

Chapter 1

Background to Electronic Materials

1.1 Introduction to Semiconductors¹

NOTE: This module is adapted from the Connexions module entitled *Introduction to Semiconductors* by Bill Wilson.

If we only had to worry about simple conductors, life would not be very complicated, but on the other hand we wouldn't be able to make computers, CD players, cell phones, i-Pods and a lot of other things which we have found to be useful. We will now move on, and talk about another class of conductors called semiconductors.

In order to understand semiconductors and in fact to get a more accurate picture of how metals, or normal conductors actually work, we really have to resort to quantum mechanics. Electrons in a solid are very tiny objects, and it turns out that when things get small enough, they no longer exactly following the classical "Newtonian" laws of physics that we are all familiar with from everyday experience. It is not the purpose of this course to teach quantum mechanics, so what we are going to do instead is describe the results which come from looking at the behavior of electrons in a solid from a quantum mechanical point of view.

Solids (at least the ones we will be talking about, and especially semiconductors) are crystalline materials, which means that they have their atoms arranged in a ordered fashion. We can take silicon (the most important semiconductor) as an example. Silicon is a group 14(IV) element, which means it has four electrons in its outer or valence shell. Silicon crystallizes in a structure called the diamond crystal lattice, shown in Figure 1.1. Each silicon atom has four covalent bonds, arranged in a tetrahedral formation about the atom center.

¹This content is available online at <<http://cnx.org/content/m33647/1.3/>>.

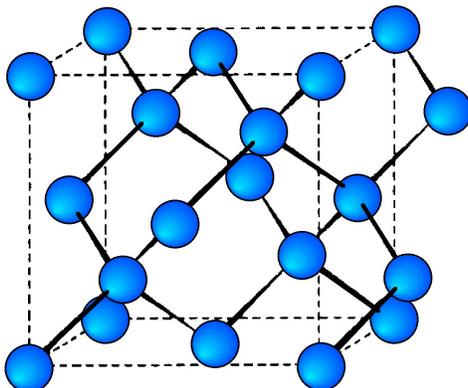


Figure 1.1: Crystal structure of silicon.

In two dimensions, we can schematically represent a piece of single-crystal silicon as shown in Figure 1.2. Each silicon atom shares its four valence electrons with valence electrons from four nearest neighbors, filling the shell to 8 electrons, and forming a stable, periodic structure. Once the atoms have been arranged like this, the outer valence electrons are no longer strongly bound to the host atom. The outer shells of all of the atoms blend together and form what is called a band. The electrons are now free to move about within this band, and this can lead to electrical conductivity as we discussed earlier.

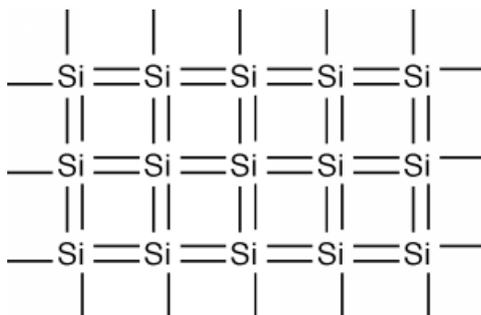


Figure 1.2: A 2-D representation of a silicon crystal.

This is not the complete story however, for it turns out that due to quantum mechanical effects, there is not just one band which holds electrons, but several of them. What will follow is a very qualitative picture of how the electrons are distributed when they are in a periodic solid, and there are necessarily some details which we will be forced to gloss over. On the other hand this will give you a pretty good picture of what is going on, and may enable you to have some understanding of how a semiconductor really works. Electrons are not only distributed throughout the solid crystal spatially, but they also have a distribution in energy as well. The potential energy function within the solid is periodic in nature. This potential function comes from the positively charged atomic nuclei which are arranged in the crystal in a regular array. A detailed

analysis of how electron wave functions, the mathematical abstraction which one must use to describe how small quantum mechanical objects behave when they are in a periodic potential, gives rise to an energy distribution somewhat like that shown in Figure 1.3.

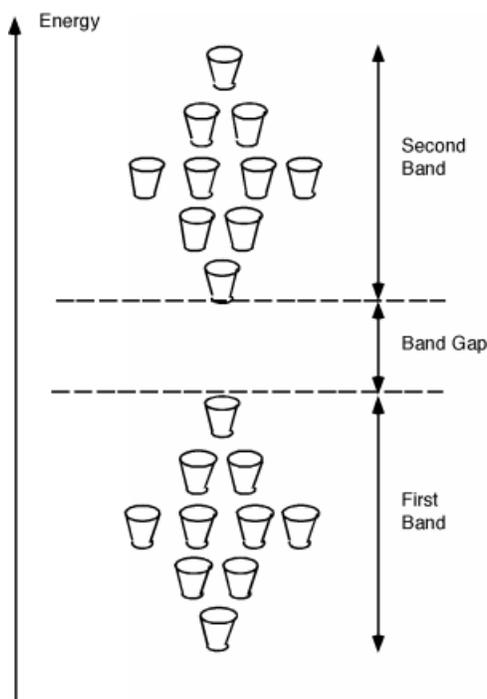


Figure 1.3: Schematic of the first two bands in a periodic solid showing energy levels and bands.

Firstly, unlike the case for free electrons, in a periodic solid, electrons are not free to take on any energy value they wish. They are forced into specific energy levels called allowed states, which are represented by the cups in Figure 1.3. The allowed states are not distributed uniformly in energy either. They are grouped into specific configurations called energy bands. There are no allowed levels at zero energy and for some distance above that. Moving up from zero energy, we then encounter the first energy band. At the bottom of the band there are very few allowed states, but as we move up in energy, the number of allowed states first increases, and then falls off again. We then come to a region with no allowed states, called an energy band gap. Above the band gap, another band of allowed states exists. This goes on and on, with any given material having many such bands and band gaps. This situation is shown schematically in Figure 1.3, where the small cups represent allowed energy levels, and the vertical axis represents electron energy.

It turns out that each band has exactly $2N$ allowed states in it, where N is the total number of atoms in the particular crystal sample we are talking about. (Since there are 10 cups in each band in the figure, it must represent a crystal with just 5 atoms in it. Not a very big crystal at all!) Into these bands we must now distribute all of the valence electrons associated with the atoms, with the restriction that we can only put one electron into each allowed state. This is the result of something called the Pauli exclusion principle. Since in the case of silicon there are 4 valence electrons per atom, we would just fill up the first two bands, and the next would be empty. If we make the logical assumption that the electrons will fill in the levels with the lowest energy first, and only go into higher lying levels if the ones below are already filled. This situation

is shown in Figure 1.4, in which we have represented electrons as small black balls with a "-" sign on them. Indeed, the first two bands are completely full, and the next is empty. What will happen if we apply an electric field to the sample of silicon? Remember the diagram we have at hand right now is an energy based one, we are showing how the electrons are distributed in energy, not how they are arranged spatially. On this diagram we can not show how they will move about, but only how they will change their energy as a result of the applied field. The electric field will exert a force on the electrons and attempt to accelerate them. If the electrons are accelerated, then they must increase their kinetic energy. Unfortunately, there are no empty allowed states in either of the filled bands. An electron would have to jump all the way up into the next (empty) band in order to take on more energy. In silicon, the gap between the top of the highest most occupied band and the lowest unoccupied band is 1.1 eV. (One eV is the potential energy gained by an electron moving across an electrical potential of one volt.) The mean free path or distance over which an electron would normally move before it suffers a collision is only a few hundred angstroms (ca. 300×10^{-8} cm) and so you would need a very large electric field (several hundred thousand V/cm) in order for the electron to pick up enough energy to "jump the gap". This makes it appear that silicon would be a very bad conductor of electricity, and in fact, very pure silicon is very poor electrical conductor.



Figure 1.4: Silicon, with first two bands full and the next empty.

A metal is an element with an odd number of valence electrons so that a metal ends up with an upper band which is just half full of electrons. This is illustrated in Figure 1.5. Here we see that one band is full, and the next is just half full. This would be the situation for the Group 13(III) element aluminum for

instance. If we apply an electric field to these carriers, those near the top of the distribution can indeed move into higher energy levels by acquiring some kinetic energy of motion, and easily move from one place to the next. In reality, the whole situation is a bit more complex than we have shown here, but this is not too far from how it actually works.

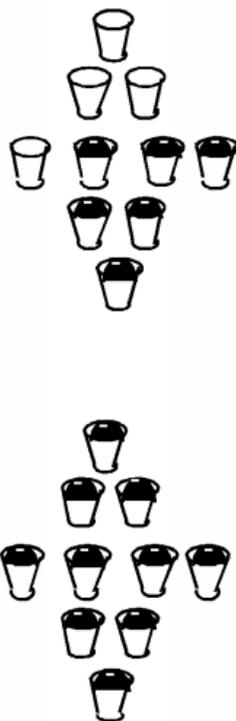


Figure 1.5: Electron distribution for a metal or good conductor.

So, back to our silicon sample. If there are no places for electrons to "move" into, then how does silicon work as a "semiconductor"? Well, in the first place, it turns out that not all of the electrons are in the bottom two bands. In silicon, unlike say quartz or diamond, the band gap between the top-most full band, the next empty one is not so large. As we mentioned above it is only about 1.1 eV. So long as the silicon is not at absolute zero temperature, some electrons near the top of the full band can acquire enough thermal energy that they can "hop" the gap, and end up in the upper band, called the conduction band. This situation is shown in Figure 1.6.

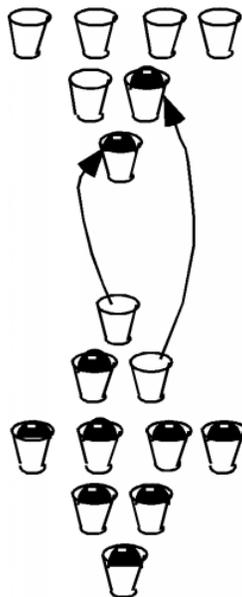


Figure 1.6: Thermal excitation of electrons across the band gap.

In silicon at room temperature, roughly 10^{10} electrons per cubic centimeter are thermally excited across the band-gap at any one time. It should be noted that the excitation process is a continuous one. Electrons are being excited across the band, but then they fall back down into empty spots in the lower band. On average however, the 10^{10} in each cm^3 of silicon is what you will find at any given instant. Now 10 billion electrons per cubic centimeter seems like a lot of electrons, but lets do a simple calculation. The mobility of electrons in silicon is about $1000 \text{ cm}^2/\text{V}\cdot\text{s}$. Remember, mobility times electric field yields the average velocity of the carriers. Electric field has units of V/cm , so with these units we get velocity in cm/s as we should. The charge on an electron is 1.6×10^{-19} coulombs. Thus from (1.1):

$$\begin{aligned}
 \sigma &= nq\mu \\
 &= 10^{10} (1.6 \times 10^{-19}) 1000 \\
 &= 1.6 \times 10^{-6} \text{ mhos/cm}
 \end{aligned}
 \tag{1.1}$$

If we have a sample of silicon 1 cm long by (1 mm x 1mm) square, it would have a resistance, (1.2), which does not make it much of a "conductor". In fact, if this were all there was to the silicon story, we could pack up and move on, because at any reasonable temperature, silicon would conduct electricity very poorly.

$$\begin{aligned}
 R &= L/\sigma A \\
 &= 1/(1.6 \times 10^{-6})(0.1)^2 \\
 &= 1.6 \times 10^6 \text{ M}\Omega
 \end{aligned}
 \tag{1.2}$$

1.2 Doped Semiconductors²

NOTE: This module is adapted from the Connexions module entitled *Doped Semiconductors* by Bill Wilson.

To see how we can make silicon a useful electronic material, we will have to go back to its crystal structure (Figure 1.7). Suppose somehow we could substitute a few atoms of phosphorus for some of the silicon atoms.

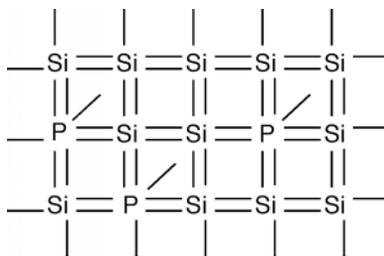


Figure 1.7: A two dimensional representation of a silicon crystal lattice "doped" with phosphorus.

If you sneak a look at the periodic table, you will see that phosphorus is a group V element, as compared with silicon which is a group 14(IV) element. What this means is the phosphorus atom has five outer or valence electrons, instead of the four which silicon has. In a lattice composed mainly of silicon, the extra electron associated with the phosphorus atom has no "mating" electron with which it can complete a shell, and so is left loosely dangling to the phosphorus atom, with relatively low binding energy. In fact, with the addition of just a little thermal energy (from the natural or latent heat of the crystal lattice) this electron can break free and be left to wander around the silicon atom freely. In our "energy band" picture, we have something like what we see in Figure 1.8. The phosphorus atoms are represented by the added cups with P's on them. They are new allowed energy levels which are formed within the "band gap" near the bottom of the first empty band. They are located close enough to the empty (or "conduction") band, so that the electrons which they contain are easily excited up into the conduction band. There, they are free to move about and contribute to the electrical conductivity of the sample. Note also, however, that since the electron has left the vicinity of the phosphorus atom, there is now one more proton than there are electrons at the atom, and hence it has a net positive charge of $1q$. We have represented this by putting a little "+" sign in each P-cup. Note that this positive charge is fixed at the site of the phosphorous atom called a *donor* since it "donates" an electron up into the conduction band, and is not free to move about in the crystal.

²This content is available online at <<http://cnx.org/content/m33703/1.2/>>.

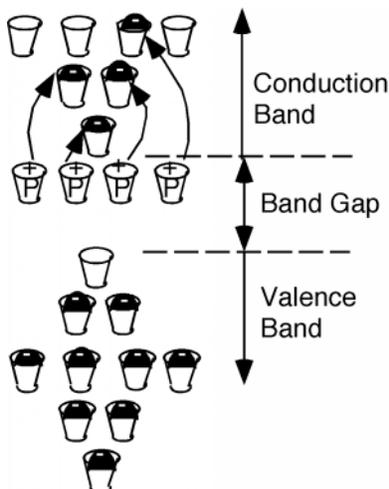


Figure 1.8: Silicon doped with phosphorus.

