EE3310 Class notes

Version: Fall 2002
These class notes were originally based on the handwritten notes of Larry Overzet. It is expected that they will be modified (improved?) as time goes on. This version was typed up by Matthew Goeckner.

Solid State Electronic Devices - EE3310

Class notes

Introduction

Homework Set 1
Streetman Chap 1 # 1,3,4,12, Chap. 2 # 2,5 Assigned 8/22/02 Due 8/29/02

Q: Why study electronic devices?
A: They are the backbone of modern technology

1) Computers.
2) Scientific instruments.
3) Cars and airplanes (sensors and actuators).
4) Homes (radios, ovens, clocks, clothes dryers, etc.).
5) Public bathrooms (Auto-on sinks and toilets).

Q: Why study the physical operation?
A: This is an engineering class. You are studying so that you know how to make better devices and tools. If you do not understand how a tool works, you cannot make a better tool. (Technicians and electricians can make a tool work but they cannot significantly improve it. They, however, are not trained to understand the basic operation of the tool.)

1) Design systems (Can you get something to work or not?).
2) Make new – improved – devices.
3) Be able to keep up with new devices.

Q: What devices will we study?
A:

1) Bulk semiconductors to resistors.
2) P-n junction diodes and Schottky diodes.
3) Field Effect Transistors (FETs) – This is the primary logic transistor!
4) Bipolar junction transistors – This is the primary ‘power’ transistor!

By course end, the students should know:

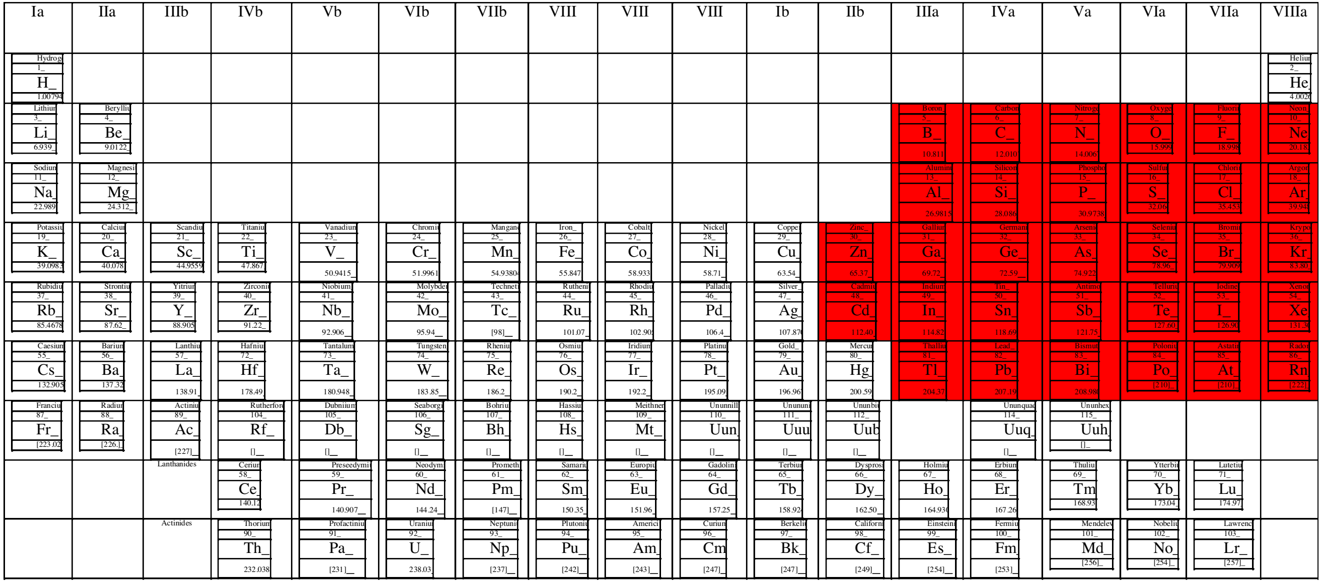
1) How these devices act.
2) Why these devices act the way they do.

3) Finally, the students should gain a “manure” detector. This can be described as the ability to judge whether or not a device should act in a given manner, i.e., if someone describes a device and says that its operational characteristics are “such and such”, the student should be able to briefly look at the situation and say “maybe” or “unlikely”. (Only a detailed study can give “absolutely” or “absolutely not”.)

Let us start the class by describing just what is a ‘semiconductor’.

1) The conductivity of semiconductors occupy the area between conductors and insulators. This implies that the conductivity can range over many orders of magnitude. Further, the conductivity of semiconductors can be adjusted through a number of means, each related to the physical properties of semiconductors. Typical methods for adjusting the conductivity of a semiconductor are:
    a. Temperature
    b. Purity (Doping)
    c. Optical excitation
    d. Electrical excitation.

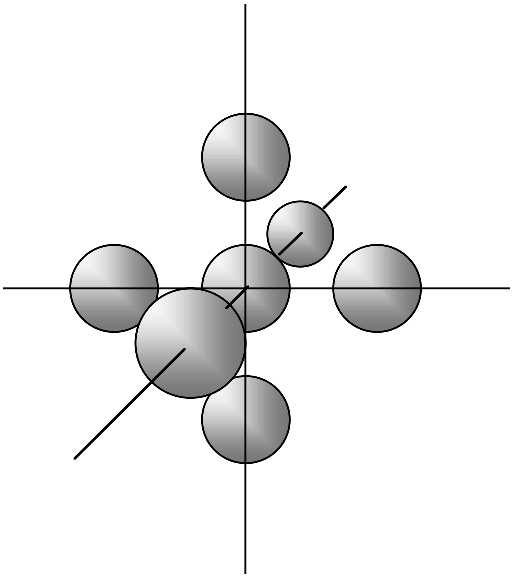
2) Materials come from



Ia	IIa	IIIb	IVb	Vb	VIb	VIIb	VIII	VIII	VIII	Ib	IIb	IIIa	IVa	Va	VIa	VIIa	VIIIa
Hydrog 1_ H_ 1.00794																	Heliu 2_ He 4.002
Lithiu 3_ Li_ 6.939_	Berylliu 4_ Be_ 9.0122_											Boron_ 5_ B_ 10.811	Carbon 6_ C_ 12.010	Nitroge 7_ N_ 14.0067	Oxyge 8_ O_ 15.999	Fluori 9_ F_ 18.998	Neon_ 10_ Ne 20.18
Sodiu 11_ Na 22.989	Magnesi 12_ Mg_ 24.312_											Alumin 13_ Al_ 26.9815	Silicon 14_ Si_ 28.086	Phospho 15_ P_ 30.9738	Sulfu 16_ S_ 32.06	Chlori 17_ Cl_ 35.453	Argon 18_ Ar 39.94
Potassiu 19_ K_ 39.0983	Calciu 20_ Ca_ 40.078	Scandiu 21_ Sc_ 44.9559	Titaniu 22_ Ti_ 47.867	Vanadium 23_ V_ 50.9415_	Chromi 24_ Cr_ 51.9961	Mangane 25_ Mn_ 54.9380	Iron_ 26_ Fe_ 55.847	Cobalt 27_ Co_ 58.933	Nickel 28_ Ni_ 58.71_	Coppe 29_ Cu_ 63.54_	Zinc_ 30_ Zn_ 65.37_	Galliu 31_ Ga 69.72_	Germani 32_ Ge_ 72.59_	Arseni 33_ As_ 74.922	Seleniu 34_ Se_ 78.96_	Bromi 35_ Br_ 79.909	Krypo 36_ Kr_ 83.80_
Rubidiu 37_ Rb_ 85.4678	Strontiu 38_ Sr_ 87.62_	Yitriu 39_ Y_ 88.905	Zirconi 40_ Zr_ 91.22_	Niobium 41_ Nb_ 92.906_	Molybde 42_ Mo_ 95.94_	Techneti 43_ Tc_ [98]__	Rutheni 44_ Ru_ 101.07_	Rhodiu 45_ Rh_ 102.90	Palladiu 46_ Pd_ 106.4_	Silver_ 47_ Ag_ 107.87	Cadmiu 48_ Cd_ 112.40_	Indium 49_ In_ 114.82	Tin_ 50_ Sn_ 118.69	Antimo 51_ Sb_ 121.75_	Telluriu 52_ Te_ 127.60_	Iodine 53_ I_ 126.90	Xenon 54_ Xe 131.3
Caesium 55_ Cs_ 132.905	Bariu 56_ Ba_ 137.32	Lanthiu 57_ La_ 138.91_	Hafniu 72_ Hf_ 178.49	Tantalum 73_ Ta_ 180.948_	Tungsten 74_ W_ 183.85_	Rheniu 75_ Re_ 186.2_	Osmiu 76_ Os_ 190.2_	Iridium 77_ Ir_ 192.2_	Platinu 78_ Pt_ 195.09	Gold_ 79_ Au_ 196.96	Mercu 80_ Hg 200.59	Thalliu 81_ Tl_ 204.37	Lead_ 82_ Pb_ 207.19	Bismut 83_ Bi_ 208.98	Poloniu 84_ Po_ [210]_	Astatin 85_ At_ [210]_	Rador 86_ Rn [222]
Franciu 87_ Fr_ [223.02	Radiu 88_ Ra [226.]_	Actiniu 89_ Ac_ [227]__	Rutherfor 104_ Rf_ []__	Dubniium 105_ Db_ []__	Seaborgi 106_ Sg_ []__	Bohriu 107_ Bh_ []__	Hassiu 108_ Hs_ []__	Meithner 109_ Mt_ []__	Ununnill 110_ Uun []__	Ununun 111_ Uuu []__	Ununbi 112_ Uub []__		Ununquad 114_ Uuq_ []__	Ununhex 115_ Uuh []_			
		Lanthanides	Ceriu 58_ Ce 140.12	Preseedymi 59_ Pr_ 140.907__	Neodymi 60_ Nd_ 144.24_	Prometh 61_ Pm_ [147]__	Samariu 62_ Sm 150.35_	Europiu 63_ Eu_ 151.96_	Gadolini 64_ Gd_ 157.25_	Terbiu 65_ Tb_ 158.92	Dysprosi 66_ Dy_ 162.50_	Holmiu 67_ Ho_ 164.93	Erbium 68_ Er_ 167.26	Thuliu 69_ Tm 168.93	Ytterbiu 70_ Yb_ 173.04_	Lutetiu 71_ Lu_ 174.97	
		Actinides	Thorium 90_ Th_ 232.038	Profactiniu 91_ Pa_ [231]__	Uraniu 92_ U_ 238.03	Neptuni 93_ Np_ [237]__	Plutoni 94_ Pu_ [242]__	Americi 95_ Am [243]__	Curiun 96_ Cm [247]_	Berkeli 97_ Bk_ [247]__	Californ 98_ Cf_ [249]__	Einsteini 99_ Es_ [254]__	Fermiu 100_ Fm [253]_	Mendelev 101_ Md_ [256]_	Nobeliu 102_ No_ [254]_	Lawrenc 103_ Lr_ [257]_	

3) In most semiconductor devices, the atoms are arranged in crystals. Again, this is because of the physical properties of the material. The structures of solid materials are described with three main categories. (This can and is further subdivided.) These categories are:
    a. Amorphous
    b. Poly crystalline
    c. Crystalline

To understand the distinction between these solid material types, we must first understand the concept of order. Order can be described as the repetition of identical structures or identical placement of atoms. An example of this would be an atom that has six nearby atoms, each 5 Å away, arranged in a pattern as such.



If one where to pick any other atom in the material and find the same arrangement, then the material would be described as having order. This order can be either Short Range Order, SRO, or Long Range Order. Short-range order is typically on the order of 100 inter atom distances or less, while long range is over distance greater than 1000 inter atom distances, with a transitional region in between.

We will now discuss each of the solid material types in turn.

Amorphous solids are such that the atoms that make up the material have some local order, i.e. SRO, but there is no Long Range Order, LRO. (Materials with no SRO or LRO are liquids.)

Crystalline solids are such that the atoms have both SRO and LRO.

Polycrystalline solids are such that there are a large number of small crystals ‘pasted’ together to make the larger piece.

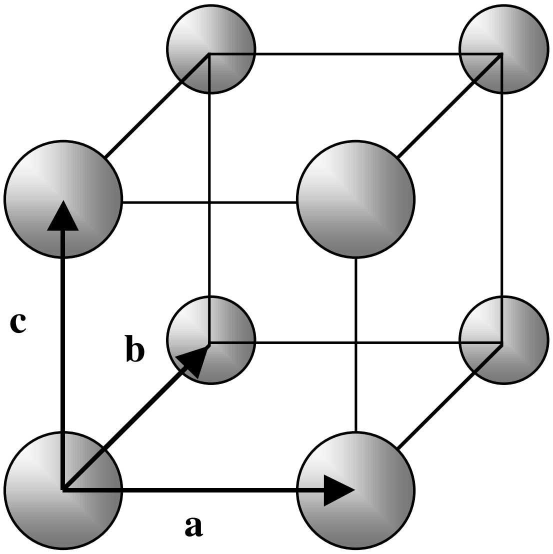
For the purposes of this class, crystals, as we have said before, are the most important of these types of solids. Because of this we need to understand crystals in more detail.
WE now need to introduce some basic concepts:

1) The crystal structure is known as the LATTICE or LATTICE STRUCTURE.
2) The locations of each of the atoms in the lattice are known as the LATTICE POINTS.
3) A UNIT CELL is a volume-enclosing group of atoms that can be used to describe the lattice by repeated translations (no rotations!). This is further restricted such that the translations of the cells must fill all of the crystalline volume and cells may not overlap. In this way, the structure is uniquely defined.
4) A PRIMITIVE CELL is the smallest possible unit cell.

Often primitive cells are not easy to work with and thus we often use slightly larger unit cells to describe the crystal. There are four of very simple – basic – unit cells that are often seen in crystalline structures.

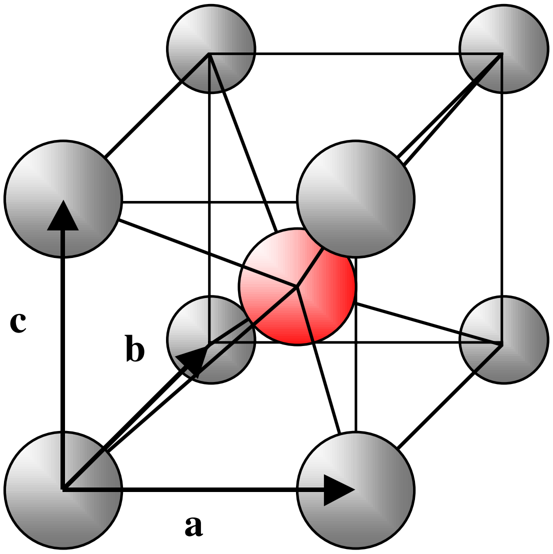
IT SHOULD BE UNDERSTOOD THAT THESE ARE NOT ALL OF THE POSSIBLE STRUCTURES. These structures are:

1) Simple Cubic, SC



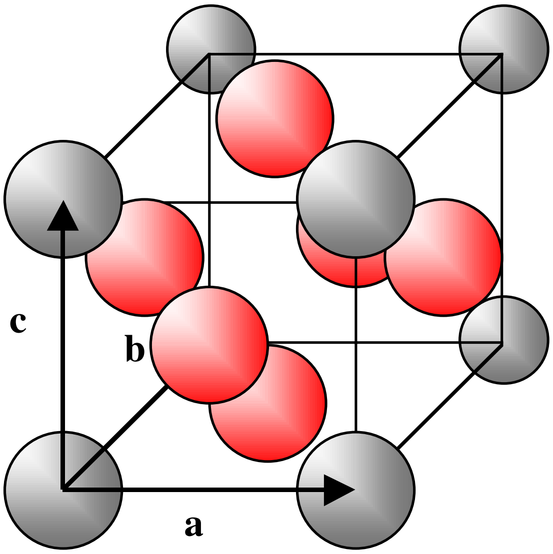
Here **a**, **b**, and **c** are the BASIS VECTORS along the edges of the standard SC cell.

2) Body Center Cubic, BCC



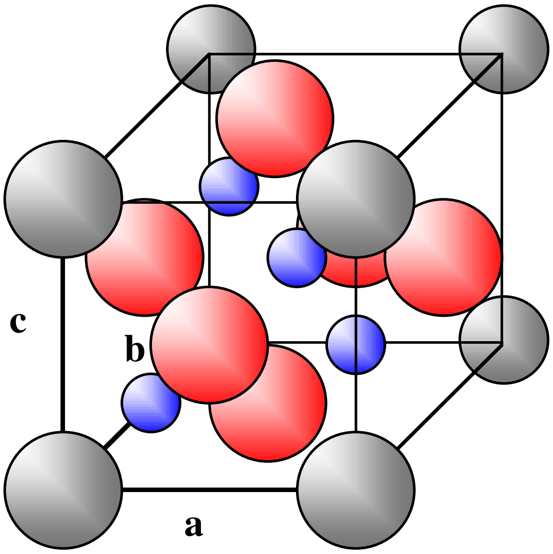
Here the ‘new’ atom is at **a**/2 + **b**/2 + **c**/2

3) Face Center Cubic, FCC



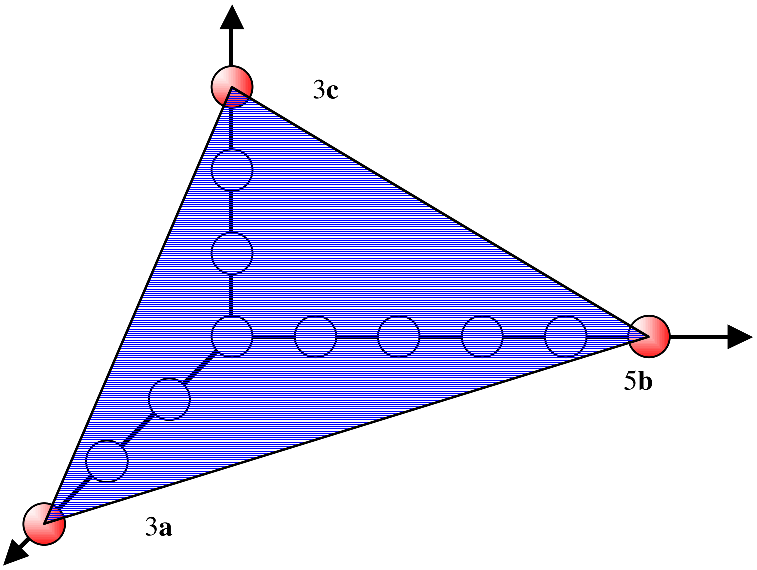
Here the 'new' atoms are at (**a**/2 + **b**/2), (**b**/2 + **c**/2), (**a**/2 + **c**/2), (**a** + **b**/2 + **c**/2), (**a**/2 + **b** + **c**/2), (**a**/2 + **b**/2 + **c**).

4) Diamond Lattice
   The diamond lattice is fairly difficult to draw. However, it is very important as it is the typical lattice found with Si, the leading material used in the semiconductor industry.



A Diamond lattice starts with a FCC and then adds <u>four</u> additional INTERAL atoms at locations **a**/4 + **b**/4 + **c**/4, away from each of the atoms.

Now that we have described a few of the simple crystal types, we need to figure out how to describe a location in the crystal. We could use our basis vectors, a, b and c, but it has been found that this is not the most advantageous description. For that we turn to MILLER INDICES. Miller Indices define both planes in the crystal and the direction normal to said plane. As we know, all planes are defined by three points. Thus, one can pick three Lattice points in the crystal and hence define a plane. From these three points, we can find an origin that is such that travel from the origin to each lattice point is only along one basis vector and the distance is an integer multiple of that same basis vector. Thus our points are located at i**a**, j**b** and k**c**, where i, j and k are integers.



To determine Miller Indices does the following:

1) Determine the proper origin and the associated integers, i, j k.
2) Invert i, j and k. Thus (i,j,k) => (1/i,1/j,1/k). In our picture above we find that (3,5,3) goes to (1/3,1/5,1/3).
3) Next, one finds the least common multiple of i, j and k and use this to multiple the fractions. In our picture above that multiple is 15. Thus we find (1/3,1/5,1/3) goes to (5,3,5). This is our Miller Index.

If one of the integers is negative, it is denoted with a 'bar' over the number. Thus (-5,3,5) is written $(\bar{5},3,5)$

Often, multiple planes are equivalent. These are denoted with curly brackets {}. In a SC lattice, all faces appear to be the same. Thus {1,0,0} ⇔ (1,0,0), (0,1,0), (0,0,1), $(\bar{1},0,0)$, $(0,\bar{1},0)$, $(0,0,\bar{1})$.

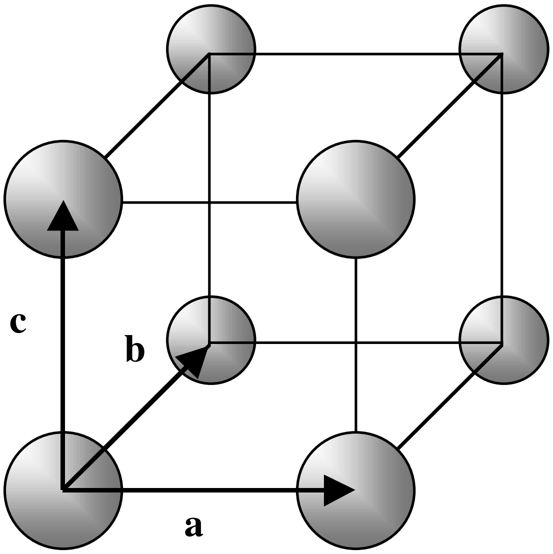
In addition to the planes, we might also be interested in a vector, i.e. moving in a given direction for a given distance. In this field, vectors are denoted with square brackets, []. Thus a vector V= 1.5**a**+1**b** => [1.5,1,0] or more commonly [3,2,0], since we always want to move from lattice point to lattice point. Equivalent vectors are given with angle brackets, <>.

Of interest is that the plane given by (x,y,z) has a normal of [x,y,z].

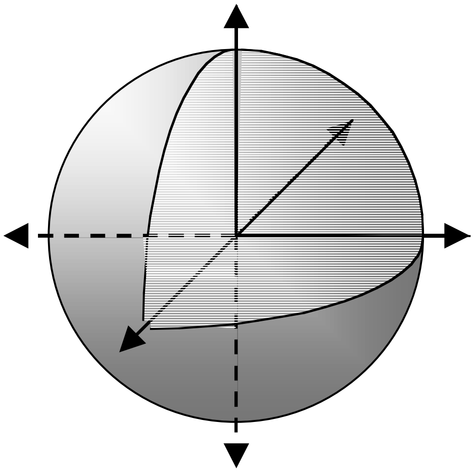
Example Problem:

Q: What fraction of a SC Lattice can be filled by the atoms?
A: Let us assume that the atoms are perfect hard spheres. This is an approximation known as the "HARD SPHERE" approximation. Further let us assume that the atoms are touching their nearest

neighbor. This is known as the "HARD PACK" approximation. Now each of the sides of the SC have a length of a. ('a' is not to be confused with the vector '**a**'.) Thus the volume of the cube is a^3. Now we need to determine how much of each atom is inside the cubic volume. For this we need to look at our picture of the SC lattice.



Let us look in more detail at the atom at the origin.



We see that 1/8 of each atom is inside the cube. Thus the total volume of atoms in the cube is 8*1/8*volume of an atom = $\frac{4}{3}\pi r^3 = \frac{4}{3}\pi\left(\frac{a}{2}\right)^3 = \frac{1}{6}\pi a^3$. This means that the fraction of the volume filled by the atoms is $\frac{1}{6}\pi \approx 0.52 = 52\%$.

Chapter 2 Carrier Modeling

Read Sections 9.1 and 9.2

The late 1800s and the early 1900s set the stage for modern electronic devices. A number of experiments showed that classical mechanics was not a good model for processes on the very small scale. Among these experiments were the following:

1) Light passed through two slits clearly shows an interference pattern. This means that light must be treated as a wave. However, light hitting a metal surface causes the ejection of an electron, which indicates a particle nature for light. Further, it was found that the energy of the ejected electrons depends only on the frequency of the incident light and not the amount of light.
2) Electrons passed through two slits clearly show an interference pattern but they had clearly been found to be particles.
3) In 1911, Rutherford established that atoms were made of ‘solid’ core of protons and neutron surrounded by a much larger shell of electrons. For example Hydrogen has a proton at the center with a electron orbiting it. However, classic electromagnetism combined with classical mechanics implies that the electron must continue to lose energy (through radiation of electromagnetic waves – light) and collapse to the center of the atom. Clearly this was not happening.
4) A spectrum of radiation (light) is observed to come from heated objects that did not follow standard electromagnetism. [This radiation is known as ‘blackbody’ radiation.] A theory based on the wave nature of light was not able to account for this – in fact the theory predicted what was known as ultraviolet catastrophe – where by the amount of energy given off in the UV went to infinity.
5) Hydrogen atoms (and all other atoms and molecules) were found to give off light at well-defined frequencies. Further these frequencies exhibited a interesting series of patterns that did not follow any known model of the nature of physical matter.
6) Electrons shot through a magnetic field were observed to have an associated magnetic field. Further this field could be either ‘up’ of ‘down’ but no place in between.

A rapid series of new models were developed which began to explain these observed phenomena.

1) 1901 Planck assumed that processes occurred in steps, ‘Quanta’, and thus was able to accurately predict Blackbody radiation.
2) 1905 – Einstein successfully explained the photoelectric effect using a particle nature for light.
3) 1913 – Bohr explained the spectra of the Hydrogen atom by assuming a quantized nature for the orbit of electrons around atoms.
4) 1922 – Compton showed that photons can be scattered off of electrons
5) 1924 – Pauli showed that some ‘particles’ are such that they cannot occupy the same location at the same time (The Pauli exclusion principle).
6) 1925 – deBroglie showed that matter such as electrons and atoms exhibited a wave-like property as well as the standard particle-like property. $p = h/\lambda = \hbar k$, where p is the momentum, h is constant (Planck’s Constant), λ is the wavelength, k is the wavenumber $2\pi/\lambda$ and $\hbar = h/2\pi$.
7) 1926 – Schrodinger came up with a wave-based version of Quantum Mechanics.

