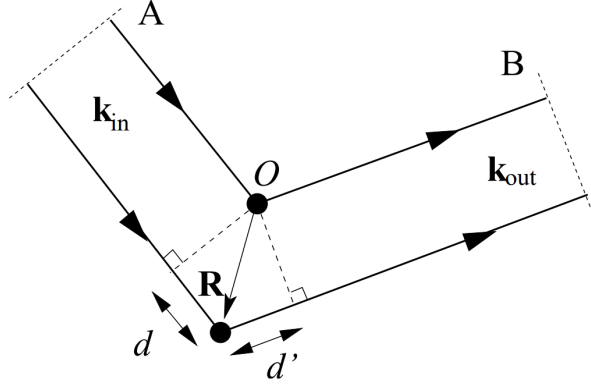
Figure 4: For constructive interference to occur, the waves must have the same phase on line B.

Clearly, there is a phase difference when the waves reach the unit cells:

$$\phi_{in}(\mathbf{R}) - \phi_{in}(0) = \mathbf{k}_{in} \cdot \mathbf{R}.$$

For the waves to arrive at line B with the same phase, we can think of the waves starting from B (just as A) with a wavevector $-\mathbf{k}_{out}$. This gives another phase difference:

$$\phi_{out}(\mathbf{R}) - \phi_{out}(0) = -\mathbf{k}_{out} \cdot \mathbf{R}.$$

Therefore, the total phase difference is the sum of the aforementioned two:

$$\Delta\phi(\mathbf{R}) = (\mathbf{k}_{in} - \mathbf{k}_{out}) \cdot \mathbf{R} \equiv -\mathbf{q} \cdot \mathbf{R}.$$

Thus, when a constructive interference occurs, the phase difference must be a multiple of 2π, or, in other words,

$$\exp(-i\mathbf{q} \cdot \mathbf{R}) = 1.$$

Note that the sign in the exponent doesn't matter now: $\exp(-i\mathbf{q} \cdot \mathbf{R}) = \exp(i\mathbf{q} \cdot \mathbf{R}) = 1$.

Isn't that suspiciously familiar? It must be. Therefore, we arrive at the Laue condition for constructive interference:

$$\mathbf{q} = \mathbf{G}. \tag{12}$$

2.3   Bragg Scattering & Ewald Sphere

Given the elastic condition and the Laue condition, we realize that waves are scattered in special directions by a crystal. This phenomenon is called **Bragg diffraction**. Combining the two conditions, we may write

$$2\mathbf{k}_{in} \cdot \mathbf{G} + |\mathbf{G}|^2 = 0. \tag{13}$$

However, if $\mathbf{G}$ is a reciprocal lattice vector, then so is $-\mathbf{G}$. We can as well write

$$2\mathbf{k}_{in} \cdot \mathbf{G} = |\mathbf{G}|^2. \tag{14}$$

Geometrically, we can make sure both conditions are satisfied through the Ewald method.
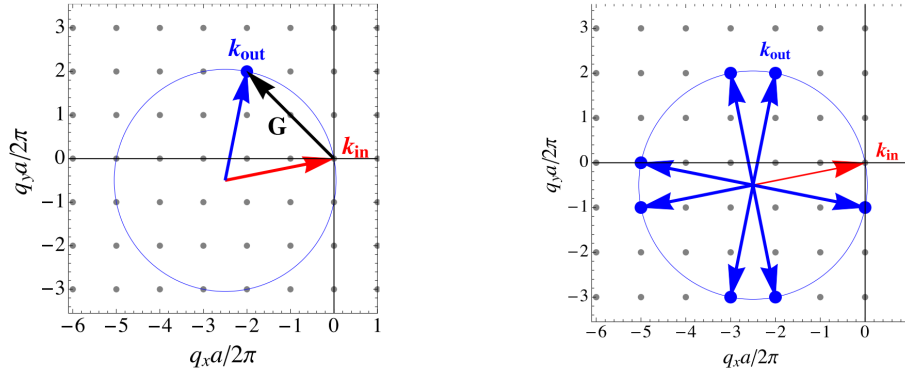
Figure 5: Take the top of the vector $\mathbf{k}_{in}$ at a point on the reciprocal lattice. A possible $\mathbf{k}_{out}$ is found if the sphere intersects another point on the reciprocal lattice.

2.4   Bragg Planes & Miller Indices

Bragg diffraction gives the reciprocal lattice vectors, $\mathbf{G}$, as the only allowed scattering wavevectors. This helps reclaim our definition of reciprocal lattice. The equation

$$\mathbf{G} \cdot \mathbf{R} = 2\pi m, \quad m \in \mathbb{Z}, \tag{15}$$

is also an equation for planes. We call these planes as **Bragg planes**.

A Bragg plane, which is a plane in real space, is completely specified if we specify $\mathbf{G}$.

Miller indices are a set of three integers (hkl) that uniquely label a family of Bragg planes in a crystal lattice. They are defined as the reciprocals of the fractional intercepts that a plane makes with the crystallographic axes, reduced to the smallest set of integers with no common factor.

2.5 Bragg's Law from Bragg Diffraction

Consider $\mathbf{g} = h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3$ with relatively prime h_1, h_2, and h_3. Consider the associated Bragg planes with equation $\mathbf{g} \cdot \mathbf{R} = 2\pi m$. This is a family of parallel planes. Consider a particular $\mathbf{r}$ parallel to $\mathbf{g}$. The equation then yields

$$rg = 2\pi m.$$

Therefore, the distance between each plane is given by

$$d = \frac{2\pi}{|\mathbf{g}|}. \tag{16}$$
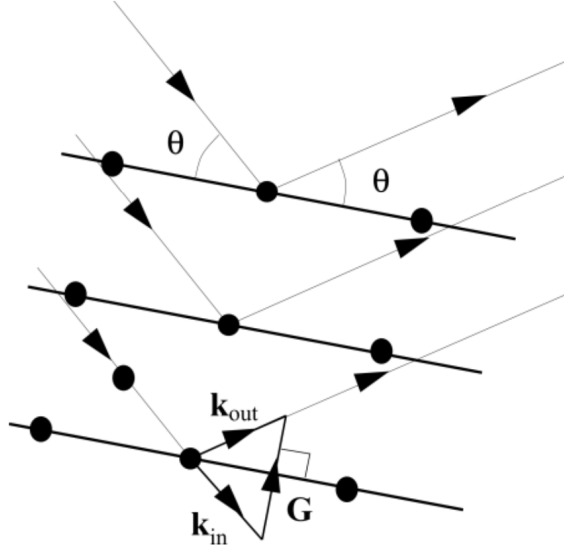
Figure 6: Bragg planes.

If scattering happens in this direction, we have the scattered wavevector as a integer multiple of $\mathbf{g}$, because its relatively prime indices suggest that it is the smallest vector in this direction. So, $\mathbf{q} = \mathbf{G} = N\mathbf{g}$. Then, by $\lambda = 2\pi/|\mathbf{k}_{in}|$,

$$2\mathbf{k}_{in} \cdot \mathbf{G} = |\mathbf{G}|^2$$

$$2|\mathbf{k}_{in}|\sin\theta = N|\mathbf{g}|$$

$$2\frac{2\pi}{|\mathbf{g}|}\sin\theta = N\frac{2\pi}{|\mathbf{k}_{in}|}.$$

We've reclaimed Bragg's law:

$$2d\sin\theta = N\lambda. \tag{17}$$

3 Lattice Vibrations & Phonons

3.1 Harmonic Approximation

Around equilibrium, the first derivative of the potential is zero. Therefore, perturbations lead to a nonzero second derivative. This explains the name of harmonic approximation. Consider $u = r - r_{eq}$.

$$V(r) = V(r_{eq}) + \frac{1}{2}V''(r_{eq})u^2 + \dots \tag{18}$$

Therefore,

$$F = -\frac{\mathrm{d}V}{\mathrm{d}r} = -V''(r_{eq})u, \tag{19}$$

a standard expression for the restoring force.

3.2 One-dimensional Monatomic Chain

For a 1D lattice, consider particle mass m with integer index n, Hooke constant κ, and equilibrium spacing a. Therefore, the equilibrium positions are $x_n^{eq} = na$. Let u_n be the displacement for the equilibrium:

$$u_n(t) \equiv x_n(t) - na. \tag{20}$$

> We assume that the interaction is dominated by the interaction between nearest neighbors.

3.2.1 Equation of Motion & Debye Frequency

The force on mass n due to mass $n+1$ is

$$F_{n,n+1} = \kappa\left(u_{n+1} - u_n\right). \tag{21}$$

Then, by Newton's third law, the force on mass n due to mass $n-1$ is

$$F_{n,n-1} = -F_{n-1,n} = \kappa\left(u_{n-1} - u_n\right). \tag{22}$$

Therefore, the total force on mass n is given by

$$m\ddot{x}_n(t) = F_{n,n+1} + F_{n,n-1} = \kappa\left(u_{n+1} - 2u_n + u_{n-1}\right). \tag{23}$$

Written in terms of u_n, we have

$$\frac{\mathrm{d}^2 u_n}{\mathrm{d}t^2} = \omega_D^2 \left(u_{n+1} - 2u_n + u_{n-1} \right),\tag{24}$$

where we define **Debye frequency** ω_D as the characteristic frequency of the chain's oscillations:

$$\omega_D \equiv \sqrt{\frac{\kappa}{m}}.\tag{25}$$

As these masses share the same parameters and the same equation of motion, we should view this oscillation as collective motion.

3.2.2 Propagating Wave Solutions & First Brillouin Zone (FBZ)

It's time to make some educated guesses. We try

$$u_{nk}(t) = \tilde{u}_k \exp(ikna - i\omega t).\tag{26}$$

This trial solution has a definite wavevector k, and we will find it correspond to a frequency as well, $\omega = \omega_k$. The relationship between ω and k is often called the **dispersion relation** or **frequency spectrum**. Anyway, to make a general solution, we should sum over different k:

$$u_n(t) = \mathrm{Re}\left[\sum_k u_{nk}(t) \right] = \mathrm{Re}\left[\sum_k \tilde{u}_k \exp(ikna - i\omega t) \right].\tag{27}$$

Before we move on, we must note that trial solutions with wavevector k and $k + 2\pi h/a$, $h \in \mathbb{Z}$ are not independent:

$$\exp\left[i\left(k + \frac{2\pi h}{a} \right) na - i\omega t \right] = \mathrm{e}^{i 2\pi n h} \exp(ikna - i\omega t) = \exp(ikna - i\omega t).$$

Therefore, to make sure we have independent solutions, we restrict k to a range of $2\pi/a$:

$$-\frac{\pi}{a} < k \leq \frac{\pi}{a}.\tag{28}$$

This is called the **first Brillouin zone** (FBZ) of the one-dimensional reciprocal space.

As for 1D, $G = 2\pi h/a$, we have $2\pi/a$ as the smallest nonzero reciprocal lattice vector. This makes the FBZ a Wigner-Seitz unit cell of this reciprocal lattice in k-space.

13

3.2.3 Dispersion Relation

Plugging our educated guesses in, we have

$$-\omega^2 u_{nk}(t) = \omega_D^2 u_{nk}(t) \left(e^{ika} + e^{-ika} - 2 \right).$$

For non-trivial solutions, we require $u_{nk}(t) \neq 0$, so

$$\omega^2 = 2\omega_D^2 \left(1 - \cos ka \right) = 4\omega_D^2 \sin^2 \left(\frac{ka}{2} \right).$$

Taking the square root, we get the dispersion relation for the monatomic chain:

$$\omega = 2\omega_D \left| \sin \left(\frac{ka}{2} \right) \right|, \qquad -\frac{\pi}{a} < k \leq \frac{\pi}{a}. \tag{29}$$

An astute reader may find that we've neglected the negative root. This is because after taking a real part, $\omega \to -\omega$ is equivalent to $k \to -k$. As the FBZ is symmetric with respect to the origin, the negative root is thus redundant.
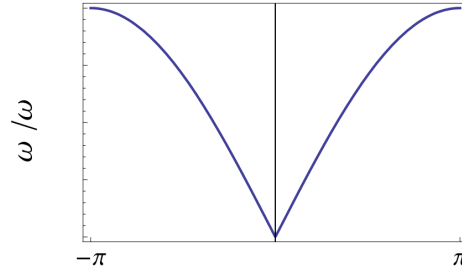
Figure 7: The dispersion relation ω plotted against k. This plot is not professional nor ethical.