How many phosphorus atoms do we need to significantly change the resistance of our silicon? Suppose we wanted our 1 mm by 1 mm square sample to have a resistance of one ohm as opposed to more than 60 M Ω . Turning the resistance equation around we get, (1.3). And hence, if we continue to assume an electron mobility of 1000 cm²/volt.sec, (1.4).

$$\begin{aligned}
 \sigma &= L/RA \\
 &= 1 \Omega/1 \times (0.1)^2 \\
 &= 100 \text{ mho/cm}
 \end{aligned}
 \tag{1.3}$$

$$\begin{aligned}
 n &= \sigma/q\mu \\
 &= 100/(1.6 \times 10^{-19})1000 \\
 &= 6.25 \times 10^{17} \text{ cm}^3
 \end{aligned}
 \tag{1.4}$$

Now adding more than 6×10^{17} phosphorus atoms per cubic centimeter might seem like a lot of phosphorus, until you realize that there are almost 10^{24} silicon atoms in a cubic centimeter and hence only one in every 1.6 million silicon atoms has to be changed into a phosphorus one to reduce the resistance of the sample from several 10s of M Ω down to only one Ω . This is the real power of semiconductors. You can make dramatic changes in their electrical properties by the addition of only minute amounts of impurities. This process is called *doping* the semiconductor. It is also one of the great challenges of the semiconductor manufacturing industry, for it is necessary to maintain fantastic levels of control of the impurities in the material in order to predict and control their electrical properties.

Again, if this were the end of the story, we still would not have any calculators, cell phones, or stereos, or at least they would be very big and cumbersome and unreliable, because they would have to work using vacuum tubes. We now have to focus on the few "empty" spots in the lower, almost full band (called the *valence band*.) We will take another view of this band, from a somewhat different perspective. I must confess at this point that what I am giving you is even further from the way things really work, then the "cups at different energies" picture we have been using so far. The problem is, that in order to do things right, we have to get involved in momentum phase-space, a lot more quantum mechanics, and generally a bunch

of math and concepts we don't really need in order to have some idea of how semiconductor devices work. What follows below is really intended as a motivation, so that you will have some feeling that what we state as results, is actually reasonable.

Consider Figure 1.9. Here we show all of the electrons in the valence, or almost full band, and for simplicity show one missing electron. Let's apply an electric field, as shown by the arrow in the figure. The field will try to move the (negatively charged) electrons to the left, but since the band is almost completely full, the only one that can move is the one right next to the empty spot, or *hole* as it is called.

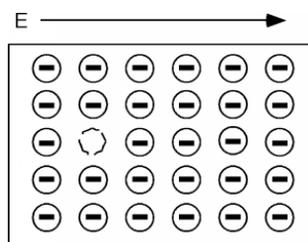


Figure 1.9: Band full of electrons, with one missing.

One thing you may be worrying about is what happens to the electrons at the ends of the sample. This is one of the places where we are getting a somewhat distorted view of things, because we should really be looking in momentum, or wave-vector space rather than "real" space. In that picture, they magically drop off one side and "reappear" on the other. This doesn't happen in real space of course, so there is no easy way we can deal with it.

A short time after we apply the electric field we have the situation shown in Figure 1.10, and a little while after that we have Figure 1.11. We can interpret this motion in two ways. One is that we have a net flow of negative charge to the left, or if we consider the effect of the aggregate of all the electrons in the band we could picture what is going on as a single positive charge, moving to the right. This is shown in Figure 1.12. Note that in either view we have the same net effect in the way the total net charge is transported through the sample. In the mostly negative charge picture, we have a net flow of negative charge to the left. In the single positive charge picture, we have a net flow of positive charge to the right. Both give the same sign for the current!

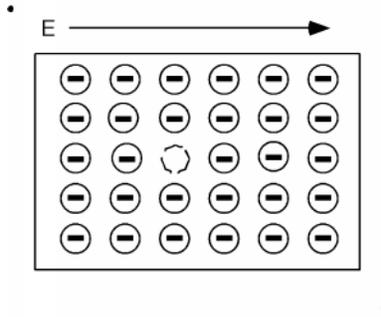


Figure 1.10: Motion of the "missing" electron with an electric field.

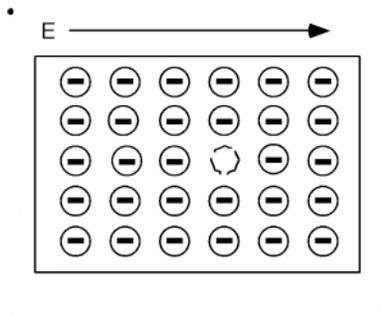


Figure 1.11: Further motion of the "missing electron" spot.

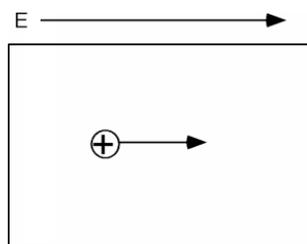


Figure 1.12: Motion of a "hole" due to an applied electric field.

Thus, it turns out, we can consider the consequences of the empty spaces moving through the co-ordinated motion of electrons in an almost full band as being the motion of positive charges, moving wherever these empty spaces happen to be. We call these charge carriers "holes" and they too can add to the total conduction of electricity in a semiconductor. Using ρ to represent the density (in cm^{-3}) of spaces in the valence band and μ_e and μ_h to represent the mobility of electrons and holes respectively (they are usually not the same) we can modify to give the conductivity σ , when both electrons' holes are present, (1.5).

$$\sigma = nq\mu_e + \rho q\mu_h \quad (1.5)$$

How can we get a sample of semiconductor with a lot of holes in it? What if, instead of phosphorus, we dope our silicon sample with a group III element, say boron? This is shown in Figure 1.13. Now we have some missing orbitals, or places where electrons could go if they were around. This modifies our energy picture as follows in Figure 1.14. Now we see a set of new levels introduced by the boron atoms. They are located within the band gap, just a little way above the top of the almost full, or valence band. Electrons in the valence band can be thermally excited up into these new allowed levels, creating empty states, or holes, in the valence band. The excited electrons are stuck at the boron atom sites called *acceptors*, since they "accept" an electron from the valence band, and hence act as fixed negative charges, localized there. A semiconductor which is doped predominantly with acceptors is called *p-type*, and most of the electrical conduction takes place through the motion of holes. A semiconductor which is doped with donors is called *n-type*, and conduction takes place mainly through the motion of electrons.

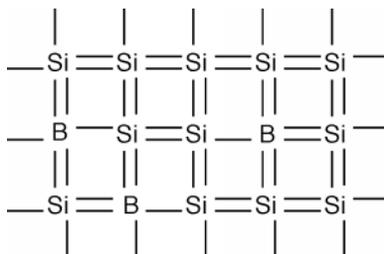


Figure 1.13: A two dimensional representation of a silicon crystal lattice doped with boron.

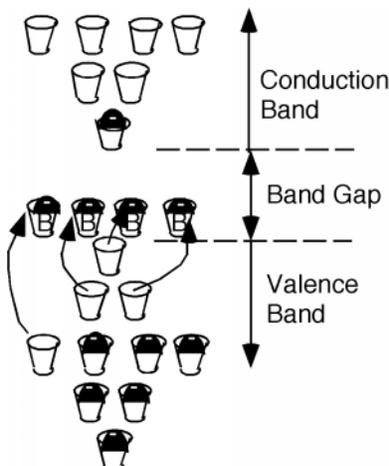


Figure 1.14: P-type silicon, due to boron acceptors.

In n-type material, we can assume that all of the phosphorous atoms, or *donors*, are fully ionized when they are present in the silicon structure. Since the number of donors is usually much greater than the native, or intrinsic electron concentration, ($\approx 10^{10} \text{ cm}^{-3}$), if N_d is the density of donors in the material, then n , the electron concentration, $\approx N_d$. If an electron deficient material such as boron is present, then the material is called *p-type* silicon, and the hole concentration is just $\approx N_a$ the concentration of *acceptors*, since these atoms "accept" electrons from the valence band.

If both donors and acceptors are in the material, then whichever one has the higher concentration wins out. This is called compensation. If there are more donors than acceptors then the material is n-type and $n \approx N_d - N_a$. If there are more acceptors than donors then the material is p-type and $p \approx N_a - N_d$. It should be noted that in most compensated material, one type of impurity usually has a much greater (several orders of magnitude) concentration than the other, and so the subtraction process described above usually does not change things very much, e.g., $10^{18} - 10^{16} \approx 10^{18}$.

One other fact which you might find useful is that, again, because of quantum mechanics, it turns out that the product of the electron and hole concentration in a material must remain a constant. In silicon at room temperature:

$$np \equiv n_i^2 \approx 10^{20} \text{ cm}^{-3} \quad (1.6)$$

Thus, if we have an n-type sample of silicon doped with 10^{17} donors per cubic centimeter, then n , the electron concentration is just p , the hole concentration, is $10^{20}/10^{17} = 10^3 \text{ cm}^{-3}$. The carriers which dominate a material are called *majority carriers*, which would be the electrons in the above example. The other carriers are called *minority carriers* (the holes in the example) and while 10^3 might not seem like much compared to 10^{17} the presence of minority carriers is still quite important and can not be ignored. Note that if the material is undoped, then it must be that $n = p$ and $n = p = 10^{10}$.

The picture of "cups" of different allowed energy levels is useful for gaining a pictorial understanding of what is going on in a semiconductor, but becomes somewhat awkward when you want to start looking at devices which are made up of both n and p type silicon. Thus, we will introduce one more way of describing what is going on in our material. The picture shown in Figure 1.15 is called a band diagram. A **band diagram** is just a representation of the energy as a function of position with a semiconductor device. In a band diagram, positive energy for electrons is upward, while for holes, positive energy is downwards. That

is, if an electron moves upward, its potential energy increases just as a with a normal mass in a gravitational field. Also, just as a mass will "fall down" if given a chance, an electron will move down a slope shown in a band diagram. On the other hand, holes gain energy by moving downward and so they have a tendency to "float" upward if given the chance - much like a bubble in a liquid. The line labeled E_c in Figure 1.15 shows the edge of the conduction band, or the bottom of the lowest unoccupied allowed band, while E_v is the top edge of the valence, or highest occupied band. The band gap, E_g for the material is obviously $E_c - E_v$. The dotted line labeled E_f is called the *Fermi level* and it tells us something about the chemical equilibrium energy of the material, and also something about the type and number of carriers in the material. More on this later. Note that there is no zero energy level on a diagram such as this. We often use either the Fermi level or one or other of the band edges as a reference level on lieu of knowing exactly where "zero energy" is located.

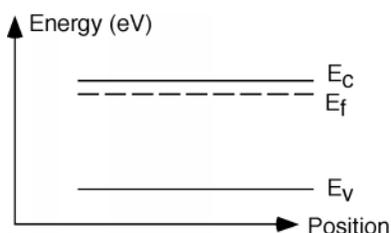


Figure 1.15: An electron band-diagram for a semiconductor.

The distance (in energy) between the Fermi level and either E_c and E_v gives us information concerning the density of electrons and holes in that region of the semiconductor material. The details, once again, will have to be begged off on grounds of mathematical complexity. It turns out that you can say:

$$n = N_c e^{\left(\frac{E_c - E_f}{kT}\right)} \quad (1.7)$$

$$p = N_v e^{\left(\frac{E_f - E_v}{kT}\right)} \quad (1.8)$$

Both N_c and N_v are constants that depend on the material you are talking about, but are typically on the order of 10^{19} cm^{-3} . The expression in the denominator of the exponential is just Boltzman's constant ($8.63 \times 10^{-5} \text{ eV/K}$), k , times the temperature T of the material (in absolute temperature or Kelvin). At room temperature $kT = 1/40$ of an electron volt. Look carefully at the numerators in the exponential. Note first that there is a minus sign in front, which means the bigger the number in the exponent, the fewer carriers we have. Thus, the top expression says that if we have n-type material, then E_f must not be too far away from the conduction band, while if we have p-type material, then the Fermi level, E_f must be down close to the valence band. The closer E_f gets to E_c the more electrons we have. The closer E_f gets to E_v , the more holes we have. Figure 1.15 therefore must be for a sample of n-type material. Note also that if we know how heavily a sample is doped (i.e., we know what N_d is) and from the fact that $n \approx N_d$ we can use to find out how far away the Fermi level is from the conduction band, (1.9).

$$E_c - E_f = kT \ln \left(\frac{N_c}{N_d} \right) \quad (1.9)$$

To help further in our ability to picture what is going on, we will often add to this band diagram, some small signed circles to indicate the presence of mobile electrons and holes in the material. Note that the electrons are spread out in energy. From our "cups" picture we know they like to stay in the lower energy states if possible, but some will be distributed into the higher levels as well. What is distorted here is the scale. The band-gap for silicon is 1.1 eV, while the actual spread of the electrons would probably only be a few tenths of an eV, not nearly as much as is shown in Figure 1.16. Lets look at a sample of p-type material, just for comparison. Note that for holes, increasing energy goes *down* not up, so their distribution is inverted from that of the electrons. You can kind of think of holes as bubbles in a glass of soda or beer, they want to float to the top if they can. Note also for both n and p-type material there are also a few "minority" carriers, or carriers of the opposite type, which arise from thermal generation across the band-gap.

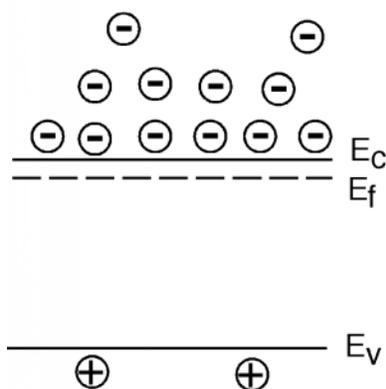


Figure 1.16: Band diagram for an n-type semiconductor.

1.3 Diffusion³

NOTE: This module is adapted from the Connexions module entitled *Diffusion* by Bill Wilson.