8) 1927 – Heisenberg showed that you could not know both the time and energy or the momentum and position perfectly at the same instant. Specifically

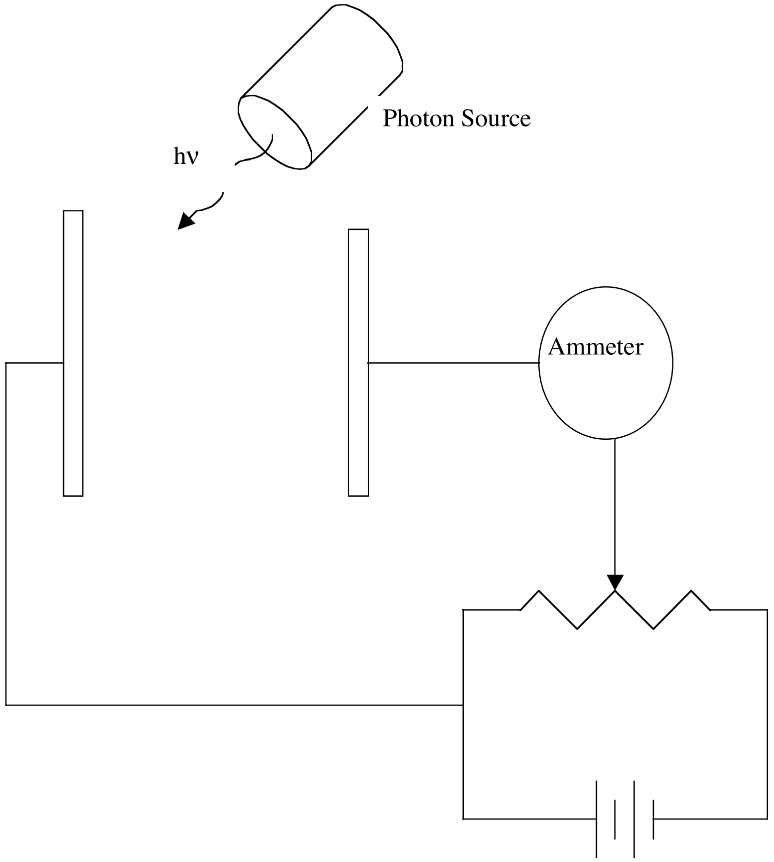
$$\Delta p \Delta x \geq \hbar$$

$$\Delta E \Delta t \geq \hbar$$

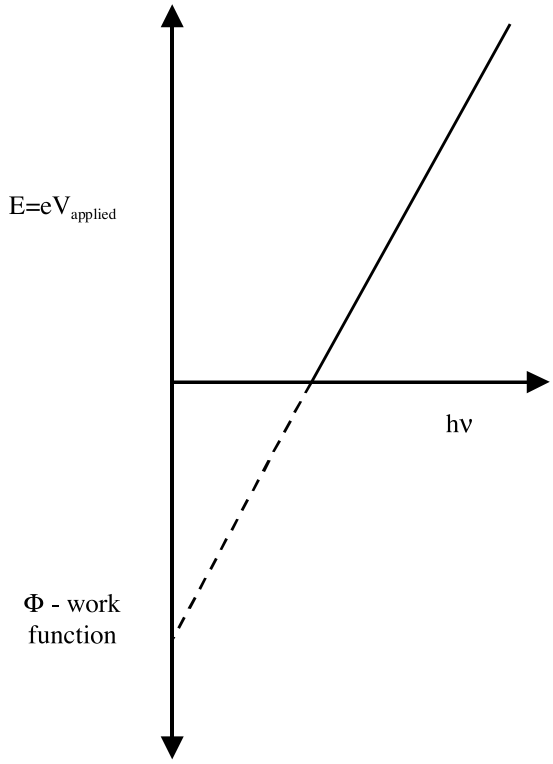
9) Etc.

We will look at three of these in a little detail so that you the students have a little understanding of the principles involved.

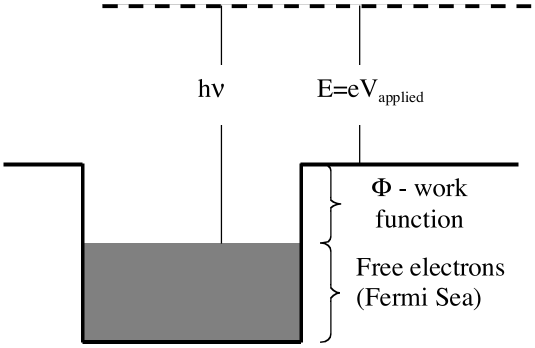
The Photoelectric effect



It is found that the electrons emitted from can be stopped from reaching the collector plate by applying a bias to the collector plate. If one plots the bias required to stop all of the electrons, one finds a very simple curve



Einstein explained this by hypothesizing that light is made up of localized bundles of electromagnetic energy called photons. Each of these photons had the same amount of energy, namely hν, where ν is the frequency of the light and h is a constant, the slope of the line, known as Planck's constant. Sommerfield later proposed a model of a conductor that looks like



Thus, one finds that the electrons in the metal are 'stuck' in a potential energy well. The photons then supply all of their energy to a single electron. The electron uses the first part of the energy to overcome the potential energy well, and the rest remains as kinetic energy.

Bohr model of the Hydrogen atom

Bohr's model of the Hydrogen atom was perhaps the first 'true' quantum model. It does a wonderful job of predicting the then measured frequency of light emitted from an atom. (It misses some 'splitting' of the lines that later improvements to the experiments found and later improved versions of the model deal with correctly.)

The basis of the model is that the path integrated angular momentum of the electron, while in orbit around an atom, is in discrete states that vary as integer multiples of h. Namely,

$$\begin{aligned} p_\theta &= mvr \\ &= nh/2\pi \\ &= n\hbar \\ &\Downarrow \end{aligned}$$

$$v = \frac{n\hbar}{mr}$$

We now have two other equations to work with

The energy of the electron

$$\begin{aligned} E &= K.E. + P.E. \\ &= \tfrac{1}{2}mv^2 - \frac{e^2}{Kr} \end{aligned}$$

The centripetal force on the electron

$$F_{centripetal} = \frac{mv^2}{r} = \frac{e^2}{Kr^2}$$

$$\Downarrow$$

$$r = \frac{e^2}{Kmv^2}$$

$$\Downarrow$$

$$r_n = \frac{K\hbar^2 n^2}{me^2}$$

From this we note that r is a function of n. For n = 1, ‘ground’ state, we find

$$r_1 = a_0 = \frac{K\hbar^2}{me^2} = 0.529\text{Å}$$

where a_0 is the Bohr radius and is the smallest radius at which the electron orbits the proton in the Hydrogen atom. Finally plugging both velocity and radius into our energy equation we find the energy of the electron,

$$
\begin{aligned}
E &= K.E.+P.E. = \tfrac{1}{2}m(v)^2 - \frac{e^2}{Kr} \\
&= \tfrac{1}{2}m\left(\frac{n\hbar}{mr}\right)^2 - \frac{e^2}{Kr} \\
&= \tfrac{1}{2}\frac{n^2\hbar^2}{mr^2} - \frac{e^2}{Kr} \\
&= \tfrac{1}{2}\frac{n^2\hbar^2}{m\left(\frac{K\hbar^2 n^2}{me^2}\right)^2} - \frac{e^2}{K\left(\frac{K\hbar^2 n^2}{me^2}\right)} \\
&= \tfrac{1}{2}\frac{me^4}{K^2\hbar^2 n^2} - \frac{me^4}{K^2\hbar^2 n^2} \\
&= -\tfrac{1}{2}\frac{me^4}{K^2\hbar^2 n^2}
\end{aligned}
$$

We see that the total energy of the electron is 'quantized' with the smaller quantum number having more energy. Again, we can look at the ground state, n=1, and find

$$
\begin{aligned}
E_1 &= R \\
&= -\frac{me^4}{2K^2\hbar^2} \\
&= -13.56eV
\end{aligned}
$$

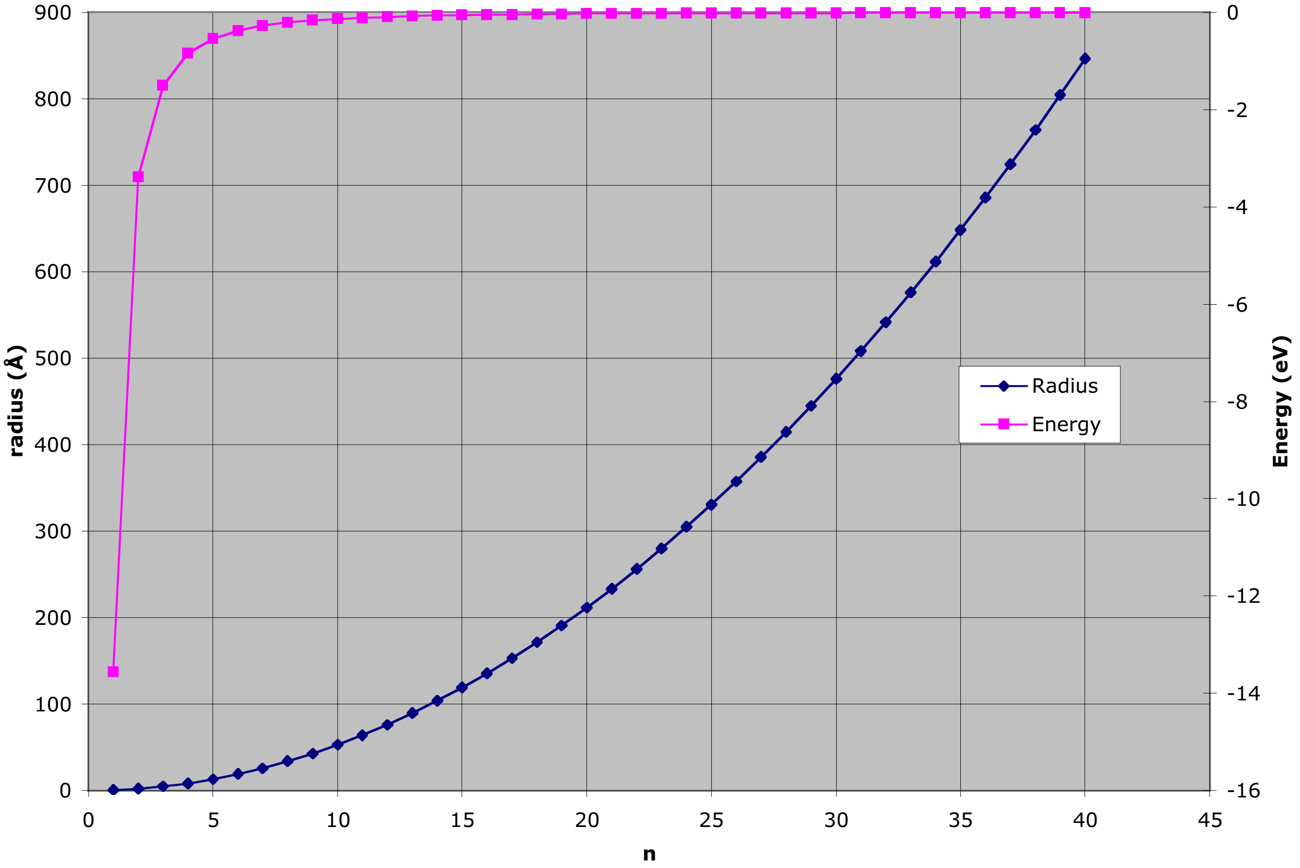
where R is the Rydberg constant and is also the amount energy required to remove an electron from a Hydrogen atom. (This is a processes known as ionization.) This ionization potential 'exactly' matches the experimentally measured ionization energy. The energy emitted/gained between the states is 'exactly' the energy of the photons emitted/adsorbed. (Better experiment showed that the model was not perfect but very close.)

We can extend this model somewhat by assuming that the binding (electric) potential is due to all of the charges **inside** the outer shell. Then we get

$$
\begin{aligned}
r_n &= \frac{K\hbar^2 n^2}{Zme^2} \\
&= 0.529\text{Å}n^2 \quad Z = 1 \\
E_{Bohr} &= -\tfrac{1}{2}\frac{Z^2 me^4}{K^2\hbar^2 n^2} \\
&= -13.56\text{eV}/n^2 \quad Z = 1
\end{aligned}
$$

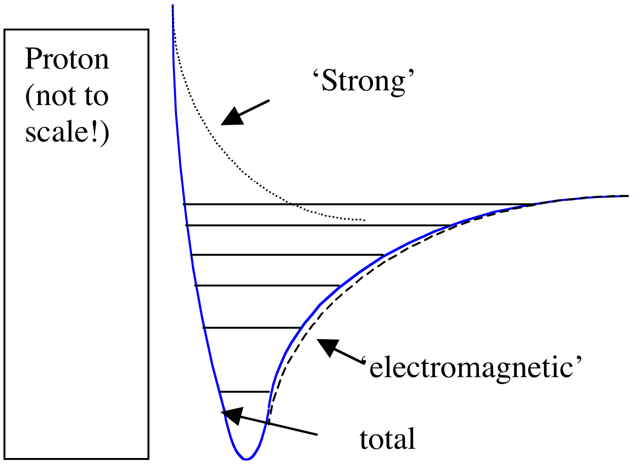
where Z is the number of protons less the number of non-outer shell electrons.

We can now graphically look at the energy and radius as a function of n



If we look at the potential well the electron is trapped in, we see that the higher the energy, the higher the expected radius.

In a true Hydrogen atom, the electron is trapped between the repulsive 'strong force' and the attractive electromagnetic force. The potential well that is created between these forces looks like

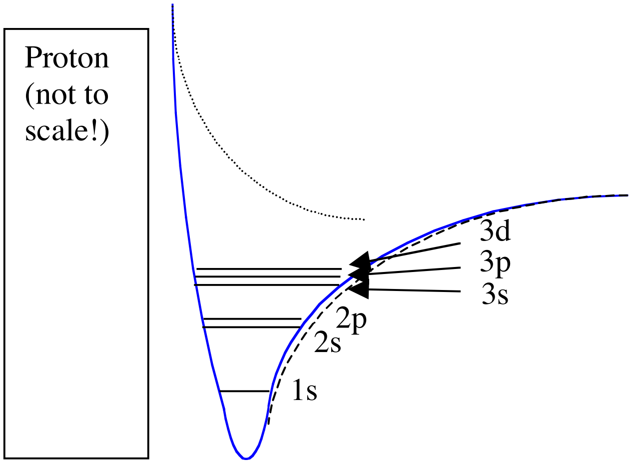


The one major item that Bohr's model missed is a splitting of the levels, or 'shells'. This splitting is due to a splitting in the allowed angular momentum and particle spin (internal angular momentum) in each shell. Thus we find each shell given by a label n has an allowed set of angular momentums, given by a labels l, and labels m, as well as spin given by label s.

The overall requirements are

$n \geq 1$
$L \leq n-1$
$-L \leq m \leq L$
$s = \pm 1/2$

The label 'l' is often replaced with l=0 => 's', l=2 => 'p', l=3 => 'd', l=4 => 'f', (and then follow the alphabet). Thus an electron in shell n=3, l=3 can be labeled 3d. The higher the quantum numbers n and L, the higher the energy. This means that our picture of the potential well now looks like



We can have up to 2(2L+1) electrons in that state because of the possible m's and 's's. We often add a superscript to our label to tell us how many electrons are in a given state thus

3d => $3d^5$
or $3d^2$
etc.

Usually the lowest energy states are the first to fill -

This is in fact why the periodic table is the shape that it is. The Noble gases are on the right hand side and have completely filled – or closed – outer shells. The element on the farthest left will have a [noble]ns^1 configuration, i.e. [He]$2s^1$ is Lithium while [Ne]$3s^1$ is Sodium (Na).

At the close of the 1920, two versions of full fledge Quantum Mechanics were proposed, a wave version of QM by Schrödinger and a particle version, employing matrices, by Heisenberg. These are equivalent yet different and can be used to independently solve problems. For what little QM we do do, we will predominately use Schrödinger's version is this class.

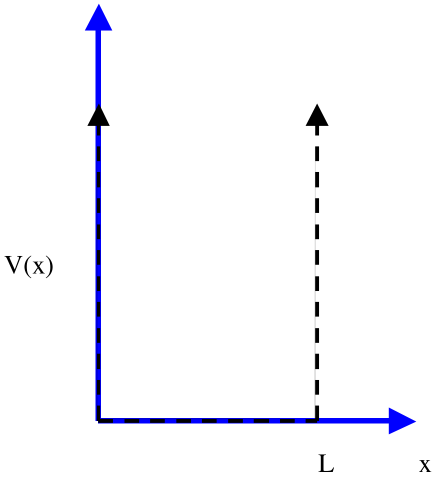
$$K.E + P.E. = E$$

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V\right)\Psi(\mathbf{r},t) = -\frac{\hbar}{j}\partial_t\Psi(\mathbf{r},t)$$

where $\Psi(\mathbf{r},t)$ is the 'wave function', a probability function for the particle/wave, $\partial_t = \frac{\partial}{\partial t}$, *etc* and $\nabla^2 = \partial_x^2 + \partial_y^2 + \partial_z^2$. We note that the wave function implies that we can not know exactly when and where a 'particle' is located at. At best we get a general idea of were it might be. This is a very important concept as it leads to the idea of tunneling, which is very important in some modern devices.

The unfortunate part of QM is that the equations are very hard to solve for any real physical system. The Hydrogen atom has been explored in detail this way but it would take us most of the semester to go through these calculations. As we are interested in understanding devices instead, we now look at some approximate models that will give us a feel for what is physically happening.

The first model of a physical system that we will look at using Schrödinger' equation will be a square well potential. We do this for two reasons, 1) it is a very simple mimic of the Hydrogen atom and 2) it is very similar to real devices that we can build. The potential is such that it is



$$V(x) = \begin{cases} 0 & 0 \le x \le L \\ \infty & elsewhere \end{cases}$$

We can now apply this to Schrödinger's equation

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V\right)\Psi(\mathbf{r},t) = -\frac{\hbar}{j}\partial_t \Psi(\mathbf{r},t)$$

but

$$\Psi(\mathbf{r},t) = \psi(\mathbf{r})\phi(t)$$

so that

$$\frac{1}{\psi(\mathbf{r})}\left(-\frac{\hbar^2}{2m}\nabla^2 + V\right)\psi(\mathbf{r}) = -\frac{\hbar}{j\phi(t)}\partial_t \phi(t) = \mathrm{E(nergy)} = \mathrm{constant}$$

(What we have just done, is known as separation of variables. It is a standard method for solving a multi-dimensional differential equation.) Thus

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + \left(V(\mathbf{r}) - E\right)\right)\psi(\mathbf{r}) = 0$$

Finally going to one dimension we find

$$\left(-\frac{\hbar^2}{2m}\partial_x^2 + \left(V(x) - E\right)\right)\psi(\mathbf{r}) = 0$$

Outside the well, the wave function must be zero, as the potential is infinite. (Or else the second derivative is infinite which is unphysical.) Thus we find

$$\left(\frac{\hbar^2}{2m}\partial_x^2 + E\right)\psi(x) = 0 \quad 0 \le x \le L$$
$$\psi(x) = 0 \qquad elsewhere$$

We start by looking at the 0 to L part and integrate twice to find

$$\frac{\hbar^2}{2m}\partial_x^2\psi(x) = -E\psi(x)$$
$$\Downarrow$$
$$\psi(x) = \psi_0 e^{\pm i\sqrt{2mE}x/\hbar}$$

or

$$\psi(x) = A_0\cos\left(\sqrt{2mE}\,x/\hbar\right) + B_0\sin\left(\sqrt{2mE}\,x/\hbar\right)$$

Now our wave function must be continuous in both zeroth and first order derivatives, so that at x = 0 we find $A_0 = 0$. (Remember $\psi(x) = 0$ elsewhere.) Now at x= L $\psi(x) = 0$ so that

$$\psi(L) = 0 = B_0\sin\left(\sqrt{2mE}\,L/\hbar\right)$$
$$\Downarrow$$
$$\frac{\sqrt{2mE}\,L}{\hbar} = n\pi \quad n = 0,1,2,...$$
$$\Downarrow$$
$$E_n = \frac{\hbar^2 n^2 \pi^2}{2mL^2}$$
$$\Downarrow$$
$$\psi(x) = B_0\sin\left(\frac{n\pi x}{L}\right)$$

Finally, we typically normalize the wavefunction to 1, so that our total probability is ‘1’. This is done by integrating

$$\text{Prop} = \int_{-\infty}^{\infty} \psi^*(x)\psi(x)\,dx \equiv 1$$

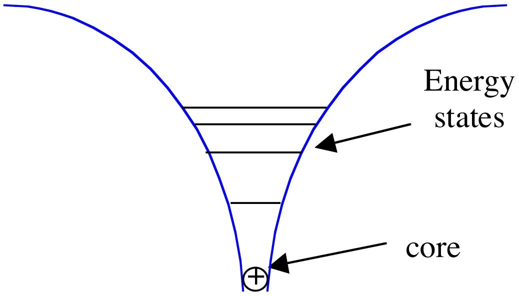
$$\Downarrow$$

$$
\begin{aligned}
1 &= \int_{-\infty}^{\infty} B_0^* \sin\left(\frac{n\pi x}{L}\right) B_0 \sin\left(\frac{n\pi x}{L}\right) dx \\
&= B_0^2 \int_0^L \sin^2\left(\frac{n\pi x}{L}\right) dx \\
&= B_0^2 \int_0^L \tfrac{1}{2}\left(1-\cos\left(\frac{2n\pi x}{L}\right)\right) dx \\
&= B_0^2 \frac{L}{2}
\end{aligned}
$$

Thus,

$$\psi(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right).$$

How is this related to our Hydrogen Atom?
First the higher the value of n the higher we move up the sides of the potential well. Now, if we look at both positive and negative direction of our potential well around the core of the Hydrogen atom, we see a shape that looks like



Where we have ignored the core area where the electron is not allowed.
This, we can model Hydrogen in a way that is very similar to the above. Further, we would expect to see that higher energy states correspond to being higher up the potential well. Because of the shape of the well, we expect the more energetic electrons to orbit at a distance further from the core. This is indeed what we see.

Day 3
Homework set 2
Chapter 3 # 4,7,8,9 Due Sept 10th, 2002

Recap

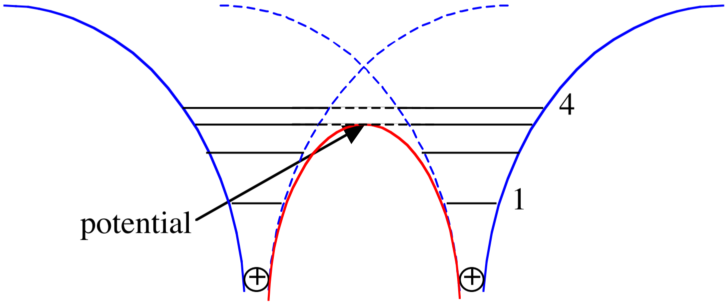
We have now learned a few of things

1) We can know thing only with so much certainty. This is governed by the Heisenberg uncertainty principle
2) We know that particles can act like waves and electromagnetic waves can act like particles. Further the wavelength/momentum relation is given by $p=h/\lambda$.
3) Quantum Mechanics does an excellent job of describing atoms as well as how individual atoms are structured.
4) The Pauli exclusion principle states that two electrons can not occupy the same state in the same location at the same time.

We will briefly look at that last item as is concerns our study of solid materials. There are few things that we need to note.

1) Atoms have discrete energy levels caused by the potential wells around the nucleus.
2) Solids are made up a large number of atoms. These atoms have energy levels as well, but the potential wells are adjusted by the fields from the nearby atoms. Here the Pauli Exclusion principle comes into play.

This leads us to a new issue. We are dealing with atoms that are in close proximity to each other. What happens in such cases? Well let's put two atoms close together and draw the total potential well. This is effectively what happens when two atoms are bonded together.



Here we see that shells 3 and 4 above in each of the atoms 'mix' with the states in the other atoms. This would imply that if both atoms had state 3 filled, then we would have two identical electrons orbiting the two atoms. This cannot happen, rather we get a small splitting of the energy states. Because the potential is lower, the energy of one of the states is typically lower while the other state may be slightly higher. In general, the total energy of the combined state 3 is lower. We known this because if the energy was higher, the combined particles would try to go to a lower state, e.g. an unbound state. Further the average potential that the electrons are sitting in is lower. (This can also be shown with QM.)

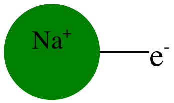
When there are a large number of atoms, say N atoms, we get an equally large number of splits in the energy band structure. Thus, it very common for a gas of a certain species to have a set of very well defined sharp spectral lines, while a solid of the same species will have very broad spectral lines.

We bring this idea up because we are dealing with solid-state devices. Thus the interaction of multiple atoms and atomic species is important to our understanding of this topic. How these atoms bond together is critical to the characteristics of the devices.

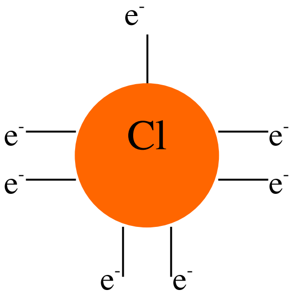
We will now examine bonds between atoms. They fall into four main categories.

1) Ionic	NaCl and all other salts
2) Metallic	Al, Na, Ag, Au, Fe, etc
3) Covalent	Si, Ge, C, etc
4) Mixed	GaAs, AlP, etc.