3.2.4 Short Wavelengths: Collective Motion

Short wavelengths correspond to large wavenumbers. At the edge of FBZ, $ka = \pm\pi \Rightarrow \lambda = 2a$, we see

$$u_n = u_{nk} \Big|_{k=\pm\pi/a} = \tilde{u}_k \cos\left(\pm\pi n - \omega_k t \right) = (-1)^n \tilde{u}_k \cos\left(\omega_k t \right). \tag{30}$$

We also see that $k = \pi/a$ and $k = -\pi/a$ correspond to the same solution. This solution represents that the even-n atoms moving in opposite direction to the odd-n atoms. This is also the highest energy mode.

3.2.5 Long Wavelength Modes

Firstly, as $k \to 0$, $\omega_k \to 0$. Therefore, this limit sorresponds to a uniform displacement: $u_n = \text{const}$.
Also, we shall notice that

$$\omega = 2\omega_D \left| \sin\left(\frac{ka}{2} \right) \right| \approx \omega_D a |k| = v|k|, \qquad |k|a \ll 1. \tag{31}$$

We say that for small k, the dispersion relation is linear.

3.3 Three-dimensional Generalization

The 1D equidistant lattice points generalizes to the <u>simple cubic</u> in 3D. So, for a simple cubic lattice,

$$\omega_{\mathbf{k}}^2 = 4\omega_D^2 \left[\sin^2\left(\frac{k_x a}{2} \right) + \sin^2\left(\frac{k_y a}{2} \right) + \sin^2\left(\frac{k_z a}{2} \right) \right], \qquad -\frac{\pi}{a} < k_i \leq \frac{\pi}{a}. \tag{32}$$

Also, the dispersion relation is linear in $|\mathbf{k}|$ near the origin.
The general solution should be written as

$$u_{\mathbf{R}}(t) = \sum_{\mathbf{k}} \tilde{u}_{\mathbf{k}} \exp\left(i\mathbf{k} \cdot \mathbf{R} - i\omega_{\mathbf{k}} t \right). \tag{33}$$

We may check that $\mathbf{k} \to \mathbf{k} + \mathbf{G}$ doesn't give an independent solution. This must be true due to the definition of reciprocal lattice.

As a reminder, the method to determine Brillouin zones is completely the same as determining Wigner-Seitz cells.

3.4 Boundary Conditions

	Box BC	Periodic (Born-von Karman) BC
Mathematical Description	$\psi(0) = \psi(L) = 0$	$\psi(x) = \psi(x + L)$
Allowed Plane Waves	$\psi(x) \sim \sin\left(\frac{n\pi}{L} \right)$	$\psi(x) \sim \exp\left(i\frac{2\pi n}{L} x \right)$
Allowed Wavevectors	$k = \frac{n\pi}{L}, \qquad n \in \mathbb{N}$	$k = \frac{2\pi n}{L}, \qquad n \in \mathbb{Z}$

Box boundary conditions require that the motion of waves on the boundary must be zero. On contrast, **periodic**, or **Born-von Karman**, boundary conditions only require that the wave motion is the same on

either ends of the wall.

In our case of monatomic chain with N atoms, $n+N$ and n refer to the same atom, so we have a periodic BC:

$$u_{n+N}(t) = u_n(t). \tag{34}$$

This means that our possible choices of wavevector k are quantized: with $L = Na$,

$$k = m\frac{2\pi}{Na}, \qquad m \in \mathbb{Z}, \qquad -\frac{\pi}{a} < k \le \frac{\pi}{a}. \tag{35}$$

This leaves N possible choices of k for either even or odd N — the number of allowed wavevectors is equal to the number of masses (unit cells) in the chain.

3.5 Density of States in Reciprocal Space

Carrying on from our allowed wavevectors in the table, we find that each wavevector takes a volume in the reciprocal space. The **density of states**, g, is the reciprocal of this volume.

> **Boundary Conditions Have No Influence on the Total Number of States**
>
> One may find $g_b = \pi/L$ for box BC and $g_p = 2\pi/L$ for periodic BC. This seems like a discrepancy, but remember that box BC only allows positive wavenumbers, i.e. half the allowed wavenumbers for periodic BC. Thus, the total number of states is the same.
>
> Boundary conditions may constrain the phase of the wavefunction, but **they will not change the spacing of the orthogonal states**.
>
> From now on, we will stick to
> $$g(k) = \frac{L}{2\pi}. \tag{36}$$

In 3D, we have

$$g(\mathbf{k}) = \frac{L_x}{2\pi} \cdot \frac{L_y}{2\pi} \cdot \frac{L_z}{2\pi} = \frac{V}{(2\pi)^3}. \tag{37}$$

Then, we may check the validity of this result. Consider the cubic Brillouin zone for a simple cubic lattice with side $2\pi/a$.

$$N_{\text{states}} = \int_{\text{FBZ}} g(\mathbf{k})\, \mathrm{d}^3\mathbf{k} = \frac{V}{(2\pi)^3}\frac{(2\pi)^3}{a^3} = \frac{V}{a^3} = N_{\text{unit cells}}. \tag{38}$$

This clearly agrees with our discovery in 1D monatomic chain, where we said "the number of modes is the same as the number of unit cells (atoms)."

In 3D, there should be $3N$ mechanical degrees of freedom. Our analysis above only accounts for one branch or polarization of possible vibrational modes. The counting analysis applies to each branch separately, and finally giving $3N$ as the total number of modes.

$$\text{Number of states for each vibrational mode} = \text{Number of unit cells} \qquad (39)$$

In the thermodynamic limit, we assume that the wavevectors are so densely spaced that we can convert a sum into an integral. In general, for any smooth function $F(\mathbf{k})$,

$$\sum_{\text{FBZ}} F(\mathbf{k}) = \sum_{\text{FBZ}} \left(\prod_{i=1}^{n} \frac{\Delta k_i}{2\pi/L} \right) F(\mathbf{k}) \equiv \frac{V}{(2\pi)^n} \int_{\text{FBZ}} \mathrm{d}^n \mathbf{k} \, F(\mathbf{k}),$$

where we've used the fact that $\Delta k = 2\pi/L$.

This is our formula to convert a sum to an integral in the FBZ:

$$\sum_{\text{FBZ}} F(\mathbf{k}) = \frac{V}{(2\pi)^n} \int_{\text{FBZ}} \mathrm{d}^n \mathbf{k} \, F(\mathbf{k}). \qquad (40)$$

3.6 Phonons

The concept of a phonon might be easier to understand through "quantizing" our classical picture.

Classical Picture	Quantum Picture
Every mode of $\mathbf{k}$ represents a SHM.	Every mode of $\mathbf{k}$ is a quantum harmonic oscillator.
Frequency $\omega_{\mathbf{k}}$	Energy $\hbar\omega_{\mathbf{k}}$ (second quantization)
Continuous energy	Discrete energy: $E_n = n\hbar\omega_{\mathbf{k}}, \quad n \in \mathbb{N}$
Collective motion	The quantized particles of collective excitations $\equiv$ phonons

Phonons v.s. Quantum Harmonic Oscillators

In a 3D crystal with N atoms, there are $3N$ degrees of freedom, each of which corresponds to a quantum harmonic oscillator. These $3N$ degrees of freedom are associated with $3N$ normal modes of vibration, and each normal mode can be quantized, leading to $3N$ phonons. Each phonon represents a quantized excitation of one of these modes for the N atoms.

To reiterate, a **phonon** is a <u>collective excitation</u> in a lattice. One must pay attention to the word collective here, as it defines the difference between a phonon and a quantum harmonic oscillator.

Also, the word excitation insinuates the fact **the ground state of the collective motion is not counted as a phonon**. This also implies that the phonon number is not conserved, and that the chemical potential for phonons is zero.

With this quantum description of vibrations of a solid, each phonon at wavevector $\mathbf{k}$ has energy $E_{\mathbf{k}} = \hbar\omega_{\mathbf{k}}$ and momentum $\mathbf{p_k} = \hbar\mathbf{k}$. However, this is not quite the true momentum. As we have restricted $\mathbf{k}$ to the FBZ, we refer to this momentum as **crystal momentum**, whose addition is in fact **addition modulo the reciprocal lattice vector**. Despite this fact, the gradient of energy with respect to crystal momentum still gives the correct group velocity.

$$\boldsymbol{\nabla}_{\mathbf{p}} E = \boldsymbol{\nabla}_{\hbar\mathbf{k}}(\hbar\omega_{\mathbf{k}}) = \boldsymbol{\nabla}_{\mathbf{k}}\omega_{\mathbf{k}} \equiv \mathbf{v}_g \tag{41}$$

3.7   Thermodynamics & Density of States in Energy Space

The phonons are clearly bosons — normal modes are independent of each other. Therefore, phonons submit to Bose-Einstein statistics: with $\mu = 0$,

$$f(E, T) = \frac{1}{e^{E/k_{\mathrm{B}}T} - 1}. \tag{42}$$

Now we can consider the energy in the thermodynamic limit, $\Delta U(T) \equiv U(T) - U(0)$:

$$\Delta U(T) = \sum_{\mathbf{k} \in \mathrm{FBZ}} f(E_{\mathbf{k}}, T) E_{\mathbf{k}} \equiv \frac{V}{(2\pi)^3} \int_{\mathrm{FBZ}} \mathrm{d}^3\mathbf{k}\, \frac{E_{\mathbf{k}}}{\exp\left(E_{\mathbf{k}}/k_{\mathrm{B}}T\right) - 1}. \tag{43}$$

An integral hideous as such should not deserve an analytical solution, and this turns out to be the truth. In the high temperature regime, we have $\exp\left(E_{\mathbf{k}}/k_{\mathrm{B}}T\right) \approx 1 + E_{\mathbf{k}}/k_{\mathrm{B}}T$, and this integral reduces to

$$\Delta U(T) \approx \frac{V}{(2\pi)^3} \int_{\mathrm{FBZ}} \mathrm{d}^3\mathbf{k}\, k_{\mathrm{B}}T = N k_{\mathrm{B}}T, \qquad k_{\mathrm{B}}T \gg E_{\mathbf{k}}, \tag{44}$$

as expected.

In the low temperature regime, first note that we may define a temperature corresponding to a frequency:

$$T_0 = \frac{\hbar\omega_0}{k_{\mathrm{B}}} \quad \text{such that} \quad \exp\left(\frac{E_0}{k_{\mathrm{B}}T}\right) = e^{T_0/T}.$$

Therefore, when $T < T_0$, this exponential blows up, and as it's in the denominator of the integrand, we say the corresponding mode $\mathbf{k}$ doesn't make a dominant contribution to the integral. Therefore, by the

dispersion relation for a simple cubic lattice,

$$\omega_{\mathbf{k}}^2 = 4\omega_D^2 \sum_i \sin^2\left(\frac{k_i a}{2}\right),$$

we know that the dominant contribution is by those with small $|\mathbf{k}|$ (large wavelengths). In this limit,

$$E_{\mathbf{k}} = \hbar\omega_{\mathbf{k}} \approx \hbar\omega_D |\mathbf{k}| a.$$

Therefore,

$$\begin{aligned}
\Delta U(T) &= \frac{V}{(2\pi)^3} \int_{\text{FBZ}} \mathrm{d}^3\mathbf{k}\, \frac{E_{\mathbf{k}}}{\exp\left(E_{\mathbf{k}}/k_{\mathrm{B}}T\right) - 1} \\
&= \frac{V}{(2\pi)^3} \int_{\text{FBZ}} 4\pi k^2\, \mathrm{d}k\, \frac{\hbar\omega_D k a}{\exp\left(\hbar\omega_D k a/k_{\mathrm{B}}T\right) - 1} \\
&\approx V \int_0^\infty \mathrm{d}k\, \frac{4\pi k^2}{(2\pi)^3} \frac{\hbar\omega_D k a}{\exp\left(\hbar\omega_D k a/k_{\mathrm{B}}T\right) - 1}, \qquad k_{\mathrm{B}}T \ll \hbar\omega_D.
\end{aligned}$$

By substitution of $x = \hbar\omega_D k a/k_{\mathrm{B}}T$, we find a standard integral. Should you meet this in the exam, improvise with mathematics. Nevertheless, we should emphasize that this gives a heat capacity $C_V \propto T^3$.