1.3.1 Introduction

Let us turn our attention to what happens to the electrons and holes once they have been injected across a forward-biased junction. We will concentrate just on the electrons which are injected into the p-side of the junction, but keep in mind that similar things are also happening to the holes which enter the n-side.

When electrons are injected across a junction, they move away from the junction region by a diffusion process, while at the same time, some of them are disappearing because they are minority carriers (electrons in basically p-type material) and so there are lots of holes around for them to recombine with. This is all shown schematically in Figure 1.17.

³This content is available online at <<http://cnx.org/content/m33730/1.2/>>.

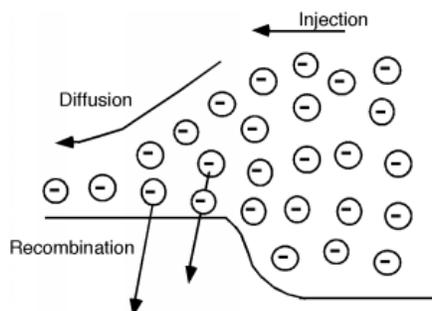


Figure 1.17: Processes involved in electron transport across a p-n junction.

1.3.2 Diffusion process quantified

It is actually fairly easy to quantify this, and come up with an expression for the electron distribution within the p-region. First we have to look a little bit at the diffusion process however. Imagine that we have a series of bins, each with a different number of electrons in them. In a given time, we could imagine that all of the electrons would flow out of their bins into the neighboring ones. Since there is no reason to expect the electrons to favor one side over the other, we will assume that exactly half leave by each side. This is all shown in Figure 1.18. We will keep things simple and only look at three bins. Imagine there are 4, 6, and 8 electrons respectively in each of the bins. After the required "emptying time," we will have a net flux of exactly one electron across each boundary as shown.

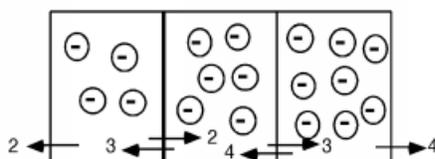


Figure 1.18: A schematic representation of a diffusion problem.

Now let's raise the number of electrons to 8, 12 and 16 respectively (Figure 1.19). We find that the net flux across each boundary is now 2 electrons per emptying time, rather than one. Note that the gradient (slope) of the concentration in the boxes has also doubled from one per box to two per box. This leads us to a rather obvious statement that the flux of carriers is proportional to the gradient of their density. This is stated formally in what is known as Fick's First Law of Diffusion, (1.10). Where D_e is simply a proportionality constant called the diffusion coefficient. Since we are talking about the motion of electrons, this diffusion flux must give rise to a current density $J_{e_{diff}}$. Since an electron has a charge $-q$ associated

with it, (1.11).

$$\text{Flux} = (-D_e) \frac{dn(x)}{dx} \quad (1.10)$$

$$J_{e\text{diff}} = qD_e \frac{dn}{dx} \quad (1.11)$$

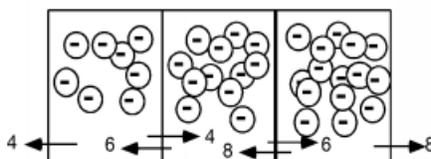


Figure 1.19: A schematic representation of a diffusion from bins.

Now we have to invoke something called the continuity equation. Imagine we have a volume (V) which is filled with some charge (Q). It is fairly obvious that if we add up all of the current density which is flowing out of the volume that it must be equal to the time rate of decrease of the charge within that volume. This idea is expressed in the formula below which uses a closed-surface integral, along with all the other integrals to follow:

$$\oint_S J dS = -\frac{dQ}{dt} \quad (1.12)$$

We can write Q as, (1.13), where we are doing a volume integral of the charge density (ρ) over the volume (V). Now we can use Gauss' theorem which says we can replace a surface integral of a quantity with a volume integral of its divergence, (1.14).

$$Q = \int_V \rho(v) dV \quad (1.13)$$

$$\oint_S J dS = \int div(J) dV \quad (1.14)$$

So, combining (1.12), (1.13) and (1.14), we have, (1.15).

$$\int div(J) dV = -\int \frac{d\rho}{dt} dV \quad (1.15)$$

Finally, we let the volume V shrink down to a point, which means the quantities inside the integral must be equal, and we have the differential form of the continuity equation (in one dimension), (1.16).

$$\begin{aligned} div(J) &= \frac{\partial J}{\partial x} \\ &= -\frac{d\rho(x)}{dt} \end{aligned} \quad (1.16)$$

1.3.3 What about the electrons?

Now let's go back to the electrons in the diode. The electrons which have been injected across the junction are called *excess minority carriers*, because they are electrons in a p-region (hence minority) but their concentration is greater than what they would be if they were in a sample of p-type material at equilibrium. We will designate them as n' , and since they could change with both time and position we shall write them as $n'(x,t)$. Now there are two ways in which $n'(x,t)$ can change with time. One would be if we were to stop injecting electrons in from the n-side of the junction. A reasonable way to account for the decay which would occur if we were not supplying electrons would be to write:

$$\frac{d}{dt}n'(x,t) = -\frac{n'(x,t)}{\tau_r} \quad (1.17)$$

Where τ_r called the minority carrier recombination lifetime. It is pretty easy to show that if we start out with an excess minority carrier concentration n_0' at $t = 0$, then $n'(x,t)$ will go as, (1.18). But, the electron concentration can also change because of electrons flowing into or out of the region x . The electron concentration $n'(x,t)$ is just $\frac{\rho(x,t)}{q}$. Thus, due to electron flow we have, (1.19).

$$n'(x,t) = n_0' e^{-\frac{t}{\tau_r}} \quad (1.18)$$

$$\begin{aligned} \frac{d}{dt}n'(x,t) &= \frac{1}{q} \frac{d\rho(x,t)}{dt} \\ &= \frac{1}{q} \text{div}(J(x,t)) \end{aligned} \quad (1.19)$$

But, we can get an expression for $J(x,t)$ from (1.11). Reducing the divergence in (1.19) to one dimension (we just have a $\frac{\partial J}{\partial x}$) we finally end up with, (1.20).

$$\frac{d}{dt}n'(x,t) = D_e \frac{d^2n'(x,t)}{dx^2} \quad (1.20)$$

Combining (1.20) and (1.17) (electrons will, after all, suffer from both recombination and diffusion) and we end up with:

$$\frac{d}{dt}n'(x,t) = D_e \frac{d^2n'(x,t)}{dx^2} - \frac{n'(x,t)}{\tau_r} \quad (1.21)$$

This is a somewhat specialized form of an equation called the ambipolar diffusion equation. It seems kind of complicated but we can get some nice results from it if we make some simply boundary condition assumptions.

1.3.3.1 Using the ambipolar diffusion equation

For anything we will be interested in, we will only look at steady state solutions. This means that the time derivative on the LHS of (1.21) is zero, and so letting $n'(x,t)$ become simply $n'(x)$ since we no longer have any time variation to worry about, we have:

$$\frac{d^2}{dx^2}n'(x) - \frac{1}{D_e\tau_r}n'(x) = 0 \quad (1.22)$$

Picking the not unreasonable boundary conditions that $n'(0) = n_0$ (the concentration of excess electrons just at the start of the diffusion region) and $n'(x) \rightarrow 0$ as $x \rightarrow \infty$ (the excess carriers go to zero when we get far from the junction) then:

$$n(x) = n_0 e^{-\frac{x}{\sqrt{D_e\tau_r}}} \quad (1.23)$$

The expression in the radical $\sqrt{D_e\tau_r}$ is called the electron diffusion length, L_e , and gives us some idea as to how far away from the junction the excess electrons will exist before they have more or less all recombined.

This will be important for us when we move on to bipolar transistors. A typical value for the diffusion coefficient for electrons in silicon would be $D_e = 25 \text{ cm}^2/\text{sec}$ and the minority carrier lifetime is usually around a microsecond. As shown in (1.24) this is not very far at all.

$$\begin{aligned} L_e &= \sqrt{D_e \tau_r} \\ &= \sqrt{25 \times 10^{-6}} \\ &= 5 \times 10^{-3} \text{cm} \end{aligned} \tag{1.24}$$

1.4 Crystal Structure⁴

1.4.1 Introduction

In any sort of discussion of crystalline materials, it is useful to begin with a discussion of crystallography: the study of the formation, structure, and properties of crystals. A crystal structure is defined as the particular repeating arrangement of atoms (molecules or ions) throughout a crystal. Structure refers to the internal arrangement of particles and not the external appearance of the crystal. However, these are not entirely independent since the external appearance of a crystal is often related to the internal arrangement. For example, crystals of cubic rock salt (NaCl) are physically cubic in appearance. Only a few of the possible crystal structures are of concern with respect to simple inorganic salts and these will be discussed in detail, however, it is important to understand the nomenclature of crystallography.

1.4.2 Crystallography

1.4.2.1 Bravais lattice

The Bravais lattice is the basic building block from which all crystals can be constructed. The concept originated as a topological problem of finding the number of different ways to arrange points in space where each point would have an identical “atmosphere”. That is each point would be surrounded by an identical set of points as any other point, so that all points would be indistinguishable from each other. Mathematician Auguste Bravais discovered that there were 14 different collections of the groups of points, which are known as Bravais lattices. These lattices fall into seven different “crystal systems”, as differentiated by the relationship between the angles between sides of the “unit cell” and the distance between points in the unit cell. The unit cell is the smallest group of atoms, ions or molecules that, when repeated at regular intervals in three dimensions, will produce the lattice of a crystal system. The “lattice parameter” is the length between two points on the corners of a unit cell. Each of the various lattice parameters are designated by the letters a , b , and c . If two sides are equal, such as in a tetragonal lattice, then the lengths of the two lattice parameters are designated a and c , with b omitted. The angles are designated by the Greek letters α , β , and γ , such that an angle with a specific Greek letter is not subtended by the axis with its Roman equivalent. For example, α is the included angle between the b and c axis.

Table 1.1 shows the various crystal systems, while Figure 1.20 shows the 14 Bravais lattices. It is important to distinguish the characteristics of each of the individual systems. An example of a material that takes on each of the Bravais lattices is shown in Table 1.2.

⁴This content is available online at <<http://cnx.org/content/m16927/1.10/>>.

| System | Axial lengths and angles | Unit cell geometry |
|--------------|--|--|
| cubic | $a = b = c, \alpha = \beta = \gamma = 90^\circ$ |  |
| tetragonal | $a = b \neq c, \alpha = \beta = \gamma = 90^\circ$ |  |
| orthorhombic | $a \neq b \neq c, \alpha = \beta = \gamma = 90^\circ$ |  |
| rhombohedral | $a = b = c, \alpha = \beta = \gamma \neq 90^\circ$ |  |
| hexagonal | $a = b \neq c, \alpha = \beta = 90^\circ, \gamma = 120^\circ$ |  |
| monoclinic | $a \neq b \neq c, \alpha = \gamma = 90^\circ, \beta \neq 90^\circ$ |  |
| triclinic | $a \neq b \neq c, \alpha \neq \beta \neq \gamma$ |  |

Table 1.1: Geometrical characteristics of the seven crystal systems.

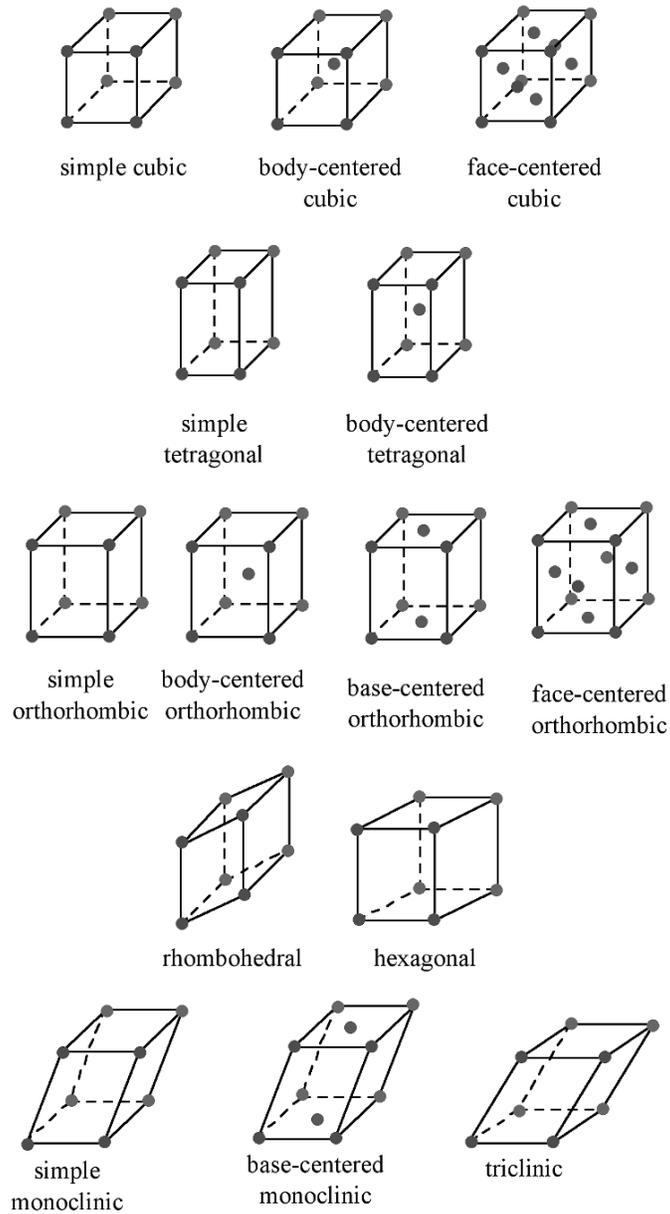


Figure 1.20: Bravais lattices.

| Crystal system | Example |
|----------------|--|
| triclinic | $\text{K}_2\text{S}_2\text{O}_8$ |
| monoclinic | As_4S_4 , KNO_2 |
| rhombohedral | Hg, Sb |
| hexagonal | Zn, Co, NiAs |
| orthorhombic | Ga, Fe_3C |
| tetragonal | In, TiO_2 |
| cubic | Au, Si, NaCl |

Table 1.2: Examples of elements and compounds that adopt each of the crystal systems.