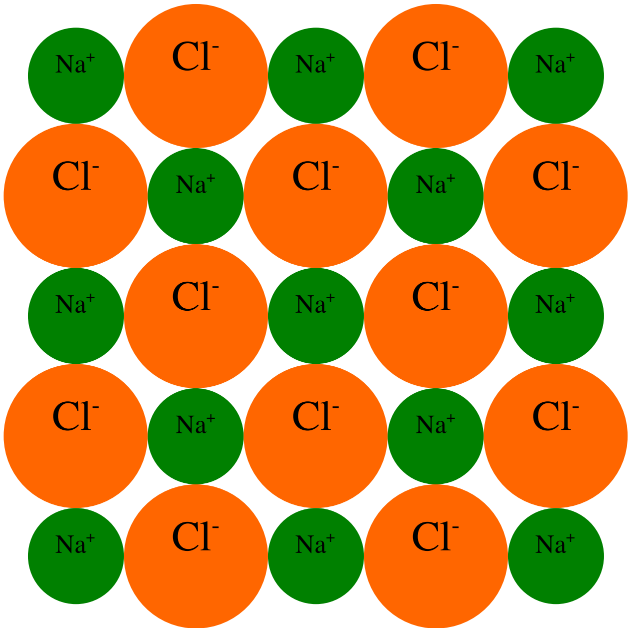
Ionic:
The first of these types of material is related to the complete transfer of an electron from one atom to another. Cl for example would like to have a closed top shell and thus it takes an electron from the Na to produce a [Ar] electron cloud. Sodium on the other hand would like to give up an electron, so to also have a closed shell, in this case [Ne]. [THESE OUTER SHELL ELECTRONS ARE KNOW AS VALANCE ELECTRONS.] Both of these acceptor/donor processes provide lower energy states. This means that the two particles Na^+ and Cl^- are electrostatically pulled together or bonded. The electrons in question, are not shared by the atoms. Picture wise, this looks like



$Na = [Ne]\ 3s^1 \Rightarrow Na^+ = [Ne]$

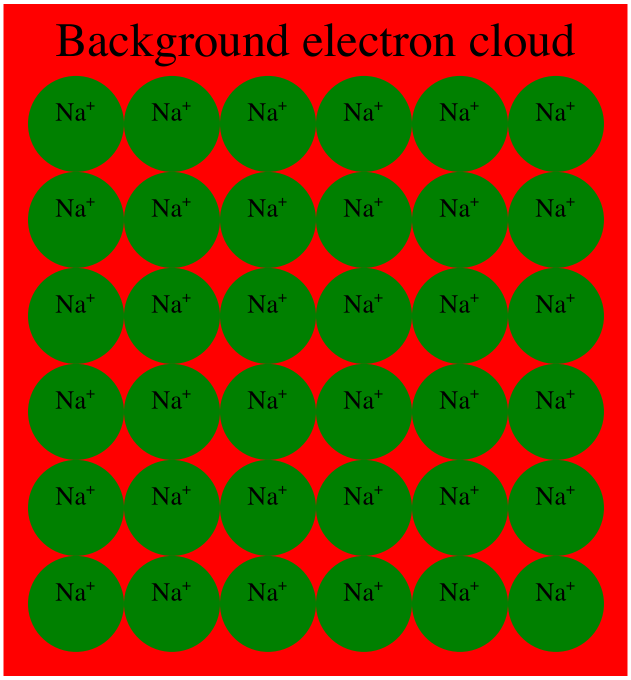


$Cl = [Ne]\ 3s^2 3p^5 \Rightarrow Cl^- = [Ne]\ 3s^2 3p^6 = [Ar]$



Metallic:

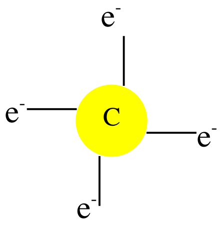
The second of these comes in two forms. The first form has only a few valance electrons in the outer orbital. These outer valance electrons thus tend to be weakly bound to the atoms and are ‘free’ to move around. An example of this type would be Sodium, Na = [Ne]$3s^1$.

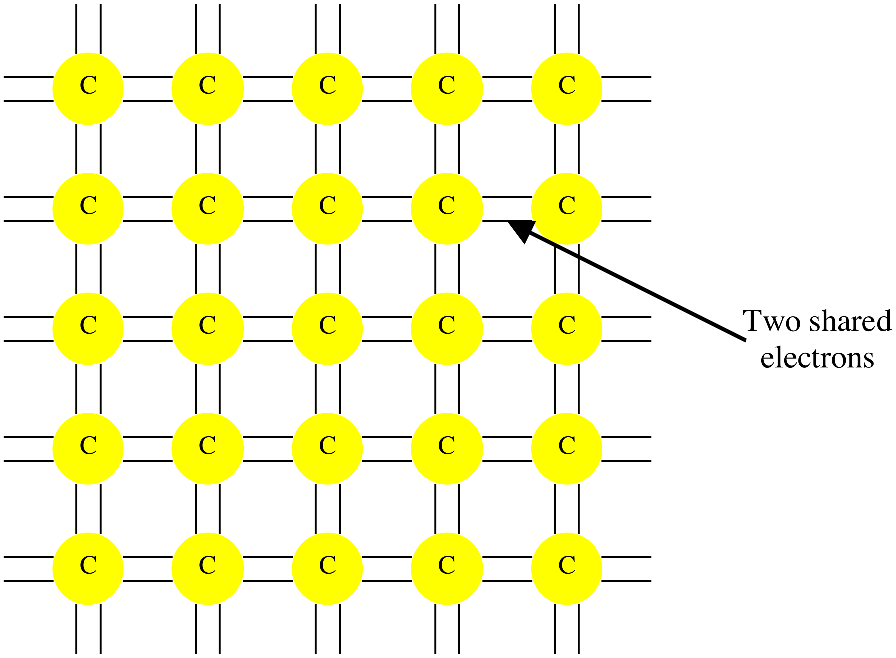


WE WILL DISCUSS THE SECOND TYPE OF METALS BELOW.

Covalent:

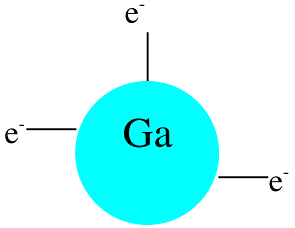
In the covalent bond, two atoms share one or more valance electrons. In this way, each atom thinks that it has a closed outer shell. Because the outer shell is closed, these materials are typically insulators – although some might also be semiconductors. (This in part depends on the size of the atoms. The smaller it is, the more likely it is to be an insulator.) An example of this is Carbon, C=[He]$2s^2 2p^2$.



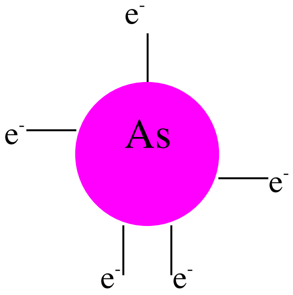


Mixed States:

Mixed states are a combination of Covalent and ionic. An example is GaAs. Picture wise, these look like

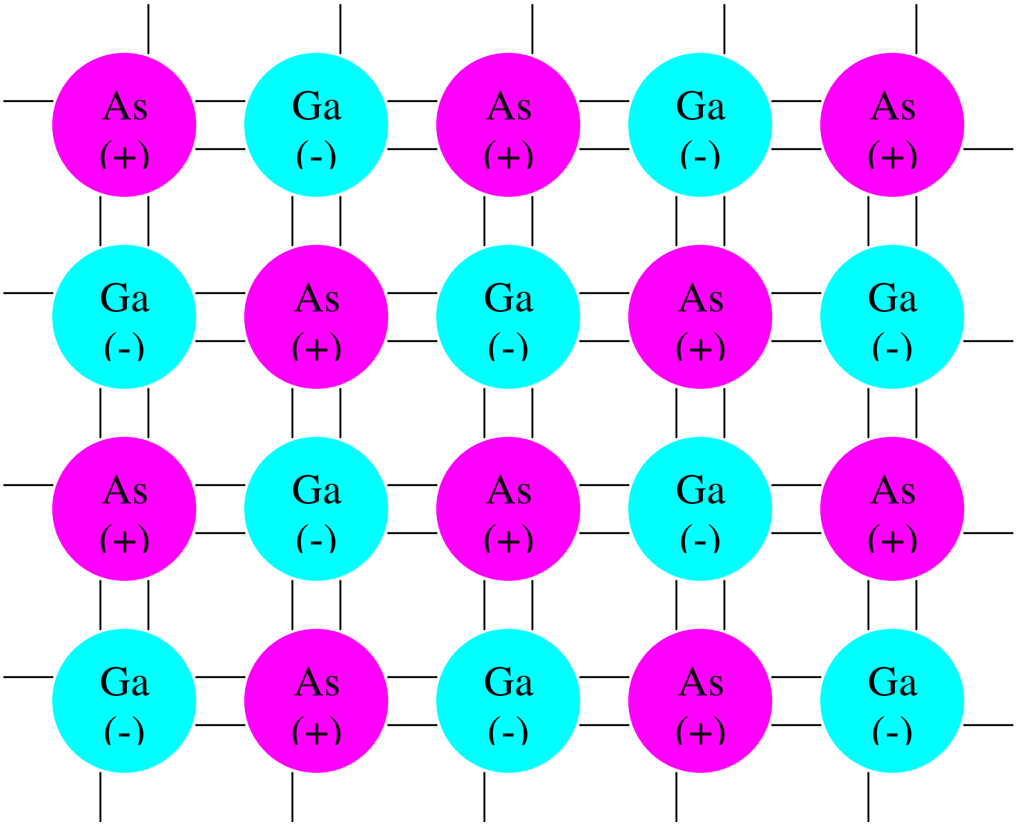


Ga = [Ar] $4s^2 3d^{10} 4p^1$



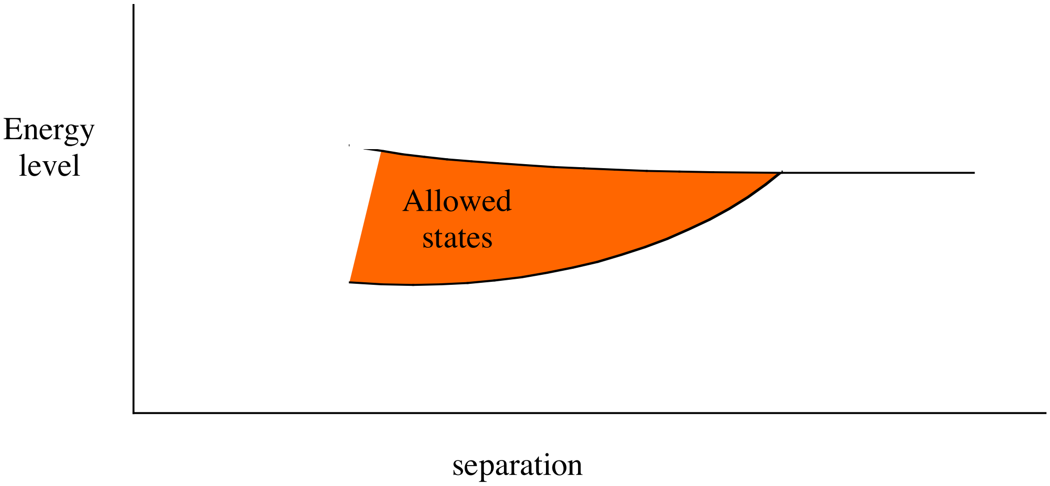
As = [Ar] $4s^2 3d^{10} 4p^3$

We see that Ga, which is column III, has three electrons in the outer shell while As, which is column V, has 5. As such the pair has 8 outer shell electrons, just enough to create a closed outer shell. Ga, it turns out wants to attract an additional electron more than As wants an additional electron. Thus one of the electrons spends more time near the Ga atom, making it partially negatively charged and the As partially positively charged. (A full electron is not transferred.) Thus, GaAs has some properties of covalent bonding and some properties of ionic bonding. The final structure looks like:

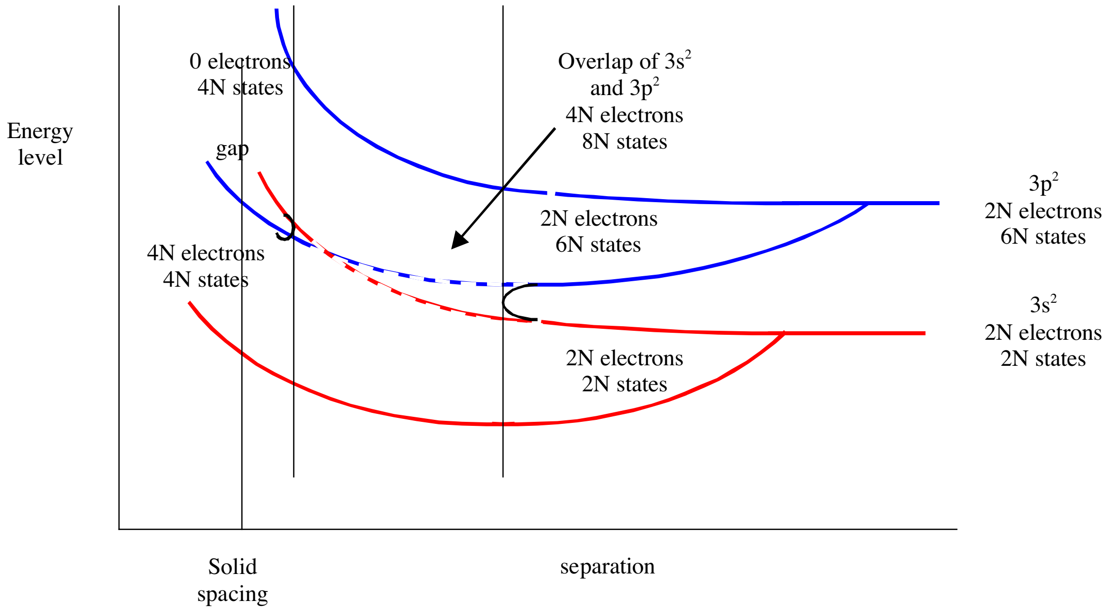


Energy Bands

We can look at this in a second fashion. The properties of materials are the out growth of the splitting of the states in atoms that are close together. If we where to take N atoms and equally space them apart then slowly move them together, we would find that the splitting of the states grows as we get closer together. Hence for a single state we might see

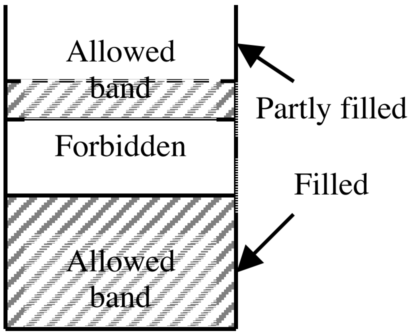


As our atoms get closer and closer, the more of the energy levels begin to split. Thus for an atomic species such as Si, $1s^2 2s^2 2p^6 3s^2 3p^2$ or [Ne] $3s^2 3p^2$, we get

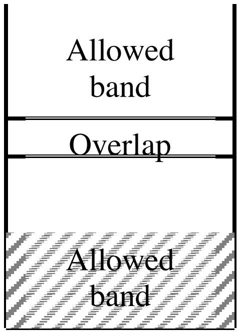


If we could vary the separation of the Si, with in this diagram we see regions that correspond to two types of metal, a semiconductor and an insulator. (Note Si has a specific separation and hence it is a semiconductor.)

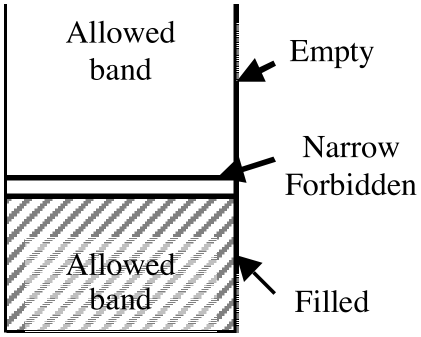
Metal type 1(seen at very large separations)



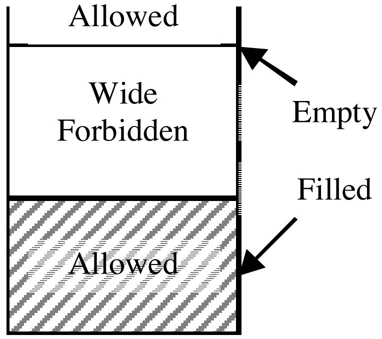
Metal type 2 (seen at large separations)



Semiconductor (seen at moderate separations)



Dielectrics (insulators) (Seen at small separations)



The allowed bands have specific names. The lower band is known as the Valance Band, as this is where all of the outer shell electrons will typically move, while the upper band is known as the conduction band as this is where we find conduction of electrons in solids. [NOTE ELECTRONS ARE NOT THE ONLY CHARGE CARRIERS IN SOLIDS. We will discuss this soon.]

End here day 3

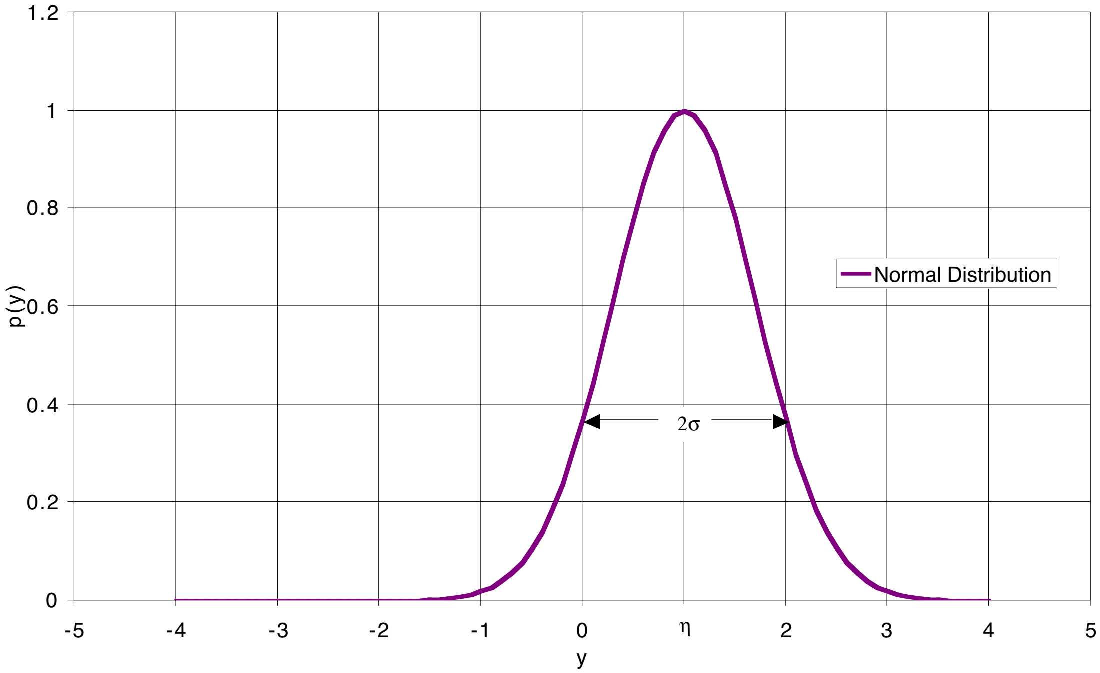
What is important for conduction to occur, is that electrons must have ready access to allowed energy states that are empty. This is in the conduction band. This is because, for the electron to move physically, it needs to have both a position and an energy state to move into. By ready access, we mean that the electron must have enough available energy, through light or random motion (Think Temperature!) to be able to make the transition. What sort of energy might we be talking about? Well, room temperature is about 1/40 eV (or 1 eV ~ 11,000 K). Thus at room temperature, we might expect a significant number of the electrons to be able to gain ~1/40 eV.

> Aside on Temperature:
> The concept of temperature is relatively simple concept. If a material has a temperature that is above 0 K then there is some random internal motion. (This is very different than directed motion where $<v>\neq 0$.) Often we find that the a materials motion is such that $<v>=0$, (velocity has direction and magnitude) while $<|v|>\neq 0$ (speed has only magnitude.) Further, because of the random statistical nature of atoms, the distribution of velocities is a 'Normal' or 'bell curve' distribution. (This is also known as a Maxwellian or Boltzman distribution.) The temperature is

a measure of the width of the distribution. Thus, the higher the temperature, the higher the variation in particle velocities.
The normal distribution is

$$p(y) = const \ \frac{1}{\sigma} \exp\left[\frac{-(y-\eta)^2}{\sigma^2}\right].$$

Here σ^2 is the population variance, σ is the population standard deviation, and η is the central value. The distribution looks like



The Maxwellian distribution in terms of velocity (in 3-D) is given by

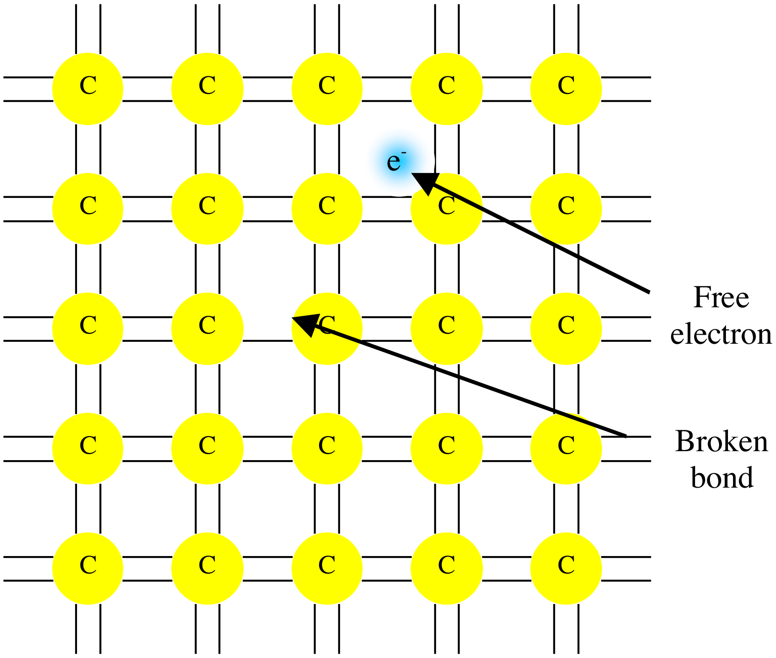
$$f(v) = n\left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left[\frac{-m(v)^2}{2kT}\right].$$

While in terms of energy, it is given by

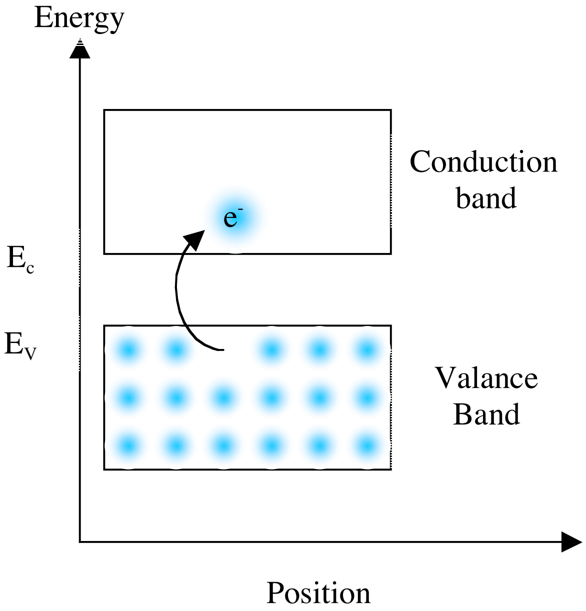
$$f(\mathcal{E}) = n\frac{1}{kT}\exp\left[\frac{-\mathcal{E}}{kT}\right]$$

We will see this again soon.

Now back to how Temperature influences conduction. We will look at C. If the material has enough internal (random) energy, some of the electrons in the covalent bonds may gain enough energy to break free. Picture wise this looks like:



In terms of the energy band, it looks like:



Now we can go back and look at the concept of temperature in terms of the fraction of electrons that can jump from the valance band to the conduction band.
We will use a Maxwellian distribution to approximate the number of electrons that might cross the band gap,

$$f(\mathcal{E}) = n\frac{1}{kT}\exp\left[\frac{-\mathcal{E}}{kT}\right],$$

we see that the higher the temperature, the greater the chance that an electron will have enough energy.

At room temperature, kT=1/40 eV. The band gap for carbon (diamond) is on the order of 3 or 4 eV. If we assume 4 for simplicity, we find that the fraction of electrons that are in the conduction band is

$$\begin{aligned}
\text{Fract} &= \int_{\mathcal{E}_g}^{\infty} \frac{f(\mathcal{E})}{n} d\mathcal{E} \\
&= \frac{1}{kT} \int_{\mathcal{E}_g}^{\infty} \exp\left[\frac{-\mathcal{E}}{kT}\right] d\mathcal{E} \\
&= \frac{1}{kT} kT \exp\left[\frac{-\mathcal{E}_g}{kT}\right] \\
&= \exp\left[\frac{-\mathcal{E}_g}{kT}\right] \\
&\approx \exp[-160] = 3.3E-70 \quad !!!!!
\end{aligned}$$

This means that if the lattice constant for diamond is 4 Å, then the number density of atoms is Number/volume = 8/(4Å)^3 = 1.25E23 Carbon/cm^3. Each carbon atom has 4 electrons in the conduction band and thus, we might expect about 1.6E-46 electrons/cm^3 in the conduction band.

UNDERSTAND THAT THE ABOVE IS A ROUGH APPROXIMATION AND WOULD ONLY TRULY APPLY TO FERMIONS THAT DO NOT 'INTERACT' (Particles that do not interact but do follow Pauli Exclusion Principle). WE WILL GET TO A MORE APPROPRIATE METHOD SHORTLY.

Carrier types and Carrier Properties

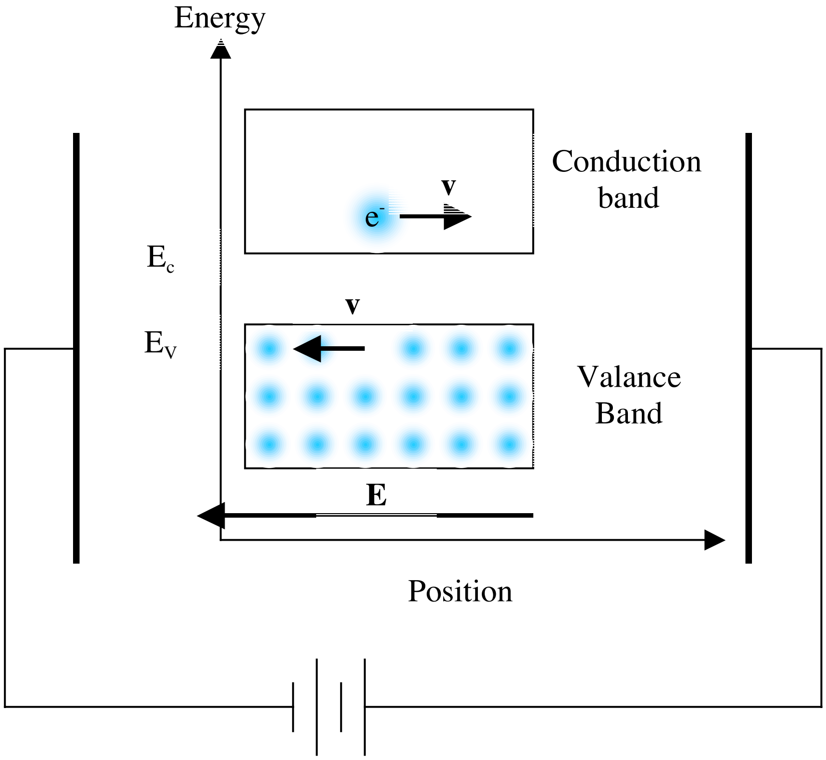
From the above picture, we see that we have moved an electron into the conduction band. This means that it can move around and thus conduct current. However, we have also produced a vacant spot in the valance band. This means that the rest of the electrons in the valance band can now move but just into the empty spot – or 'hole'. But if a valance band electron fills that hole, a new hole must be created from somewhere else. This allows conduction of current through the movement of 'holes' in the valance band.