It might be easier if we can compute integrals in the energy space. By

$$\mathrm{d}N_{\text{states}} = g(E)\,\mathrm{d}E = g(\mathbf{k})\,\mathrm{d}^3\mathbf{k} = g(k)\,\mathrm{d}k\,, \tag{45}$$

with $g(\mathbf{k}) = V/(2\pi)^3$ and $E_k = \hbar\omega_D a k$ for long-wavelength phonons,

$$g(E) = g(\mathbf{k})\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}k}\frac{\mathrm{d}k}{\mathrm{d}E} = \frac{V}{(2\pi)^3}4\pi k^2 \frac{1}{\hbar\omega_D a} = N\frac{E^2}{2\pi^2\left(\hbar\omega_D\right)^3}, \qquad E \ll \hbar\omega_D. \tag{46}$$

In general, there should also be a term d due to degeneracy. We may incorporate this in our definition of the extensive density of states $G(E) \equiv V g(E)$. Note carefully the integration limit as FBZ.

$$N_{\text{states}} = \int_{\text{FBZ}} \mathrm{d}E\, G(E) \equiv V \int_{\text{FBZ}} \mathrm{d}E\, g(E) \tag{47}$$

$$N_{\text{particles}} = V \int_{\text{FBZ}} \mathrm{d}E\, f(E)g(E), \tag{48}$$

$$\Delta U(T) = V \int_{\text{FBZ}} \mathrm{d}E\, E f(E)g(E). \tag{49}$$

The limit should be intuitive: there are no arbitrarily high energy phonons due to the finite dispersion relation. Phonons are not after all free particles; they aren't even real.

4 Electrons in Solids: Models & Band Structure

4.1 Free Electron Model (Fermi Gas Model)

- Non-interacting mobile electrons which may conduct electricity — "conduction electrons";

- Pauli exclusion principle;

- Fermi-Dirac statistics.

Fermi Energy v.s. Fermi Level

The **Fermi energy**, E_F, is defined as the energy of the highest occupied state at $T = 0\,\text{K}$.

The **Fermi level** is just a synonym for chemical potential, defined at various temperatures. You bloody physicists.

The energy for an electron with momentum $\hbar k$ is

$$E_F = \frac{\hbar^2 k^2}{2m}. \tag{50}$$

At $T = 0\,\text{K}$, the Fermi energy is equal to the Fermi level, E_F. This corresponds to the Fermi vector k_F, and these vectors form a sphere called the Fermi sphere. All the states at the Fermi level lie on the Fermi surface. The Fermi velocity is v_F is just $mv_F = \hbar k_F$.

We can calculate the number of electrons. In 3D,

$$\frac{N}{2} = \frac{4\pi k_F^2}{3} \frac{V}{(2\pi)^3}, \tag{51}$$

where the factor of 2 accounts for degeneracy. Defining the number density $n = N/V$, we have

$$k_F = \left(3\pi^2 n\right)^{1/3}, \qquad E_F = \frac{\hbar^2}{2m} \left(3\pi^2 n\right)^{2/3}. \tag{52}$$

Considering that

$$N = V \int_0^\infty \mathrm{d}E\, f(E) g(E),$$

$$n = \int_0^\infty \mathrm{d}E\, f(E) g(E). \tag{53}$$

At temperature $T = 0\,\mathrm{K}$, we know that the Fermi-Dirac distribution becomes a step function:

$$n = \int_0^{E_F} \mathrm{d}E\, g(E).$$

This should coincide with the direct calculation.

Considering the change to the Fermi-Dirac distribution as we raise the temperature, we can conclude that only the states near the Fermi level ($\sim k_\mathrm{B}T$) play a role in the physical properties of the system.

4.1.1   Interpretation

The Fermi-Dirac distribution changes most drastically with energies $\delta\epsilon \simeq k_\mathrm{B}T$ near the Fermi level. For $T = 0$, this is $k_\mathrm{B}T$ below the Fermi level. This change in Fermi-Dirac distribution of free electrons is assumed as the only difference in the material at various temperatures. Thus, these thermally excited electrons are assumed the only contributors to the heat capacity.

Let's begin with electronic contribution to the heat capacity around the Fermi level. One can see that

$$C_V \propto \delta N = \int_{E_F - k_\mathrm{B}T}^{E_F + k_\mathrm{B}T} \mathrm{d}E\, G(E_F) \sim G(E_F) k_\mathrm{B}T. \tag{54}$$

Therefore,

$$U(T) - U(0) = U(T) \simeq G(E_F) k_\mathrm{B}^2 T^2,$$

and thus

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V \approx 2 G(E_F) k_\mathrm{B}^2 T.$$

By explicitly calculating $G(E_F)$ and N in terms of E_F, we have

$$G(E_F) = \frac{3N}{2E_F}, \qquad C_V \simeq 3 N k_\mathrm{B} \frac{k_\mathrm{B}T}{E_F}. \tag{55}$$

Therefore, we have our model of heat capacity for simple electrons:

$$C_V(T) = \gamma T + A T^3, \qquad k_\mathrm{B}T \ll E_F, \tag{56}$$

where the T^3 term comes from acoustic phonons, and T, from electronic properties.

As

$$\gamma \propto \frac{1}{E_F} \propto m, \tag{57}$$

this term is often referred to as an effective mass for the electron (higher than the rest mass of electron).

4.1.2 Empty Lattice Approximation

We might consider that the free electronic dispersion repeated at all reciprocal lattice shifts would represent the solutions of electrons in metals to a first approximation. This is called the **empty lattice approximation**.

In 1D, consider

$$E = \frac{\hbar^2 \left(k + G\right)^2}{2m}, \qquad G = m\frac{2\pi}{a}, \qquad m \in \mathbb{Z}.$$
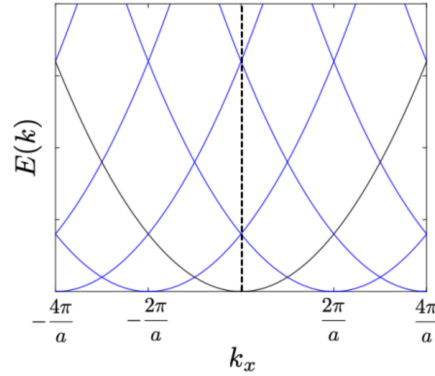
Figure 8: The free electron dispersion shifted by reciprocal lattice vectors.

Consider the intersection points, for example, between the branches k and $k + G$:

$$k^2 = (k + G)^2 \quad \Rightarrow \quad 2kG + G^2 = 0.$$

Isn't that familiar? Yes. We find that the k at intersection points are actually wavevectors that satisfy the scattering condition (Bragg points). Therefore, wave propagation cannot occur there.
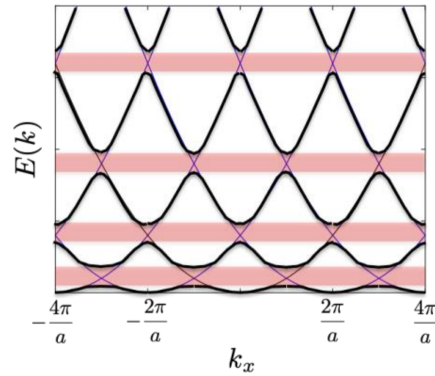
Figure 9: The Nearly free electron dispersion.

We can zoom in the lowest two bands, shown in the graph below.

Figure 10: Electronic dispersion within first 2 Brillouin zones.

We could choose a Fermi energy, $E_F^{(1)}$, within the lowest energy band. i.e., electrons fill all states up to this energy. We could draw the Fermi surface as shown in blue, where the shading shows which of the electronic states are occupied in k-space. However, if we choose a Fermi energy, $E_F^{(2)}$, within the second electron band, then the corresponding free-electron Fermi surface would have crossed the boundaries of the first Brillouin zone. Since electrons 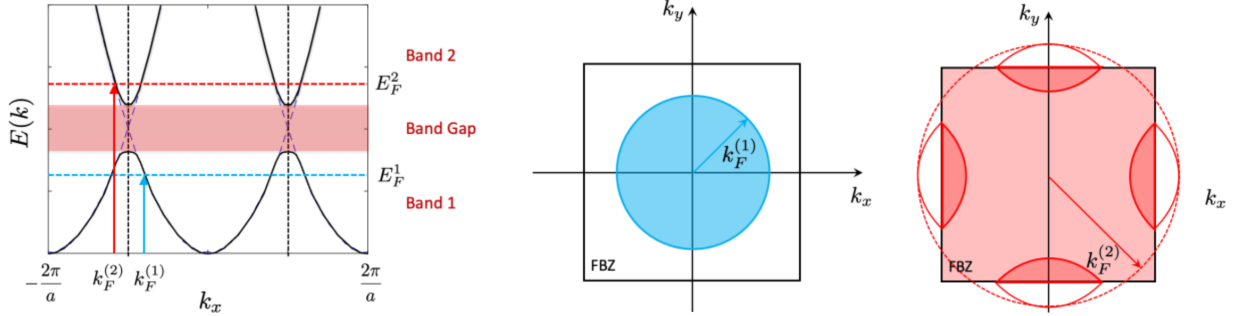in a real metal cannot propagate at the first Brillouin zone boundary due to scattering ($2kG + G^2 = 0$), the Fermi surface is distorted as shown in the right. Electrons completely fill the first band of states, thus filling the FBZ. Those states in the 2nd Brillouin zone can then be folded back into the FBZ (by shifts of G) as a new band of states, which only partially occupies regions of k-space near the zone boundaries. The partially transparent shading shows the filling of the two bands and where filled k-states of the two bands overlap.

It is very clear right now that our theory needs adjustments. Therefore, we move on to the tight binding model.

4.2 The Tight Binding Model

Having stepped from free electron model, we now realize that it's important to consider the potentials. The physics of this problem alludes to a periodic potential $V(\mathbf{r}) = V(\mathbf{r} + \mathbf{R})$, and

$$\hat{\mathcal{H}}\psi(\mathbf{r}) = \left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}).$$

4.2.1 Double Well, Tight Binding Approximation, and Hopping Energy

We assume infinite square wells at each atomic site, and restrict ourselves to the regime where the atoms are still far apart compared to the size of atomic orbitals.

For a double well in 1D, we write

$$\hat{\mathcal{H}}_{\text{double}}\psi(x) = \left[-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V_L(x) + V_R(x) \right]\psi(x) = E_\alpha\psi(x), \tag{58}$$

where $V_L(x) = V(x + a/2)$ and $V_R(x) = V(x - a/2)$ are the single wells centered at $x = -a/2$ and $x = a/2$, respectively.