The cubic lattice is the most symmetrical of the systems. All the angles are equal to 90° , and all the sides are of the same length ($a = b = c$). Only the length of one of the sides (a) is required to describe this system completely. In addition to simple cubic, the cubic lattice also includes body-centered cubic and face-centered cubic (Figure 1.20). Body-centered cubic results from the presence of an atom (or ion) in the center of a cube, in addition to the atoms (ions) positioned at the vertices of the cube. In a similar manner, a face-centered cubic requires, in addition to the atoms (ions) positioned at the vertices of the cube, the presence of atoms (ions) in the center of each of the cubes face.

The tetragonal lattice has all of its angles equal to 90° , and has two out of the three sides of equal length ($a = b$). The system also includes body-centered tetragonal (Figure 1.20).

In an orthorhombic lattice all of the angles are equal to 90° , while all of its sides are of unequal length. The system needs only to be described by three lattice parameters. This system also includes body-centered orthorhombic, base-centered orthorhombic, and face-centered orthorhombic (Figure 1.20). A base-centered lattice has, in addition to the atoms (ions) positioned at the vertices of the orthorhombic lattice, atoms (ions) positioned on just two opposing faces.

The rhombohedral lattice is also known as trigonal, and has no angles equal to 90° , but all sides are of equal length ($a = b = c$), thus requiring only by one lattice parameter, and all three angles are equal ($\alpha = \beta = \gamma$).

A hexagonal crystal structure has two angles equal to 90° , with the other angle (γ) equal to 120° . For this to happen, the two sides surrounding the 120° angle must be equal ($a = b$), while the third side (c) is at 90° to the other sides and can be of any length.

The monoclinic lattice has no sides of equal length, but two of the angles are equal to 90° , with the other angle (usually defined as β) being something other than 90° . It is a tilted parallelogram prism with rectangular bases. This system also includes base-centered monoclinic (Figure 1.20).

In the triclinic lattice none of the sides of the unit cell are equal, and none of the angles within the unit cell are equal to 90° . The triclinic lattice is chosen such that all the internal angles are either acute or obtuse. This crystal system has the lowest symmetry and must be described by 3 lattice parameters (a , b , and c) and the 3 angles (α , β , and γ).

1.4.2.2 Atom positions, crystal directions and Miller indices

1.4.2.2.1 Atom positions and crystal axes

The structure of a crystal is defined with respect to a unit cell. As the entire crystal consists of repeating unit cells, this definition is sufficient to represent the entire crystal. Within the unit cell, the atomic arrangement is expressed using coordinates. There are two systems of coordinates commonly in use, which can cause some confusion. Both use a corner of the unit cell as their origin. The first, less-commonly seen system is that of Cartesian or orthogonal coordinates (X , Y , Z). These usually have the units of Angstroms and relate to the distance in each direction between the origin of the cell and the atom. These coordinates may be

manipulated in the same fashion are used with two- or three-dimensional graphs. It is very simple, therefore, to calculate inter-atomic distances and angles given the Cartesian coordinates of the atoms. Unfortunately, the repeating nature of a crystal cannot be expressed easily using such coordinates. For example, consider a cubic cell of dimension 3.52 Å. Pretend that this cell contains an atom that has the coordinates (1.5, 2.1, 2.4). That is, the atom is 1.5 Å away from the origin in the x direction (which coincides with the *a* cell axis), 2.1 Å in the y (which coincides with the *b* cell axis) and 2.4 Å in the z (which coincides with the *c* cell axis). There will be an equivalent atom in the next unit cell along the x-direction, which will have the coordinates (1.5 + 3.52, 2.1, 2.4) or (5.02, 2.1, 2.4). This was a rather simple calculation, as the cell has very high symmetry and so the cell axes, *a*, *b* and *c*, coincide with the Cartesian axes, X, Y and Z. However, consider lower symmetry cells such as triclinic or monoclinic in which the cell axes are not mutually orthogonal. In such cases, expressing the repeating nature of the crystal is much more difficult to accomplish.

Accordingly, atomic coordinates are usually expressed in terms of fractional coordinates, (*x*, *y*, *z*). This coordinate system is coincident with the cell axes (*a*, *b*, *c*) and relates to the position of the atom in terms of the fraction along each axis. Consider the atom in the cubic cell discussion above. The atom was 1.5 Å in the *a* direction away from the origin. As the *a* axis is 3.52 Å long, the atom is ($1.5/3.52$) or 0.43 of the axis away from the origin. Similarly, it is ($2.1/3.52$) or 0.60 of the *b* axis and ($2.4/3.5$) or 0.68 of the *c* axis. The fractional coordinates of this atom are, therefore, (0.43, 0.60, 0.68). The coordinates of the equivalent atom in the next cell over in the *a* direction, however, are easily calculated as this atom is simply 1 unit cell away in *a*. Thus, all one has to do is add 1 to the x coordinate: (1.43, 0.60, 0.68). Such transformations can be performed regardless of the shape of the unit cell. Fractional coordinates, therefore, are used to retain and manipulate crystal information.

1.4.2.2.2 Crystal directions

The designation of the individual vectors within any given crystal lattice is accomplished by the use of whole number multipliers of the lattice parameter of the point at which the vector exits the unit cell. The vector is indicated by the notation [*hkl*], where *h*, *k*, and *l* are reciprocals of the point at which the vector exits the unit cell. The origination of all vectors is assumed defined as [000]. For example, the direction along the *a*-axis according to this scheme would be [100] because this has a component only in the *a*-direction and no component along either the *b* or *c* axial direction. A vector diagonally along the face defined by the *a* and *b* axis would be [110], while going from one corner of the unit cell to the opposite corner would be in the [111] direction. Figure 1.21 shows some examples of the various directions in the unit cell. The crystal direction notation is made up of the lowest combination of integers and represents unit distances rather than actual distances. A [222] direction is identical to a [111], so [111] is used. Fractions are not used. For example, a vector that intercepts the center of the top face of the unit cell has the coordinates *x* = 1/2, *y* = 1/2, *z* = 1. All have to be inversed to convert to the lowest combination of integers (whole numbers); i.e., [221] in Figure 1.21. Finally, all parallel vectors have the same crystal direction, e.g., the four vertical edges of the cell shown in Figure 1.21 all have the crystal direction [*hkl*] = [001].

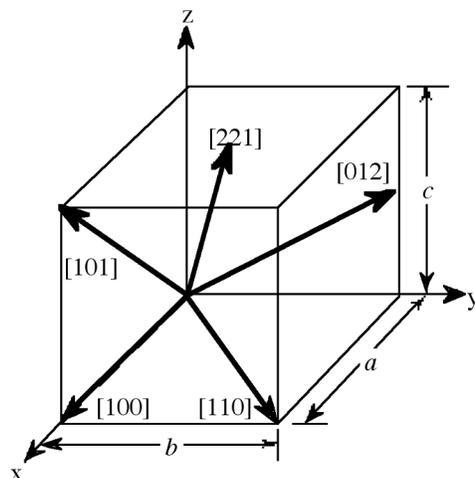


Figure 1.21: Some common directions in a cubic unit cell.

Crystal directions may be grouped in families. To avoid confusion there exists a convention in the choice of brackets surrounding the three numbers to differentiate a crystal direction from a family of direction. For a direction, square brackets $[hkl]$ are used to indicate an individual direction. Angle brackets $\langle hkl \rangle$ indicate a family of directions. A family of directions includes any directions that are equivalent in length and types of atoms encountered. For example, in a cubic lattice, the $[100]$, $[010]$, and $[001]$ directions all belong to the $\langle 100 \rangle$ family of planes because they are equivalent. If the cubic lattice were rotated 90° , the a , b , and c directions would remain indistinguishable, and there would be no way of telling on which crystallographic positions the atoms are situated, so the family of directions is the same. In a hexagonal crystal, however, this is not the case, so the $[100]$ and $[010]$ would both be $\langle 100 \rangle$ directions, but the $[001]$ direction would be distinct. Finally, negative directions are identified with a bar over the negative number instead of a minus sign.

1.4.2.2.3 Crystal planes

Planes in a crystal can be specified using a notation called Miller indices. The Miller index is indicated by the notation $[hkl]$ where h , k , and l are reciprocals of the plane with the x , y , and z axes. To obtain the Miller indices of a given plane requires the following steps:

- Step 1. The plane in question is placed on a unit cell.
- Step 2. Its intercepts with each of the crystal axes are then found.
- Step 3. The reciprocal of the intercepts are taken.
- Step 4. These are multiplied by a scalar to insure that is in the simple ratio of whole numbers.

For example, the face of a lattice that does not intersect the y or z axis would be (100) , while a plane along the body diagonal would be the (111) plane. An illustration of this along with the (111) and (110) planes is given in Figure 1.22.

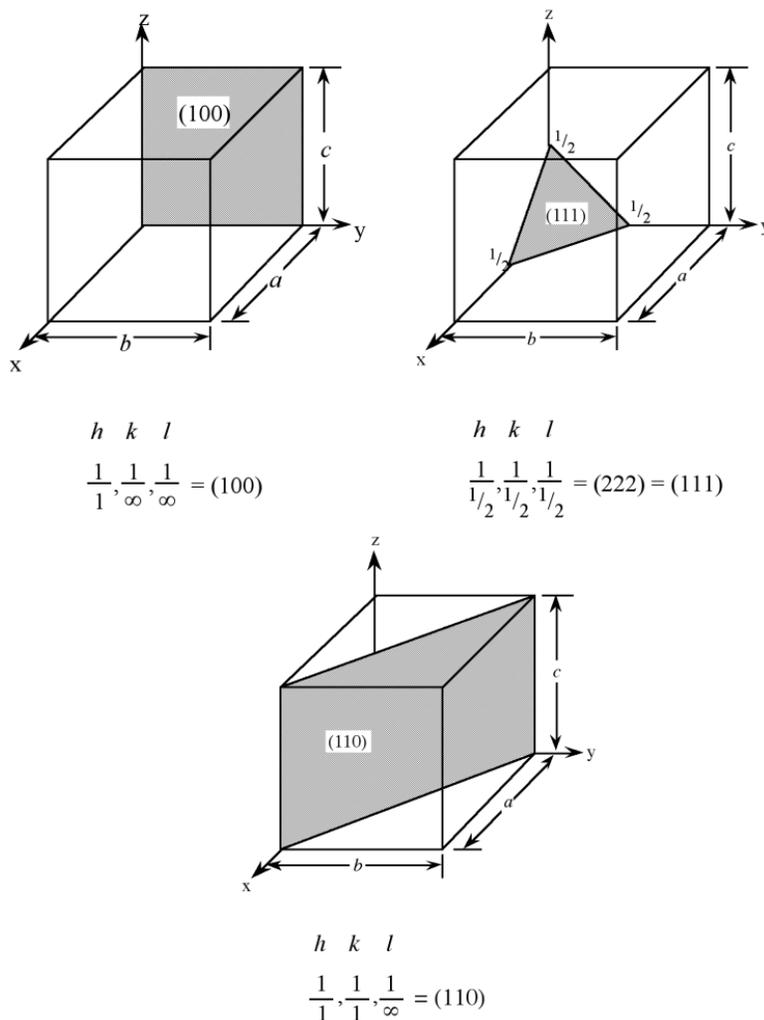


Figure 1.22: Examples of Miller indices notation for crystal planes.

As with crystal directions, Miller indices directions may be grouped in families. Individual Miller indices are given in parentheses (hkl), while braces $\{hkl\}$ are placed around the indices of a family of planes. For example, (001), (100), and (010) are all in the $\{100\}$ family of planes, for a cubic lattice.

1.4.3 Description of crystal structures

Crystal structures may be described in a number of ways. The most common manner is to refer to the size and shape of the unit cell and the positions of the atoms (or ions) within the cell. However, this information is sometimes insufficient to allow for an understanding of the true structure in three dimensions. Consideration of several unit cells, the arrangement of the atoms with respect to each other, the number of other atoms they in contact with, and the distances to neighboring atoms, often will provide a better understanding. A number of methods are available to describe extended solid-state structures. The most applicable with

regard to elemental and compound semiconductor, metals and the majority of insulators is the close packing approach.

1.4.3.1 Close packed structures: hexagonal close packing and cubic close packing

Many crystal structures can be described using the concept of close packing. This concept requires that the atoms (ions) are arranged so as to have the maximum density. In order to understand close packing in three dimensions, the most efficient way for equal sized spheres to be packed in two dimensions must be considered.

The most efficient way for equal sized spheres to be packed in two dimensions is shown in Figure 1.23, in which it can be seen that each sphere (the dark gray shaded sphere) is surrounded by, and is in contact with, six other spheres (the light gray spheres in Figure 1.23). It should be noted that contact with six other spheres the maximum possible is the spheres are the same size, although lower density packing is possible. Close packed layers are formed by repetition to an infinite sheet. Within these close packed layers, three close packed rows are present, shown by the dashed lines in Figure 1.23.

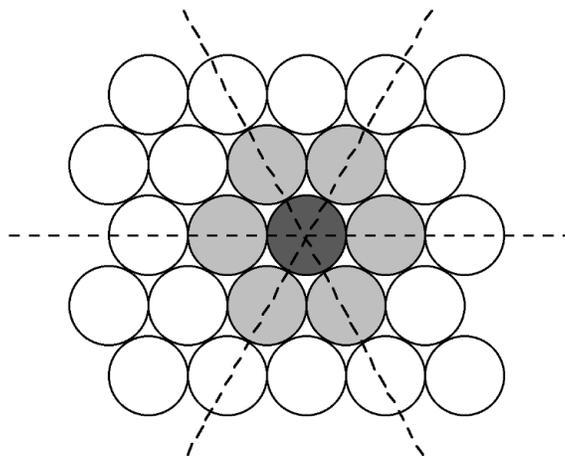


Figure 1.23: Schematic representation of a close packed layer of equal sized spheres. The close packed rows (directions) are shown by the dashed lines.