Now, if we apply an electric field to our material we get a force applied to both the conduction electron and the hole. That force is

$$\mathbf{F} = -e\mathbf{E} \quad \text{(electron)}$$

$$\mathbf{F} = +e\mathbf{E} \quad \text{(hole)}$$

Pictorially this looks like



If we were in free space, we would expect

$\mathbf{F}_n = -e\mathbf{E} = m_n\mathbf{a}$ (electron)

$\mathbf{F}_p = +e\mathbf{E} = m_p\mathbf{a}$ (hole)

where n and p stand for negative and positive or electron and hole.

This brings up a valid question, What is the mass of a hole?!?!?

To answer that question, let us first consider the electron. First, we are in a system that is clearly not free space. As the electron is accelerated toward the positive bias, it will collide with atoms. This will cause it to slow down. This means that the acceleration we see in free space will not be the acceleration we see in the material. We can solve this by defining an effective mass, m*, such that our force equation is correct. Holes will also have an effective mass, as they also move via the movement of valance band electrons. Thus we get in the material

$\mathbf{F}_n = -e\mathbf{E} = m_n^*\mathbf{a}$ (electron)

$\mathbf{F}_p = +e\mathbf{E} = m_p^*\mathbf{a}$ (hole)

[YOU MUST ALWAYS USE THE EFFECTIVE MASS IN YOUR CALCULATIONS!]
Some times $m^*>m_e$ and sometimes $m^*<m_e$. It depends on the specific system.

Holes are not real particles but they act like them. They are in effect the average movement of electrons in the valance band. Because they are easier to deal with than all of the valance electrons, we keep track of them in the valance band. In the conduction band, the number of empty spots is much larger than the number of electrons. Thus, in the conduction band, we keep track of the electrons in the conduction band and holes in the valance band.

Types of Semiconductors

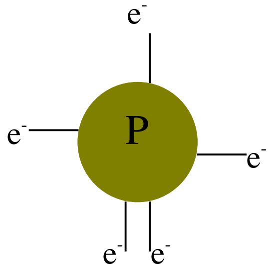
In the above discussion, we have created electrons in the conduction band and holes in the valance band by liberating an electron from its bound state in the valance band. This means that the number of holes and electrons, or p and n carriers, is identical. If we lose one, it is through recombination with a particle of the opposite type. Hence, to get rid of an electron in the conduction band we need to move it into the valance band and have it fill a hole. (This sort of process occurs all of the time. Likewise, we have new electron-hole pairs being created all of the time. In equilibrium, the number of electron-hole pairs created in a time period is equal to the number destroyed.)

Semiconductors that naturally operate with equal numbers of electrons and holes are known as INTRINSIC or pure (less common) semiconductors.

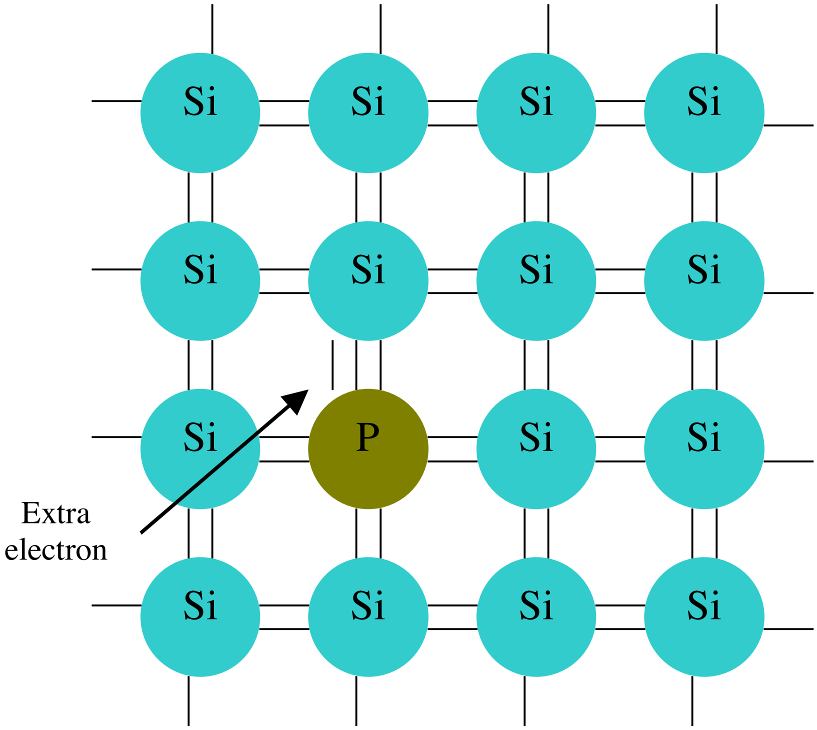
There is also a second class of semiconductors that has an unequal number of holes and electrons. These are known as EXTRINSIC or DOPED semiconductors. There are two basic types of semiconductor dopants, n-type and p-type. n-type give rise to excess negative (electrons in the conduction band) charge carriers, while p-type give rise to excess positive (holes in the valance band) charge carriers. We will discuss each of these in turn.

n-type and p-type dopants

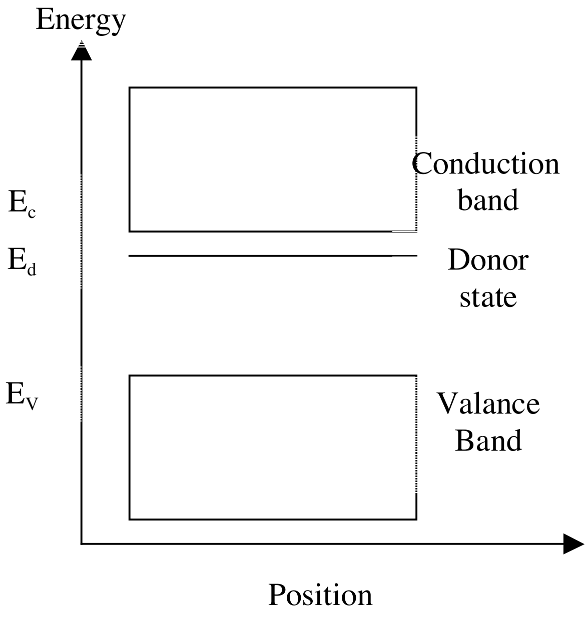
n-type dopants, also known as DONORS, operate in the following manner:
For Si technology, donors come from column V. THUS WE MOVE UP ONE COLUMN TO GET A DONOR. The three most prevalent are P {[Ne]$3s^2 3p^3$}, As {[Ar]$4s^2 3d^{10} 4p^3$} and Sb {[Kr]$5s^2 4d^{10} 5p^3$}. We will look at the valance structure of a P atom.



Now we will now place this in our Si lattice structure

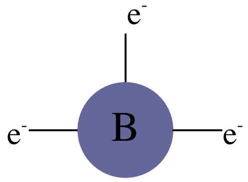


Because this electron is an extra in the covalent bonds of the Si crystal, it takes very little energy to remove it from the P atom and thus move in to the conduction band. [This can be seen from the Bohr model of the hydrogen atom.] In terms of the energy band, it looks like:

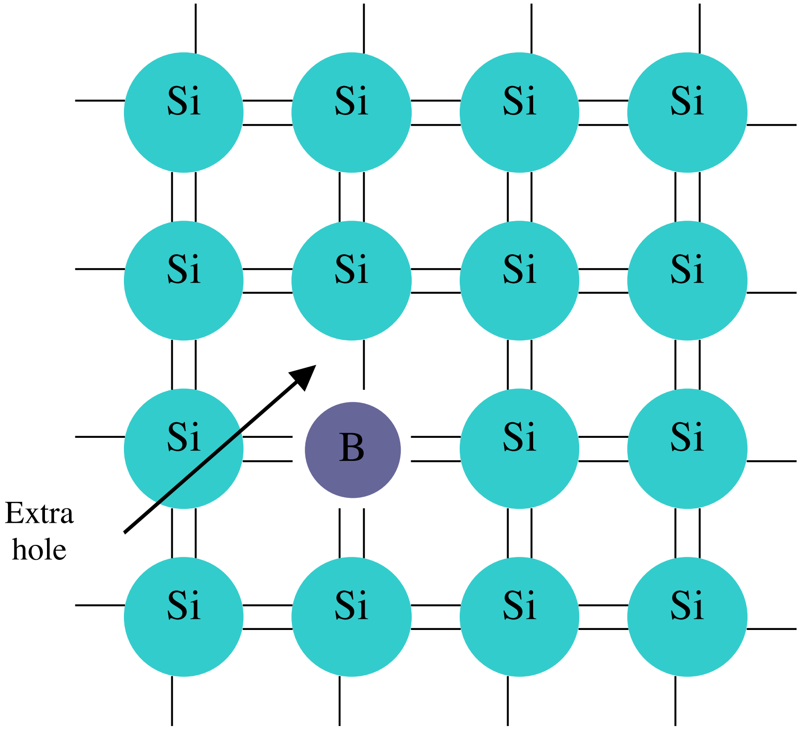


p-type dopants, also known as ACCEPTORS, operate in the following manner:

For Si technology, donors come from column III. THUS WE MOVE <u>DOWN</u> ONE COLUMN TO GET A DONOR. The most prevalent one is B {[He]$3s^2 3p^1$}. We will look at the valance structure of a B atom.

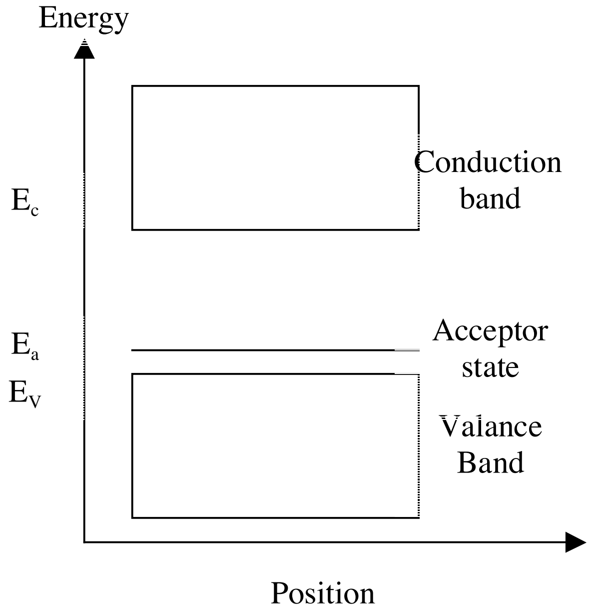


Now we will now place this in our Si lattice structure



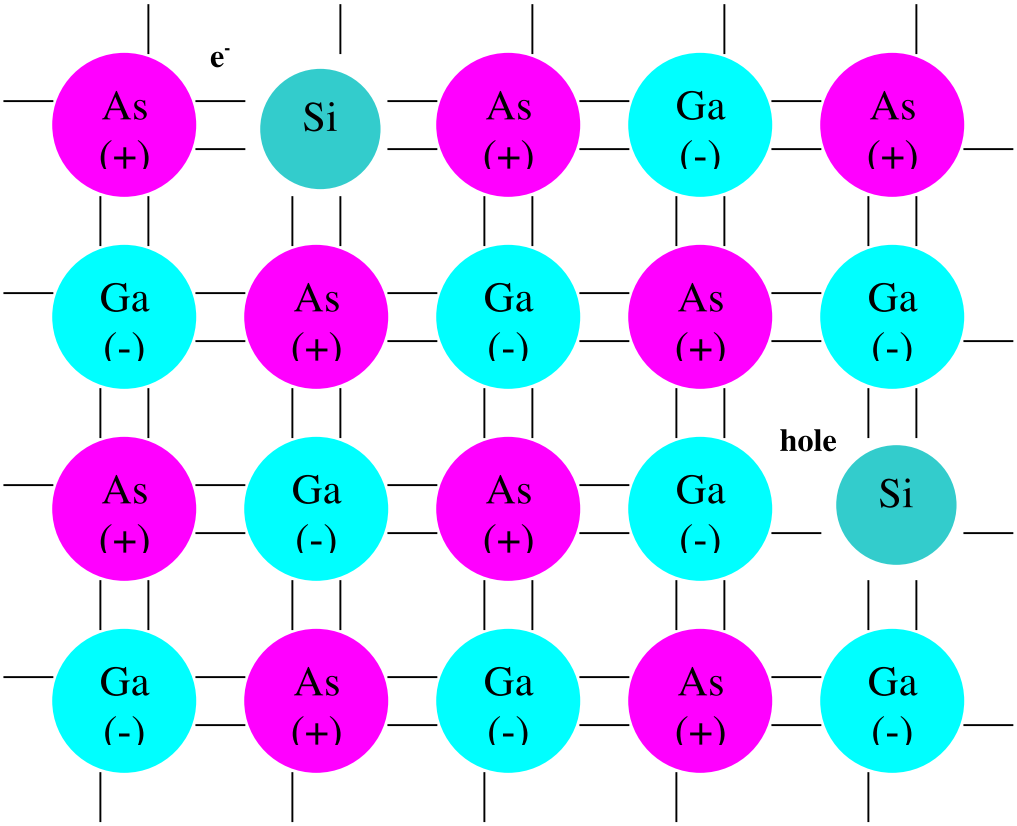
Because this hole is an extra in the covalent bonds of the Si crystal, it takes very little energy to add an electron to the B atom and thus remove in from the valance band. [This can be seen from the Bohr model of the hydrogen atom.] In terms of the energy band, it looks like:

[ONE OTHER THING TO NOTE IS THAT THE B ATOM IS SMALLER THAN THE Si ATOMS. THIS PLAYS A ROLE IN HOW THE CRYSTAL IS STRUCTURED.]



ACCEPTOR, DONORS and AMPHOTERIC atoms in III-V semiconductors.

As with Si/Ge types of technology, to create an acceptor type state, we need to add a dopant that has one fewer valance electrons than the base material. For GaAs, this means we note that our lowest column is Ga of column III and thus we need to pick something from column II, e.g. Be or Mg. This would then sit in the Ga lattice site and act as a hole. To create a Donor, we would have to pick something with one more valance electron and thus from column VI, e.g. S or Se. This would sit in an As site and act as donor. Amphoteric dopants are those that can be either acceptors or donors, depending on which lattice site they occupy. For example Si dopant replacing Ga has an extra electron and thus is a donor. On the other hand, Si on an As site is missing an electron and thus is an acceptor. Species with this dual ability are known as <u>Amphoteric dopants</u>.



[Again note difference in size of atoms.]

End day 4

Carrier Densities

To get to the concept of carrier density, we first have to understand two things

1) How the particles interact and what the appropriate probability distributions.
2) The distribution of energy states available for the charge carriers.

We will start with item 1.

Energy distributions

There are in general two types of particles.

1) Bosons
2) Fermions

These particle types are separated by whether-or-not they follow the Pauli exclusion principle. Boson do not interact and hence do not follow Pauli. [Further all Bosons have integer angular momentum states (spins), e.g. 0, 1, 2…] Photons are an example of a Boson. Since Bosons do not interact we can have as many of them in a place-time as we like. All Bosons follow BOSE-EINSTEIN statistics. [Where you think of Bosons, think of BOZO THE CLOWN. Clowns are infamous for being able to pack dozens of themselves, many more than imaginable, in very small cars.]

Fermions do interact and do follow the Pauli exclusion principle. Further Fermions all have half integer spins (1/2, 3/2 etc.). Electrons are an example of Fermions. [Since Fermions are 'Firm,' you cannot put two of them in the same place at the same time.] Fermions can be further broken into two categories,

a) Those with over overlapping wave functions – and hence always interacting
b) Those without or only briefly overlapping wave functions – and hence seldom or never interacting.

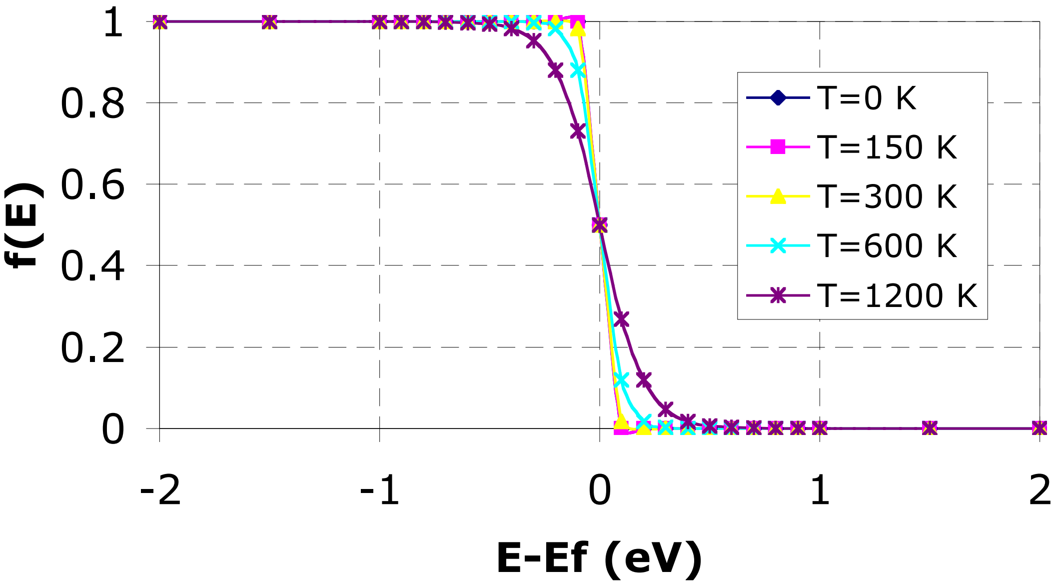
Type (b) is what we typically associate with gas molecules. We know that these follow a 'bell' shaped or normal distribution. This is known as a Maxwell-Boltzmann distribution.

$$f(\mathcal{E}) = n\frac{1}{kT}\exp\left[\frac{-\mathcal{E}}{kT}\right]$$

Type (a) is what we are dealing with in solid-state devices, electrons that are constantly interacting and that obey the Pauli exclusion principle. These particles follow a Fermi-Dirac function. The function is set by rules that are not always simple to follow all the way through – but it can be done – see the back of the book, Appendix III. [The exact derivation is not useful for the purposes of this class.]

$$f(\mathcal{E}) = \frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]}$$

We know what The Maxwell-Boltzmann distribution looks like from above. What does the Fermi-Dirac function look like? **[Note that we are not calling it a distribution!]**



This is very different than the Maxwellian distribution and it is entirely due to the fact that the particle cannot occupy the same energy state at the time. Note also that we cannot normalize the distribution such that the integral over all energies is one. This implies that we are missing a piece of the puzzle, which we will get to shortly. However, we need to look at a few other things first.

First:

For relatively large energy shifts $\Delta\mathcal{E} = \mathcal{E} - \mathcal{E}_0 \geq 3\mathrm{kT}$, we find that the Fermi-Dirac function becomes very like much the Maxwell-Boltzmann distribution.

$$
\begin{aligned}
\mathrm{f}(\mathcal{E} - \mathcal{E}_0 \geq 3\mathrm{kT}) &= \frac{1}{1 + \exp\left[\frac{(\mathcal{E} - \mathcal{E}_0)}{\mathrm{kT}}\right]} \\
&\approx \frac{1}{\exp\left[\frac{(\mathcal{E} - \mathcal{E}_0)}{\mathrm{kT}}\right]} \\
&= \exp\left[\frac{-(\mathcal{E} - \mathcal{E}_0)}{\mathrm{kT}}\right]
\end{aligned}
$$

This means that at high relative energy separations, we can approximate our function with the Maxwellian, and get reasonably good results. [This is in fact what we did with our example of conduction electrons in diamond above.]

Second:

For our situation $\mathcal{E}_0$ is known as the Fermi energy and is written $\mathcal{E}_\mathrm{F}$. Thus we have

$$f(\mathcal{E}) = \frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]}$$

$\mathcal{E}_F$ is very useful. This is true for a number of reasons.

1) When we plug $\mathcal{E}_F$ into our distribution, we find

$$f(\mathcal{E}_F) = \frac{1}{1+\exp\left[\frac{(\mathcal{E}_F-\mathcal{E}_F)}{kT}\right]}$$
$$= \frac{1}{1+1} = \frac{1}{2}$$
.

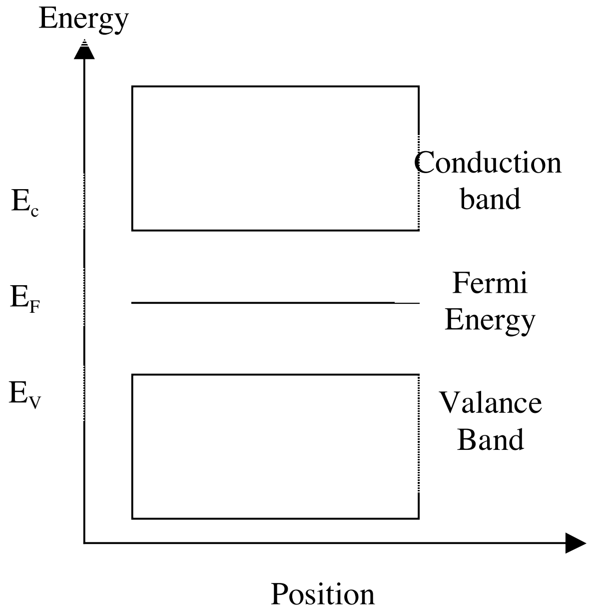
This means that there are an equal number of holes and electrons at the Fermi energy.

2) We also find that the function is 'symmetric' around the Fermi energy. By 'symmetric,' we mean that the number of holes at an energy $\mathcal{E}_x$ below the Fermi energy is equal to the number of electrons at that energy $\mathcal{E}_x$ above the Fermi energy. This can be proved very easily

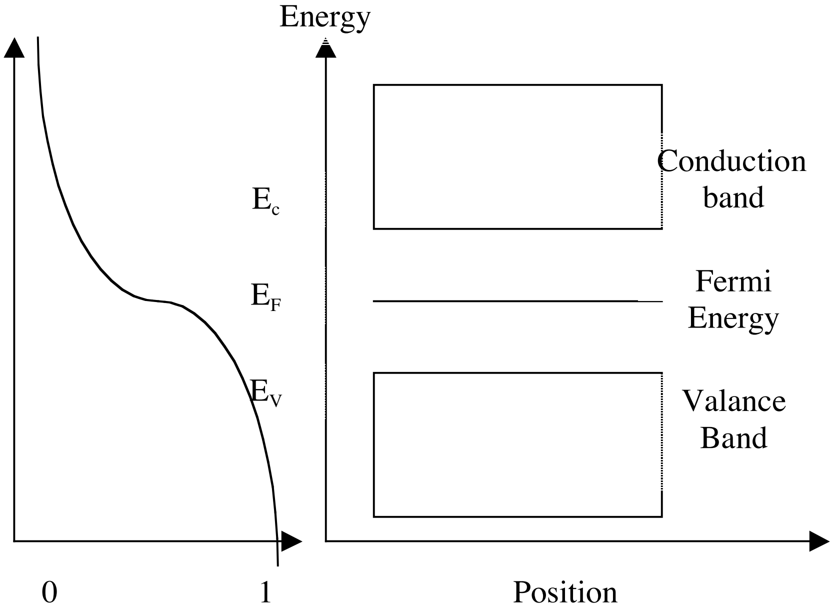
$$f(\mathcal{E}_x + \mathcal{E}_F) = \frac{1}{1+\exp\left[\frac{((\mathcal{E}_x+\mathcal{E}_F)-\mathcal{E}_F)}{kT}\right]}$$
$$= \frac{1}{1+\exp\left[\frac{((\mathcal{E}_x))}{kT}\right]}$$
$$= \frac{\exp\left[\frac{(-(\mathcal{E}_x))}{kT}\right]}{1+\exp\left[\frac{(-(\mathcal{E}_x))}{kT}\right]}$$
$$= \frac{1+\exp\left[\frac{(-(\mathcal{E}_x))}{kT}\right]}{1+\exp\left[\frac{(-(\mathcal{E}_x))}{kT}\right]} - \frac{1}{1+\exp\left[\frac{(-(\mathcal{E}_x))}{kT}\right]}$$
$$= 1 - \frac{1}{1+\exp\left[\frac{(-(\mathcal{E}_x))}{kT}\right]}$$

$$1 - f(\mathcal{E}_F - \mathcal{E}_x) = 1 - \frac{1}{1+\exp\left[\frac{((-\mathcal{E}_x+\mathcal{E}_F)-\mathcal{E}_F)}{kT}\right]}$$

We can now use this to look at where the Fermi energy should lie for an intrinsic semiconductor. By its very nature, the number of holes in the valance band must equal the number of electrons in the conduction band. For simplicity, we will assume that all of the holes and electrons are at their respective band edges, $\mathcal{E}_\text{v}$ and $\mathcal{E}_\text{c}$. From the above discussion, we find that the Fermi energy must be equi-distant from the conduction and valance band edges. (The distribution is 'symmetric'.) [THIS IS NOT QUITE CORRECT AS WE WILL SEE LATER.] Thus if we look at a energy band diagram we see:



We can further complicate the picture by adding the Fermi-Dirac function. However, in our graph of the function above, we have energy along the horizontal axis. In our energy diagrams, we have the energy along the vertical axis. Thus, we need to rotate the function to get:



This is often not done as simply knowing the temperature and the Fermi energy uniquely defines the distribution over which electrons will spread. Thus often times one only sees the Fermi energy on the band energy diagram.

Likewise, we tell something about the semiconductor by the position of the Fermi energy.

1) If there are more holes than electrons (p>n) than the Fermi energy must be closer to the valance band.
2) If there are more electrons than holes (n>p) than the Fermi energy must be closer to the conduction band.

This is entirely due to the 'symmetry' of the function and can be seen by moving the function relative to the band structure.