We suppose that the wells are widely separated such that the solution is a linear combination of the single-well solutions u_α:

$$\psi_\alpha(x) \simeq c_L u_L(x) + c_R u_R(x)$$
$$\equiv c_L u\left(x + \frac{a}{2}\right) + c_R u\left(x - \frac{a}{2}\right). \tag{59}$$

The reduction of quantum motion in the continuum, described by a continuous wavefunction, $\psi(\mathbf{r})$, to a discrete problem about the probability amplitudes, c_i, of finding an electron in each unit cell is the central simplification of the tight binding model. This method is called time-independent degenerate perturbation theory.

As we define

$$\hat{\mathcal{H}}_{\text{single}}u_{L/R}(x) = \left[-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V_{L/R}(x) \right]u_{L/R}(x) = E_\alpha u_{L/R}(x),$$

we find that

$$\hat{\mathcal{H}}_{\text{double}}u_{L/R} = \left[\hat{\mathcal{H}}_{\text{single}}^{(L/R)} + V_{R/L} \right]u_{L/R} = \left(E_\alpha + V_{R/L} \right)u_{L/R}.$$

Therefore, the double-well Schrödinger equation can be written as

$$c_L\left(E_\alpha + V_R \right)u_L + c_R\left(E_\alpha + V_L \right)u_R = E\left(c_L u_L + c_R u_R \right). \tag{60}$$

Now, by assuming normality, take the inner product with u_L and u_R respectively:

$$c_L\left(E_\alpha + \int \mathrm{d}x\, |u_L|^2 V_R \right) + c_R\left(E_\alpha \int \mathrm{d}x\, u_L^* u_R + \int \mathrm{d}x\, u_L^* u_R V_L \right) = E\left(c_L + c_R \int \mathrm{d}x\, u_L^* u_R \right),$$

$$c_R\left(E_\alpha + \int \mathrm{d}x\, |u_R|^2 V_L \right) + c_L\left(E_\alpha \int \mathrm{d}x\, u_R^* u_L + \int \mathrm{d}x\, u_R^* u_L V_R \right) = E\left(c_L \int \mathrm{d}x\, u_R^* u_L + c_R \right).$$

We assume the terms in blue and green vanish.

Invoking the symmetry of this problem, we see that the terms in orange are equal. They are referred to

as the **hopping integral** or **hopping energy**:

$$t_\alpha = -\int \mathrm{d}x\, u_R^* u_L V_R = -\int \mathrm{d}x\, u_L^* u_R V_L. \tag{61}$$

This minus sign is just the convention; the value of the hopping integral can be either positive or negative based on the parity of u_α. With all these simplifications, we reduce the Schrödinger equation into two coupled equations:

$$\begin{pmatrix} E_\alpha & -t_\alpha \\ -t_\alpha & E_\alpha \end{pmatrix} \begin{pmatrix} c_L \\ c_R \end{pmatrix} = E \begin{pmatrix} c_L \\ c_R \end{pmatrix}. \tag{62}$$

Solving for the eigenvalues, we get $E = E_\alpha \pm t_\alpha$ — an energy splitting of $2|t_\alpha|$. This result is valid if $|t_\alpha| \ll |E_\alpha - E_{\alpha \pm 1}|$.

4.2.2   One-dimensional Array

We're now prepared for an infinite number of 1D square wells. With lattice constant a, we may write our solution as a linear combination of individual single-well solutions:

$$\psi_\alpha(x) = \sum_n c_n u_{\alpha n}(x) = \sum_n c_n u_\alpha(x - na). \tag{63}$$

Plugging this into the Schrödinger equation, we get

$$\sum_m \left[E_\alpha + \sum_{k \neq m} V_k \right] c_m u_{\alpha m} = E \sum_m c_m u_{\alpha m}.$$

We assume orthonormality,

$$\int \mathrm{d}x\, u_{\alpha m}^* u_{\alpha n} = \delta_{mn},$$

and that interaction is dominated by interaction with neighbors, i.e.

$$\sum_{k \neq m} \int \mathrm{d}x\, u_{\alpha n}^* u_{\alpha m} V_k = -t_\alpha \left[\delta_{m(n-1)} + \delta_{m(n+1)} \right].$$

Taking the inner product with respect to $u_{\alpha n}$, we get

$$\sum_m c_m \left[E_\alpha \int \mathrm{d}x\, u_{\alpha n}^* u_{\alpha m} + \sum_{k \neq m} \int \mathrm{d}x\, u_{\alpha n}^* u_{\alpha m} V_k \right] = E \sum_m c_m \int \mathrm{d}x\, u_{\alpha n}^* u_{\alpha m},$$

$$\sum_m c_m \left[E_\alpha \delta_{mn} - t_\alpha \left(\delta_{m(n-1)} + \delta_{m(n+1)} \right) \right] = E \sum_m c_m \delta_{mn},$$

and, finally, the **discrete Schrödinger equation**:

$$E_\alpha c_n - t_\alpha c_{n-1} - t_\alpha c_{n+1} = E c_n. \tag{64}$$

4.2.3 Wave-like Solutions

Consider the trial solution

$$c_n = A \mathrm{e}^{ikna}, \qquad -\frac{\pi}{a} < k \leq \frac{\pi}{a}.$$

This gives us, for a particular k,

$$A \mathrm{e}^{ikna} \left(E_\alpha - t_\alpha \mathrm{e}^{-ika} - t_\alpha \mathrm{e}^{ika} \right) = E A \mathrm{e}^{ikna}.$$

As c_n cannot vanish, we have derived the associated energy $E_{\alpha k}$. Therefore,

$$\psi_k(x) = \sum_n c_n u_\alpha(x - na) = A \sum_n \mathrm{e}^{ikna} u_\alpha(x - na), \tag{65}$$

$$E_{\alpha k} = E_\alpha - 2 t_\alpha \cos(ka), \qquad -\frac{\pi}{a} < k \leq \frac{\pi}{a}. \tag{66}$$

If we require $c_{n+N} = c_n$, this will give a discrete set of allowed wavevectors.

4.2.4 Higher Dimensions

By the linear combination,

$$\psi(\mathbf{r}) = \sum_{\mathbf{R}} c_{\mathbf{R}} u_\alpha(\mathbf{r} - \mathbf{R}). \tag{67}$$

Only the nearest neighbors contribute to the hopping integral, so the discrete Schrödinger equation is

$$E_\alpha c_{\mathbf{R}} - t_\alpha \sum_{\boldsymbol{\delta}} c_{\mathbf{R}+\boldsymbol{\delta}} = E c_{\mathbf{R}}, \tag{68}$$

where the set of vectors $\boldsymbol{\delta}$ contains all the vectors joining $\mathbf{R}$ to its nearest neighbors.

The trial solution is

$$c_{\mathbf{R}} = A e^{i\mathbf{k}\cdot\mathbf{R}}, \qquad \mathbf{k} \in \text{FBZ}. \tag{69}$$

Plugging this trial solution gives

$$A e^{i\mathbf{k}\cdot\mathbf{R}} \left(E_\alpha - t_\alpha \sum_{\boldsymbol{\delta}} e^{i\mathbf{k}\cdot\boldsymbol{\delta}} \right) = A e^{i\mathbf{k}\cdot\mathbf{R}} E.$$

Therefore,

$$\psi(\mathbf{r}) = A \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} u_\alpha(\mathbf{r} - \mathbf{R}), \tag{70}$$

$$E = E_\alpha - t_\alpha \sum_{\boldsymbol{\delta}} e^{i\mathbf{k}\cdot\boldsymbol{\delta}}, \qquad \mathbf{k} \in \text{FBZ}. \tag{71}$$

As the total number of allowed wavevectors in the FBZ is equal to the number of unit cells, N unit cells would imply N distinct eigenstates, and thus $2N$ electrons. We say that each tight-binding band is full when there are 2 electrons per unit cell.

4.2.5 Bloch's Theorem

For any eigenstate $\psi(\mathbf{r})$ of a periodic system,

$$\psi_{\alpha k}(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}} \psi_{\alpha k}(\mathbf{r}), \qquad \mathbf{k} \in \text{FBZ}, \tag{72}$$

where α is the band index.

For electrons in a periodic potential, we should restrict $\mathbf{R}$ to the lattice vectors of the crystal. An equivalent statement to Bloch's theorem is

$$\psi_{\alpha k}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \phi_{\alpha k}(\mathbf{r}), \qquad \mathbf{k} \in \text{FBZ}. \tag{73}$$

A simple corollary is

$$|\psi_{\alpha k}(\mathbf{r} + \mathbf{R})|^2 = |\psi_{\alpha k}(\mathbf{r})|^2. \tag{74}$$

4.3 Semi-classical Theory of Bloch Electron Transport

4.3.1 Crystal Momentum & Momentum of Electrons

Consider Bloch's solution,

$$\psi_{\alpha k}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\phi_{\alpha k}(\mathbf{r}).$$

Electron momentum is the eigenvalue yielded by the momentum operator:

$$\hat{\mathbf{p}}\psi_{\alpha k} = \hbar\mathbf{k}e^{i\mathbf{k}\cdot\mathbf{r}}\phi_{\alpha k} - i\hbar e^{i\mathbf{k}\cdot\mathbf{r}}\boldsymbol{\nabla}_{\mathbf{k}}\phi_{\alpha k} = e^{i\mathbf{k}\cdot\mathbf{r}}\left(\hat{\mathbf{p}} + \hbar\mathbf{k}\right)\phi_{\alpha k}(\mathbf{r}).$$

Clearly, $\hat{\mathbf{p}}\phi_{\alpha k} \neq 0$, so the momentum of electrons does not equal to the crystal momentum.

4.3.2 Effective Mass of Electrons

Clearly, we see

$$\hat{\mathbf{p}}^2\psi_{\alpha k} = e^{i\mathbf{k}\cdot\mathbf{r}}\left(\hat{\mathbf{p}} + \hbar\mathbf{k}\right)^2\phi_{\alpha k}(\mathbf{r}).$$

Therefore, dropping the $e^{i\mathbf{k}\cdot\mathbf{r}}$ term, we have

$$\left[\frac{\left(\hat{\mathbf{p}} + \hbar\mathbf{k}\right)^2}{2m} + V(\mathbf{r})\right]\phi_{\alpha k}(\mathbf{r}) = E_\alpha\phi_{\alpha k}(\mathbf{r}).$$

Consider a small perturbation, $\mathbf{k} \to \mathbf{k} + \mathbf{q}$:

$$\left(\hat{\mathbf{p}} + \hbar\mathbf{k} + \hbar\mathbf{q}\right)^2 = \left(\hat{\mathbf{p}} + \hbar\mathbf{k}\right)^2 + 2\hbar\mathbf{q}\cdot\left(\hat{\mathbf{p}} + \hbar\mathbf{k}\right) + \hbar^2\mathbf{q}^2.$$

The electron energy may also be expanded:

$$E_\alpha(\mathbf{k} + \mathbf{q}) \approx E_\alpha(\mathbf{k}) + \mathbf{q}\cdot\boldsymbol{\nabla}_{\mathbf{k}}E_\alpha(\mathbf{k}) + \frac{1}{2}\sum_{ij}\frac{\partial^2 E_\alpha}{\partial k_i \partial k_j}q_i q_j + \ldots.$$

Equating the first-order terms, we get

$$\frac{\hbar}{m}\left(\hat{\mathbf{p}} + \hbar\mathbf{k}\right) = \boldsymbol{\nabla}_{\mathbf{k}}E_\alpha(\mathbf{k}). \tag{75}$$

As we know the RHS as the group velocity, this gives us a way to estimate the group velocity. Consider

$$\boldsymbol{\nabla}_{\mathbf{k}}E_\alpha(\mathbf{k})\phi_{\alpha k}(\mathbf{r}) = \frac{\hbar}{m}\left(\hat{\mathbf{p}} + \hbar\mathbf{k}\right)\phi_{\alpha k}(\mathbf{r}) = \frac{\hbar}{m}e^{-i\mathbf{k}\cdot\mathbf{r}}\left(\hat{\mathbf{p}}\psi_{\alpha k}\right),$$

so

$$\int \mathrm{d}V \, \phi_{\alpha k}^* \boldsymbol{\nabla}_{\mathbf{k}} E_\alpha(\mathbf{k}) \phi_{\alpha k} = \frac{\hbar}{m} \int \mathrm{d}V \, \phi_{\alpha k}^* \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{r}} \hat{\mathbf{p}} \psi_{\alpha k} = \frac{\hbar}{m} \int \mathrm{d}V \, \psi_{\alpha k}^* \hat{\mathbf{p}} \psi_{\alpha k}.$$

Therefore, we get

$$\hbar^{-1} \langle \boldsymbol{\nabla}_{\mathbf{k}} E_\alpha \rangle = \langle \mathbf{v}_g \rangle = \frac{\langle \mathbf{p} \rangle}{m}. \tag{76}$$

For our effective mass, consider the second-order terms. In 1D,

$$\frac{\hbar^2}{m} = \frac{\mathrm{d}^2 E_\alpha}{\mathrm{d}k^2}.$$

We call this the effective mass (m^*) of electrons:

$$m^* = \hbar^2 \left[\frac{\mathrm{d}^2 E_\alpha}{\mathrm{d}k^2} \right]^{-1}. \tag{77}$$

For our free electron theory, $E = \hbar^2 k^2 / 2m$, and $m = m^*$.