The most efficient way for equal sized spheres to be packed in three dimensions is to stack close packed layers on top of each other to give a close packed structure. There are two simple ways in which this can be done, resulting in either a hexagonal or cubic close packed structures.

1.4.3.1.1 Hexagonal close packed

If two close packed layers A and B are placed in contact with each other so as to maximize the density, then the spheres of layer B will rest in the hollow (vacancy) between three of the spheres in layer A. This is demonstrated in Figure 1.24. Atoms in the second layer, B (shaded light gray), may occupy one of two possible positions (Figure 1.24a or b) but not both together or a mixture of each. If a third layer is placed on top of layer B such that it exactly covers layer A, subsequent placement of layers will result in the following sequence ...ABABAB.... This is known as hexagonal close packing or *hcp*.

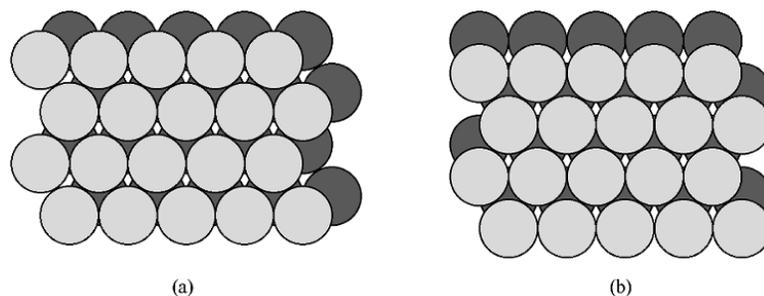


Figure 1.24: Schematic representation of two close packed layers arranged in A (dark grey) and B (light grey) positions. The alternative stacking of the B layer is shown in (a) and (b).

The hexagonal close packed cell is a derivative of the hexagonal Bravais lattice system (Figure 1.20) with the addition of an atom inside the unit cell at the coordinates $(\frac{1}{3}, \frac{2}{3}, \frac{1}{2})$. The basal plane of the unit cell coincides with the close packed layers (Figure 1.25). In other words the close packed layer makes-up the $\{001\}$ family of crystal planes.

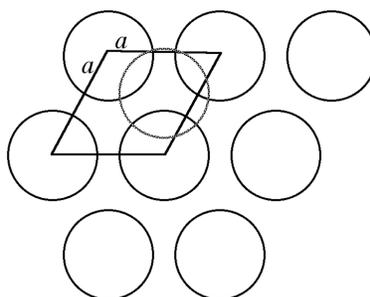


Figure 1.25: A schematic projection of the basal plane of the hcp unit cell on the close packed layers.

The “packing fraction” in a hexagonal close packed cell is 74.05%; that is 74.05% of the total volume is occupied. The packing fraction or density is derived by assuming that each atom is a hard sphere in contact with its nearest neighbors. Determination of the packing fraction is accomplished by calculating the number of whole spheres per unit cell (2 in hcp), the volume occupied by these spheres, and a comparison with the total volume of a unit cell. The number gives an idea of how “open” or filled a structure is. By comparison, the packing fraction for body-centered cubic (Figure 1.20) is 68% and for diamond cubic (an important semiconductor structure to be described later) is 34%.

1.4.3.1.2 Cubic close packed: face-centered cubic

In a similar manner to the generation of the hexagonal close packed structure, two close packed layers are stacked (Figure 1.23) however, the third layer (C) is placed such that it does not exactly cover layer A,

while sitting in a set of troughs in layer B (Figure 1.26), then upon repetition the packing sequence will be ...ABCABCABC.... This is known as cubic close packing or *ccp*.

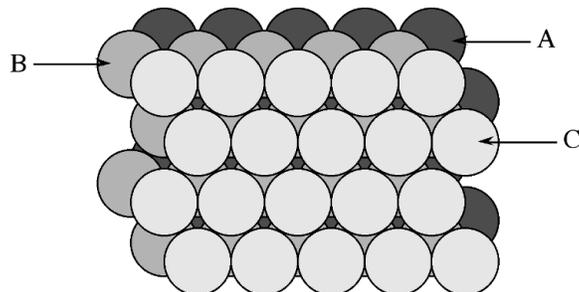


Figure 1.26: Schematic representation of the three close packed layers in a cubic close packed arrangement: A (dark grey), B (medium grey), and C (light grey).

The unit cell of cubic close packed structure is actually that of a face-centered cubic (*fcc*) Bravais lattice. In the *fcc* lattice the close packed layers constitute the $\{111\}$ planes. As with the *hcp* lattice packing fraction in a cubic close packed (*fcc*) cell is 74.05%. Since face centered cubic or *fcc* is more commonly used in preference to cubic close packed (*ccp*) in describing the structures, the former will be used throughout this text.

1.4.3.2 Coordination number

The coordination number of an atom or ion within an extended structure is defined as the number of nearest neighbor atoms (ions of opposite charge) that are in contact with it. A slightly different definition is often used for atoms within individual molecules: the number of donor atoms associated with the central atom or ion. However, this distinction is rather artificial, and both can be employed.

The coordination numbers for metal atoms in a molecule or complex are commonly 4, 5, and 6, but all values from 2 to 9 are known and a few examples of higher coordination numbers have been reported. In contrast, common coordination numbers in the solid state are 3, 4, 6, 8, and 12. For example, the atom in the center of body-centered cubic lattice has a coordination number of 8, because it touches the eight atoms at the corners of the unit cell, while an atom in a simple cubic structure would have a coordination number of 6. In both *fcc* and *hcp* lattices each of the atoms have a coordination number of 12.

1.4.3.3 Octahedral and tetrahedral vacancies

As was mentioned above, the packing fraction in both *fcc* and *hcp* cells is 74.05%, leaving 25.95% of the volume unfilled. The unfilled lattice sites (interstices) between the atoms in a cell are called interstitial sites or vacancies. The shape and relative size of these sites is important in controlling the position of additional atoms. In both *fcc* and *hcp* cells most of the space within these atoms lies within two different sites known as octahedral sites and tetrahedral sites. The difference between the two lies in their “coordination number”, or the number of atoms surrounding each site. Tetrahedral sites (vacancies) are surrounded by four atoms arranged at the corners of a tetrahedron. Similarly, octahedral sites are surrounded by six atoms which make-up the apices of an octahedron. For a given close packed lattice an octahedral vacancy will be larger than a tetrahedral vacancy.

Within a face centered cubic lattice, the eight tetrahedral sites are positioned within the cell, at the general fractional coordinate of $(n/4, n/4, n/4)$ where $n = 1$ or 3 , e.g., $(1/4, 1/4, 1/4)$, $(1/4, 1/4, 3/4)$, etc. The

octahedral sites are located at the center of the unit cell ($1/2, 1/2, 1/2$), as well as at each of the edges of the cell, e.g., ($1/2, 0, 0$). In the hexagonal close packed system, the tetrahedral sites are at ($0, 0, 3/8$) and ($1/3, 2/3, 7/8$), and the octahedral sites are at ($1/3, 1/3, 1/4$) and all symmetry equivalent positions.

1.4.3.4 Important structure types

The majority of crystalline materials do not have a structure that fits into the one atom per site simple Bravais lattice. A number of other important crystal structures are found, however, only a few of these crystal structures are those of which occur for the elemental and compound semiconductors and the majority of these are derived from *fcc* or *hcp* lattices. Each structural type is generally defined by an archetype, a material (often a naturally occurring mineral) which has the structure in question and to which all the similar materials are related. With regard to commonly used elemental and compound semiconductors the important structures are diamond, zinc blende, Wurtzite, and to a lesser extent chalcopyrite. However, rock salt, β -tin, cinnabar and cesium chloride are observed as high pressure or high temperature phases and are therefore also discussed. The following provides a summary of these structures. Details of the full range of solid-state structures are given elsewhere.

1.4.3.4.1 Diamond Cubic

The diamond cubic structure consists of two interpenetrating face-centered cubic lattices, with one offset $1/4$ of a cube along the cube diagonal. It may also be described as face centered cubic lattice in which half of the tetrahedral sites are filled while all the octahedral sites remain vacant. The diamond cubic unit cell is shown in Figure 1.27. Each of the atoms (e.g., C) is four coordinate, and the shortest interatomic distance (C-C) may be determined from the unit cell parameter (a).

$$\text{C-C} = \frac{a\sqrt{3}}{4} \approx 0.422 a \quad (1.25)$$

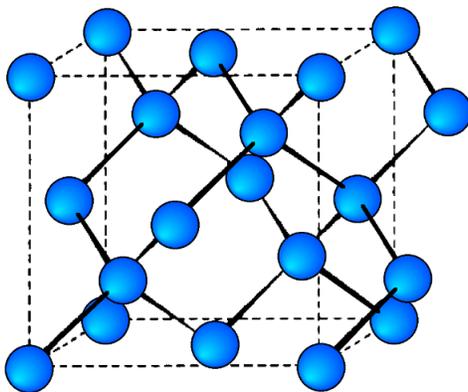


Figure 1.27: Unit cell structure of a diamond cubic lattice showing the two interpenetrating face-centered cubic lattices.

1.4.3.4.2 Zinc blende

This is a binary phase (ME) and is named after its archetype, a common mineral form of zinc sulfide (ZnS). As with the diamond lattice, zinc blende consists of the two interpenetrating *fcc* lattices. However, in zinc blende one lattice consists of one of the types of atoms (Zn in ZnS), and the other lattice is of the second type of atom (S in ZnS). It may also be described as face centered cubic lattice of S atoms in which half of the tetrahedral sites are filled with Zn atoms. All the atoms in a zinc blende structure are 4-coordinate. The zinc blende unit cell is shown in Figure 1.28. A number of inter-atomic distances may be calculated for any material with a zinc blende unit cell using the lattice parameter (a).

$$\text{Zn-S} = \frac{a\sqrt{3}}{4} \approx 0.422 a \quad (1.26)$$

$$\text{Zn-Zn} = \text{S-S} = \frac{a}{\sqrt{2}} \approx 0.707 a \quad (1.27)$$

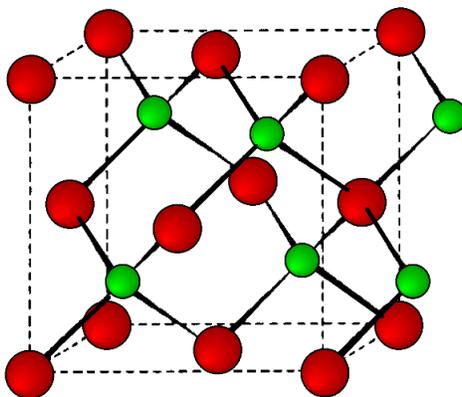


Figure 1.28: Unit cell structure of a zinc blende (ZnS) lattice. Zinc atoms are shown in green (small), sulfur atoms shown in red (large), and the dashed lines show the unit cell.

1.4.3.4.3 Chalcopyrite

The mineral chalcopyrite CuFeS_2 is the archetype of this structure. The structure is tetragonal ($a = b \neq c$, $\alpha = \beta = \gamma = 90^\circ$), and is essentially a superlattice on that of zinc blende. Thus, is easiest to imagine that the chalcopyrite lattice is made-up of a lattice of sulfur atoms in which the tetrahedral sites are filled in layers, ...FeCuCuFe..., etc. (Figure 1.29). In such an idealized structure $c = 2a$, however, this is not true of all materials with chalcopyrite structures.

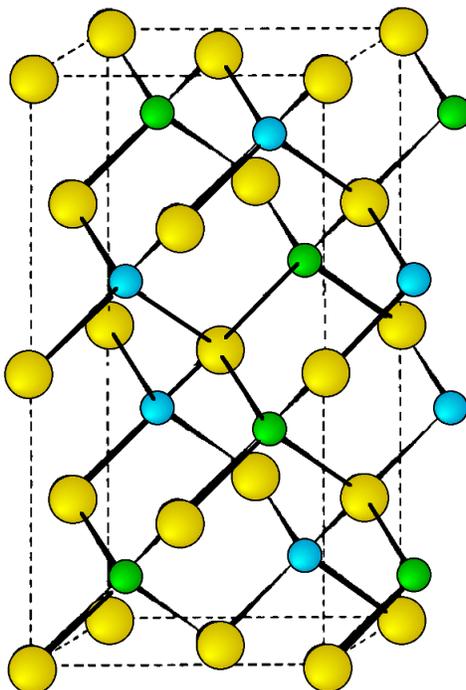


Figure 1.29: Unit cell structure of a chalcopyrite lattice. Copper atoms are shown in blue, iron atoms are shown in green and sulfur atoms are shown in yellow. The dashed lines show the unit cell.

1.4.3.4.4 Rock salt

As its name implies the archetypal rock salt structure is NaCl (table salt). In common with the zinc blende structure, rock salt consists of two interpenetrating face-centered cubic lattices. However, the second lattice is offset $1/2a$ along the unit cell axis. It may also be described as face centered cubic lattice in which all of the octahedral sites are filled, while all the tetrahedral sites remain vacant, and thus each of the atoms in the rock salt structure are 6-coordinate. The rock salt unit cell is shown in Figure 1.30. A number of inter-atomic distances may be calculated for any material with a rock salt structure using the lattice parameter (a).

$$\text{Na-Cl} = \frac{a}{2} \approx 0.5 a \quad (1.28)$$

$$\text{Na-Na} = \text{Cl-Cl} = \frac{a}{\sqrt{2}} \approx 0.707 a \quad (1.29)$$

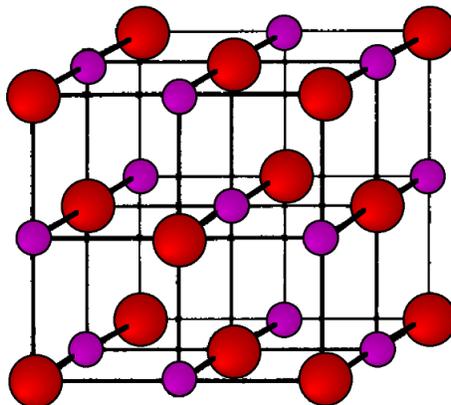


Figure 1.30: Unit cell structure of a rock salt lattice. Sodium ions are shown in purple (small spheres) and chloride ions are shown in red (large spheres).