STATE DENSITY

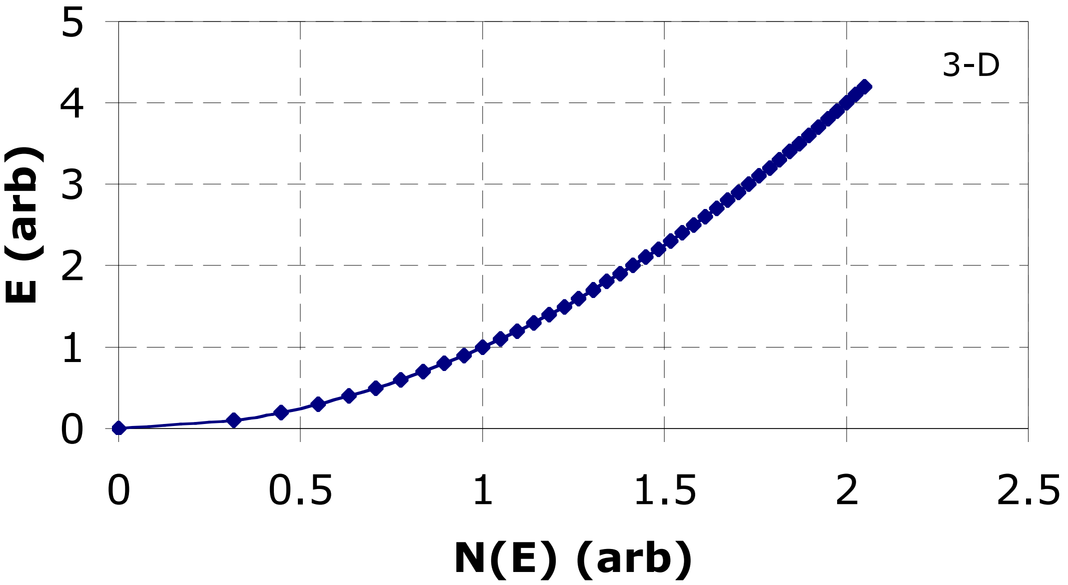
Up to this point we have ignored the fact that the Fermi-Dirac function cannot be normalized to a physically reasonable number. The reason that this is so, is because we have not included the fact that we have a set of discrete states that the electrons can occupy. The density of possible states is determined by quantum mechanical rules that are not simple to follow all the way through – but one can come up with a convincing argument as to why it has to be what it is – see the back of the book, Appendix IV. [The exact derivation is not useful for the purposes of this class.] This density is:

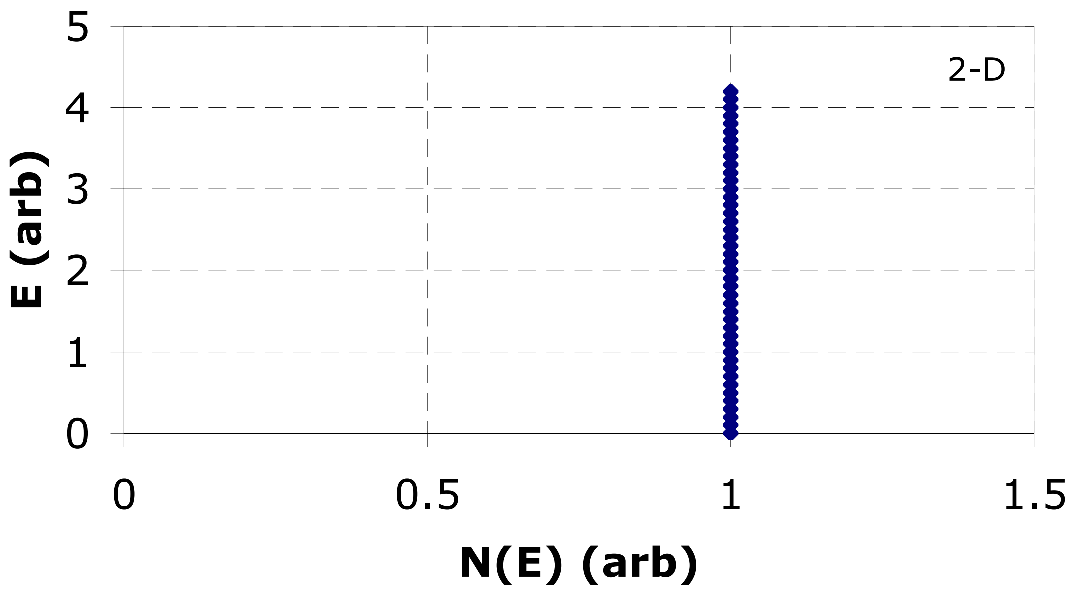
$$N(\mathcal{E})d\mathcal{E} = \frac{2}{(2\pi)^3} 4\pi k^2 dk = \frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2} \mathcal{E}^{1/2} d\mathcal{E} \qquad \text{3-D}$$

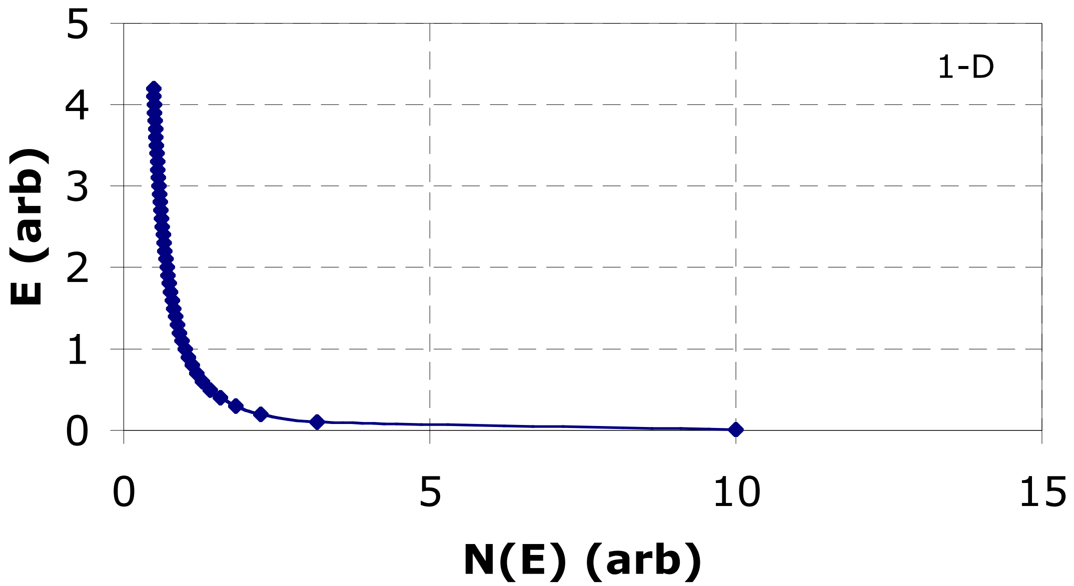
$$N(\mathcal{E})d\mathcal{E} = \frac{2}{(2\pi)^2} 2\pi k dk = \frac{m^*}{\pi\hbar^2} d\mathcal{E} \qquad \text{2-D}$$

$$N(\mathcal{E})d\mathcal{E} = \frac{2}{(2\pi)^1} 2dk = \frac{\sqrt{2m^*}}{\pi\hbar} \mathcal{E}^{-1/2} d\mathcal{E} \qquad \text{1-D}$$

Note that $N(\mathcal{E}) = g(\mathcal{E})$ in some books. These functions look like:


3-D
E (arb)
N(E) (arb)
0
1
2
3
4
5
0
0.5
1
1.5
2
2.5





If we take into account that the energies under consideration are relative to the band edges, we find slightly different densities in the conduction and valance bands. (The difference is a simple sign flip in the square-root of the energy term.) For three dimensions they are

$$N_c(\mathcal{E})d\mathcal{E} = \frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}-\mathcal{E}_c}\,d\mathcal{E} \quad \text{Conduction band}$$

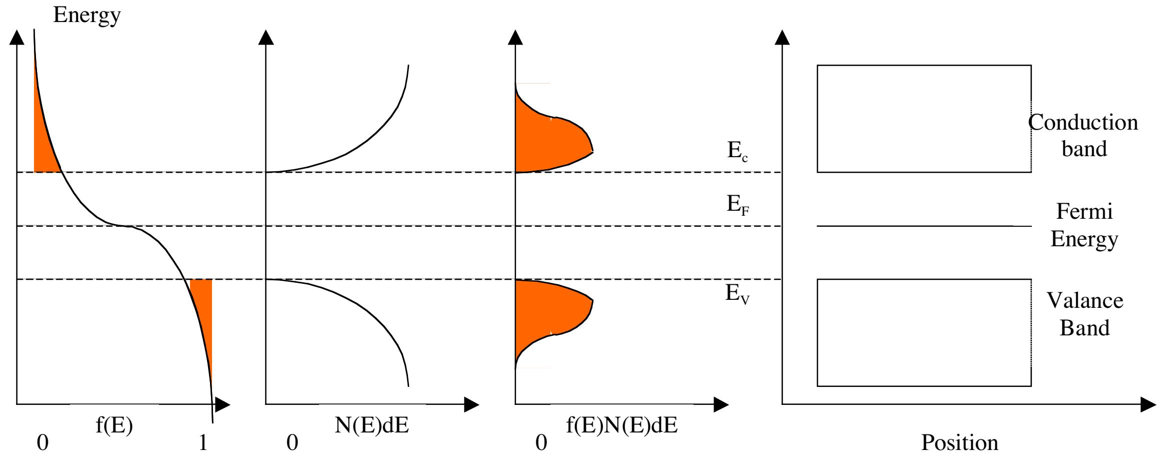
$$N_v(\mathcal{E})d\mathcal{E} = \frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}_v-\mathcal{E}}\,d\mathcal{E} \quad \text{Valance band}$$

Now to get the distribution of states that are fill (electrons) or empty (holes) we need to multiply the Fermi-Dirac function with the state distribution function.

$$n(\mathcal{E})d\mathcal{E} = f(\mathcal{E})N_c(\mathcal{E})d\mathcal{E} = \frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]}\frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}-\mathcal{E}_c}\,d\mathcal{E}$$ Electrons in the Conduction band

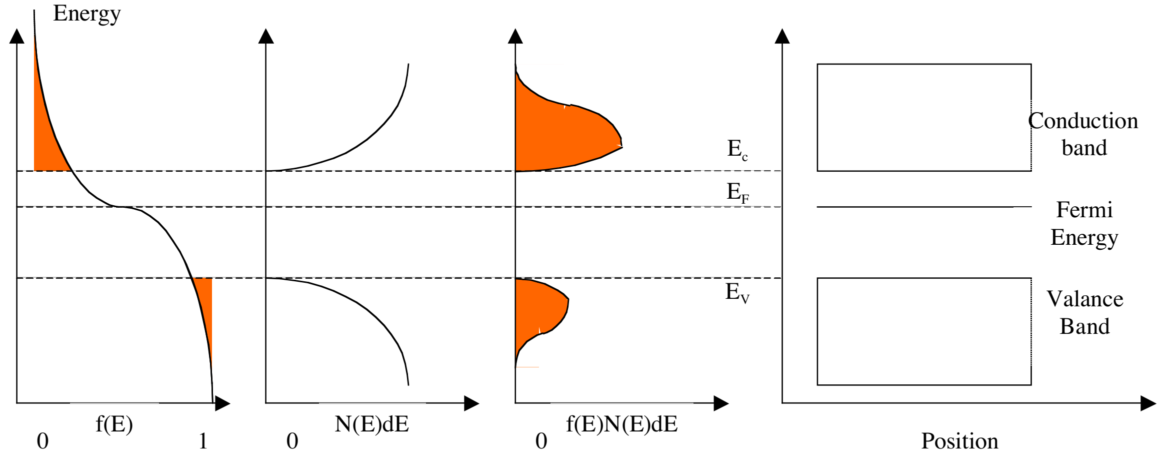
$$p(\mathcal{E})d\mathcal{E} = (1-f(\mathcal{E}))N_v(\mathcal{E})d\mathcal{E} = \left(1-\frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]}\right)\frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}_v-\mathcal{E}}\,d\mathcal{E}$$ Holes in the Valance band

We can look at these again in terms of our band structures:

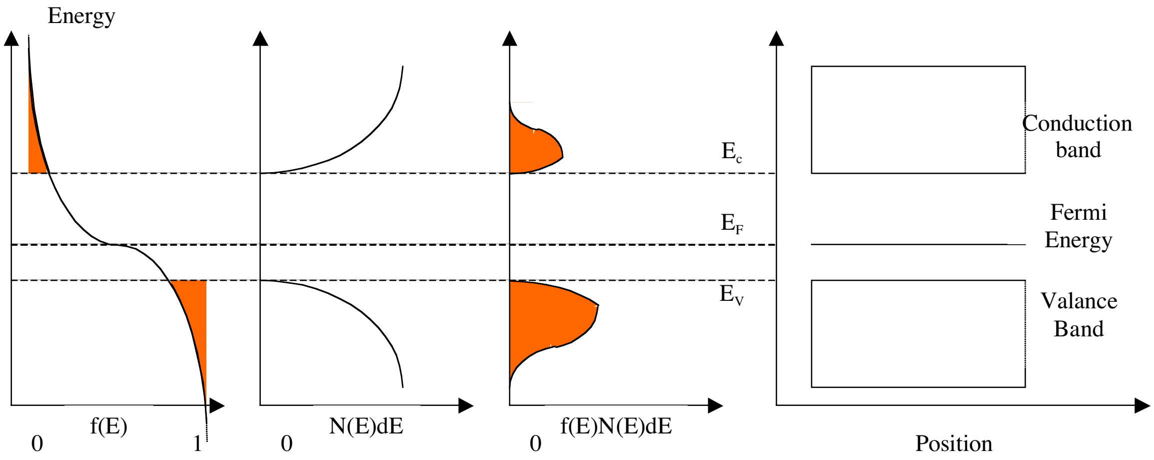


If we move the Fermi Energy up or down we get very different results

Up ⇔ more electrons – p-type dopant



Down ⇔ more holes – n-type dopant



We note that the charge carriers tend to bunch around the band edge.

To get the total number of electrons and holes, we simply integrate over the whole range of energies in each band.

$n_0 = \int_{lower}^{upper} f(\mathcal{E})N_c(\mathcal{E})d\mathcal{E}$ Electrons in the Conduction band

$p_0 = \int_{lower}^{upper} (1-f(\mathcal{E}))N_v(\mathcal{E})d\mathcal{E}$ Holes in the Valance band

Here n_0 and p_0 represent the numbers in thermal equilibrium.

The above equations are difficult to work with – and in fact cannot be solved analytically. We can however use a simplifying assumption, that we have relatively large energy shifts $\Delta\mathcal{E} = \mathcal{E} - \mathcal{E}_0 \geq 3kT$ the equations become significantly easier to deal with.

$$n = \int_{\mathcal{E}_c}^{\infty} n(\mathcal{E})d\mathcal{E} = \int_{\mathcal{E}_c}^{\infty} f(\mathcal{E})N_c(\mathcal{E})d\mathcal{E} = \int_{\mathcal{E}_c}^{\infty} \frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]} \frac{\sqrt{2}}{\pi^2}\left(\frac{m_n^*}{\hbar^2}\right)^{3/2} \sqrt{\mathcal{E}-\mathcal{E}_c}\,d\mathcal{E}$$

$$\approx \int_{\mathcal{E}_c}^{\infty} \exp\left[\frac{-(\mathcal{E}-\mathcal{E}_F)}{kT}\right] \frac{\sqrt{2}}{\pi^2}\left(\frac{m_n^*}{\hbar^2}\right)^{3/2} \sqrt{\mathcal{E}-\mathcal{E}_c}\,d\mathcal{E}$$

$$\approx 2\left(\frac{m_n^* kT}{2\pi\hbar^2}\right)^{3/2} \exp\left[\frac{-(\mathcal{E}_c-\mathcal{E}_F)}{kT}\right]$$

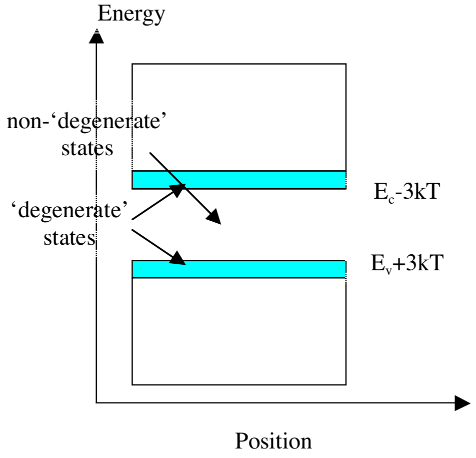
$$= N_c \exp\left[\frac{-(\mathcal{E}_c-\mathcal{E}_F)}{kT}\right]$$

Likewise

$$p \approx 2\left(\frac{m_p^* kT}{2\pi\hbar^2}\right)^{3/2} \exp\left[\frac{(\mathcal{E}_v - \mathcal{E}_F)}{kT}\right]$$

$$= N_v \exp\left[\frac{(\mathcal{E}_v - \mathcal{E}_F)}{kT}\right]$$

N_c and N_v are known as the effective density of conduction and valance band states.



A final way in which we can write these are in terms of the 'intrinsic' energy, E_i, and the 'intrinsic' density, n_i. The intrinsic energy is the energy half way between the conduction and the valance band. (In reality, it is the Fermi energy for the intrinsic material and hence it only has to lie close to the mid energy.) The density is found from the hole/electron densities at that energy.

$$n_i = N_v \exp\left[\frac{(\mathcal{E}_v - \mathcal{E}_i)}{kT}\right] \Rightarrow n = n_i \exp\left[\frac{(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right]$$

$$n_i = N_c \exp\left[\frac{-(\mathcal{E}_c - \mathcal{E}_i)}{kT}\right] \Rightarrow p = n_i \exp\left[\frac{-(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right]$$

or we can multiply the two forms together to get

$$n_i^2 = N_v N_c \exp\left[\frac{(\mathcal{E}_v - \mathcal{E}_i)}{kT}\right] \exp\left[\frac{-(\mathcal{E}_c - \mathcal{E}_i)}{kT}\right]$$

$$= N_v N_c \exp\left[\frac{-\mathcal{E}_G}{kT}\right]$$

$$= np$$

Additional ideas from Streetman and Banerjee

We have been drawing pictures of the energy band structure, showing position on the horizontal axis and energy on the vertical axis. First we note that position is by its nature, 3-D. However, we also need to realize that the momentum of an electron is independent of the location. (Or at least is potentially independent of the location.) This means that we can add three additional coordinates, one each for the momentums in the three different directions. In QM parlance, the total energy of the system is given by

$$E = \frac{1}{2}m\mathbf{v}^2 + V(r,\mathbf{v})$$

but

$$\mathbf{p} = m\mathbf{v}$$

so

$$E = \frac{\mathbf{p}^2}{2m} + V(r,\mathbf{p})$$

When this is used to operate on a wave function (our electron's probability function) we get

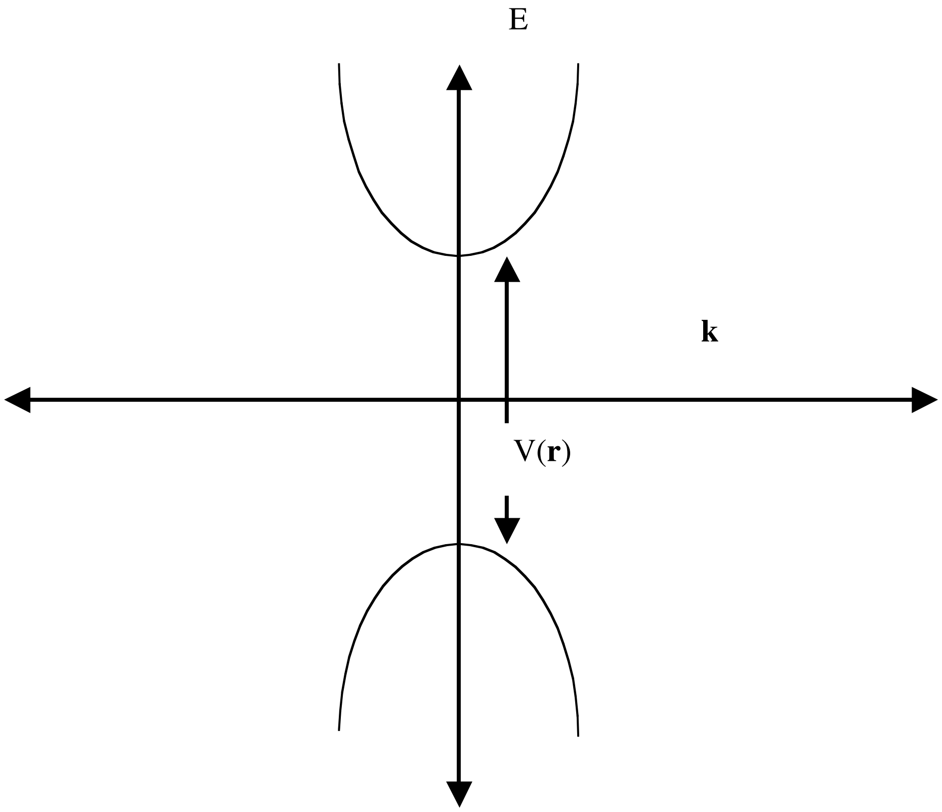
$$E = \frac{\mathbf{k}^2}{2m} + V(r,\mathbf{k})$$

where the vector **k** is known as the 'eigenvalue' of the 'eigenstate'. [It is just a fixed vector quantity that depends on the state that the electron is in.] If there is no potential shift due to the momentum then we get

$$E - V(r) = \frac{\mathbf{k}^2}{2m}$$

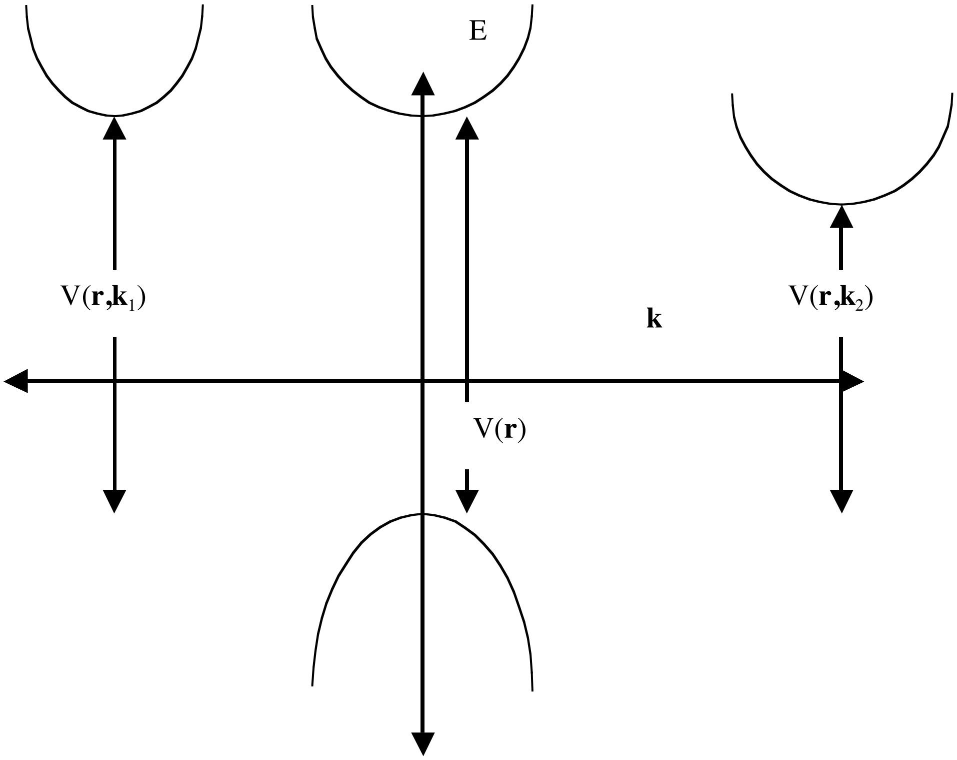
We can plot this and find that the energy-momentum curve acts like a parabola in energy-momentum space. In other words (pictures?):

Direct Band Structure



Sometimes, we get a potential shift due to the momentum, causing a shift in our parabolic curve.

Indirect Band Structure



In general we can and often do have peaks and valleys in our E-**k** space plots. They can and often do become very ugly, including having/giving different m* for each of the valleys. Streetman and Banerjee show a picture along those lines, which I will not try to duplicate here.

This however brings up an interesting issue. The electrons in the conduction band typically occupy the lower part of the band – or in reality any valley in the conduction band. One can understand this by think about what happens to a bunch of balls randomly drop on an area of small hills. From our understanding of classical mechanics, we would fully expect that the balls to be somehow distributed among the valleys, with the largest number in the deepest/widest/lowest valley. (A very deep but very skinny valley may not allow any balls to enter, while a very wide but very shallow valley may not trap many balls either – so balance must be reached.)

Fortunately we can use our model derived above to arrive at the number of electrons in the different valleys in the conduction band. We will show how this is done by example.

Example problem
We start this making a few simple assumptions.

1) All objects have some amount of kinetic energy.
2) All objects have some amount of momentum.
3) The effective mass in each of these valleys can and often is different.

We know from above that the density of states is given by:

$$N_c(\mathcal{E})d\mathcal{E} = \frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}-\mathcal{E}_c}\,d\mathcal{E}$$

Let us assume that we have two valleys. Number one due to the [1,0,0] planes and number two due to the [1,1,1] planes. Thus we have for valley one has 2 equivalent planes and thus:

$$\begin{aligned} N_{c1}(\mathcal{E})d\mathcal{E} &= \sum_{\text{planes}}\frac{\sqrt{2}}{\pi^2}\left(\frac{m_1^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}-\mathcal{E}_{c1}}\,d\mathcal{E} \\ &= 2\frac{\sqrt{2}}{\pi^2}\left(\frac{m_1^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}-\mathcal{E}_{c1}}\,d\mathcal{E} \end{aligned}$$

,

while valley two has 8 equivalent planes and thus for valley two

$$\begin{aligned} N_{c2}(\mathcal{E})d\mathcal{E} &= \sum_{\text{planes}}\frac{\sqrt{2}}{\pi^2}\left(\frac{m_2^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}-\mathcal{E}_{c2}}\,d\mathcal{E} \\ &= 8\frac{\sqrt{2}}{\pi^2}\left(\frac{m_2^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}-\mathcal{E}_{c2}}\,d\mathcal{E} \end{aligned}$$

Multiplying by the Fermi function and integrating (assuming non-degenerate) we find

$$n_{01} = 2\left[2\left(\frac{2\pi m_{n1}^* kT}{\hbar^2}\right)^{3/2}\exp\left(\frac{-(\mathcal{E}_{c1}-\mathcal{E}_F)}{kT}\right)\right]$$

$$n_{02} = 8\left[2\left(\frac{2\pi m_{n2}^* kT}{\hbar^2}\right)^{3/2}\exp\left(\frac{-(\mathcal{E}_{c2}-\mathcal{E}_F)}{kT}\right)\right]$$

we can now ask at what temperature they will have the same number of electrons. We find this by setting $n_{01} = n_{02}$. A few algebra steps gives (provided I have done these steps correctly!):

$$kT = \frac{\mathcal{E}_{c2}-\mathcal{E}_{c1}}{\frac{3}{2}\ln\left[\frac{4m_{n2}^*}{m_{n1}^*}\right]}$$

Start lecture 5
Homework Set 3 Chapter 3: 10, 11, 12
Note: A good web site to look at: http://jas2.eng.buffalo.edu/applets/

At this point we should probably stop at look at what we have done.