4.3.3 Semi-classical Equations of Motion

By semi-classical, we mean that we use Newton's second law with the mass and velocity replaced by effective mass and group velocity just found.

The infinitesimal work is

$$\mathrm{d}W = \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \mathbf{F} \cdot \mathbf{v}_g \, \mathrm{d}t.$$

Another interpretation of work is through the work-energy theorem:

$$\mathrm{d}W = \boldsymbol{\nabla}_{\mathbf{k}} E_\alpha \cdot \mathrm{d}\mathbf{k} = \hbar \mathbf{v}_g \cdot \mathrm{d}\mathbf{k}.$$

Therefore, by equating $\mathrm{d}W$ we have

$$\mathbf{F} = \hbar \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t}. \tag{78}$$

This is claimed to be remarkable, showing that the effect of an external force on an electron in the crystal is to change the crystal momentum, $\hbar\mathbf{k}$.

We shall use this result to re-derive the form of effective mass:

$$F = m^* \frac{\mathrm{d}v_g}{\mathrm{d}t} = \frac{m^*}{\hbar} \frac{\mathrm{d}}{\mathrm{d}t} \frac{\mathrm{d}E_\alpha}{\mathrm{d}k} = \frac{m^*}{\hbar} \frac{\mathrm{d}^2 E_\alpha}{\mathrm{d}k^2} \frac{\mathrm{d}k}{\mathrm{d}t},$$

and, thus,

$$F = \frac{m^*}{\hbar^2} \frac{\mathrm{d}^2 E_\alpha}{\mathrm{d}k^2} F \quad \Rightarrow \quad m^* = \hbar^2 \left(\frac{\mathrm{d}^2 E_\alpha}{\mathrm{d}k^2} \right)^{-1}.$$
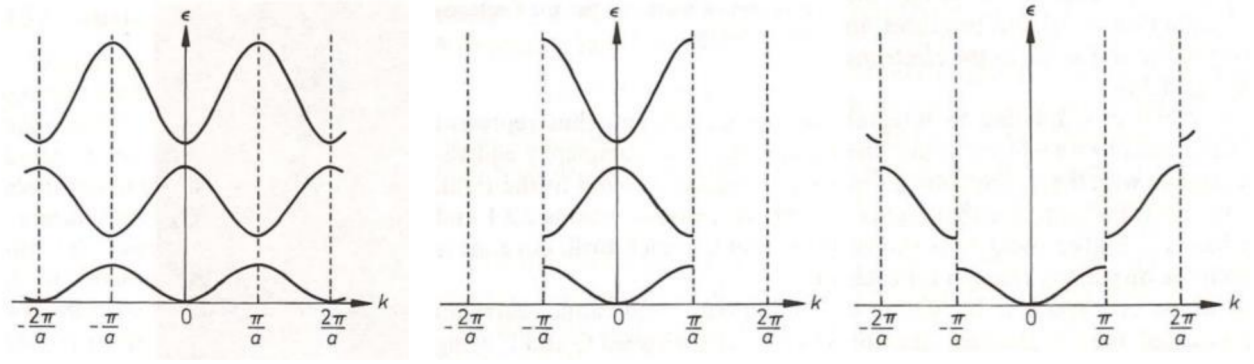
4.4 Different Zone Schemes

Figure 11: Periodic zone, reduced zone, and extended zone.

4.5 Parabolic Band Approximation

By 1D tight binding dispersion relation,

$$E = E_\alpha - 2t_\alpha \cos(ka) \approx E_\alpha - 2t_\alpha + t_\alpha k^2 a^2.$$

Using the effective mass,

$$m^* = \hbar^2 \left(\frac{\mathrm{d}^2 E_\alpha}{\mathrm{d}k^2} \right)^{-1} \approx \frac{\hbar^2}{2t_\alpha a^2},$$

we have

$$E \approx E_\alpha - 2t_\alpha + \frac{\hbar^2 k^2}{2m^*}. \tag{79}$$

Therefore, the larger a band's curvature, the heavier the effective mass.

4.6 Metals, Insulators, and Semiconductors

In the case of a partially filled band, the Fermi level lies within that band. Filled bands carry no current. Conduction may also occur when two bands overlap in energy, such that electrons could be shared between two partially filled bands, rather than a single filled one. Any situation where electronic states are available near the Fermi level will allow electrons to easily move between nearby states and thus conduct.
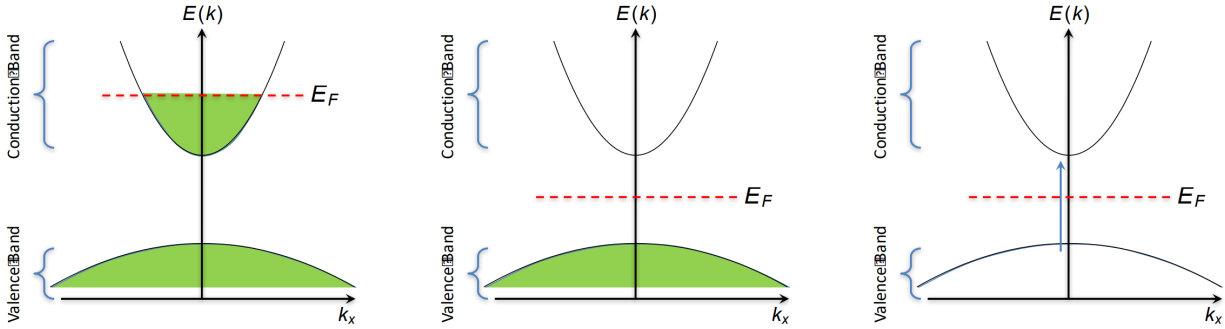

Figure 12: Typical electron occupancy in metals (left), insulators (middle), and semiconductors (right).

For an electron to be excited, it must be present in its original state (E_{V0}, where V is for valence), and there must be a hole in the state (E_{C0}, where C is for conduction) it is being excited to. Therefore, i.e. $1 - f(E_{V0}) = f(E_{C0})$. Substitution into the Fermi-Dirac occupancy function shows that this can only happen if the Fermi level lies at the midpoint of the band gap:

$$\frac{\exp\left[(E_{V0} - E_F)/k_B T\right]}{\exp\left[(E_{V0} - E_F)/k_B T\right] + 1} = \frac{1}{\exp\left[(E_{C0} - E_F)/k_B T\right] + 1},$$

the equality only possible if E_F is the average of E_{V0} and E_{C0}. This explains the position of E_F in insulators.

Semiconductors, like insulators, usually have no partially occupied bands at $T = 0\,\text{K}$. However, semiconductors have bandgaps that are small enough for a small proportion of electrons to partially occupy the next highest band by thermal excitation. In this case, empty electronic states are available in both the mostly occupied and mostly empty bands. Such materials are called semiconductors simply because the few electrons available for conduction make them poor conductors.

4.7 Holes

The effective mass of electrons can be negative, as we have seen in the shape of valence bands. To solve this problem, we introduce **holes** as the missing electrons.

In a full band, the net crystal momentum of electrons should be zero, as they are not really moving

anywhere. Therefore,

$$\sum_i \hbar k_i = 0.$$

If an electron with momentum $\hbar k_e$ is missing, then the net momentum of the band becomes $-\hbar k_e$, and we say this is the momentum of the hole, $k_h = -k_e$.

Take the energy zero to be the top of the valence band. The lower the electron is in the band, the more energy it takes to remove it. (Lower electron energy $\equiv$ higher hole energy.) Therefore, we have

$$E_h(\mathbf{k}_e) = -E_e(\mathbf{k}_e). \tag{80}$$

As bands are usually symmetric, we have $E_h(\mathbf{k}_h) = E_h(-\mathbf{k}_h) = -E_e(\mathbf{k}_e)$.

Note that these particles share the same group velocity:

$$\mathbf{v}_{ge} = \frac{1}{\hbar}\boldsymbol{\nabla}_{\mathbf{ke}}E_e(\mathbf{k}_e) = -\frac{1}{\hbar}\boldsymbol{\nabla}_{-\mathbf{kh}}E_h(\mathbf{k}_h) = \mathbf{v}_{gh}.$$

5 Electronic Properties & Charge Carriers

5.1 Density of States for Parabolic Bands

We know

$$
\begin{aligned}
g(E) &= \frac{1}{V} G(E) = \frac{1}{V} G(\mathbf{k}) \frac{d\mathbf{k}}{dk} \frac{dk}{dE} \\
&= 2 \cdot \frac{1}{(2\pi)^3} 4\pi k^2 \frac{m}{\hbar^2 k} \\
&= \frac{1}{2\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \sqrt{E}.
\end{aligned}
$$

For parabolic bands with dispersion relation

$$
E \approx E_0 + \frac{\hbar^2 k^2}{2m^*},
$$

we substitute E by $E - E_0$ to get the density of states. If this ever perplexes you, think of the work beginning with k.

$$
g_C(E) = \frac{1}{2\pi^2} \left(\frac{2m_C^*}{\hbar^2} \right)^{3/2} \sqrt{E - E_{C0}} \tag{81}
$$

$$
g_V(E) = \frac{1}{2\pi^2} \left(\frac{2m_V^*}{\hbar^2} \right)^{3/2} \sqrt{E_{V0} - E} \tag{82}
$$

5.2 Carrier Numbers in Insulators & Semiconductors

Electron density in conduction band:

$$
n = \int_{E_{C0}}^{\infty} dE \, g(E) f(E). \tag{83}
$$

That this integral extends to infinity does not matter, as Fermi-Dirac statistics ensure most of the electrons present at the bottom of the conduction band.

Hole density in valence band:

$$
p = \int_0^{E_{V0}} dE \, g(E) \left[1 - f(E) \right]. \tag{84}
$$

The same approximation of limits also works here:

$$
p \simeq \int_{-\infty}^{E_{V0}} dE \, g(E) \left[1 - f(E) \right].
$$

Plugging in the density of states, we get the hideous Fermi-Dirac integrals. These integrals are neither

possible nor necessary to solve analytically. However, we may use Boltzmann approximation to solve them: with $E - E_F \gg k_{\mathrm{B}}T$,

$$n \approx N_C \exp\left(-\frac{E_{C0} - E_F}{k_{\mathrm{B}}T}\right), \qquad N_C = 2\left(\frac{m_c^* k_{\mathrm{B}}T}{2\pi\hbar^2}\right)^{3/2} \tag{85}$$

$$p \approx N_V \exp\left(-\frac{E_F - E_{V0}}{k_{\mathrm{B}}T}\right), \qquad N_V = 2\left(\frac{m_v^* k_{\mathrm{B}}T}{2\pi\hbar^2}\right)^{3/2}. \tag{86}$$

We've defined the important quantities N_C and N_V, called the **effective carrier densities**.