1.4.3.4.5 Cinnabar

Cinnabar, named after the archetype mercury sulfide, HgS , is a distorted rock salt structure in which the resulting cell is rhombohedral (trigonal) with each atom having a coordination number of six.

1.4.3.4.6 Wurtzite

This is a hexagonal form of the zinc sulfide. It is identical in the number of and types of atoms, but it is built from two interpenetrating *hcp* lattices as opposed to the *fcc* lattices in zinc blende. As with zinc blende all the atoms in a wurtzite structure are 4-coordinate. The wurtzite unit cell is shown in Figure 1.31. A number of inter atomic distances may be calculated for any material with a wurtzite cell using the lattice parameter (a).

$$\text{Zn-S} = a\sqrt{3}/8 = 0.612a = \frac{3c}{8} = 0.375c \quad (1.30)$$

$$\text{Zn-Zn} = \text{S-S} = a = 1.632c \quad (1.31)$$

However, it should be noted that these formulae do not necessarily apply when the ratio a/c is different from the ideal value of 1.632.

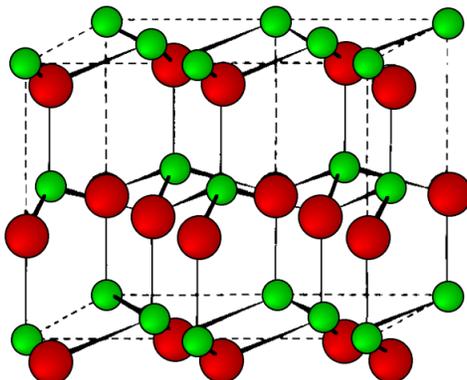


Figure 1.31: Unit cell structure of a wurtzite lattice. Zinc atoms are shown in green (small spheres), sulfur atoms shown in red (large spheres), and the dashed lines show the unit cell.

1.4.3.4.7 Cesium Chloride

The cesium chloride structure is found in materials with large cations and relatively small anions. It has a simple (primitive) cubic cell (Figure 1.20) with a chloride ion at the corners of the cube and the cesium ion at the body center. The coordination numbers of both Cs^+ and Cl^- , with the inner atomic distances determined from the cell lattice constant (a).

$$\text{Cs-Cl} = \frac{a\sqrt{3}}{2} \approx 0.866 a \quad (1.32)$$

$$\text{Cs-Cs} = \text{Cl-Cl} = a \quad (1.33)$$

1.4.3.4.8 β -Tin.

The room temperature allotrope of tin is β -tin or white tin. It has a tetragonal structure, in which each tin atom has four nearest neighbors ($\text{Sn-Sn} = 3.016 \text{ \AA}$) arranged in a very flattened tetrahedron, and two next nearest neighbors ($\text{Sn-Sn} = 3.175 \text{ \AA}$). The overall structure of β -tin consists of fused hexagons, each being linked to its neighbor via a four-membered Sn_4 ring.

1.4.4 Defects in crystalline solids

Up to this point we have only been concerned with ideal structures for crystalline solids in which each atom occupies a designated point in the crystal lattice. Unfortunately, defects ordinarily exist in equilibrium between the crystal lattice and its environment. These defects are of two general types: point defects and extended defects. As their names imply, point defects are associated with a single crystal lattice site, while extended defects occur over a greater range.

1.4.4.1 Point defects: “too many or too few” or “just plain wrong”

Point defects have a significant effect on the properties of a semiconductor, so it is important to understand the classes of point defects and the characteristics of each type. Figure 1.32 summarizes various classes of native point defects, however, they may be divided into two general classes; defects with the wrong number of atoms (deficiency or surplus) and defects where the identity of the atoms is incorrect.

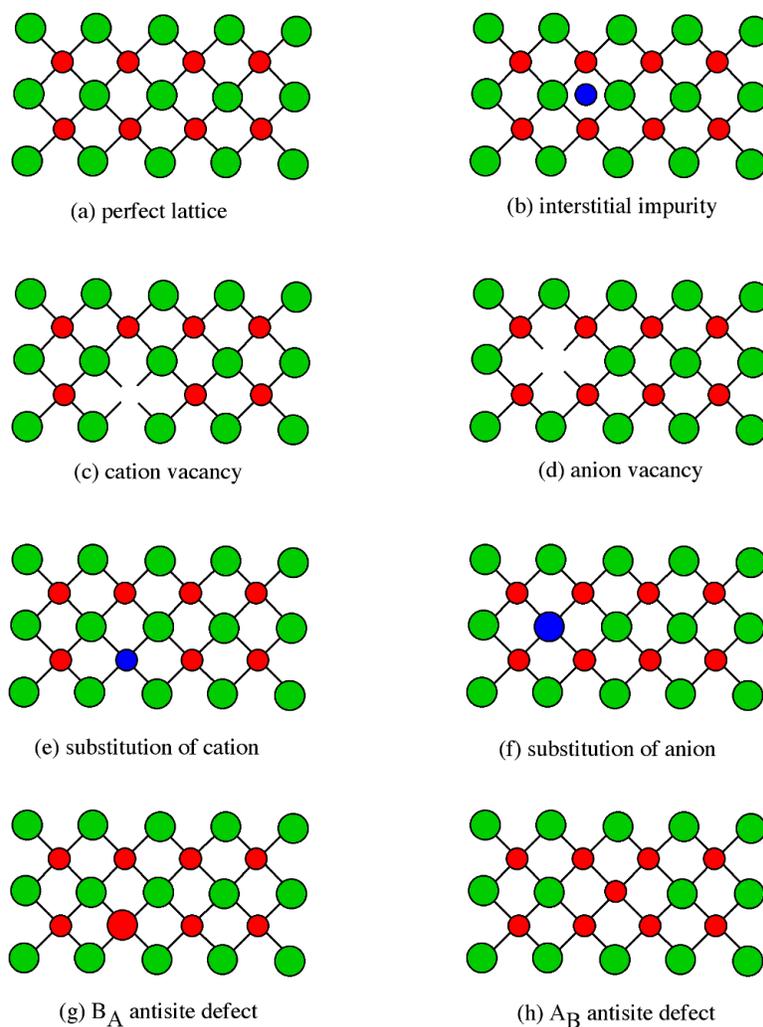


Figure 1.32: Point defects in a crystal lattice.

1.4.4.1.1 Interstitial Impurity

An interstitial impurity occurs when an extra atom is positioned in a lattice site that should be vacant in an ideal structure (Figure 1.32b). Since all the adjacent lattice sites are filled the additional atom will have to squeeze itself into the interstitial site, resulting in distortion of the lattice and alteration in the local

electronic behavior of the structure. Small atoms, such as carbon, will prefer to occupy these interstitial sites. Interstitial impurities readily diffuse through the lattice via interstitial diffusion, which can result in a change of the properties of a material as a function of time. Oxygen impurities in silicon generally are located as interstitials.

1.4.4.1.2 Vacancies

The converse of an interstitial impurity is when there are not enough atoms in a particular area of the lattice. These are called vacancies. Vacancies exist in any material above absolute zero and increase in concentration with temperature. In the case of compound semiconductors, vacancies can be either cation vacancies (Figure 1.32c) or anion vacancies (Figure 1.32d), depending on what type of atom are “missing”.

1.4.4.1.3 Substitution

Substitution of various atoms into the normal lattice structure is common, and used to change the electronic properties of both compound and elemental semiconductors. Any impurity element that is incorporated during crystal growth can occupy a lattice site. Depending on the impurity, substitution defects can greatly distort the lattice and/or alter the electronic structure. In general, cations will try to occupy cation lattice sites (Figure 1.32e), and anion will occupy the anion site (Figure 1.32f). For example, a zinc impurity in GaAs will occupy a gallium site, if possible, while a sulfur, selenium and tellurium atoms would all try to substitute for an arsenic. Some impurities will occupy either site indiscriminately, e.g., Si and Sn occupy both Ga and As sites in GaAs.

1.4.4.1.4 Antisite Defects

Antisite defects are a particular form of substitution defect, and are unique to compound semiconductors. An antisite defect occurs when a cation is misplaced on an anion lattice site or vice versa (Figure 1.32g and h). Depending on the arrangement these are designated as either A_B antisite defects or B_A antisite defects. For example, if an arsenic atom is on a gallium lattice site the defect would be an As_{Ga} defect. Antisite defects involve fitting into a lattice site atoms of a different size than the rest of the lattice, and therefore this often results in a localized distortion of the lattice. In addition, cations and anions will have a different number of electrons in their valence shells, so this substitution will alter the local electron concentration and the electronic properties of this area of the semiconductor.

1.4.4.2 Extended Defects: Dislocations in a Crystal Lattice

Extended defects may be created either during crystal growth or as a consequence of stress in the crystal lattice. The plastic deformation of crystalline solids does not occur such that all bonds along a plane are broken and reformed simultaneously. Instead, the deformation occurs through a dislocation in the crystal lattice. Figure 1.33 shows a schematic representation of a dislocation in a crystal lattice. Two features of this type of dislocation are the presence of an extra crystal plane, and a large void at the dislocation core. Impurities tend to segregate to the dislocation core in order to relieve strain from their presence.

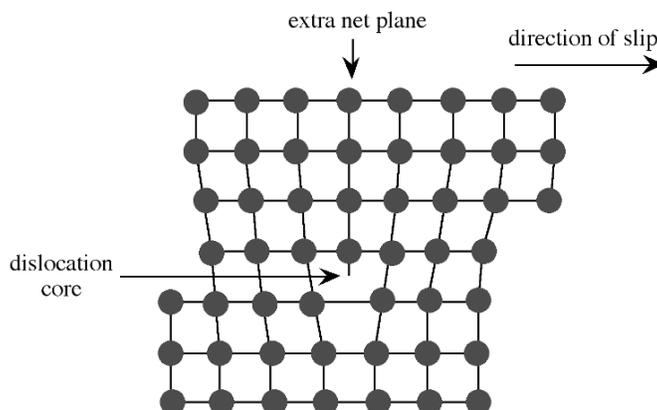


Figure 1.33: Dislocation in a crystal lattice.

1.4.5 Epitaxy

Epitaxy, is a transliteration of two Greek words *epi*, meaning "upon", and *taxis*, meaning "ordered". With respect to crystal growth it applies to the process of growing thin crystalline layers on a crystal substrate. In epitaxial growth, there is a precise crystal orientation of the film in relation to the substrate. The growth of epitaxial films can be done by a number of methods including molecular beam epitaxy, atomic layer epitaxy, and chemical vapor deposition, all of which will be described later.

Epitaxy of the same material, such as a gallium arsenide film on a gallium arsenide substrate, is called homoepitaxy, while epitaxy where the film and substrate material are different is called heteroepitaxy. Clearly, in homoepitaxy, the substrate and film will have the identical structure, however, in heteroepitaxy, it is important to employ where possible a substrate with the same structure and similar lattice parameters. For example, zinc selenide (zinc blende, $a = 5.668 \text{ \AA}$) is readily grown on gallium arsenide (zinc blende, $a = 5.653 \text{ \AA}$). Alternatively, epitaxial crystal growth can occur where there exists a simple relationship between the structures of the substrate and crystal layer, such as is observed between Al_2O_3 (100) on Si (100). Whichever route is chosen a close match in the lattice parameters is required, otherwise, the strains induced by the lattice mismatch results in distortion of the film and formation of dislocations. If the mismatch is significant epitaxial growth is not energetically favorable, causing a textured film or polycrystalline untextured film to be grown. As a general rule of thumb, epitaxy can be achieved if the lattice parameters of the two materials are within about 5% of each other. For good quality epitaxy, this should be less than 1%. The larger the mismatch, the larger the strain in the film. As the film gets thicker and thicker, it will try to relieve the strain in the film, which could include the loss of epitaxy or the growth of dislocations. It is important to note that the $\langle 100 \rangle$ directions of a film must be parallel to the $\langle 100 \rangle$ direction of the substrate. In some cases, such as Fe on MgO, the [111] direction is parallel to the substrate [100]. The epitaxial relationship is specified by giving first the plane in the film that is parallel to the substrate [100].

1.4.6 Bibliography

- *International Tables for X-ray Crystallography*. Vol. IV; Kynoch Press: Birmingham, UK (1974).
- B. F. G. Johnson, in *Comprehensive Inorganic Chemistry*, Pergamon Press, Vol. 4, Chapter 52 (1973).
- A. R. West, *Solid State Chemistry and its Applications*, Wiley, New York (1984).

1.5 Structures of Element and Compound Semiconductors⁵

1.5.1 Introduction

A single crystal of either an elemental (e.g., silicon) or compound (e.g., gallium arsenide) semiconductor forms the basis of almost all semiconductor devices. The ability to control the electronic and opto-electronic properties of these materials is based on an understanding of their structure. In addition, the metals and many of the insulators employed within a microelectronic device are also crystalline.

1.5.2 Group IV (14) elements

Each of the semiconducting phases of the group IV (14) elements, C (diamond), Si, Ge, and α -Sn, adopt the diamond cubic structure (Figure 1.34). Their lattice constants (a , Å) and densities (ρ , g/cm³) are given in Table 1.3.

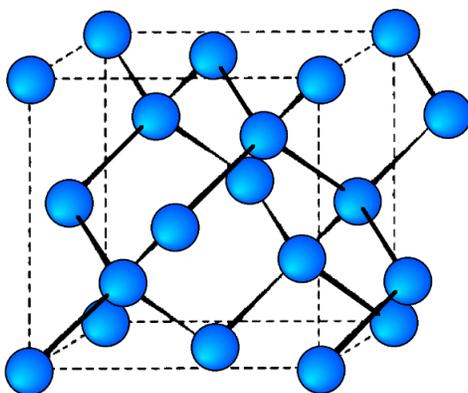


Figure 1.34: Unit cell structure of a diamond cubic lattice showing the two interpenetrating face-centered cubic lattices.

| Element | Lattice parameter, a (Å) | Density (g/cm ³) |
|---------------------|----------------------------|------------------------------|
| carbon (diamond) | 3.56683(1) | 3.51525 |
| silicon | 5.4310201(3) | 2.319002 |
| germanium | 5.657906(1) | 5.3234 |
| tin (α -Sn) | 6.4892(1) | 7.285 |

Table 1.3: Lattice parameters and densities (measured at 298 K) for the diamond cubic forms of the group IV (14) elements.