First, we found that the electrons (holes) where distributed over a range of energies. How they are distributed is given by the Fermi Function:

$$f(\mathcal{E}) = \frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]}$$

Second, we found that the electrons (holes) can only occupy certain states, as determined by QM. The available state densities are given by:

$$N_c(\mathcal{E})d\mathcal{E} = \frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2} \sqrt{\mathcal{E}-\mathcal{E}_c}\,d\mathcal{E} \qquad \mathcal{E} > \mathcal{E}_c \qquad \text{Conduction band}$$

$$N_v(\mathcal{E})d\mathcal{E} = \frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2} \sqrt{\mathcal{E}_v-\mathcal{E}}\,d\mathcal{E} \qquad \mathcal{E} < \mathcal{E}_v \qquad \text{Valance band}$$

Third, we found that multiplying the functions gives us our electron (hole) energy distributions:

$$n(\mathcal{E})d\mathcal{E} = f(\mathcal{E})N_c(\mathcal{E})d\mathcal{E} = \frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]}\frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2} \sqrt{\mathcal{E}-\mathcal{E}_c}\,d\mathcal{E} \qquad \text{Electrons in the Conduction band}$$

$$p(\mathcal{E})d\mathcal{E} = (1-f(\mathcal{E}))N_v(\mathcal{E})d\mathcal{E} = \left(1-\frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]}\right)\frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2} \sqrt{\mathcal{E}_v-\mathcal{E}}\,d\mathcal{E} \qquad \text{Holes in the Valance band}$$

We however, are only interested in the total number of electrons in the conduction band (holes in the valance band) and hence we integrate over all allowed states. These integrals are none trivial but provided the Fermi energy is not close (>3kT away) to either the conduction or valance band we can approximate the number densities in equilibrium as

$$n_0 \approx N_c \exp\left[\frac{-(\mathcal{E}_c-\mathcal{E}_F)}{kT}\right]$$

$$p_0 \approx N_v \exp\left[\frac{(\mathcal{E}_v-\mathcal{E}_F)}{kT}\right]$$

where

$$N_c = 2\left(\frac{m_n^* kT}{2\pi\hbar^2}\right)^{3/2}$$

$$N_v = 2\left(\frac{m_p^* kT}{2\pi\hbar^2}\right)^{3/2}$$

or

$$N_{v,c} = 2\left(\frac{m_{p,n}^* kT}{2\pi\hbar^2}\right)^{3/2}$$

$$= 2\left(\frac{m_e kT}{2\pi\hbar^2}\right)^{3/2}\left(\frac{m_{p,n}^*}{m_e}\right)^{3/2}$$

$$= 2.510\text{E}19 \ \text{cm}^{-3}\left(\frac{m_{p,n}^*}{m_e}\right)^{3/2} \quad \text{at} \ \ 300\text{K}$$

At this point we realized that we have defined things in very general terms. However, we can modify our semiconductor by adding dopants, etc. and so we came up with a new set of values that are just for the basic (intrinsic) material, $E_F \Rightarrow E_i$, etc.

$$n_i = N_v \exp\left[\frac{(\mathcal{E}_v - \mathcal{E}_i)}{kT}\right] \Rightarrow n_0 = n_i \exp\left[\frac{(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right]$$

$$n_i = N_c \exp\left[\frac{-(\mathcal{E}_c - \mathcal{E}_i)}{kT}\right] \Rightarrow p_0 = n_i \exp\left[\frac{-(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right]$$

or we can multiply the two forms together to get

$$n_i^2 = N_v N_c \exp\left[\frac{(\mathcal{E}_v - \mathcal{E}_i)}{kT}\right]\exp\left[\frac{-(\mathcal{E}_c - \mathcal{E}_i)}{kT}\right]$$

$$= N_v N_c \exp\left[\frac{-\mathcal{E}_G}{kT}\right]$$

$$= n_0 p_0$$

eliminating the need to even know $\mathcal{E}_c$ and $\mathcal{E}_v$.

Are we missing anything? We need to know n_0 and p_0 to be able to do anything.

We know how to get N_v and N_c from the temperature, m_p^* and m_n^*. All three of these are measurable values. We have ways to measure $\mathcal{E}_G$, which we will discuss latter in the class, while $\mathcal{E}_c$ and $\mathcal{E}_v$ have been eliminated from our equations. Thus the only things that we do not have are our Fermi and Intrinsic energies. How do we get them?

To do this, we need to look at our system again. When we have an intrinsic material, we expect that at 0 K all of the energy sites in the valance band will be filled and all of the sites in the conduction band will be empty. As the temperature is increased, some of the electrons will jump from the valance band to the conduction band, creating electron-hole pairs. This however means that the number of electrons is always equal to the number of holes. This concept is known as charge neutrality. It is guided by more than just counting, it is also guided by the fact that any large separation of charges will lead to strong electric fields that tend to pull the charges back together again. Thus, we set

$$n_0 = p_0$$

Now plugging in our distribution functions, and noting for this system $\mathcal{E}_F = \mathcal{E}_i$ we find

$$n_{0i} = N_c \exp\left[\frac{-(\mathcal{E}_c - \mathcal{E}_i)}{kT}\right] = p_{0i} = N_v \exp\left[\frac{(\mathcal{E}_v - \mathcal{E}_i)}{kT}\right]$$

eliminating terms, we get:

$$\begin{aligned}
\frac{N_c}{N_v} &= \frac{\exp\left[\frac{(\mathcal{E}_v - \mathcal{E}_i)}{kT}\right]}{\exp\left[\frac{-(\mathcal{E}_c - \mathcal{E}_i)}{kT}\right]} \\
&= \exp\left[\frac{(\mathcal{E}_v + \mathcal{E}_c - 2\mathcal{E}_i)}{kT}\right] \\
&\Downarrow \\
\mathcal{E}_i &= \frac{(\mathcal{E}_v + \mathcal{E}_c)}{2} + \frac{kT}{2}\ln\left(\frac{N_v}{N_c}\right) \\
&= \frac{(\mathcal{E}_v + \mathcal{E}_c)}{2} + \frac{kT}{2}\ln\left(\left(\frac{m_p^*}{m_n^*}\right)^{3/2}\right)
\end{aligned}$$

This means that the intrinsic energy lies very near the mid gap energy, with a slight offset due to the effective mass ratio of the electrons and holes. This offset is typically very small.

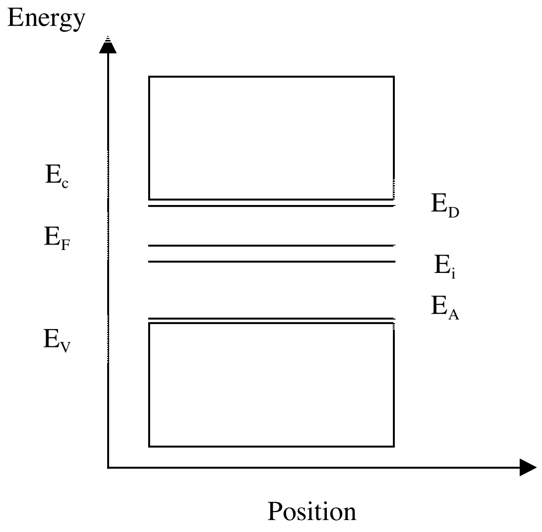
Now, all we need is the Fermi energy. We understand that the Fermi energy is set by energy at which we would expect to have the same number of electrons and holes. If we are dealing with an intrinsic material we have a way to get at that number. However, we are often dealing with a material that has been doped and hence has either excess holes (p-type) or excess electrons (n-type). We need to understand how these dopants affect the Fermi energy in order to understand how to calculate the Fermi energy.

The distinction is that we have added either donors or acceptors. However, we should still have charge neutrality – the electric field is a powerful force! This means that we need to add up our positive charges and set them equal to our negative charges.

$$p_0 + N_D^+ = n_0 + N_A^-$$

where N_D^+ and N_A^- are the number of ionized donors and ionized acceptors respectfully. (Think about this for a minute. A donor that is ionized has given up an electron, which is now moving in the conduction band. An acceptor that is ionized has pulled an electron out of the valance band, leaving a hole to move in the conduction band. Neither ion is able to move and hence neither is a charge carrier.)

For donor and acceptor states, Fermi-Dirac statistics determine whether or not the state is filled. If we look at an energy diagram, we see:



We note

1) that $E_D > E_F \Rightarrow f(E_D) \gg 1 \Leftrightarrow$ almost all donors are ionized positively
2) that $E_A < E_F \Rightarrow 1-f(E_D) \gg 1 \Leftrightarrow$ almost all donors are ionized negatively.

Thus we assert that the number of donor ions and the number of acceptor ions are that same as the number of donors and the number of acceptors. WE ARE IN EFFECT ASSERTING THAT THE SYSTEM IS NON-DEGENERATE!

This gives us two equations,

$$p_0 + N_D = n_0 + N_A$$

$$n_i^2 = n_0 p_0$$

Combining them gives

$$\frac{n_i^2}{n_0} + N_D^+ = n_0 + N_A^-$$

$$\Downarrow$$

$$0 = n_0^2 + n_0\left(N_A^- - N_D^+\right) - n_i^2$$

or

$$\frac{n_i^2}{p_0} + N_A = p_0 + N_D$$

$$\Downarrow$$

$$0 = p_0^2 + p_0\left(N_D - N_A\right) - n_i^2$$

These are both quadratic equations that we can easily solve:

$$p_0 = \frac{(N_A - N_D)}{2} + \frac{1}{2}\sqrt{(N_D - N_A)^2 + 4n_i^4}$$
$$n_0 = \frac{(N_D - N_A)}{2} + \frac{1}{2}\sqrt{(N_A - N_D)^2 + 4n_i^4}$$

The two solutions with the negative signs do not make physical sense as this will give us negative densities for the charge carriers. Thus we find

$$p_0 = \frac{(N_A - N_D)}{2} + \frac{1}{2}\sqrt{(N_D - N_A)^2 + 4n_i^4} = n_i \exp\left[\frac{-(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right]$$
$$n_0 = \frac{(N_D - N_A)}{2} + \frac{1}{2}\sqrt{(N_A - N_D)^2 + 4n_i^4} = n_i \exp\left[\frac{(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right]$$

We can now use these to get the Fermi energy is terms of measurable quantities.

Finally, there are a few simplifications that are often made to the above equations:

1) $N_A, N_D = 0$ => intrinsic $n_0 = p_0 = n_i$.
2) $N_D - N_A >> n_i$ => Doped n-type.
   Then

   $$\sqrt{(N_D - N_A)^2 + 4n_i^4} \approx \sqrt{(N_D - N_A)^2} = |N_D - N_A| = N_D - N_A$$
   $$\Downarrow$$
   $$n_0 = \frac{(N_D - N_A)}{2} + \frac{(N_D - N_A)}{2} = N_D - N_A$$

   and

   $$p_0 = \frac{(N_A - N_D)}{2} + \frac{(N_D - N_A)}{2} \approx 0$$

   Noting the $p_0 n_0$ is still equal to n_i^2 we still get

   $$p_0 = \frac{n_i^2}{n_0}$$

   which is small compared to n_0. This says that the electrons are our majority carriers while the holes are our minority carriers.
3) $N_A - N_D >> n_i$ => Doped p-type.
   Then almost copying the above math we get

$$\sqrt{(N_D - N_A)^2 + 4n_i^4} \approx \sqrt{(N_D - N_A)^2} = |N_D - N_A| = N_A - N_D$$

$$\Downarrow$$

$$n_0 \approx \frac{(N_D - N_A)}{2} + \frac{(N_A - N_D)}{2} \approx 0$$

and

$$p_0 \approx \frac{(N_A - N_D)}{2} + \frac{(N_A - N_D)}{2} = N_A - N_D$$

Noting the p_0n_0 is still equal to n_i^2 we still get

$$n_0 = \frac{n_i^2}{p_0}$$

which is small compared to p_0. This says that the holes are our majority carriers while the electrons are our minority carriers.

4) $N_A - N_D \sim n_i$ => we have to use the full equations.

Example:

Question

We have Si at room temperature (30 °C or 300 K) doped with 10^{16} cm^{-3} B atoms. What is:

1) n_i
2) p_0, n_0
3) $E_i - E_F$
4) $E_F - E_v$
5) Draw the band structure

Answer

1) We can look up the intrinsic density from table 3.17 (page 89) in Streetman. It is $n_i = 1.5 \text{ X } 10^{10}$ cm^{-3}.
2) The doping level is:
   $N_A = 10^{16}$ cm^{-3}.
   $N_D = 0$.
   N_A-N_D >> n_i. Thus we can use the approximation:
   $p_0 = N_A$-$N_D = 10^{16}$ cm^{-3}.
   $n_0 = n_i^2/p_0$ = 2.25 E 20/10 E 16 cm^{-3} = 2.25 E 4 cm^{-3}.
3) We can now get $E_i - E_F$ from

$$p_0 = n_i \exp\left[\frac{-(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right] \text{ or}$$

$$\mathcal{E}_i - \mathcal{E}_F = -kT\ln\left(\frac{n_i}{p_0}\right)$$

$$= -0.026\text{eV } \ln\left(\frac{1.5\text{E}10}{1\text{E}16}\right)$$

$$= 0.35 \text{ eV}$$

4) We can now get $E_F - E_v$ from

$$p_0 = N_v \exp\left[\frac{-(\mathcal{E}_F - \mathcal{E}_V)}{kT}\right]$$ or

$$\mathcal{E}_F - \mathcal{E}_v = kT\ln\left(\frac{N_v}{p_0}\right)$$ Noting that

$$N_{v,c} = 2.510E19 \ \text{cm}^{-3}\left(\frac{m_{p,n}^*}{m_e}\right)^{3/2}$$, and $m_p^* = \left(m_{lh}^{3/2} + 2m_{hh}^{3/2}\right)^{2/3} = 0.81 \ m_e$ (see appendix III)

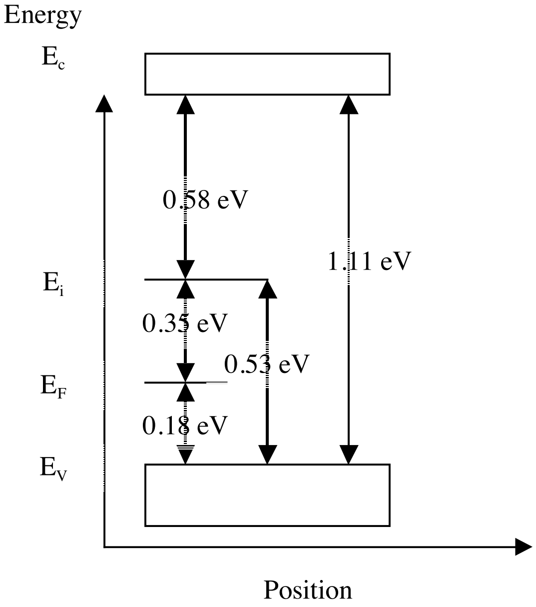
[Note that this mass ratio comes from addition of the three p_0s (one for each direction) that have the same exponential terms. This is equivalent to the 'averaging' we have done above.] we find

$N_v = 1.02E19 \ \text{cm}^{-3}$

We can now plug this into our above equation to get:

$\mathcal{E}_F - \mathcal{E}_v = 0.180 \ \text{eV}$

5)



Carrier action – motion, recombination and generation

Carrier action is driven by a large number of processes. We can look at the in a general sense by looking at the densities of our charge carriers in a small volume. The first thing that we should ask is how does this density change in a given time? Hence, we want to know:

$$\frac{dn}{dt} = ? \quad \text{and} \quad \frac{dp}{dt} = ?$$

We can work with either of these equations to begin to understand the process. The first thing to realize is that the density can change either by production/destruction of the carriers or by carriers moving into or out of the volume. We can take this into account by expanding our derivative.

$$\begin{aligned}\frac{dn}{dt} &= \frac{\partial n}{\partial t} + \frac{\partial n}{\partial x}\frac{dx}{dt} + \frac{\partial n}{\partial y}\frac{dy}{dt} + \frac{\partial n}{\partial z}\frac{dz}{dt} \\ &= \frac{\partial n}{\partial t} + \frac{\partial n}{\partial x}v_x + \frac{\partial n}{\partial y}v_y + \frac{\partial n}{\partial z}v_z \\ &= \frac{\partial n}{\partial t} + \mathbf{v}\,\Sigma\nabla n\end{aligned}$$

This is known as the continuity equation. It describes conservation of particles. If particles are not conserved, and here that is possible – we can have electron-hole recombination etc. – then the right-hand side of the equation is just the production/destruction.

$$\frac{dn}{dt} = \frac{\partial n}{\partial t} + \mathbf{v}\,\Sigma\nabla n = f_{\text{production/destruction}}$$

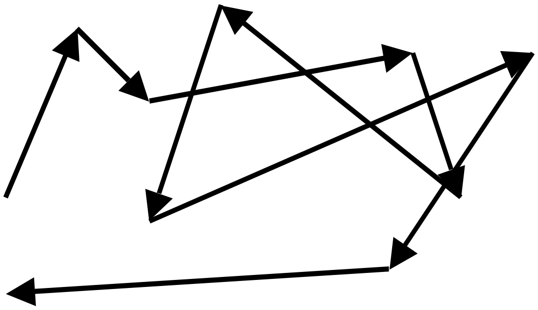
The first term on the right-hand side describes the local gain/loss of the carriers, while the second term describes the flow into or out of a volume. Note that gain/loss is not the same as production/destruction. With this understanding, we can begin to look at how the charges move, as well as how they are produced/destroyed. (There is a similar equation describing conservation of energy.) We will start with current flow.

Current flow

At this point, we can calculate n_0, p_0, E_i, E_F, etc but how does that relate to our real use of semiconducting materials, current flow? It is after all the flow of current, and the ability to turn that flow on and off that makes semiconductors useful. To answer that question, we need to know how electrons and holes move in the lattice.

There are three useful terms which describe how our charge carriers move in the lattice. They are conductivity (σ), mobility (μ) and Diffusion (D). We will now discuss where each of these comes from and their physical meaning.

Without any external force, we expect our charge carriers to have a random spread of energy and to be randomly spread across the device. This randomness is attributed to a Brownian motion type of process. Brownian motion is the observed motion of dust particles in the air. If one where to sit in a room with a bright light, one can observe the motion of dust in the air. This motion is random in nature with dust particles moving at a variety of speeds in a variety of directions. An individual dust particle might move in one direction for a while then turn around and head in the opposite direction latter. (This is also known as a 'random walk' or a 'drunken sailor random walk'.) Pictorially this looks like:



For dust particles in air, the collisions are between air molecules and the dust. For our charge carriers, the collisions are with impurities in the lattice and the lattice atoms. Which of these two scattering centers is more important depends on the temperature.

Now let us apply an electric field to the device. This means that we now have a force on the charge carriers.

$$\mathbf{F} = q\mathbf{E}$$

assuming that the electric field is only in the x-direction, we find:

$$\begin{aligned}\mathbf{F} &= q\mathbf{E} = qE_x\hat{\mathbf{x}} \\ &= m\frac{dv_x}{dt} \\ &= \frac{dp_x}{dt}\end{aligned}$$

The total change in momentum is then simply the sum over all particles, giving

$$\begin{aligned}\frac{dP_{total}}{dt} &= \sum_{i=1}^{n}\frac{dp_i}{dt} \\ &= \sum_{i=1}^{n} qE_{i-(internal)} \\ &= nqE_{average\ or\ external}\end{aligned}$$

Just looking at this equation would imply that the electrons/holes will accelerate ‘forever’. However, we don’t see this sort of phenomena – think of the current in a wire – rather we see a constant velocity. (What we have described should apply to a metal wire as well as our semiconductor.) Additionally, what we really see is an average over all velocities. This means that we have to have a way to slow down the electrons/holes. That way is through scattering with our atoms. In this, we can think about throwing a BB at wall – the wall does not notice but the BB stops or bounces in the opposite direction. We can describe this deceleration using the average time it takes an electron/hole to lose the average particle momentum.

$$\frac{d\langle p_{loss}\rangle}{dt} = -\frac{\langle p\rangle}{\tau} = -\frac{P_{total}}{n\tau}$$

where the average is just the total momentum divided by the number of particles.

At some point we will balance the acceleration due to the electric field with the deceleration due to collisions. Thus,

$$\frac{d\langle p_{\text{total}}\rangle}{dt} = \frac{d\langle p_{\text{gain}}\rangle}{dt} + \frac{d\langle p_{\text{loss}}\rangle}{dt} = 0$$
$$\Downarrow$$
$$qE - \frac{\langle p\rangle}{\tau} = 0$$

rearranging this, and writing the momentum in terms of the average velocity, we get:

$$\langle p\rangle = \tau qE = m^{*}\langle v\rangle$$
$$\Downarrow$$
$$\begin{aligned}\langle v\rangle &= \frac{\tau q}{m^{*}}E \\ &= \mu E\end{aligned}$$

where μ is our mobility of the electron or hole. Note that in the book, q = electron charge, -e. Here q is simply a place holder and requires the insertion of –e for electrons and +e for holes. To get the current flow in the correct direction, see below, the mobility is always given as a positive number. Thus we find

$$\mu_n = -\frac{\langle v\rangle}{E} = \frac{\tau|q|}{m_n^{*}}$$
$$\mu_p = \frac{\langle v\rangle}{E} = \frac{\tau|q|}{m_p^{*}}$$

This is all well and good but we still need to current flowing through our device.
For a single species of charge carriers, this is simply:

$$\begin{aligned}\mathbf{J} &= nq\langle \mathbf{v}\rangle \\ &= nq\mu\mathbf{E} = nq\frac{\tau|q|}{m^{*}}\mathbf{E} \\ &= \sigma\mathbf{E}\end{aligned}$$

where σ is the conductivity. If we have both charge carriers

$$\begin{aligned}\mathbf{J} &= nq\langle \mathbf{v}\rangle \\ &= q\left(n_0\mu_n + p_0\mu_p\right)\mathbf{E} \\ &= \sigma\mathbf{E} \\ &= \mathbf{E}/\rho\end{aligned}$$

where ρ is the resistivity.

At this point, we want to complete this analysis. Most notably, we need to get the total current and the total resistance. (Remember, we measure things in bulk not in differential areas.)

The total current is

$$\mathbf{I} = \int_{\substack{\text{cross} \\ \text{sectional} \\ \text{area}}} \mathbf{J}d\mathbf{a}$$

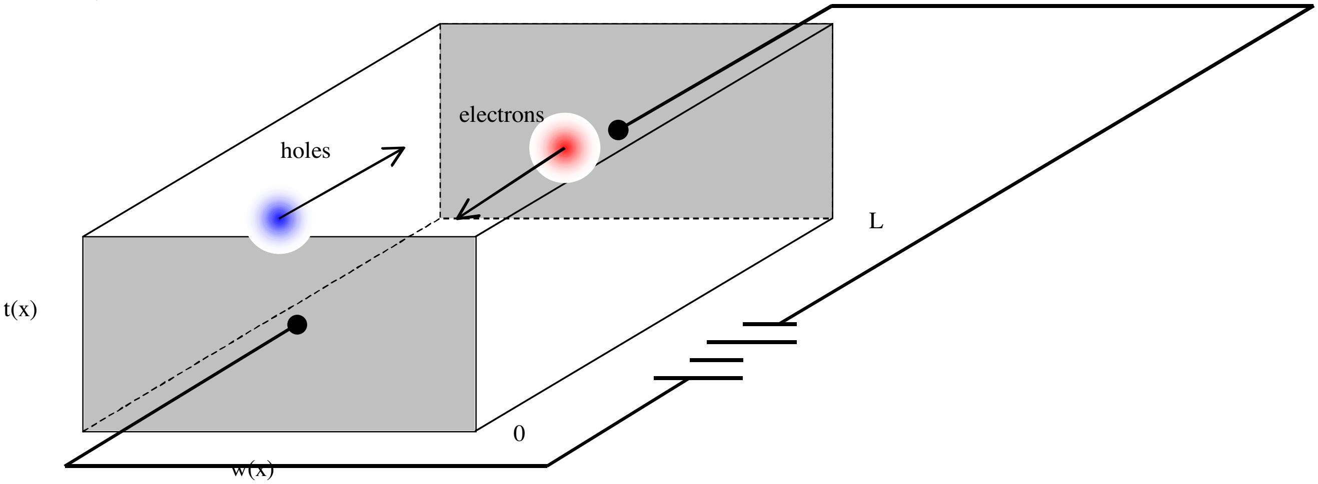
The resistance is

$$\begin{aligned} R &= \int_0^L dR(x) \\ &= \int_0^L \frac{dV(x)}{I(x)} \\ &= \int_0^L \frac{Edx}{J(x)A(x)} \\ &= \int_0^L \frac{\rho(x)}{A(x)}dx \end{aligned}$$

Assuming constant values, which is typical, we get

$$\begin{aligned} R &= \frac{\rho L}{A} \\ &= \frac{L}{\sigma A} \end{aligned}$$

Pictorially, this looks like:



Electrons are inserted at x = L and removed at x = 0.
Holes are inserted at x = 0 and removed at x = L.

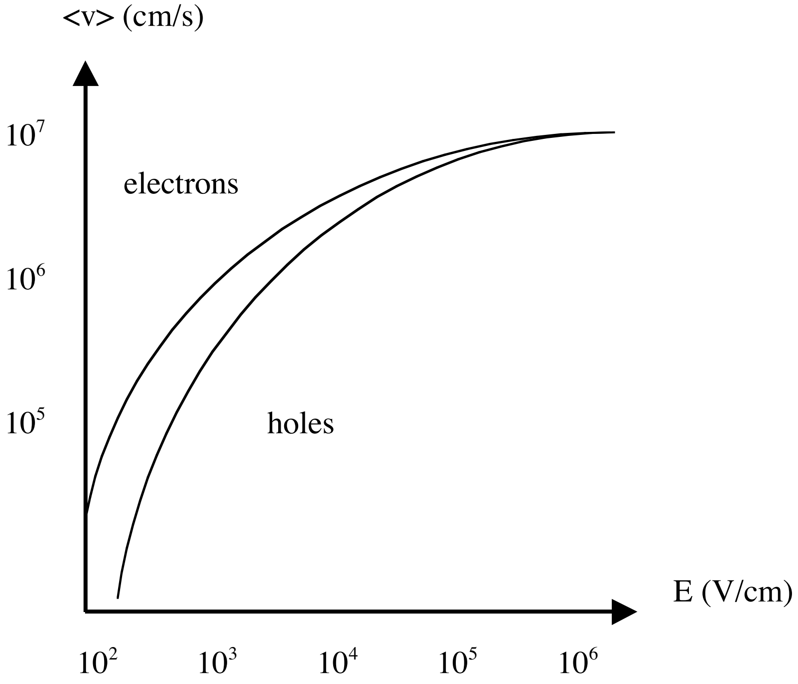
So now we need to ask just what does the mobility mean?

$$\mu = \frac{\tau|q|}{m^*}$$

We know what the charge and the mass are, but what is tau?
Tau depends on the number of scattering centers and the temperature of the lattice.