5.3 Intrinsic & Extrinsic Semiconductors

5.3.1 Introduction

Property	Intrinsic Semiconductor	Extrinsic Semiconductor
Material Purity	Pure	Doped with impurities (dopants)
Majority Carriers	Electrons = Holes	Electrons or Holes
Fermi Level Position	In the middle of the band gap	Near conduction or valence band
Conductivity	Low (thermally activated)	High (due to dopants)

Since an intrinsic semiconductor has a completely filled band at $T = 0\,\mathrm{K}$, for $T > 0$, an electron excited to the conduction band must also create one hole in the valence band. This implies

$$n = p = n_i, \tag{87}$$

where n_i is called the **intrinsic carrier concentration**:

$$np = n_i^2 = N_C N_V \exp\left(-\frac{\Delta E}{k_{\mathrm{B}}T}\right). \tag{88}$$

This quantity is independent of the Fermi level, dependent on the difference between energy of two bands, $\Delta E = E_{C0} - E_{V0}$. Thus, n_i is truly an intrinsic property of a solid.

Given $n = p$, it is possible to calculate the Fermi level:

$$N_C \exp\left(-\frac{E_{C0} - E_F}{k_{\mathrm{B}}T}\right) = N_V \exp\left(-\frac{E_F - E_{V0}}{k_{\mathrm{B}}T}\right) \quad \Rightarrow \quad E_F = \frac{E_{C0} + E_{V0}}{2} + \frac{3}{4}k_{\mathrm{B}}T \ln\left(\frac{m_v^*}{m_c^*}\right). \tag{89}$$

Boosting the concentration of electrons leads to an **n-type semiconductor**, while boosting conduction

by holes leads to a **p-type semiconductor**.

> Providing the doping concentration is low, we may assume the original band dispersion but with additional carriers.

5.3.2 Probability of Dopant Carrier Ionization

If we are to calculate the number of excited electrons and holes in a doped semiconductor, we must be able to account for the possibility that the number of electrons or holes can change. This is because dopant carriers could be bound to ions instead of being available to conduct.

By the grand canonical ensemble,

$$p(E, N) = A \exp \left(-\frac{E - N E_F}{k_B T} \right), \tag{90}$$

we see that an electron is either ionized ($E = 0$, $N = 0$) or bound (spin-up or spin-down). Supposing that the donor level is E_d away from the conduction band bottom E_{C0}, the bound state probability is

$$p(E = E_{C0} - E_d, N = 1) = A \exp \left(-\frac{E_{C0} - E_d - E_F}{k_B T} \right).$$

Therefore, conservation of probability gives

$$A + 2A \exp \left(-\frac{E_{C0} - E_d - E_F}{k_B T} \right) = 1 \quad \Rightarrow \quad A = p(E = 0, N = 0) = \frac{1}{1 + 2 e^{-(E_{C0} - E_d - E_F)/k_B T}}.$$

Supposing we have a donor atom concentration of N_D, then the concentration of ionized donor atoms is

$$N_D^+ = N_D p(E = 0, N = 0) = \frac{N_D}{1 + 2 e^{-(E_{C0} - E_d - E_F)/k_B T}} = \frac{N_D N_C}{N_C + 2n e^{E_d/k_B T}}. \tag{91}$$

We can do the same with acceptor atoms and thus get

$$N_A^- = \frac{N_A N_V}{N_V + 2p e^{E_a/k_B T}}. \tag{92}$$

5.3.3 Fermi Level of Extrinsic Semiconductors & The Law of Mass Action

In extrinsic semiconductors, we will always want the situation where $N_D, N_A \gg n_i$, otherwise what is the point introducing the impurities? Therefore, we have $n \approx N_D^+$ and $p \approx N_A^-$. These two equations are actually quadratic and solvable.

By $N_D^+ \approx n \approx N_C e^{-(E_{C0}-E_F)/k_\mathrm{B}T}$,

$$E_{C0} - E_F = k_\mathrm{B}T \ln\left(\frac{N_C}{N_D^+}\right). \tag{93}$$

As donors become ionized, the Fermi level moves towards the conduction band. Also, for acceptors, we have

$$E_F - E_{V0} = k_\mathrm{B}T \ln\left(\frac{N_V}{N_A^-}\right). \tag{94}$$

As acceptors become ionized, the Fermi level moves towards the valence band.

We should also notice that as as one of the carriers increase, the other must decrease due to

$$np = n_i^2 \left[= N_C N_V \exp\left(-\frac{E_{C0}-E_{V0}}{k_\mathrm{B}T}\right) = N_C N_V \exp\left(-\frac{E_g}{k_\mathrm{B}T}\right)\right]. \tag{95}$$

This push and pull of the extrinsic carrier densities is known as **the law of mass action**.

It is possible with very high doping to shift the Fermi level so far that it eventually enters one of the bands. Under this condition, we will essentially create a metal, although it is often referred to as **degenerately doped semiconductor**.

5.4 The Drude Model & Its Failure

In this purely classical model by Drude, we assume that electrons collide with ions at a rate τ^{-1}.

If carriers scatter, Drude assumed they return to thermal equilibrium immediately, and thus do not contribute to the total momentum change.

Under these assumptions, the momentum after time δt is

$$\langle \mathbf{p}(t+\delta t)\rangle = (\langle \mathbf{p}(t)\rangle + \delta \mathbf{p})\left(1 - \frac{\delta t}{\tau}\right) = \langle \mathbf{p}(t)\rangle + \left(\mathbf{F} - \frac{\langle \mathbf{p}\rangle}{\tau}\right)\delta t + \mathcal{O}(\delta t^2).$$

Therefore,

$$\mathbf{F} = \frac{\langle \mathbf{p}\rangle}{\tau} + \frac{\mathrm{d}\langle p\rangle}{\mathrm{d}t}. \tag{96}$$

In the steady state, $\langle \dot{\mathbf{p}}\rangle = 0$,

$$\mathbf{v}_d \equiv \langle \mathbf{v}\rangle = \frac{\mathbf{F}\tau}{m} = \frac{q\mathbf{E}\tau}{m}, \tag{97}$$

where $\mathbf{v}_d$ is defined as the **drift velocity**. Through the current density we can define **conductivity**:

$$\mathbf{J} = qn\mathbf{v}_d = \frac{q^2 n\tau}{m}\mathbf{E} \equiv \sigma\mathbf{E}. \tag{98}$$

While the scattering time in Drude's theory is reasonably consistent, the mean free path is completely wrong. Sommerfeld's free-electron quantum theory addresses this problem– the maximum velocity is not related to thermal motion, but to the Fermi velocity. Secondly, the quantum theory of electronic states in a solid yields Bloch wavefunctions, which are not scattered by a perfect lattice of atoms at all. There must therefore be a different cause for carrier scattering in solids. Finally, band theory addresses why only the valence electrons contribute to conduction — carriers in full bands cannot conduct.

5.5   Semi-Classical Theory of Charge Transport & Bloch Oscillations

We now use Drude's formalism adjusted to include the quantum mechanical parameters of effective mass and velocity. In 1D,

$$\hbar\frac{\mathrm{d}k}{\mathrm{d}t} = qE,$$

and the solution shows that k increases linearly with t.

In the crystal, we have

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

so, while k increases linearly with time,

$$v_g = \frac{\mathrm{d}\omega}{\mathrm{d}k}$$

oscillates back and forth. There is effectively no current under this electric field.

The force of the electric field on electrons causes a change in crystal momentum, so this change must have been restored by other entities within the system: phonons.
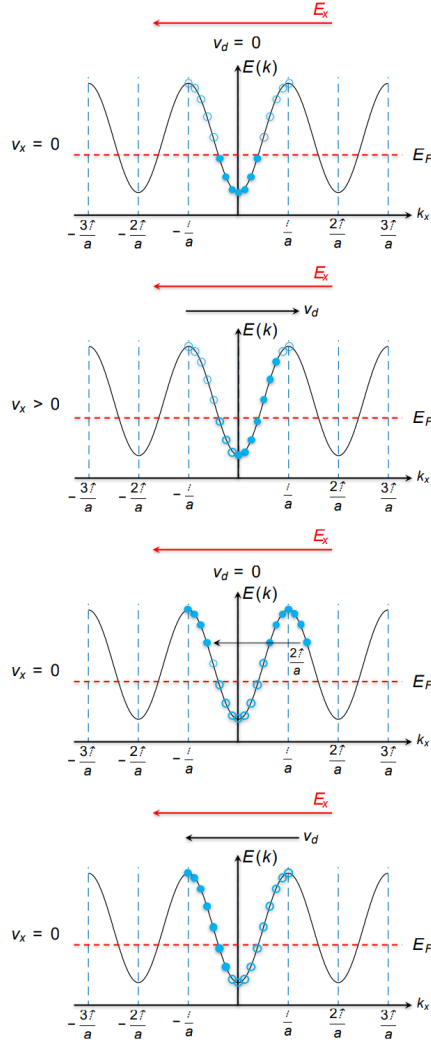
Figure 13: Four processes of Bloch oscillations.

The Sommerfeld-Drude model of conduction is actually the free-electron theory that describes conduction in metals.

With the application of an external field, all electrons will be accelerated; however, they will eventually be scattered by phonons, and this is most likely to occur by electrons losing energy and occupying states on the opposite side of the Brillouin zone.

We define the conductivity by replacing the mass with effective mass:

$$\sigma = \frac{q^2 n \tau}{m^*}.\tag{99}$$

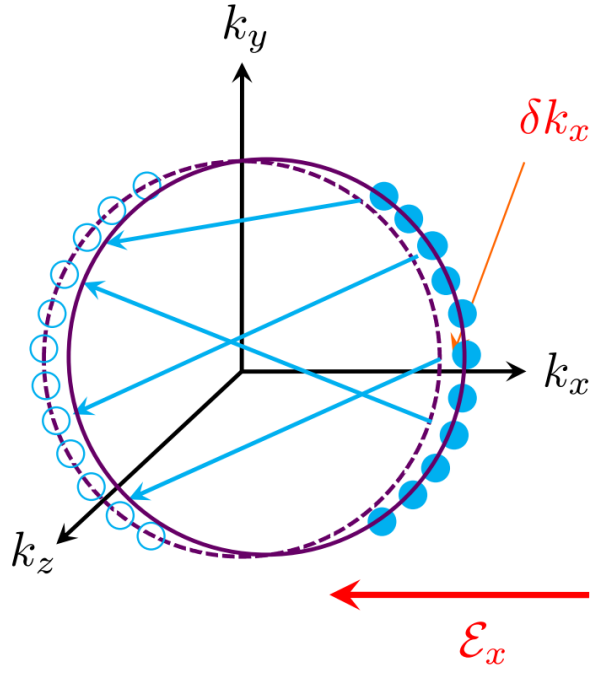It's important to realize that each partially filled band will contribute to the conductivity.

Figure 14: Electron scattering by phonons.

Carrier mobility, μ, is a quantity that relates the drift velocity to the electric field:

$$\mathbf{v}_d = \mu\mathbf{E}. \tag{100}$$

It's conventional for the mobility to be positively valued. Also, for simplicity, we often assume $\tau_n = \tau_p = \tau$.

6 Semiconductor Devices

6.1 pn-Junction

A **pn-junction** is a boundary between n-type and p-type extrinsic semiconductor regions.