As would be expected the lattice parameter increase in the order $C < Si < Ge < \alpha$ -Sn. Silicon and germanium form a continuous series of solid solutions with gradually varying parameters. It is worth noting the high degree of accuracy that the lattice parameters are known for high purity crystals of these elements. In addition, it is important to note the temperature at which structural measurements are made, since the

⁵This content is available online at <<http://cnx.org/content/m23905/1.6/>>.

lattice parameters are temperature dependent (Figure 1.35). The lattice constant (a), in Å, for high purity silicon may be calculated for any temperature (T) over the temperature range 293 - 1073 K by the formula shown below.

$$a_T = 5.4304 + 1.8138 \times 10^{-5} (T - 298.15 \text{ K}) + 1.542 \times 10^{-9} (T - 298.15 \text{ K})^2$$

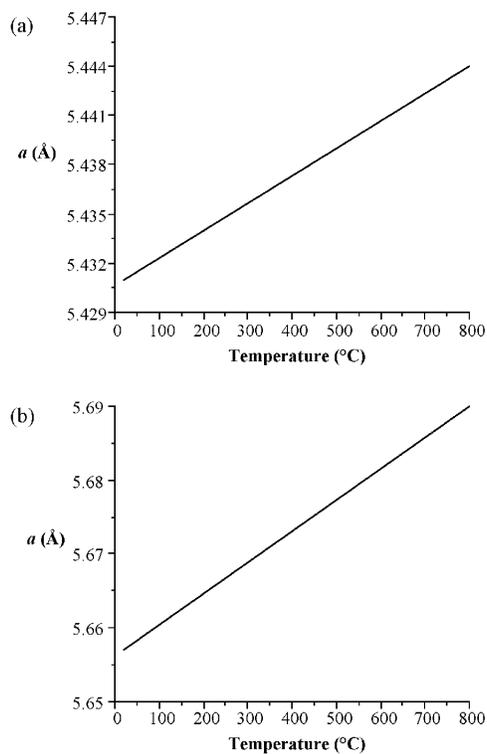


Figure 1.35: Temperature dependence of the lattice parameter for (a) Si and (b) Ge.

Even though the diamond cubic forms of Si and Ge are the only forms of direct interest to semiconductor devices, each exists in numerous crystalline high pressure and meta-stable forms. These are described along with their interconversions, in Table 1.4.

| Phase | Structure | Remarks |
|--------|------------------------|---|
| Si I | diamond cubic | stable at normal pressure |
| Si II | grey tin structure | formed from Si I or Si V above 14 GPa |
| Si III | cubic | metastable, formed from Si II above 10 GPa |
| Si IV | hexagonal | |
| Si V | unidentified | stable above 34 GPa, formed from Si II above 16 GPa |
| Si VI | hexagonal close packed | stable above 45 GPa |
| Ge I | diamond cubic | low-pressure phase |
| Ge II | β -tin structure | formed from Ge I above 10 GPa |
| Ge III | tetragonal | formed by quenching Ge II at low pressure |
| Ge IV | body centered cubic | formed by quenching Ge II to 1 atm at 200 K |

Table 1.4: High pressure and metastable phases of silicon and germanium.

1.5.3 Group III-V (13-15) compounds

The stable phases for the arsenides, phosphides and antimonides of aluminum, gallium and indium all exhibit zinc blende structures (Figure 1.36). In contrast, the nitrides are found as wurtzite structures (e.g., Figure 1.37). The structure, lattice parameters, and densities of the III-V compounds are given in Table 1.5. It is worth noting that contrary to expectation the lattice parameter of the gallium compounds is smaller than their aluminum homolog; for GaAs $a = 5.653 \text{ \AA}$; AlAs $a = 5.660 \text{ \AA}$. As with the group IV elements the lattice parameters are highly temperature dependent; however, additional variation arises from any deviation from absolute stoichiometry. These effects are shown in Figure 1.38.

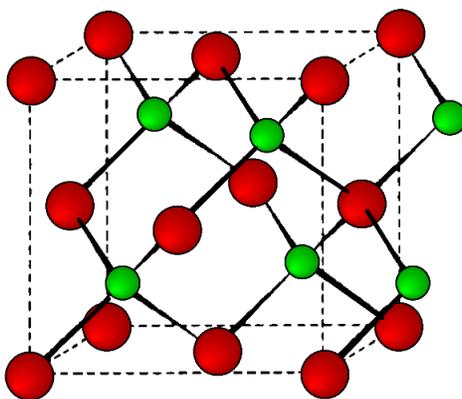


Figure 1.36: Unit cell structure of a zinc blende (ZnS) lattice. Zinc atoms are shown in green (small), sulfur atoms shown in red (large), and the dashed lines show the unit cell.

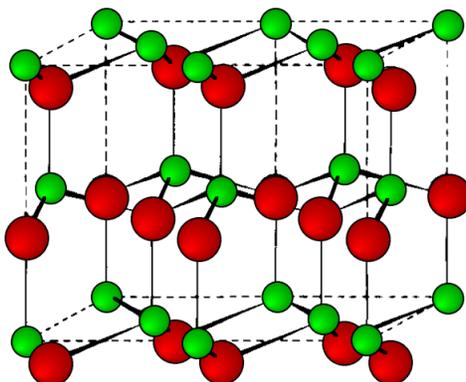


Figure 1.37: Unit cell structure of a wurtzite lattice. Zinc atoms are shown in green (small), sulfur atoms shown in red (large), and the dashed lines show the unit cell.

| Compound | Structure | Lattice parameter (Å) | Density (g/cm ³) |
|----------|-------------|----------------------------|------------------------------|
| AlN | wurtzite | $a = 3.11(1), c = 4.98(1)$ | 3.255 |
| AlP | zinc blende | $a = 5.4635(4)$ | 2.40(1) |
| AlAs | zinc blende | $a = 5.660$ | 3.760 |
| AlSb | zinc blende | $a = 6.1355(1)$ | 4.26 |
| GaN | wurtzite | $a = 3.190, c = 5.187$ | |
| GaP | zinc blende | $a = 5.4505(2)$ | 4.138 |
| GaAs | zinc blende | $a = 5.65325(2)$ | 5.3176(3) |
| InN | wurtzite | $a = 3.5446, c = 5.7034$ | 6.81 |
| InP | zinc blende | $a = 5.868(1)$ | 4.81 |
| InAs | zinc blende | $a = 6.0583$ | 5.667 |
| InSb | zinc blende | $a = 6.47937$ | 5.7747(4) |

Table 1.5: Lattice parameters and densities (measured at 298 K) for the III-V (13-15) compound semiconductors. Estimated standard deviations given in parentheses.

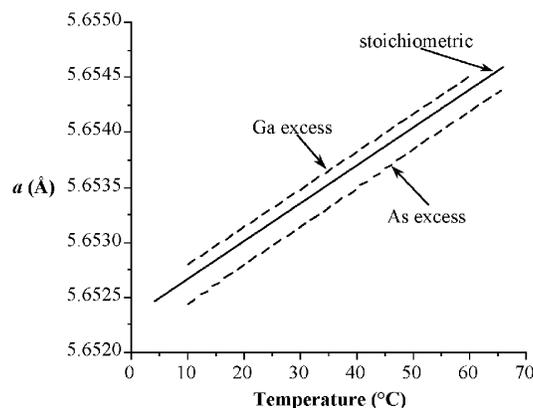


Figure 1.38: Temperature dependence of the lattice parameter for stoichiometric GaAs and crystals with either Ga or As excess.

The homogeneity of structures of alloys for a wide range of solid solutions to be formed between III-V compounds in almost any combination. Two classes of ternary alloys are formed: $\text{III}_x\text{-III}_{1-x}\text{-V}$ (e.g., $\text{Al}_x\text{-Ga}_{1-x}\text{-As}$) and $\text{III-V}_{1-x}\text{-V}_x$ (e.g., $\text{Ga-As}_{1-x}\text{-P}_x$). While quaternary alloys of the type $\text{III}_x\text{-III}_{1-x}\text{-V}_y\text{-V}_{1-y}$ allow for the growth of materials with similar lattice parameters, but a broad range of band gaps. A very important ternary alloy, especially in optoelectronic applications, is $\text{Al}_x\text{-Ga}_{1-x}\text{-As}$ and its lattice parameter (a) is directly related to the composition (x).

$$a = 5.6533 + 0.0078 x$$

Not all of the III-V compounds have well characterized high-pressure phases. However, in each case where a high-pressure phase is observed the coordination number of both the group III and group V element increases from four to six. Thus, AlP undergoes a zinc blende to rock salt transformation at high pressure above 170 kbar, while AlSb and GaAs form orthorhombic distorted rock salt structures above 77 and 172 kbar, respectively. An orthorhombic structure is proposed for the high-pressure form of InP (>133 kbar). Indium arsenide (InAs) undergoes two-phase transformations. The zinc blende structure is converted to a rock salt structure above 77 kbar, which in turn forms a β -tin structure above 170 kbar.

1.5.4 Group II-VI (12-16) compounds

The structures of the II-VI compound semiconductors are less predictable than those of the III-V compounds (above), and while zinc blende structure exists for almost all of the compounds there is a stronger tendency towards the hexagonal wurtzite form. In several cases the zinc blende structure is observed under ambient conditions, but may be converted to the wurtzite form upon heating. In general the wurtzite form predominates with the smaller anions (e.g., oxides), while the zinc blende becomes the more stable phase for the larger anions (e.g., tellurides). One exception is mercury sulfide (HgS) that is the archetype for the trigonal cinnabar phase. Table 1.6 lists the stable phase of the chalcogenides of zinc, cadmium and mercury, along with their high temperature phases where applicable. Solid solutions of the II-VI compounds are not as easily formed as for the III-V compounds; however, two important examples are $\text{ZnS}_x\text{Se}_{1-x}$ and $\text{Cd}_x\text{Hg}_{1-x}\text{Te}$.

| Compound | Structure | Lattice parameter (Å) | Density (g/cm ³) |
|----------|-------------|------------------------|------------------------------|
| ZnS | zinc blende | $a = 5.410$ | 4.075 |
| | wurtzite | $a = 3.822, c = 6.260$ | 4.087 |
| ZnSe | Zinc blende | $a = 5.668$ | 5.27 |
| ZnTe | Zinc blende | $a = 6.10$ | 5.636 |
| CdS | wurtzite | $a = 4.136, c = 6.714$ | 4.82 |
| CdSe | wurtzite | $a = 4.300, c = 7.011$ | 5.81 |
| CdTe | Zinc blende | $a = 6.482$ | 5.87 |
| HgS | cinnabar | $a = 4.149, c = 9.495$ | |
| | Zinc blende | $a = 5.851$ | 7.73 |
| HgSe | Zinc blende | $a = 6.085$ | 8.25 |
| HgTe | Zinc blende | $a = 6.46$ | 8.07 |

Table 1.6: Lattice parameters and densities (measured at 298 K) for the II-VI (12-16) compound semiconductors.

The zinc chalcogenides all transform to a cesium chloride structure under high pressures, while the cadmium compounds all form rock salt high-pressure phases (Figure 1.39). Mercury selenide (HgSe) and mercury telluride (HgTe) convert to the mercury sulfide archetype structure, cinnabar, at high pressure.

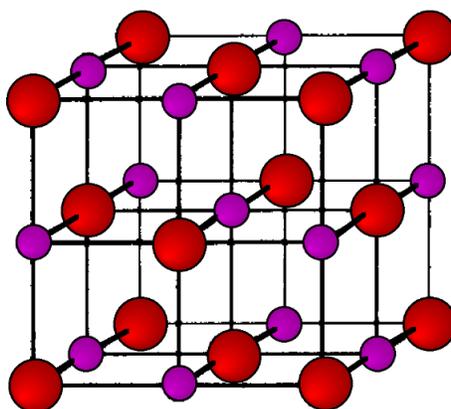


Figure 1.39: Unit cell structure of a rock salt lattice. Sodium ions are shown in purple and chloride ions are shown in red.

1.5.5 I-III-VI₂ (11-13-16) compounds

Nearly all I-III-VI₂ compounds at room temperature adopt the chalcopyrite structure (Figure 1.40). The cell constants and densities are given in Table 1.7. Although there are few reports of high temperature or

high-pressure phases, AgInS_2 has been shown to exist as a high temperature orthorhombic polymorph ($a = 6.954$, $b = 8.264$, and $c = 6.683 \text{ \AA}$), and AgInTe_2 forms a cubic phase at high pressures.