1) The more scattering centers (dopant or defects) the more often collisions will occur. Hence, the t is small and so is μ.
2) The higher the temperature => the more lattice movement => the more scattering => the smaller t => smaller μ.

At low electric fields, i.e. low velocities, the collision rate is constant. Thus μ ~ constant. However, at higher velocities, the collision rate increases, until increasing the electric field does not significantly increase the velocity. A graph of V vs E looks like:



At higher electric fields, the collision start to heat the lattice, causing more collisions. At some point most of the energy starts to get dissipated as heating of the lattice and not as kinetic energy of the electrons.

At this point, we have found that an electric field can cause current to flow. But what does this mean for the shape of our energy bands?
The answer is relatively straight forward, it bends them. We can show this very simply. Our bands are based on the energy of the electrons. Namely:

$$\text{Energy} = \text{k.E.} + \text{P.E.}$$

How do you show how much is k.E. and how much is P.E.?

Adding E_g energy to an electron at E_v – the edge of the valance band – will put the electron in the conduction band but with no additional energy, i.e. k.E. Any additional energy goes to k.E. It also turns out that the P.E. must have a reference energy, which will call E_{ref}. Thus, the potential energy for the electrons must be

$$\text{P.E.} = E_c - E_{ref}$$

and the k.E. must be

$k.E. = E - E_c$

Likewise holes have a

$P.E. = E_{ref} - E_v$ and $k.E. = E_v - E$ Note: increasing energy is in the opposite direction for electrons (up) and holes (down) in our energy diagrams.

Now

$$\mathbf{E} = -\nabla V = -\partial_x V \hat{\mathbf{x}} \text{ (in 1-D)}$$

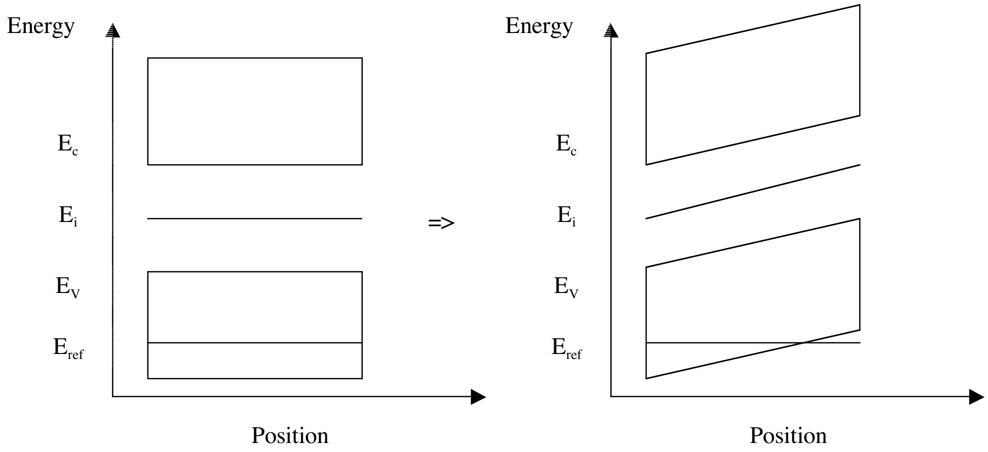
but

$$E_c - E_{ref} = -qV$$
$$\Downarrow$$
$$V = -\frac{1}{q}\left(E_c - E_{ref}\right)$$

so

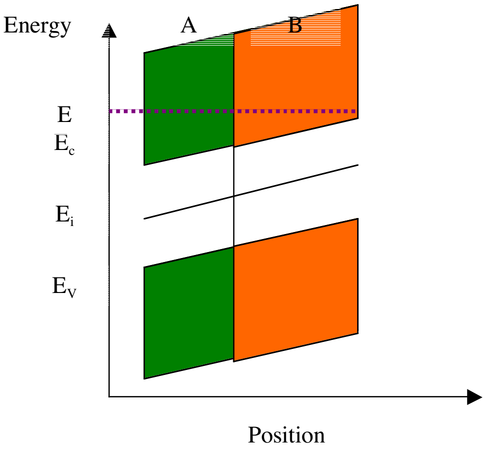
$$\mathbf{E} = -\nabla V$$
$$\Downarrow$$
$$\begin{aligned} E_x &= -\partial_x V \text{ (in 1-D)} \\ &= q\partial_x\left(E_c - E_{ref}\right) \\ &= q\partial_x\left(E_c\right) \\ &\quad \text{or} \\ &= q\partial_x\left(E_v\right) \\ &\quad \text{or} \\ &= q\partial_x\left(E_i\right) \end{aligned}$$

noting that E_v, E_c and E_i are uniquely related. What does this do to our bands?



At this point let us assume that our system is set up such that the electric field exists across the device but we have not connected the wires at the ends of the semiconductor. In essence, we have our

semiconductor sitting inside a capacitor. At this point, we can show that under these conditions, the Fermi energy does not change. (This assumes that there is no current flow. If there is a current flow then E_F does change!)
Assuming that there is no current flow, there is still transfer of charge across the 'device' – it is just that the net or total charge transfer has to be zero. Let us divide our sample into two parts, part A and part B.



For an electron (or hole) to move from A to B, an open spot needs to exist in B at the same energy.. Thus the number of electrons jumping from A to B is going to be proportional to the number of electrons (holes) in A times the number open stops in B.

$$\text{rate from A to B } \propto N_A f_A(E)\,\Sigma N_B\left(1-f_B(E)\right)$$

The rate from B to A has to be similar.

$$\text{rate from B to A } \propto N_A\left(1-f_A(E)\right)\Sigma N_B f_B(E)$$

Because we do not have a net current flow, these must be equal and thus

$$\text{rate from B to A } = \text{ rate from A to B}$$

$$\Downarrow$$

$$N_A f_A(E)\,\Sigma N_B\left(1-f_B(E)\right) = N_A\left(1-f_A(E)\right)\Sigma N_B f_B(E)$$

$$\Downarrow$$

$$f_A(E)\,\Sigma\left(1-f_B(E)\right) = \left(1-f_A(E)\right)\Sigma f_B(E)$$

$$\Downarrow$$

$$\left(f_A(E)-f_A(E)f_B(E)\right) = \left(f_B(E)-f_A(E)f_B(E)\right)$$

$$\Downarrow$$

$$f_A(E) = f_B(E)$$

$$\Downarrow$$

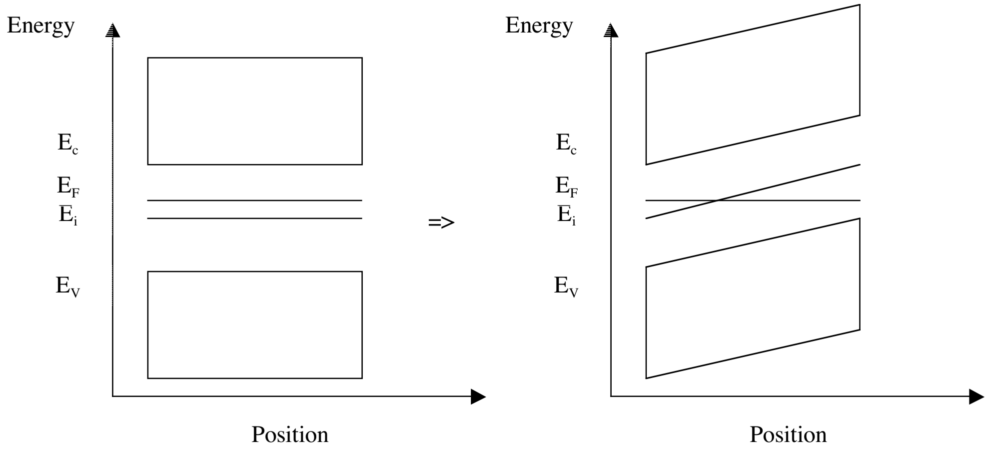
$$\left(1+e^{\left(E-E_{FA}\right)/kT}\right)^{-1} = \left(1+e^{\left(E-E_{FB}\right)/kT}\right)^{-1}$$

$$\Downarrow$$

$$E_{FA} = E_{FB}$$

Showing that at equilibrium the Fermi energy is a constant.

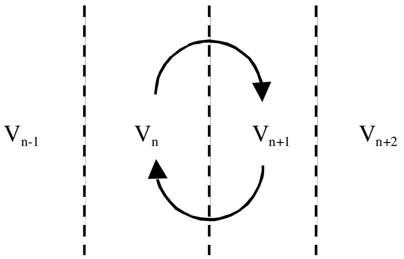
Because of this, we can use E_F as the reference energy. Thus by applying a bias, we can make our material go from a p-type to an n-type material as we move across the 'device' (or vise-a-versa).



We now have bent bands and thus move electrons/holes but we have no flow! Does this make any sense?

To understand why this might be true, we need to look at the world again. Let us consider a bowl of water that is standing on a table. Now let us put a drop of food coloring into the center. (This is a fun game for small kids and those of us whom would like to be young still!) What happens to the color? It spreads out slowly and somewhat randomly. In effect we have all of the dye molecules undergoing Brownian motion – each following a random path away from the initial position. The average over all particles is given its own name – 'Diffusion'. Like many random processes the 'average' is reasonably well behaved and fairly easy to model mathematically. [There is a whole science that studies random processes, known as Chaos Theory.]

In general, we can divide any volume into very thin slices, each of width <L>, where <L> is the average collision length of our particles. Let us look at the transfer of particles between two volumes, V_n and V_{n+1}. (This model applies not only to our electrons and holes but also to just about any particle transfer)



When a particle undergoes a collision, it will likely be scattered into the next volume. (This is because the width of each volume is so small, the particle will not collide again until it has moved <L>, or the width of the cell – putting it into the next volume.) A particle in V_n has an equal chance of moving into V_{n-1} as V_{n+1}. We now need to determine the number of particles per unit area that move from V_n to V_{n+1} in a given time. This is known as the 'Flux,' Γ. Let us fix the position of all of the particles except those in volume V_n. On average, all of the particles in V_n will undergo a scattering collision in time τ, where τ is the average scattering collision time. The number of particles in V_n is simply the density of particles times the volume. From this we get:

$$\text{Total flow}_{V_n \to V_{n+1}} = \frac{\frac{1}{2}\left(n_n A_n \langle L \rangle\right)}{\tau}$$

where A_n is the cross sectional area of volume V_n. To get the flux, we just need to divide by the area.

$$\Gamma_{V_n \to V_{n+1}} = \frac{\frac{1}{2}\left(n_n \langle L \rangle\right)}{\tau}$$

Now let us release the particles in V_{n+1}.

$$\begin{aligned}\Gamma_{V_n \leftrightarrow V_{n+1}} &= \Gamma_{V_n \to V_{n+1}} - \Gamma_{V_{n+1} \to V_n} \\ &= \frac{\frac{1}{2}\left(n_n \langle L \rangle\right)}{\tau} - \frac{\frac{1}{2}\left(n_{n+1} \langle L \rangle\right)}{\tau} \\ &= \frac{1}{2}\frac{\left(n_n - n_{n+1}\right)\langle L \rangle}{\tau}\end{aligned}$$

Now, we note that <L> is very small. Thus, in 3-D

$$\frac{\partial n}{\partial x} \approx -\frac{\left(n_n - n_{n+1}\right)}{\langle L \rangle}$$

and thus!

$$\begin{aligned}\Gamma &= -\frac{1}{2}\frac{\langle L \rangle^2}{\tau}\frac{\partial n}{\partial x} \quad \text{in 1-D or} \\ &= -\frac{1}{2}\frac{\langle L \rangle^2}{\tau}\nabla n \quad \text{in 3-D} \\ &= \text{-}D\nabla n\end{aligned}$$

where D is the diffusion coefficient. Note that D is a positive number and thus expect the particles to – on average – flow from the higher density to the lower density volumes. (This is true for any diffusion process.)

We have these net fluxes for both electrons and holes.

$$\Gamma_n = \text{-}D_n \nabla n$$

$$\Gamma_p = \text{-}D_p \nabla p$$

Because of this, we have net currents due to diffusion.

$$\mathbf{J}_n = -q_n \Gamma_n = q_n D_n \nabla n$$

$$\mathbf{J}_p = q_p \Gamma_p = \Gamma_p = \text{-}q_p D_p \nabla p$$

$$\Downarrow$$

$$\mathbf{J}_{\text{diff}} = \mathbf{J}_n + \mathbf{J}_p = q_n D_n \nabla n - q_p D_p \nabla p$$

Now let us go back to our situation in which we have an electric field dragging the electrons and holes across our ‘device’. For that we calculated a current of:

$$
\begin{aligned}
\mathbf{J}_E &= nq\langle \mathbf{v} \rangle \\
&= q\left(n\mu_n + p\mu_p\right)\mathbf{E} \\
&= q\left(n\frac{\tau|q_n|}{m_n^*} + p\frac{\tau|q_p|}{m_p^*}\right)\mathbf{E}
\end{aligned}
$$

(We got this in our last class period!)
Looking at each species independently we get

$$\mathbf{J}_p = q_p p\mu_p\mathbf{E} - q_p D_p \nabla p$$

$$\mathbf{J}_n = q_n n\mu_n\mathbf{E} + q_n D_n \nabla n$$

$$\mathbf{J}_{total} = \mathbf{J}_p + \mathbf{J}_n$$

Now in our system that we started out with the total current has to be zero. (We don't have connections to the wires!) Thus,

$$\mathbf{J}_{total} = \mathbf{J}_p + \mathbf{J}_n = 0$$

In fact, because of particle conservation, this equation must hold for each species independently.

$$\mathbf{J}_n = q_n n\mu_n\mathbf{E} + q_n D_n \nabla n = 0$$

$$\Downarrow$$

$$n\mu_n\mathbf{E} = -D_n \nabla n$$

$$\Downarrow$$

$$\mathbf{E} = \frac{-D_n \nabla n}{n\mu_n}$$

and

$$\mathbf{J}_p = q_p p\mu_p\mathbf{E} - q_p D_p \nabla p = 0$$

$$\Downarrow$$

$$p\mu_p\mathbf{E} = D_p \nabla p$$

$$\Downarrow$$

$$\mathbf{E} = \frac{D_p \nabla p}{p\mu_p}$$

Now what are ∇n and ∇p?

$$n = n_i e^{\left[-\left(E_i - E_F\right)/kT\right]}$$

$$\Downarrow$$

$$
\begin{aligned}
\nabla n &= -\frac{n_i}{kT} e^{\left[\left(E_i - E_F\right)/kT\right]} \nabla E_i \\
&= \frac{n}{kT} \nabla E_i
\end{aligned}
$$

$$p = n_i e^{[(E_i - E_F)/kT]}$$

$$\Downarrow$$

$$\nabla p = \frac{n_i}{kT} e^{[(E_i - E_F)/kT]} \nabla E_i$$

$$= \frac{p}{kT} \nabla E_i$$

So, what is ∇E_i….???? (We have gone a long way to get here!) Well we already came up with this during our last class period.

$$\mathbf{E} = -\nabla V$$

$$\Downarrow$$

$$E_x = -\partial_x V \text{ (in 1-D)}$$

$$= q\partial_x (E_c - E_{ref})$$

$$= q\partial_x (E_c)$$

or

$$= q\partial_x (E_v)$$

or

$$= q\partial_x (E_i)$$

so

$$\mathbf{E} = q\nabla(E_i)$$

SO!

$$\mathbf{E} = \frac{-D_n}{n\mu_n} \nabla n$$

$$= \frac{-D_n}{n\mu_n} \left(-\frac{n}{kT} q_n \mathbf{E} \right)$$

$$\Downarrow$$

$$D_n = \mu_n \frac{kT}{q_n}$$

and

$$\mathbf{E} = \frac{D_p}{p\mu_p}\nabla p$$

$$= \frac{D_p}{p\mu_p}\left(\frac{p}{kT}q_p\mathbf{E}\right)$$

$$\Downarrow$$

$$D_p = \mu_p \frac{kT}{q_p}$$

These are known as the Einstein relations

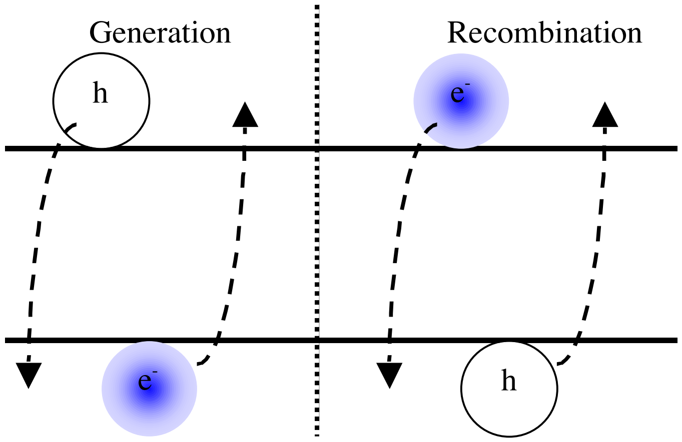
At this point we are able to move our charge carriers around by electric fields, giving rise to average velocities and hence mobilities. The charge carriers also move around on their own, through the random scattering based process known as diffusion. Further, we know that the holes and electrons are the result of another random process, energy sharing through collisional processes. This leads us to an interesting question, can we understand the generation and loss of individual electrons/holes and if so, will that help our understanding of how electronic devices work?

Homework Set 4
Streetman: Chap 4 # 1,2,3,4,5.6,7. Assigned 9/26/02 Due 10/3/02

Electron-hole loss and gain – Recombination and Generation
Loss of electron and holes is in some ways the simplest of the two processes that we want to look at. We know that an electron cannot 'just evaporate'. Charge and internal spin are conserved quantities. Electrons have both (q = –1.6E–19 C, s = ±1/2). Likewise, except under very special circumstances, mass is also a conserved quantity. Electrons clearly have that as well (even if it is a very small 9.11E–31 kg). This means that to get rid of an electron, we must put it someplace else. Well let us think for a minute. Our 'electrons' are just those electrons that have made it up into the conduction band and our 'holes' are just places in the valance band that do not have electrons – remember 'holes' are an artificial construct that makes the math a lot easier. Thus, the only obvious someplace else is for a electron to 'recombine' with a hole, i.e. an electron in the conduction band loses energy and drops into the valance band, filling a hole. Hence when we lose an electron, we also lose a hole.

Gain of electrons and holes must also follow the same pattern as lose – if we gain and electron, we have to gain a hole. Again, this is due to the fact that electrons have charge, spin and mass.

We find that the creation/destruction of electrons and holes happens at the same time. Because of this, we refer to them as an **Electron-hole pairs, or EHP**. We can look at this process graphically.



Side note: This creation/destruction that we are considering is only for a 'device' that has already been built. Doping of semiconductor material can and does independently change the relative concentrations of electrons and holes. Thus by adding an n-type dopant, we can add an electron without adding a hole. Once the 'device' is built, however, gain or loss of holes and electrons is only through the process we are describing here.

We can now look at how fast we might lose EHPs. We know that for an electron to move to the valance band, a hole must exist there. Likewise, for a hole to move into the conduction band (same process as above, just looking at it the other way around) an electron must be in the conduction band. Thus the loss rate must be proportional to both the number of electrons and the number of holes. Hence

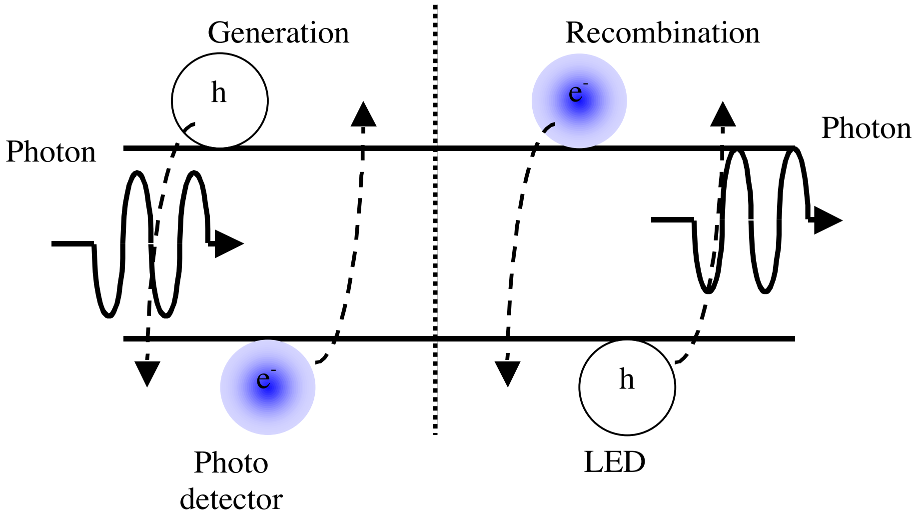
$$\begin{aligned} r &= \alpha_r n_0 p_0 \\ &= \alpha_r n_i^2 \end{aligned}$$

where r is the recombination rate and α_r is the proportionality constant. Under thermal equilibrium conditions, the lose rate must be equal to the thermal generation rate, $g_{thermal}$. Hence

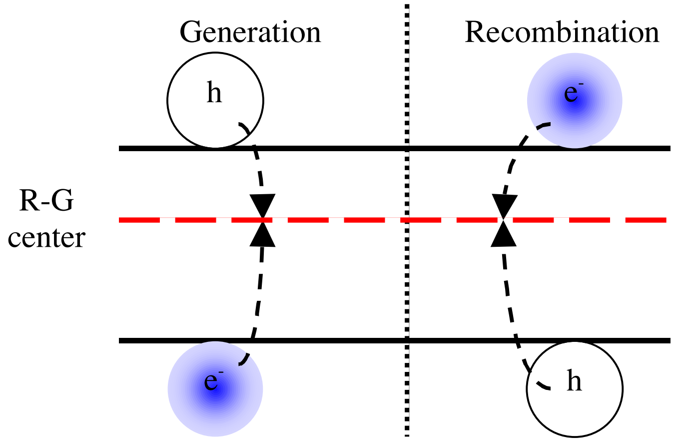
$$\begin{aligned} g_{thermal} = r &= \alpha_r n_0 p_0 \\ &= \alpha_r n_i^2 \end{aligned}$$

This means that if we raise the temperature and hence raise $g_{thermal}$, we find that the intrinsic density also increases. This is simply what we have found with the Fermi function, so it is not a big surprise.

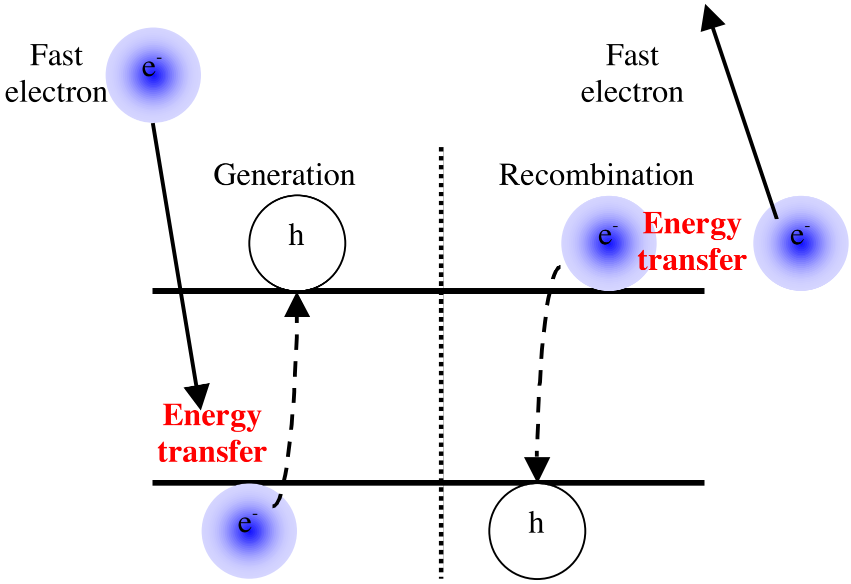
At this point we need to look for those processes that might give rise to recombination or generation of EHPs. Some example include:



Direct – Band to Band



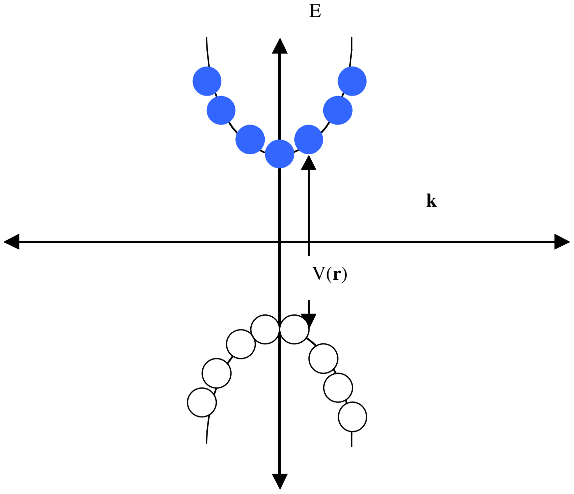
Indirect – Requires a R-G Center to change momentum. (Si and Ge are like this.)



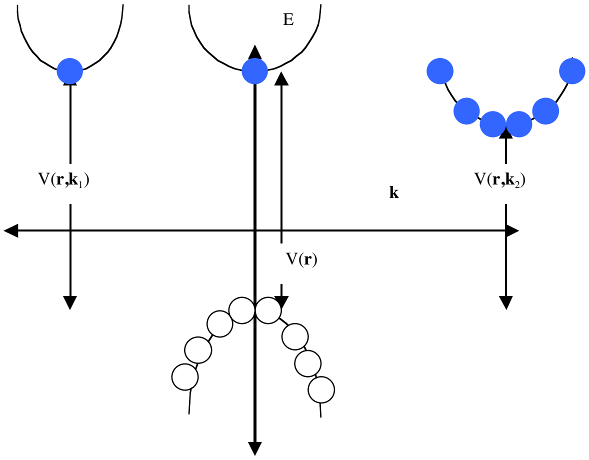
Auger (This is pronounced O'jhay – a French name – not auger as in corn auger) – or impact ionization

The process that dominates depends on the conditions and the material type. To determine which is most important, we need to better understand the second and third of these three processes because they have the additional requirement that momentum must be transferred for the process to occur.