6.1.1 Construction

- Isolated extrinsic semiconductors;

- Inhomogeneous carrier distributions diffuse;

- Fixed dopant ions establish space charge and electric field to counteract diffusion;

- In equilibrium, diffusion = drift. This leaves behind a **depletion region** of very few carriers.
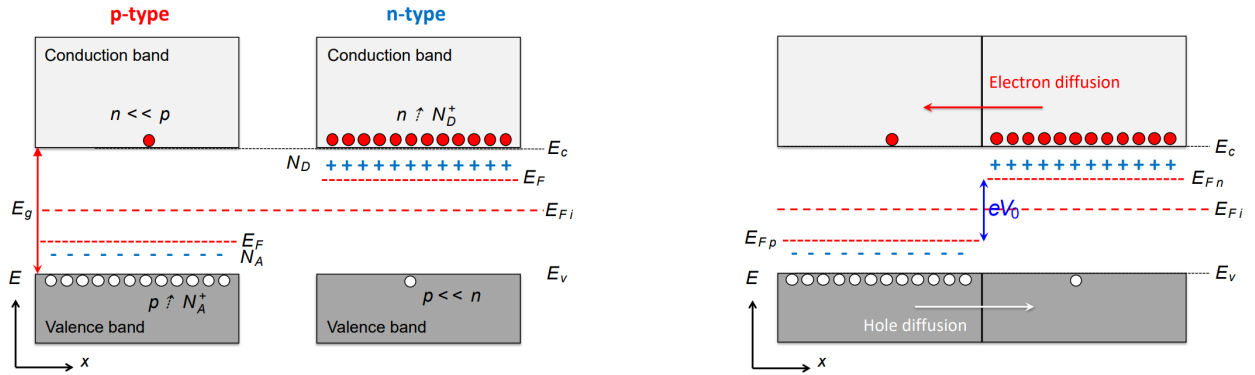

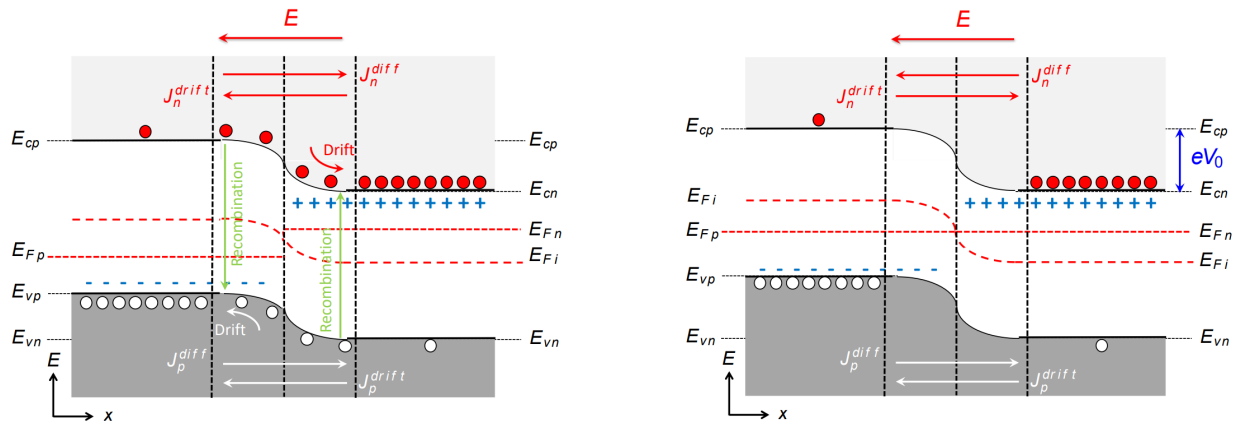Figure 15: The first two steps described above.

Figure 16: The last two steps described above.

If the same underlying semiconductor is used throughout, we call this a **homo-junction**.

6.1.2 Electrostatics

In the depletion region, there are few carriers, so we expect net charge accumulation due to ions: negative charge in p-region and positive charge in n-region.

> Depletion-region approximation: we assume a step-like charge distribution.

$$\rho(x) = \begin{cases} -eN_A & -x_p < x < 0, \\ eN_D & 0 < x < x_n. \end{cases} \tag{101}$$

By Gauss's law,

$$\frac{dE_x}{dx} = \frac{\rho(x)}{\epsilon}, \tag{102}$$

with boundary conditions $E(-x_p) = E(x_n) = 0$ and continuity at $x = 0$,

$$E(x) = \begin{cases} -\dfrac{eN_A}{\epsilon}(x + x_p) & -x_p < x < 0, \\ \dfrac{eN_D}{\epsilon}(x - x_n) & 0 < x < x_n. \end{cases} \tag{103}$$

Therefore, the continuity condition requires

$$N_A x_p = N_D x_n. \tag{104}$$

Integrating again, we get the electric potential. Requiring $V(-x_p) = 0$ and $V(x_n) = V_0$,

$$V(x) = \begin{cases} \dfrac{eN_A}{2\epsilon}(x + x_p)^2 & -x_p < x < 0, \\ -\dfrac{eN_D}{2\epsilon}(x - x_n)^2 + V_0 & 0 < x < x_n. \end{cases} \tag{105}$$

Continuity of electric potential at $x = 0$ requires:

$$V_0 = \frac{e}{2\epsilon}\left(N_D x_n^2 + N_A x_p^2\right). \tag{106}$$

We may also determine the band edge energies:

$$E(x) = E_0 - eV(x). \tag{107}$$

With these quantities, we can determine the width of the depletion region. By $N_A x_p = N_D x_n$, we can

41

express W in terms of either x_n or x_p and, thus, V_0:

$$W = \sqrt{\frac{2\epsilon}{e} V_0 \frac{N_A + N_D}{N_A N_D}}. \tag{108}$$
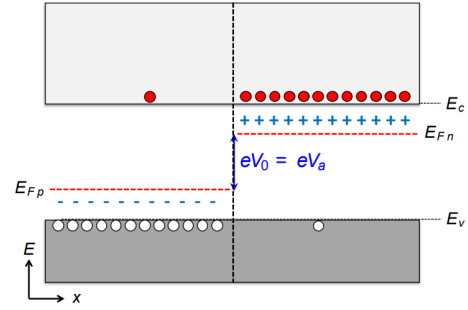
6.2 Bias

The bias to a pn-junction is an external voltage, V_a. We define **forward bias** to be when the built-in voltage step of the pn-junction is reduced. The effect of bias on the width of the depletion region is through a simple substitution $V_0 \to V_0 - V_a$.
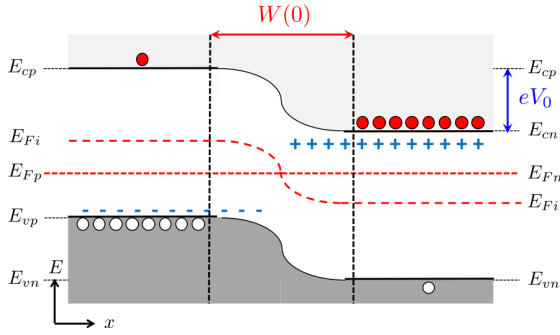
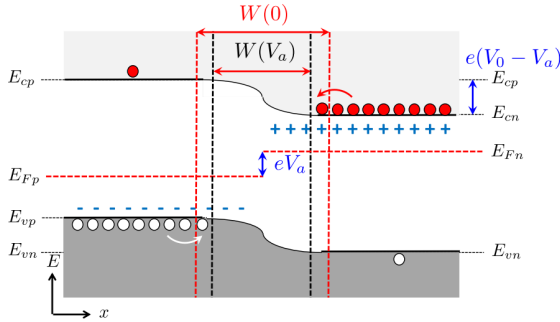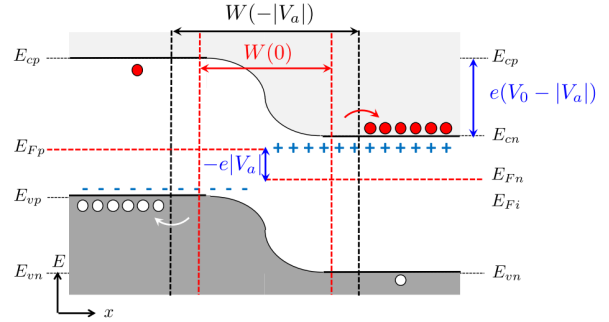Figure 17: The four key bias conditions for a pn-junction.

6.3 Drift, Diffusion, and Einstein's Relation

The diffusion current due to density gradients is

$$J_c^{\text{diffusion}} = q D_c \frac{\mathrm{d}c}{\mathrm{d}x}, \tag{109}$$

where c is for carrier, and D is the diffusion coefficient. As a reminder, the drift current is

$$J_c^{\text{drift}} = qcv_d = qc \left(\frac{q}{e} \mu_c E \right). \tag{110}$$

The definition of mobility ensures that J^{drift} has the same sign for electrons and holes.

We need to consider both carriers in both semiconductors: n_n for electrons in n-type region (majority), p_n for holes in n-type region (minority), p_n for electrons in p-type region (minority), and p_p for holes in p-type region (majority).

Weak current approximation: $n_n \approx N_D$ and $p_p \approx N_A$.

In this approximation, we only need to consider the minority carriers. Since the drift current depends on the carrier density, we assume

$$J^{\text{diffusion}} \gg J^{\text{drift}} \tag{111}$$

as well.

Kirchhoff's law requires conservation of current:

$$J_n(x) + J_p(x) = \text{constant}, \tag{112}$$

where we define the currents J as the sum of diffusion and drift currents.

Plugging the expression for $n(x)$, we have

$$J_n(x) = en(x)\mu_n E(x) + eD_n \frac{\mathrm{d}n}{\mathrm{d}x} = en(x) \left[\mu_n E(x) - \frac{D_n}{k_B T} \frac{\mathrm{d}E_C(x)}{\mathrm{d}x} \right].$$

The variation of the conduction band energy defines an electric potential that electrons are subjected to: $E_C(x) = -eV(x)$, and $\mathrm{d}E_C(x)/\mathrm{d}x = eE'(x)$. Thus, we find

$$J_n(x) = en(x) \left[\mu_n E(x) - \frac{eD_n}{k_B T} E'(x) \right].$$

When there's no external field, we expect $E(x) = E'(x)$ such that no net current flows. This gives **Einstein's Relation**:

$$\frac{\mu_n}{e} = \frac{D_n}{k_B T}. \tag{113}$$

43

6.4 Quasi-equilibrium

Excess minority carriers will undergo generation and recombination processes that drive the carrier densities back towards equilibrium. The minority carrier lifetime, τ, is the average time that excess minority carriers survive in a sea of majority carriers. Phenomenologically,

$$\frac{\mathrm{d}p_n}{\mathrm{d}t} = -\frac{\Delta p_n}{\tau}, \qquad \frac{\mathrm{d}n_p}{\mathrm{d}t} = -\frac{\Delta n_p}{\tau}. \tag{114}$$

Intuitively, the change to the density of minority carriers depends on both current and recombination. Thus, we write the **current continuity equation**:

$$\frac{\partial n_p}{\partial t} = \frac{1}{e}\frac{\partial J_n}{\partial x} - \frac{\Delta n_p}{\tau_n}. \tag{115}$$

In weak current approximation, $J_n \approx eD_n\,\partial n_p/\partial x$ such that

$$\frac{\partial \Delta n_p}{\partial t} = D_n\frac{\partial^2 \Delta n_p}{\partial x^2} - \frac{\Delta n_p}{\tau_n}.$$

In steady state, LHS $= 0$, and we get a quasi-equilibrium condition described by the diffusion equation: (only valid in weak current approximation)

$$D_n\frac{\mathrm{d}^2 \Delta n_p}{\mathrm{d}x^2} = \frac{\Delta n_p}{\tau_n}. \tag{116}$$

If electrons are being injected into the p-type material, there must also be some hole current driving the majority carriers through the p-type material. This is evident by the Kirchhoff's law.

