Chapter 2

Device Fundamentals

2.1 Introduction to Bipolar Transistors[1]

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

Let's leave the world of two terminal devices (which are all called diodes by the way; diode just means two-terminals) and venture into the much more interesting world of three terminals. The first device we will look at is called the *bipolar transistor*. Consider the structure shown in Figure 2.1:

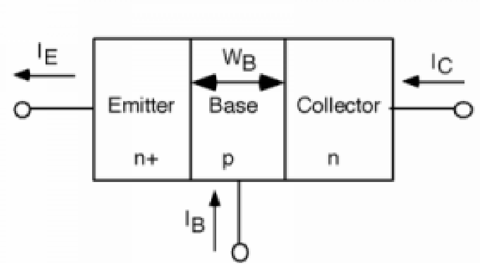


Figure 2.1: Structure of a npn bipolar transistor.

The device consists of three layers of silicon, a heavily doped n-type layer called the emitter, a moderately doped p-type layer called the base, and third, more lightly doped layer called the collector. In a biasing (applied DC potential) configuration called *forward active biasing*, the emitter-base junction is forward biased, and the base-collector junction is reverse biased. Figure 2.2 shows the biasing conventions we will use. Both bias voltages are referenced to the base terminal. Since the base-emitter junction is forward biased, and since the base is made of p-type material, V_{EB} must be negative. On the other hand, in order to reverse bias the base-collector junction V_{CB} will be a positive voltage.

---

[1] This content is available online at <http://cnx.org/content/m33798/1.2/>.

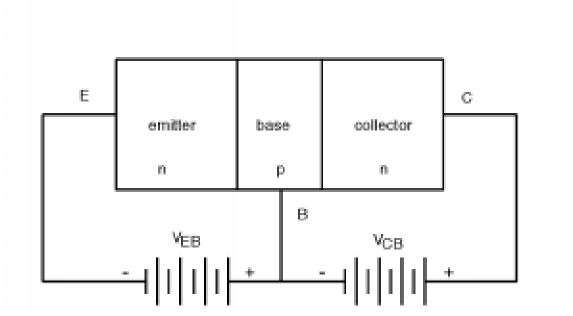


Figure 2.2: Forward active biasing of a npn bipolar transistor.

Now, let's draw the band-diagram for this device. At first this might seem hard to do, but we know what forward and reverse biased band diagrams look like, so we'll just stick one of each together. We show this in Figure 2.3, which is a very busy figure, but it is also very important, because it shows all of the important features in the operation the transistor. Since the base-emitter junction is forward biased, electrons will go from the (n-type) emitter into the base. Likewise, some holes from the base will be injected into the emitter.

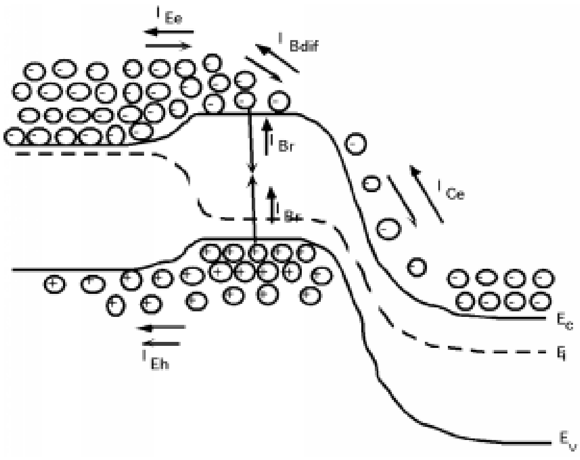


Figure 2.3: Band diagram and carrier fluxes in a bipolar transistor.

In Figure 2.3, we have two different kinds of arrows. The open arrows which are attached to the carriers, show us which way the carrier is moving. The solid arrows which are labeled with some kind of subscripted I, represent current flow. We need to do this because for holes, motion and current flow are in the same direction, while for electrons, carrier motion and current flow are in opposite directions.

Just as we saw in the last chapter, the electrons which are injected into the base diffuse away from the emitter-base junction towards the (reverse biased) base-collector junction. As they move through the base, some of the electrons encounter holes and recombine with them. Those electrons which do get to the base-collector junction run into a large electric field which sweeps them out of the base and into the collector. They "fall" down the large potential drop at the junction.