In chapter 2, Streetman introduced the concept of direct and indirect band structures.



Direct Band Structure



Indirect Band Structure

Here the electrons and hole lie right on top of the lines. (There are no other acceptable (k,E) states. Remember that the horizontal axis is momentum and not position.)

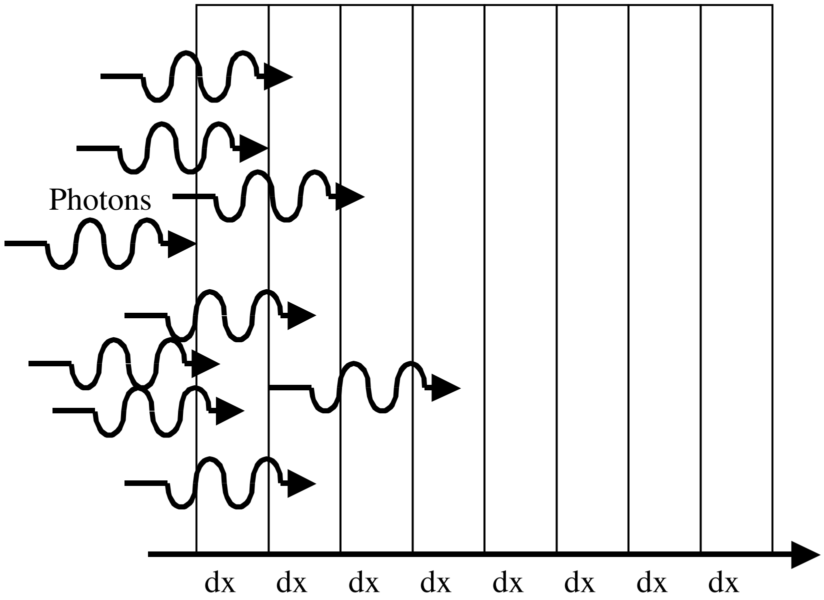
Why are these important? Well it turns out that while photons can carry away energy, hν, they cannot carry momentum, $h\mathbf{k}/2\pi$. Thus there are typically very few electrons that can recombine directly with holes by the simple emission of photons. (There still might be a few but they are of very limited number.)

On the other hand, photon absorption can and does continue to occur. While the photon does not bring in momentum, it can cause an electron to jump from the valance band to the conduction band provided it has more energy then is necessary to overcome the direct energy band gap, E_a. This new electron in the conduction band then will undergo collisions with the lattice, transferring momentum, to reach the minimum energy in the conduction band. (These collisions serve to heat up the lattice.) Under certain conditions, the electron can still make the jump, provided the photon energy is greater than E_g. This occurs because the photon excited the electron into a very brief temporary virtual state. (Remember Heizenberg!) Provided the electron then rapidly (instantaneously!) undergoes a momentum transfer collision, it can be indirectly transferred to the conduction band. If there is no collision in the length of time of the virtual state then the electron decays back to the valance band giving up the adsorbed photon. Because this is a three-body problem, i.e. the photon, electron and scattering center must all be there at the same time, it occurs less frequently – but it still can and does occur. Thus we find:

$$h\nu \begin{cases} < E_g & \text{no transfer} \\ \geq E_g, < E_a & \text{limited transfer} \\ \geq E_a & \text{significant transfer} \end{cases}$$

So, how much light is absorbed?

Let us set up an experiment that looks like:



A certain fraction of the photons will be absorbed in any given length dx. We can measure how many by 'measuring' the intensity of the light. Thus

$$\begin{aligned} &I \propto \#\ \text{of photons} \\ &\Downarrow \\ dI &= I(x+dx) - I(x) \\ &\propto \Delta\#\ \text{of photons} = -\#\ \text{of photons absorbed} \\ &= -(\text{fraction of photons absorbed})\ \text{X}\ (\text{I per photon})\ \text{X}\ (\text{initial \# of photons}) \\ &\Downarrow \\ dI &= -\gamma I dx \end{aligned}$$

we of course can integrate this equation to arrive at

$$I = I_0 e^{-\gamma x}$$

where I_0 is the initial intensity of the light.

This light produces EHPs as it is absorbed, which means that we have now increased the EHP generation rate. Thus we should expect n and p to increase.

We can find out how much using a simple 'conservation law'.

$$\frac{dn}{dt} = \frac{dp}{dt}$$
$$= \text{Gen rate - loss rate}.$$

We know from above that the thermal generation rate is equal to the recombination rate.

$g_{thermal} = r = \alpha_r np$

although now we are not in thermal equilibrium and thus

$n \neq n_0$

$p \neq p_0$

$pn \neq n_i^2$

plugging this into the above equation, we find that

$$\frac{dn}{dt} = \frac{dp}{dt}$$
$$= \underbrace{\alpha_r n_i^2}_{\substack{\text{thermal} \\ \text{gen rate}}} + \text{light gen rate} - \underbrace{\alpha_r np}_{\substack{\text{thermal} \\ \text{loss rate}}}$$

At this point, let us turn off the light. The light generation rate goes immediately to zero but the density of holes and electrons do not immediately go to zero. Thus we find

$$\frac{dn}{dt} = \frac{dp}{dt}$$
$$= \alpha_r n_i^2 - \alpha_r n(t)p(t)$$

So what are $n(t)$ and $p(t)$? Obviously we can divide it into a constant part that is due to equilibrium densities and a part that is due to our extraneous sources – such as the light shining on the 'device'. Taking those assumptions, we find

$n(t) = n_0 + \Delta n(t)$

$p(t) = p_0 + \Delta p(t)$

Further, as we are not adding dopant, so

$\Delta n(t) = \Delta p(t)$.

These are known as the excess carrier concentrations (caused by our light). We can now plug these into our time-dependent equation.

$$\frac{dn}{dt} = \frac{d(n_0 + \Delta n(t))}{dt} = \frac{d\Delta n(t)}{dt}$$

$$= \frac{dp}{dt} = \frac{d(p_0 + \Delta p(t))}{dt} = \frac{d\Delta p(t)}{dt}$$

$$= \alpha_r n_i^2 \ - \ \alpha_r (n_0 + \Delta n(t))(p_0 + \Delta p(t))$$

$$= \alpha_r n_i^2 \ - \ \alpha_r (n_0 p_0 + p_0 \Delta n(t) + n_0 \Delta p(t) + \Delta p(t) \Delta n(t))$$

$$= \alpha_r n_i^2 \ - \ \alpha_r \left(n_i^2 + p_0 \Delta n(t) + n_0 \Delta p(t) + \Delta p(t) \Delta n(t)\right)$$

$$= - \ \alpha_r (p_0 \Delta n(t) + n_0 \Delta p(t) + \Delta p(t) \Delta n(t))$$

$$= - \ \alpha_r (p_0 \Delta n(t) + n_0 \Delta n(t) + \Delta n(t) \Delta n(t))$$

$$= - \ \alpha_r \left((p_0 + n_0) \Delta n + \Delta n^2\right)$$

$$\Downarrow$$

$$\frac{d\Delta n}{dt} = - \ \alpha_r \left((p_0 + n_0) \Delta n + \Delta n^2\right)$$

or

$$\frac{d\Delta p}{dt} = - \ \alpha_r \left((p_0 + n_0) \Delta p + \Delta p^2\right)$$

These equations are tough to solve analytically. However, it is often the case that we have a **low level of injection** (creation) of charges. By this we mean that the increase in charge carriers is small compared to the total number of charge carriers.

$$\Delta n, \Delta p << n_0 + p_0$$

or

$$\Delta n^2, \Delta p^2 << \Delta n(n_0 + p_0), \Delta p(n_0 + p_0)$$

Using these, our equations collapse to

$$\frac{d\Delta n}{dt} = - \ \alpha_r \left((p_0 + n_0) \Delta n\right)$$

and

$$\frac{d\Delta p}{dt} = - \ \alpha_r \left((p_0 + n_0) \Delta p\right)$$

These are clearly solvable.

$$\Delta n = \Delta n(t = 0) e^{-t/\tau}$$

$$\Delta p = \Delta n(t = 0) e^{-t/\tau}$$

where

$$\tau = \frac{1}{\alpha_r (p_0 + n_0)}$$

Now like everything else, we break this down into a result for intrinsic, n-type and p-type materials. For intrinsic materials we get the above equations. For n-type and p-type, we get some simplifications and hopefully a better understanding of what is doing on.

n-type => $n_0 >> p_0$.

$$\tau = \frac{1}{\alpha_r(p_0 + n_0)} = \frac{1}{\alpha_r(n_0)}$$

Now

$$\tau \Rightarrow \tau_p$$

To understand this label, we need to look at the total density of each type of charge carriers

$$n(t) = n_0 + \Delta n(t) \approx n_0$$

$$p(t) = p_0 + \Delta p(t) \approx \Delta p(t) \quad \text{is possible}$$

therefore tau has to apply to the p density and not really to the n density.

We can do the same thing **for p-type materials** to arrive at

$$\tau_n = \frac{1}{\alpha_r(p_0 + n_0)} = \frac{1}{\alpha_r(p_0)}$$

where

$$n(t) = n_0 + \Delta n(t) \approx \Delta n(t) \quad \text{is possible}$$

$$p(t) = p_0 + \Delta p(t) \approx p_0$$

> We can do the same thing with indirect recombination (or trapping) or other similar processes. (We may or may not get to it in this class but you should understand that other processes are possible.) For trapping we get
>
> $$\tau_n = \frac{1}{c_n N_t}$$
>
> $$\tau_p = \frac{1}{c_p N_t}$$
>
> where c_n and c_p are the 'capture coefficients' and N_t is the trap density. Multiple such processes occurring at the same time will result in a tau that is the sum of all of the taus.

At this point, we want to go back to our system that we had been shining light on. We will turn on the light and leave it on. After a little time, the system will come in to a new steady-state density condition. (I do not want to call it 'equilibrium' as that is used for n_0 and p_0. We clearly will not be at those densities.) Let us go back to our basic equations.

$$\frac{dn}{dt} = \frac{dp}{dt}$$

$$= \underbrace{\alpha_r n_i^2}_{\substack{\text{thermal} \\ \text{gen rate}}} + \text{light gen rate} - \underbrace{\alpha_r np}_{\substack{\text{thermal} \\ \text{loss rate}}}$$

When we are in steady state, the time derivatives become zero. Thus,

$$\frac{dn}{dt} = \frac{dp}{dt} = 0$$

$$= \alpha_r n_i^2 + g_{opt} - \alpha_r np$$

$$= \alpha_r n_i^2 + g_{opt} - \alpha_r (n_0 + \Delta n)(p_0 + \Delta p)$$

As before, we will assume that

$$\Delta n = \Delta p$$

so that

$$0 = \alpha_r n_i^2 + g_{opt} - \alpha_r (n_0 + \Delta n)(p_0 + \Delta n)$$

$$= \alpha_r n_i^2 + g_{opt} - \alpha_r (n_0 p_0 + \Delta n p_0 + n_0 \Delta n + \Delta n \Delta n)$$

$$= \alpha_r n_i^2 + g_{opt} - \alpha_r \left(n_i^2 + \Delta n p_0 + n_0 \Delta n + \Delta n \Delta n\right)$$

$$= g_{opt} - \alpha_r \left(\Delta n (p_0 + n_0) + \Delta n^2\right)$$

$$\Downarrow$$

$$g_{opt} = \alpha_r \left(\Delta n (p_0 + n_0) + \Delta n^2\right)$$

Again, we assume low-level injection. (This assumption will hold for all but a very few devices.) That is

$$\Delta n, \Delta p << n_0 + p_0$$

or

$$\Delta n^2, \Delta p^2 << \Delta n (n_0 + p_0), \Delta p (n_0 + p_0)$$

hence

$$\Delta n = \Delta p = g_{opt} \frac{1}{\alpha_r (p_0 + n_0)} = g_{opt} \tau$$

So what is g_{opt}? It is simply the number of photons that are absorbed per volume per time; hence the number of EHP created per volume per time. We calculated this earlier.

Well we have now set up a situation in which one of our fundamental equations no longer 'works' for the conditions of our 'device'. We had derived:

$$p_0 = n_i \exp\left[\frac{-(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right]$$

$$n_0 = n_i \exp\left[\frac{(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right]$$

but we no longer have densities that are at the equilibrium densities! Now, we have

$$p = p_0 + \Delta p = n_i \exp\left[\frac{-(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right] + \Delta p$$

$$= ?$$

$$n = n_0 + \Delta n = n_i \exp\left[\frac{(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right] + \Delta n$$

$$= ?$$

At this point, we can define new two 'Fermi' energies, F_p and F_n, to get the same form of equation as we had before. These are known as **Quasi-Fermi energies**.

$$p \equiv n_i \exp\left[\frac{-(F_p - \mathcal{E}_i)}{kT}\right]$$

$$n \equiv n_i \exp\left[\frac{(F_n - \mathcal{E}_i)}{kT}\right]$$

Example:

We have a sample of GaAs (which is a direct semiconductor and hence we do need to worry about traps). Let us dope the 'device' such that:
$N_a=10^{14}$ cm^{-3}, $n_i \sim 10^{6}$ cm^{-3}, $\tau_n \sim 10^{-8}$ sec, $g_{opt} = 10^{20}$ EHP cm^{-3} s^{-1}, $E_{gap} = 1.43$ eV.

1) Find n_0, p_0
2) Find Δn, Δp
3) Find F_p and F_n
4) Minimum light energy (in W/cm^3) to do this.

Answer:

1) $N_a >> n_i \Rightarrow p_0 = N_a = 10^{14}$ cm^{-3}.
   $n_0 = n_i^2/p_0 = 10^{-2}$ cm^{-3}.
2) WE know

$$\Delta n = \Delta p$$

$$g_{opt} = \alpha_r\left(\Delta n(p_0 + n_0) + \Delta n^2\right)$$

$$\tau = \frac{1}{\alpha_r(p_0 + n_0)}$$

$$\alpha_r = \frac{1}{\tau(p_0 + n_0)}$$

$$= \frac{1}{10^{-8}\left(10^{14} + 10^{-2}\right)} \approx 10^{-6} \ \text{cm}^3/\text{s}$$

=>

We can now plug this into the second equation to get

$$\begin{aligned} g_{\text{opt}} &= \alpha_r\left(\Delta n(p_0+n_0)+\Delta n^2\right) \\ &= 10^{-6}\left(\Delta n\left(10^{14}+10^{-2}\right)+\Delta n^2\right) \\ &= 10^{20} \end{aligned}$$

plugging this into the solution to the quadratic equation we get
$\Delta n = \Delta p = 10^{12}\text{cm}^{-3}$ $\left(\approx g_{\text{opt}}\tau \text{ because } \Delta n \ll p_0+n_0\right)$

3) Now to get the quasi-Fermi Energies

$$n \equiv n_i \exp\left[\frac{\left(F_n-\mathcal{E}_i\right)}{kT}\right] \qquad\qquad p \equiv n_i \exp\left[\frac{-\left(F_p-\mathcal{E}_i\right)}{kT}\right]$$

$$\Downarrow \qquad\qquad\qquad\qquad \Downarrow$$

$$\begin{aligned} \left(F_n-\mathcal{E}_i\right) &= -kT\ln\left(n_i/n\right) \\ &= -kT\ln\left(n_i/\left(n_0+\Delta n\right)\right) \\ &\approx 0.357 \text{ eV} \end{aligned} \qquad \begin{aligned} \left(F_p-\mathcal{E}_i\right) &= kT\ln\left(n_i/p\right) \\ &= kT\ln\left(n_i/\left(p_0+\Delta p\right)\right) \\ &\approx -0.477 \text{ eV} \end{aligned}$$

4) We can get the minimum energy by assuming that we are only providing sufficient energy for the electrons to jump across the band gap. Hence, we need all of the photons to have the band energy. This implies

$$\begin{aligned} \text{Energy} &= 10^{20}\frac{\text{EHP}}{\text{cm}^3\text{s}} * 1.43\frac{\text{eV}}{\text{EHP}} * 1.6\text{X}10^{-19}\frac{\text{J}}{\text{eV}} \\ &= 23\frac{\text{Watts}}{\text{cm}^3} \end{aligned}$$

To gauge how much light this is, an incandescent light is about 5% efficient. Thus a 100 W light bulb puts out about 5 W of light energy. This would mean that you would have to have approximately 5 100 W light bulbs of light deposited in each cm^3 – a lot of light!

Continuity Equation

Now we need to back to the continuity equation. Remember that it describes conservation of particles. If particles are not conserved, and here that is possible – we can have electron-hole recombination etc. – then the right-hand side of the equation is just the production/destruction.

$$\frac{dn}{dt} = \frac{\partial n}{\partial t} + \mathbf{v}\,\Sigma\nabla n = f_{\text{production/destruction}}$$

$$\frac{dp}{dt} = \frac{\partial p}{\partial t} + \mathbf{v}\,\Sigma\nabla p = f_{\text{production/destruction}}$$

Let us rearrange the equation

$$\Downarrow$$

$$\frac{\partial n}{\partial t} = -\mathbf{v}\,\Sigma\nabla n + f_{\text{production/destruction}}$$

$$\frac{\partial p}{\partial t} = -\mathbf{v}\,\Sigma\nabla p + f_{\text{production/destruction}}$$

Noting that the first term on the right-hand side is simply the spatial derivative of the current density, we get

⇓

$$\frac{\partial n}{\partial t} = -\frac{1}{q}\nabla \Sigma q\mathbf{v}n + f_{\text{production/destruction}}$$

$$= \frac{1}{q_n}\nabla \Sigma \mathbf{J}_n + f_{\text{production/destruction}}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \Sigma q\mathbf{v}p + f_{\text{production/destruction}}$$

$$= -\frac{1}{q_p}\nabla \Sigma \mathbf{J}_p + f_{\text{production/destruction}}$$

We can also begin to fill in terms for creation/destruction of our charge carriers. First, we have optical excitation – hence creation. Then we have thermal excitation and recombination. Thus, we get

⇓

$$\frac{\partial n}{\partial t} = \frac{1}{q_n}\nabla \Sigma \mathbf{J}_n + \overbrace{g_{opt}}^{\substack{\text{optical}\\ \text{excitation}}} + \overbrace{\alpha_R n_i^2}^{\substack{\text{thermal}\\ \text{excitation}}} - \overbrace{\alpha_R n(t)p(t)}^{\text{recombination}}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q_p}\nabla \Sigma \mathbf{J}_p + \overbrace{g_{opt}}^{\substack{\text{optical}\\ \text{excitation}}} + \overbrace{\alpha_R n_i^2}^{\substack{\text{thermal}\\ \text{excitation}}} - \overbrace{\alpha_R n(t)p(t)}^{\text{recombination}}$$

Now, however, we have shown that the last two terms can be combined to leave a simpler single term

$$\frac{\partial n}{\partial t} = \frac{1}{q_n}\nabla \Sigma \mathbf{J}_n + g_{opt} + \alpha_R n_i^2 - \alpha_R \left(n_0 + \Delta n\right)\left(p_0 + \Delta p\right)$$

$$= \frac{1}{q_n}\nabla \Sigma \mathbf{J}_n + g_{opt} - \Delta n \alpha_R \left(n_0 + p_0\right)$$

$$= \frac{1}{q_n}\nabla \Sigma \mathbf{J}_n + g_{opt} - \frac{\Delta n}{\tau}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q_p}\nabla \Sigma \mathbf{J}_p + g_{opt} + \alpha_R n_i^2 - \alpha_R \left(n_0 + \Delta n\right)\left(p_0 + \Delta p\right)$$

$$= -\frac{1}{q_p}\nabla \Sigma \mathbf{J}_p + g_{opt} - \Delta p \alpha_R \left(n_0 + p_0\right)$$

$$= -\frac{1}{q_p}\nabla \Sigma \mathbf{J}_p + g_{opt} - \frac{\Delta p}{\tau}$$

At this point, we can now put in what we have derived for our currents

$$\mathbf{J}_n = q_n n \mu_n \mathbf{E} + q_n D_n \nabla n$$

$$\mathbf{J}_p = q_p p \mu_p \mathbf{E} - q_p D_p \nabla p$$

Plugging these in gives

$$\Downarrow$$

$$\frac{\partial n}{\partial t} = \frac{1}{q_n}\nabla\,\Sigma\left(q_n n\mu_n\mathbf{E} + q_n D_n\nabla n\right) + g_{opt} - \frac{\Delta n}{\tau}$$

$$= n\mu_n\nabla\,\Sigma\mathbf{E} + D_n\nabla^2 n + g_{opt} - \frac{\Delta n}{\tau}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q_p}\nabla\,\Sigma\left(q_p n\mu_p\mathbf{E} - q_p D_p\nabla p\right) + g_{opt} - \frac{\Delta p}{\tau}$$

$$= -p\mu_p\nabla\,\Sigma\mathbf{E} + D_p\nabla^2 p + g_{opt} - \frac{\Delta p}{\tau}$$

finally, we need to look at the left-hand side of the equation and note that the density is made up of the equilibrium part and the injected part and further note that the equilibrium part is a constant; thus

$$\Downarrow$$

$$\frac{\partial \Delta n}{\partial t} = n\mu_n\nabla\,\Sigma\mathbf{E} + D_n\nabla^2\Delta n + g_{opt} - \frac{\Delta n}{\tau}$$

$$\frac{\partial \Delta p}{\partial t} = -p\mu_p\nabla\,\Sigma\mathbf{E} + D_p\nabla^2\Delta p + g_{opt} - \frac{\Delta p}{\tau}$$

At this point we can use this rather ugly equation to gain a little more information. This new information will come in handy later. The new information is based on the following thought. We now have an equation that links shift in position with lifetimes of our carriers. Can we determine how far one of these carriers might travel before they are destroyed through recombination? Let us look at the simplest case. Let us assume that we have a **constant** source of our carriers at a single point (perhaps where a wire is attached to the material) and that we are just allowing them to move through the material on their own. Hence we are in steady-state and we are not applying any electric field. Further, we do not have any source of new carriers in the material, hence $g_{opt} = 0$. Thus

$$\Downarrow$$

$$\frac{\partial \Delta n}{\partial t} = 0 = -D_n\nabla^2\Delta n - \frac{\Delta n}{\tau}$$

$$\frac{\partial \Delta p}{\partial t} = 0 = +D_p\nabla^2\Delta p - \frac{\Delta p}{\tau}$$

We can solve these equations to get

$$\Delta n = D_n\tau\nabla^2\Delta n$$

$$= C_1 e^{-|r|/\sqrt{D_n\tau}} + C_2 e^{+|r|/\sqrt{D_n\tau}}$$

$$= C_1 e^{-|r|/L_n} + C_2 e^{+|r|/L_n}$$

$$\Delta p = -D_p\tau\nabla^2\Delta p$$

$$= C_3 e^{-|r|/\sqrt{D_p\tau}} + C_4 e^{+|r|/\sqrt{D_p\tau}}$$

$$= C_3 e^{-|r|/L_p} + C_4 e^{+|r|/L_p}$$

where $|r|$ is the distance from the point source to wherever we are looking. We can determine the constant by looking at the results at both $|r| = 0$ and $|r| = \infty$.

$$\begin{aligned}\Delta n\big|_{|r|=0} &= \Delta n \\ &= C_1 + C_2 \\ \Delta n\big|_{|r|=\infty} &= 0 \\ &= C_1 0 + C_2 \infty \\ &\Rightarrow C_1 = \Delta n,\ C_2 = 0\end{aligned}$$

$$\begin{aligned}\Delta p\big|_{|r|=0} &= \Delta p \\ &= C_3 + C_4 \\ \Delta p\big|_{|r|=\infty} &= 0 \\ &= C_3 0 + C_4 \infty \\ &\Rightarrow C_3 = \Delta p,\ C_4 = 0\end{aligned}$$

Thus, we have found that

$$\begin{aligned}n(r) &= n_0 + \Delta n(r) \\ &= n_0 + \Delta n(r=0)e^{-r/L_n} \\ p(r) &= p_0 + \Delta p(r) \\ &= p_0 + \Delta p(r=0)e^{-r/L_p}\end{aligned}$$

This means that our excess concentration of carriers drops off exponentially as we move away from the source of the carriers.

Now, we can do back and plug this into our equation for the current. This will let us know how fast the charge carriers move from one place to the next.

$$\begin{aligned}\mathbf{J}_n &= q_n D_n \nabla n \\ &= q_n D_n \nabla\left(n_0 + \Delta n(r=0)e^{-r/L_n}\right) \\ &= -q_n \frac{1}{L_n} D_n \Delta n(r=0)e^{-r/L_n} \\ &= -q_n \frac{1}{L_n} D_n \Delta n(r)\end{aligned}$$

$$\begin{aligned}\mathbf{J}_p &= -q_p D_p \nabla p \\ &= -q_p D_p \nabla\left(p_0 + \Delta p(r=0)e^{-r/L_p}\right) \\ &= q_p \frac{1}{L_p} D_p \Delta p(r=0)e^{-r/L_p} \\ &= q_p \frac{1}{L_p} D_p \Delta p(r)\end{aligned}$$

The final thing that we want to look at before we move on to real devices is what happens to our Quasi-Fermi levels. Remember that we have defined them as a convenience in our equations.

$$p \equiv n_i \exp\left[\frac{-(F_p - \mathcal{E}_i)}{kT}\right]$$

$$n \equiv n_i \exp\left[\frac{(F_n - \mathcal{E}_i)}{kT}\right]$$

Plugging these into our full equations for the currents,

$$\begin{aligned}
\mathbf{J}_n &= q_n n \mu_n \mathbf{E} + q_n D_n \nabla n \\
&= q_n n \mu_n \mathbf{E} + q_n D_n \nabla n_i \exp\left[\frac{(F_n - \mathcal{E}_i)}{kT}\right] \\
&= q_n n \mu_n \mathbf{E} + \frac{q_n D_n}{kT} n_i \exp\left[\frac{(F_n - \mathcal{E}_i)}{kT}\right] \nabla(F_p - \mathcal{E}_i) \\
&= q_n n \mu_n \mathbf{E} + \mu_n n(\nabla F_n - q_n \mathbf{E}) \\
&= \mu_n n \nabla F_n
\end{aligned}$$

$$\begin{aligned}
\mathbf{J}_p &= q_p p \mu_p \mathbf{E} - q_p D_p \nabla p \\
&= \mu_n p \nabla F_p
\end{aligned}$$