6.5 Quasi-Fermi Levels

The term quasi here reflects the fact that an applied bias V_a drives the pn-junction out of equilibrium. Defining electron quasi-Fermi level E_{Fn} and hole quasi-Fermi level E_{Fp}, we have

$$qV_a = E_{Fn} - E_{Fp}, \tag{117}$$

$$n_n = N_C \exp\left(-\frac{E_C - E_{Fn}}{k_{\mathrm{B}}T}\right), \qquad p_p = N_V \exp\left(-\frac{E_{Fp} - E_V}{k_{\mathrm{B}}T}\right). \tag{118}$$
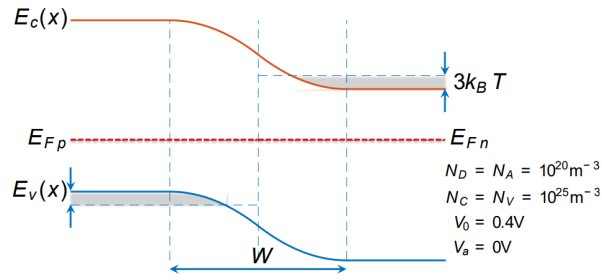
Multiplying the two quantities together gives **the law of the junction**:

$$n(x)p(x) = n_i^2 \exp\left(\frac{E_{Fn} - E_{Fp}}{k_{\mathrm{B}}T}\right) = n_i^2 \exp\left(\frac{qV_a}{k_{\mathrm{B}}T}\right). \tag{119}$$
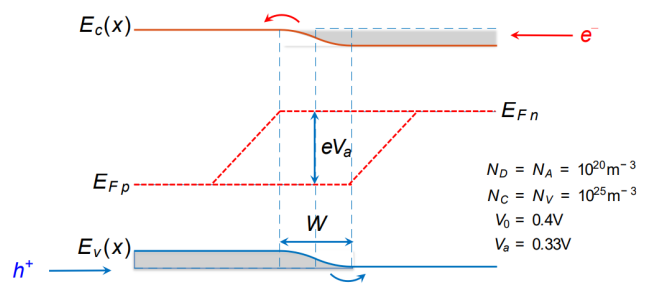
It looks like that the law of mass action is describing the equilibrium condition, where $E_{Fn} = E_{Fp}$. Far from the junction, this equality holds, and this law reduces to the law of mass action.

6.6 Current and Carrier Density Under Bias
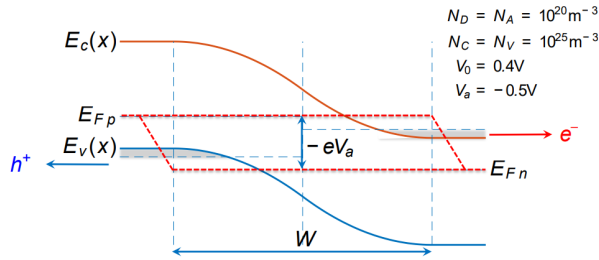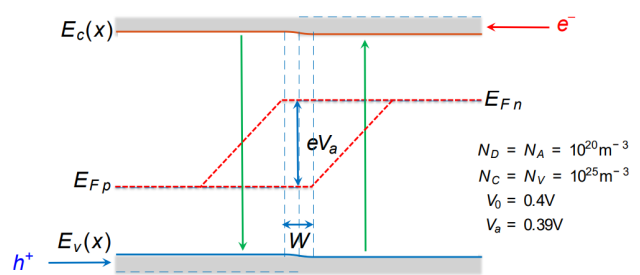
Figure 18: The available carriers are shaded.

Remember that electrons have lower energy when they fall to the bottom of the conduction band, but holes have lower energy when they rise to the top of the valence band. (It is useful to think of holes as tending to bubble up in energy.)

In forward bias, we add electrons to the n-side and holes to the p-side. This decreases the depletion region width as well as the junction's potential step.

6.7 The Shockley Ideal Diode Equation

> We will ignore generation nor recombination inside the depletion region.

By $\Delta n_p(x) = n_p(x) - n_{p0}$ and the boundary condition that

$$n_p(-x_p)N_A = n_{p0}N_A \exp\left(\frac{eVa}{k_{\mathrm{B}}T}\right),$$

the diffusion equation gives

$$\Delta n_p(x) = n_{p0}\left[\exp\left(\frac{eV_a}{k_{\mathrm{B}}T}\right) - 1\right]\exp\left(\frac{x + x_p}{L_n}\right),$$

with $L_n = \sqrt{D_n\tau_n}$, valid for only $x \leq -x_p$.

In the weak current approximation,

$$J_n(x) \approx eD_n\frac{\mathrm{d}n_p}{\mathrm{d}x} = e\frac{D_n n_{p0}}{L_n}\left[\exp\left(\frac{eV_a}{k_{\mathrm{B}}T}\right) - 1\right]\exp\left(\frac{x + x_p}{L_n}\right).$$

Therefore,

$$J_{n0}(V_a) = e\frac{D_n n_{p0}}{L_n}\left[\exp\left(\frac{eV_a}{k_{\mathrm{B}}T}\right) - 1\right], \qquad J_{p0}(V_a) = e\frac{D_p p_{n0}}{L_p}\left[\exp\left(\frac{eV_a}{k_{\mathrm{B}}T}\right) - 1\right].$$

Summing the current, we find the **Shockley ideal diode equation**:

$$J(V_a) = J_0\left[\exp\left(\frac{eV_a}{k_{\mathrm{B}}T}\right) - 1\right], \qquad J_0 = e\left(\frac{D_n n_{p0}}{L_n} + \frac{D_p p_{n0}}{L_p}\right) = en_i^2\left(\frac{D_n}{N_A L_n} + \frac{D_p}{N_D L_p}\right). \quad (120)$$

6.8 Modified Shockley Equation with Photocurrents

Outside of the space charge region, unless a bias is applied to the pn-Junction, *e-h* pairs will eventually recombine, either emitting photons, or simply generating heat via phonons. However, *e-h* pairs generated in the space charge region experience the junction's built-in field; they are split apart and can be extracted as a **photocurrent**, I_p. Notice that this photocurrent is negative with respect to our definition of forward bias. In the presence of light, Shockley's equation becomes:

$$I(V_a) = I_0\left[\exp\left(\frac{eV_a}{k_{\mathrm{B}}T}\right) - 1\right] - I_p. \tag{121}$$

7 All the Perplexing Quantities & Their Relations

7.1 Fermi Level & Fermi Energy

Well, let's first repeat with Fermi level and Fermi energy.

> **Fermi Energy v.s. Fermi Level**
>
> The **Fermi energy**, E_F, is defined as the energy of the highest occupied state at $T = 0\,\text{K}$.
>
> The **Fermi level** is just a synonym for chemical potential, defined at various temperatures. You bloody physicists.

7.2 Quantities in Doped Semiconductors

Note the difference between doped semiconductors (extrinsic) and pn-junction. An extrinsic semiconductor can only be either n- or p-type, and two of opposite types combined together make a pn-junction.

Electron Donor	Electron Acceptor	Description
N_D	N_A	Number of D / A atoms
N_D^+	N_A^-	Number of ionized D / A atoms
E_D	E_A	Energy Level of D / A
$n,\, p$		Number density of electrons / holes
n_i		Intrinsic carrier concentration
E_F		Fermi level

Description	Expression
Free carrier number density	$n = N_C \exp\left[-(E_C - E_F)/k_{\mathrm{B}}T\right],\, p = N_V \exp\left[-(E_F - E_V)/k_{\mathrm{B}}T\right]$
Intrinsic semiconductor	$n = p = n_i$
Intrinsic carrier concentration	$np = n_i^2$
Electrical neutrality	$n + N_A^- = p + N_D^+$
Complete ionization approximation (n)	$n = p + N_D^+ \approx N_D^+ = N_D,\, p = n_i^2/N_D$
Complete ionization approximation (p)	$p = n + N_A^- \approx N_A^- = N_A,\, n = n_i^2/N_A$
Shift in Fermi level (n)	$E_F - E_i = k_{\mathrm{B}}T \ln\left(n/n_i\right)$
Shift in Fermi level (p)	$E_i - E_F = k_{\mathrm{B}}T \ln\left(p/n_i\right)$

N_C and N_V are referred to as **effective carrier densities**, and E_i stands for the Fermi level in the associated intrinsic semiconductor. All equations in black are only valid in thermal equilibrium.

As for the last two equations, let's reiterate that the equations

$$n = N_C \exp\left(-\frac{E_C - E_F}{k_{\mathrm{B}}T}\right), \qquad p = N_V \exp\left(-\frac{E_F - E_V}{k_{\mathrm{B}}T}\right) \tag{122}$$

hold in thermal equilibrium and are results from Fermi-Dirac statistics. Therefore, they are valid regardless of the doping. Therefore, for an intrinsic semiconductor,

$$n_i = n(\text{intrinsic}) = N_C \exp\left(-\frac{E_C - E_i}{k_{\mathrm{B}}T}\right).$$

When doped,

$$n(\text{extrinsic}) = N_C \exp\left(-\frac{E_C - E_F}{k_{\mathrm{B}}T}\right).$$

Therefore,

$$\ln\left(\frac{n}{n_i}\right) = \frac{E_i - E_F}{k_{\mathrm{B}}T}.$$

The same equation can also be derived for p. The result is intuitive: a negatively doped semiconductor has a Fermi level closer to the conduction band, and vice versa.

7.3   Quantities in pn-junction

n-type Material	p-type Material	Description
n_n	n_p	Electron number density
p_n	p_p	Hole number density
E_{Fn}	E_{Fp}	Fermi level
E_{Cn}	E_{Cp}	Conduction band minimum
E_{Vn}	E_{Vp}	Valence band maximum
x_n	x_p	Depletion region width
V_0		Built-in potential
V_a		Applied bias
E_{Fi}		Intrinsic Fermi level

Thing have become more twisted by now. n_n and p_p are majority carriers, while n_p and p_n are minority carriers. In equilibrium with no bias, we have

$$E_{Fp} = E_{Fn}. \tag{123}$$

While the word Fermi level seems overly abstract, it is much easier if it goes by chemical potential. If the chemical potential is not equalized over the junction, there will be a net current for the electrons to flow to where chemical potential is lower.

Bias Type	p-type Connection	n-type Connection
Forward	Anode $(+)$	Cathode $(-)$
Reverse	Cathode $(-)$	Anode $(+)$

Considering the results, a forward bias helps holes to flow to the n-type side and electrons to flow to the p-type side. This reduces the depletion region.

Description	Expression
Quasi-Fermi Level & Applied Bias	$qV_a = E_{Fn} - E_{Fp}$
Majority Carrier Concentration	$n_n = N_C \exp\left(-\dfrac{E_C - E_{Fn}}{k_\mathrm{B}T}\right),\ p_p = N_V \exp\left(-\dfrac{E_{Fp} - E_V}{k_\mathrm{B}T}\right)$
Law of the Junction	$n(x)p(x) = n_i^2 \exp\left(\dfrac{E_{Fn} - E_{Fp}}{k_\mathrm{B}T}\right) = n_i^2 \exp\left(\dfrac{qV_a}{k_\mathrm{B}T}\right)$
Depletion Region Width	$W = \sqrt{\dfrac{2\epsilon}{e}(V_0 - V_a)\dfrac{N_A + N_D}{N_A N_D}}$

The built-in potential is actually easy to derive. Remembering that

$$n_n p_n = n_p p_p = n_i^2$$

in thermal equilibrium, when combining the materials, we have

$$eV_0 = E_{Fn} - E_{Fp},$$

and thus

$$eV_0 = k_\mathrm{B}T \ln\left(\frac{n_n p_p}{n_i^2}\right) = k_\mathrm{B}T \ln\left(\frac{n_n}{n_p}\right).$$

Invoking the complete ionization approximation,

$$V_0 = \frac{k_\mathrm{B}T}{e}\left(\frac{N_D N_A}{n_i^2}\right). \tag{124}$$

7.4 Each tight-binding band is full when there are 2 electrons per unit cell.

All is said in the name of this section.