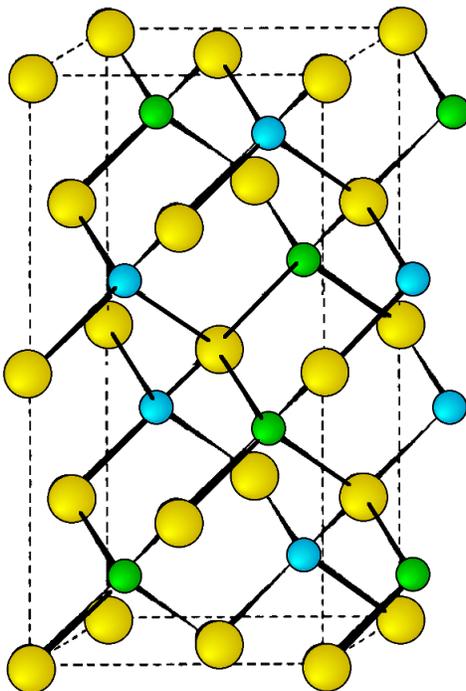


Figure 1.40: Unit cell structure of a chalcopyrite lattice. Copper atoms are shown in blue, iron atoms are shown in green and sulfur atoms are shown in yellow. The dashed lines show the unit cell.

| Compound | Lattice parameter a (Å) | Lattice parameter c (Å) | Density (g.cm ³) |
|---------------------|-------------------------|-------------------------|------------------------------|
| CuAlS ₂ | 5.32 | 10.430 | 3.45 |
| CuAlSe ₂ | 5.61 | 10.92 | 4.69 |
| CuAlTe ₂ | 5.96 | 11.77 | 5.47 |
| CuGaS ₂ | 5.35 | 10.46 | 4.38 |
| CuGaSe ₂ | 5.61 | 11.00 | 5.57 |
| CuGaTe ₂ | 6.00 | 11.93 | 5.95 |
| CuInS ₂ | 5.52 | 11.08 | 4.74 |
| CuInSe ₂ | 5.78 | 11.55 | 5.77 |
| CuInTe ₂ | 6.17 | 12.34 | 6.10 |
| AgAlS ₂ | 6.30 | 11.84 | 6.15 |
| AgGaS ₂ | 5.75 | 10.29 | 4.70 |
| AgGaSe ₂ | 5.98 | 10.88 | 5.70 |
| AgGaTe ₂ | 6.29 | 11.95 | 6.08 |
| AgInS ₂ | 5.82 | 11.17 | 4.97 |
| AgInSe ₂ | 6.095 | 11.69 | 5.82 |
| AgInTe ₂ | 6.43 | 12.59 | 6.96 |

Table 1.7: Chalcopyrite lattice parameters and densities (measured at 298 K) for the I-III-VI compound semiconductors. Lattice parameters for tetragonal cell.

Of the I-III-VI₂ compounds, the copper indium chalcogenides (CuInE₂) are certainly the most studied for their application in solar cells. One of the advantages of the copper indium chalcogenide compounds is the formation of solid solutions (alloys) of the formula CuInE_{2-x}E'_x, where the composition variable (x) varies from 0 to 2. The CuInS_{2-x}Se_x and CuInSe_{2-x}Te_x systems have also been examined, as has the CuGa_yIn_{1-y}S_{2-x}Se_x quaternary system. As would be expected from a consideration of the relative ionic radii of the chalcogenides the lattice parameters of the CuInS_{2-x}Se_x alloy should increase with increased selenium content. Vegard's law requires the lattice constant for a linear solution of two semiconductors to vary linearly with composition (e.g., as is observed for Al_xGa_{1-x}As), however, the variation of the tetragonal lattice constants (a and c) with composition for CuInS_{2-x}S_x are best described by the parabolic relationships.

$$a = 5.532 + 0.0801 x + 0.0260 x^2$$

$$c = 11.156 + 0.1204 x + 0.0611 x^2$$

A similar relationship is observed for the CuInSe_{2-x}Te_x alloys.

$$a = 5.783 + 0.1560 x + 0.0212 x^2$$

$$c = 11.628 + 0.3340 x + 0.0277 x^2$$

The large difference in ionic radii between S and Te (0.37 Å) prevents formation of solid solutions in the CuInS_{2-x}Te_x system, however, the single alloy CuInS_{1.5}Te_{0.5} has been reported.

1.5.6 Orientation effects

Once single crystals of high purity silicon or gallium arsenide are produced they are cut into wafers such that the exposed face of these wafers is either the crystallographic {100} or {111} planes. The relative structure of these surfaces are important with respect to oxidation, etching and thin film growth. These processes are orientation-sensitive; that is, they depend on the direction in which the crystal slice is cut.

1.5.6.1 Atom density and dangling bonds

The principle planes in a crystal may be differentiated in a number of ways, however, the atom and/or bond density are useful in predicting much of the chemistry of semiconductor surfaces. Since both silicon and gallium arsenide are *fcc* structures and the $\{100\}$ and $\{111\}$ are the only technologically relevant surfaces, discussions will be limited to *fcc* $\{100\}$ and $\{111\}$.

The atom density of a surface may be defined as the number of atoms per unit area. Figure 1.41 shows a schematic view of the $\{111\}$ and $\{100\}$ planes in a *fcc* lattice. The $\{111\}$ plane consists of a hexagonal close packed array in which the crystal directions within the plane are oriented at 60° to each other. The hexagonal packing and the orientation of the crystal directions are indicated in Figure 1.41b as an overlaid hexagon. Given the intra-planar inter-atomic distance may be defined as a function of the lattice parameter, the area of this hexagon may be readily calculated. For example in the case of silicon, the hexagon has an area of 38.30 \AA^2 . The number of atoms within the hexagon is three: the atom in the center plus $1/3$ of each of the six atoms at the vertices of the hexagon (each of the atoms at the hexagons vertices is shared by three other adjacent hexagons). Thus, the atom density of the $\{111\}$ plane is calculated to be 0.0783 \AA^{-2} . Similarly, the atom density of the $\{100\}$ plane may be calculated. The $\{100\}$ plane consists of a square array in which the crystal directions within the plane are oriented at 90° to each other. Since the square is coincident with one of the faces of the unit cell the area of the square may be readily calculated. For example in the case of silicon, the square has an area of 29.49 \AA^2 . The number of atoms within the square is 2: the atom in the center plus $1/4$ of each of the four atoms at the vertices of the square (each of the atoms at the corners of the square are shared by four other adjacent squares). Thus, the atom density of the $\{100\}$ plane is calculated to be 0.0678 \AA^{-2} . While these values for the atom density are specific for silicon, their ratio is constant for all diamond cubic and zinc blende structures: $\{100\}:\{111\} = 1:1.155$. In general, the fewer dangling bonds the more stable a surface structure.

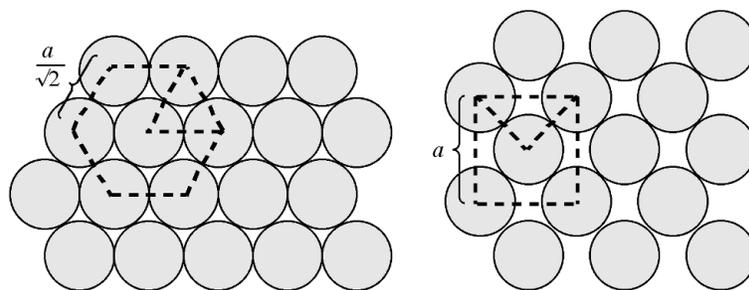


Figure 1.41: Schematic representation of the $\{111\}$ and $\{100\}$ faces of a face centered cubic (*fcc*) lattice showing the relationship between the close packed rows.

An atom inside a crystal of any material will have a coordination number (n) determined by the structure of the material. For example, all atoms within the bulk of a silicon crystal will be in a tetrahedral four-coordinate environment ($n = 4$). However, at the surface of a crystal the atoms will not make their full complement of bonds. Each atom will therefore have less nearest neighbors than an atom within the bulk of the material. The missing bonds are commonly called dangling bonds. While this description is not particularly accurate it is, however, widely employed and as such will be used herein. The number of dangling bonds may be defined as the difference between the ideal coordination number (determined by the bulk crystal structure) and the actual coordination number as observed at the surface.

Figure 1.42 shows a section of the $\{111\}$ surfaces of a diamond cubic lattice viewed perpendicular to the

{111} plane. The atoms within the bulk have a coordination number of four. In contrast, the atoms at the surface (e.g., the atom shown in blue in Figure 1.42) are each bonded to just three other atoms (the atoms shown in red in Figure 1.42), thus each surface atom has one dangling bond. As can be seen from Figure 1.43, which shows the atoms at the {100} surface viewed perpendicular to the {100} plane, each atom at the surface (e.g., the atom shown in blue in Figure 1.43) is only coordinated to two other atoms (the atoms shown in red in Figure 1.43), leaving two dangling bonds per atom. It should be noted that the same number of dangling bonds are found for the {111} and {100} planes of a zinc blende lattice. The ratio of dangling bonds for the {100} and {111} planes of all diamond cubic and zinc blende structures is $\{100\}:\{111\} = 2:1$. Furthermore, since the atom densities of each plane are known then the ratio of the dangling bond densities is determined to be: $\{100\}:\{111\} = 1:0.577$.

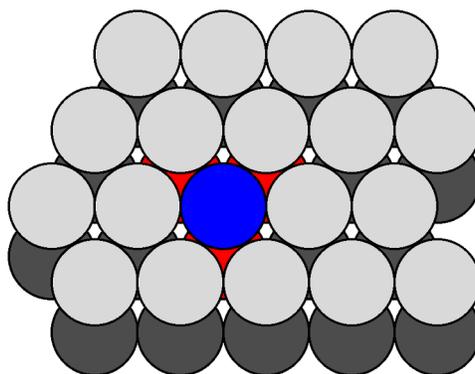


Figure 1.42: A section of the {111} surfaces of a diamond cubic lattice viewed perpendicular to the {111} plane.

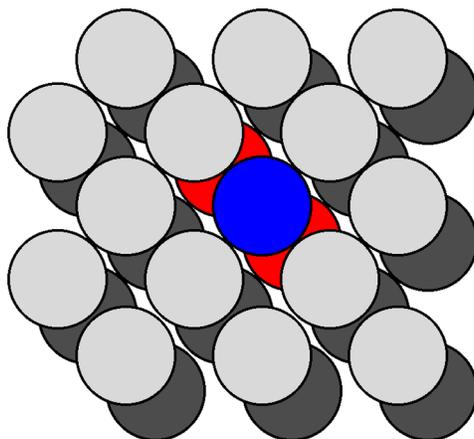


Figure 1.43: A section of the $\{100\}$ surface of a diamond cubic lattice viewed perpendicular to the $\{100\}$ plane.

1.5.6.2 Silicon

For silicon, the $\{111\}$ planes are closer packed than the $\{100\}$ planes. As a result, growth of a silicon crystal is therefore slowest in the $\langle 111 \rangle$ direction, since it requires laying down a close packed atomic layer upon another layer in its closest packed form. As a consequence $\langle 111 \rangle$ Si is the easiest to grow, and therefore the least expensive.

The dissolution or etching of a crystal is related to the number of broken bonds already present at the surface: the fewer bonds to be broken in order to remove an individual atom from a crystal, the easier it will be to dissolve the crystal. As a consequence of having only one dangling bond (requiring three bonds to be broken) etching silicon is slowest in the $\langle 111 \rangle$ direction. The electronic properties of a silicon wafer are also related to the number of dangling bonds.

Silicon microcircuits are generally formed on a single crystal wafer that is diced after fabrication by either sawing part way through the wafer thickness or scoring (scribing) the surface, and then physically breaking. The physical breakage of the wafer occurs along the natural cleavage planes, which in the case of silicon are the $\{111\}$ planes.

1.5.6.3 Gallium arsenide

The zinc blende lattice observed for gallium arsenide results in additional considerations over that of silicon. Although the $\{100\}$ plane of GaAs is structurally similar to that of silicon, two possibilities exist: a face consisting of either all gallium atoms or all arsenic atoms. In either case the surface atoms have two dangling bonds, and the properties of the face are independent of whether the face is gallium or arsenic.

The $\{111\}$ plane also has the possibility of consisting of all gallium or all arsenic. However, unlike the $\{100\}$ planes there is a significant difference between the two possibilities. Figure 1.36 shows the gallium arsenide structure represented by two interpenetrating *fcc* lattices. The $[111]$ axis is vertical within the plane of the page. Although the structure consists of alternate layers of gallium and arsenic stacked along the $[111]$ axis, the distance between the successive layers alternates between large and small. Assigning arsenic as the parent lattice the order of the layers in the $[111]$ direction is As-Ga-As-Ga-As-Ga, while in the $\left[\begin{smallmatrix} - \\ - \\ - \end{smallmatrix} \right]$

direction the layers are ordered, Ga-As-Ga-As-Ga-As (Figure 1.44). In silicon these two directions are of course identical. The surface of a crystal would be either arsenic, with three dangling bonds, or gallium, with one dangling bond. Clearly, the latter is energetically more favorable. Thus, the (111) plane shown in Figure 1.44 is called the (111) Ga face. Conversely, the $\left[\bar{1}\bar{1}\bar{1}\right]$ plane would be either gallium, with three dangling bonds, or arsenic, with one dangling bond. Again, the latter is energetically more favorable and the $\left[\bar{1}\bar{1}\bar{1}\right]$ plane is therefore called the (111) As face.

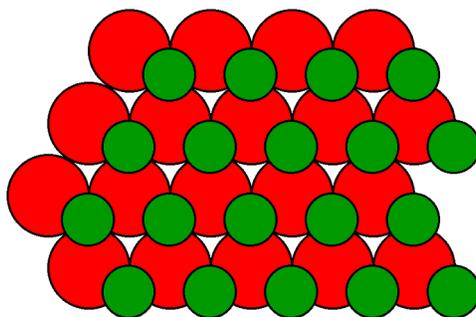


Figure 1.44: The (111) Ga face of GaAs showing a surface layer containing gallium atoms (green) with one dangling bond per gallium and three bonds to the arsenic atoms (red) in the lower layer.

The (111) As is distinct from that of (111) Ga due to the difference in the number of electrons at the surface. As a consequence, the (111) As face etches more rapidly than the (111) Ga face. In addition, surface evaporation below 770 °C occurs more rapidly at the (111) As face.

1.5.7 Bibliography

- M. Baublitz and A. L. Ruoff, *J. Appl. Phys.*, 1982, **53**, 6179.
- J. C. Jamieson, *Science*, 1963, **139**, 845.
- C. C. Landry, J. Lockwood, and A. R. Barron, *Chem. Mater.*, 1995, **7**, 699.
- M. Robbins, J. C. Phillips, and V. G. Lambrecht, *J. Phys. Chem. Solids*, 1973, **34**, 1205.
- D. Sridevi and K. V. Reddy, *Mat. Res. Bull.*, 1985, **20**, 929.
- Y. K. Vohra, S. T. Weir, and A. L. Ruoff, *Phys. Rev. B*, 1985, **31**, 7344.
- W. M. Yin and R. J. Paff, *J. Appl. Phys.*, 1973, **45**, 1456.