These effects are all seen in Figure 2.3, with arrows representing the various currents which are associated with each of the carriers fluxes. I_{Ee} represents the current associated with the electron injection into the base, i.e., it points in the opposite direction from the motion of the electrons, since electrons have a negative charge. I_{Eh} represents the current associated with holes injection into the emitter from the base. I_{Br} represents recombination current in the base, while I_{Ce} represents the electron current going into the collector. It should be easy for you to see that:

$$I_E = I_{\mathrm{Ee}} + I_{\mathrm{Eh}} \tag{2.1}$$

$$I_B = I_{\mathrm{Eh}} + I_{Br} \tag{2.2}$$

$$I_C = I_{Ce} \tag{2.3}$$

In a "good" transistor, almost all of the current across the base-emitter junction consists of electrons being injected into the base. The transistor engineer works hard to design the device so that very little emitter current is made up of holes coming from the base into the emitter. The transistor is also designed so that almost all of those electrons which are injected into the base make it across to the base-collector reverse-biased junction. Some recombination is unavoidable, but things are arranged so as to minimize this effect.

2.2 Basic MOS Structure[2]

NOTE: This module is adapted from the Connexions module entitled *Basic MOS Structure* by Bill Wilson.

Figure 2.4 shows the basic steps necessary to make the MOS structure. It will help us in our understanding if we now rotate our picture so that it is pointing sideways in our next few drawings. Figure 2.5 shows the rotated structure. Note that in the p-silicon we have positively charged mobile holes, and negatively charged, fixed acceptors. Because we will need it later, we have also shown the band diagram for the semiconductor below the sketch of the device. Note that since the substrate is p-type, the Fermi level is located down close to the valance band.

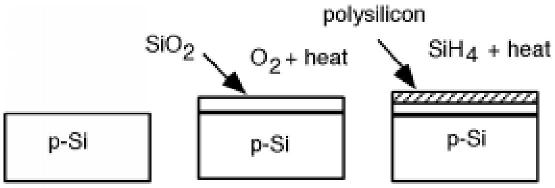


Figure 2.4: Formation of the metal-oxide-semiconductor (MOS) structure.

---

[2] This content is available online at <http://cnx.org/content/m33799/1.1/>.

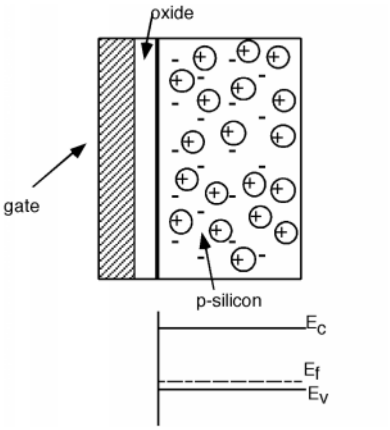


Figure 2.5: Basic metal-oxide-semiconductor (MOS) structure.

Let us now place a potential between the gate and the silicon substrate. Suppose we make the gate negative with respect to the substrate. Since the substrate is p-type, it has a lot of mobile, positively charged holes in it. Some of them will be attracted to the negative charge on the gate, and move over to the surface of the substrate. This is also reflected in the band diagram shown in Figure 2.6. Remember that the density of holes is exponentially proportional to how close the Fermi level is to the valence band edge. We see that the band diagram has been bent up slightly near the surface to reflect the extra holes which have accumulated there.

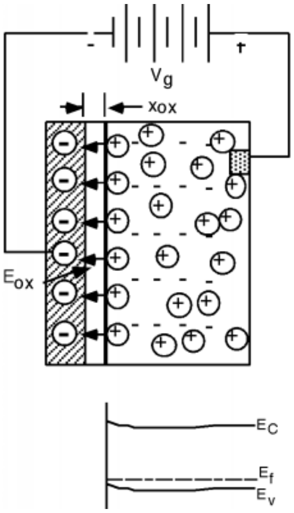


Figure 2.6: Applying a negative gate voltage to a basic metal-oxide-semiconductor (MOS) structure.

An electric field will develop between the positive holes and the negative gate charge. Note that the gate and the substrate form a kind of parallel plate capacitor, with the oxide acting as the insulating layer in-between them. The oxide is quite thin compared to the area of the device, and so it is quite appropriate to assume that the electric field inside the oxide is a uniform one. (We will ignore fringing at the edges.) The integral of the electric field is just the applied gate voltage V_g. If the oxide has a thickness x_{ox} then since E_{ox} is uniform, it is given by, (2.4).

$$E_{\text{ox}} = \frac{V_g}{x_{\text{ox}}} \tag{2.4}$$

If we focus in on a small part of the gate, we can make a little "pill" box which extends from somewhere in the oxide, across the oxide/gate interface and ends up inside the gate material someplace. The pill-box will have an area Δs. Now we will invoke Gauss' law which we reviewed earlier. Gauss' law simply says that the surface integral over a closed surface of the displacement vector D (which is, of course, ε x E) is equal to the total charge enclosed by that surface. We will assume that there is a surface charge density -Q_g Coulombs/cm^2 on the surface of the gate electrode (Figure 2.7). The integral form of Gauss' Law is just:

$$\oint \epsilon_{ox} E dS = Q_{\text{encl}} \tag{2.5}$$

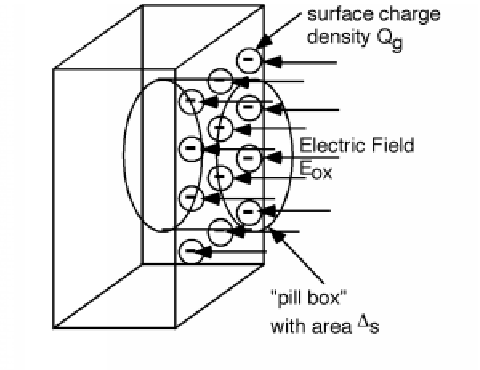


Figure 2.7: Finding the surface charge density.

Note that we have used $\varepsilon_{\text{ox}}E$ in place of D. In this particular set-up the integral is easy to perform, since the electric field is uniform, and only pointing in through one surface - it terminates on the negative surface charge inside the pill-box. The charge enclosed in the pill box is just -($Qg\Delta s$), and so we have (keeping in mind that the surface integral of a vector pointing into the surface is negative), (2.6), or (2.7).

$$\begin{aligned} \oint \epsilon_{ox} E dS &= -\left(\epsilon_{ox} E_{ox} \Delta\left(s\right)\right) \\ &= -\left(Q_g \Delta\left(s\right)\right) \end{aligned} \tag{2.6}$$

$$\epsilon_{\text{ox}} E_{\text{ox}} = Q_g \tag{2.7}$$

Now, we can use (2.4) to get (2.8) or (2.9).

$$\frac{\epsilon_{\text{ox}} V_g}{x_{\text{ox}}} = Q_g \tag{2.8}$$

$$\frac{Q_g}{V_g} = \frac{\epsilon_{\text{ox}}}{x_{\text{ox}}} \equiv c_{\text{ox}} \tag{2.9}$$

The quantity c_{ox} is called the oxide capacitance. It has units of Farads/cm^2, so it is really a capacitance per unit area of the oxide. The dielectric constant of silicon dioxide, ε_{ox}, is about 3.3 x 10^{-13} F/cm. A typical oxide thickness might be 250 Å (or 2.5 x 10^{-6} cm). In this case, c_{ox} would be about 1.30 x 10^{-7} F/cm^2. The units we are using here, while they might seem a little arbitrary and confusing, are the ones most commonly used in the semiconductor business.

The most useful form of (2.9) is when it is turned around, (2.10), as it gives us a way to find the charge on the gate in terms of the gate potential. We will use this equation later in our development of how the MOS transistor really works.

$$Q_g = c_{\text{ox}} V_g \tag{2.10}$$

It turns out we have not done anything very useful by apply a negative voltage to the gate. We have drawn more holes there in what is called an accumulation layer, but that is not helping us in our effort to create a layer of electrons in the MOSFET which could electrically connect the two n-regions together.

Let's turn the battery around and apply a positive voltage to the gate (Figure 2.8). Actually, let's take the battery out for now, and just let V_g be a positive value, relative to the substrate which will tie to ground. Making V_g positive puts positive Q_g on the gate. The positive charge pushes the holes away from the region under the gate and uncovers some of the negatively-charged fixed acceptors. Now the electric field points the other way, and goes from the positive gate charge, terminating on the negative acceptor charge within the silicon.

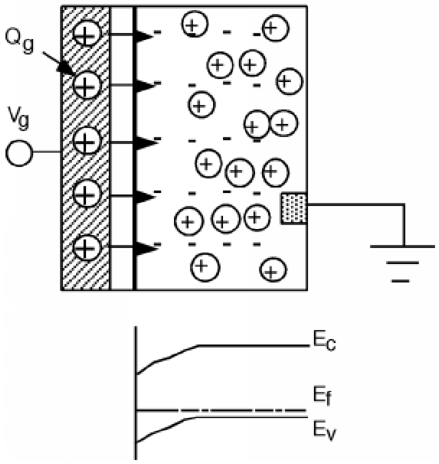


Figure 2.8: Increasing the voltage extends the depletion region further into the device.

The electric field now extends into the semiconductor. We know from our experience with the p-n junction that when there is an electric field, there is a shift in potential, which is represented in the band diagram by bending the bands. Bending the bands down (as we should moving towards positive charge) causes the valence band to pull away from the Fermi level near the surface of the semiconductor. If you remember the expression we had for the density of holes in terms of E_v and E_f it is easy to see that indeed, (2.11), there is a depletion region (region with almost no holes) near the region under the gate. (Once E_f - E_v gets large with respect to kT, the negative exponent causes $p \to 0$.)

$$p = N_v e^{-\frac{E_f - E_v}{kT}} \tag{2.11}$$

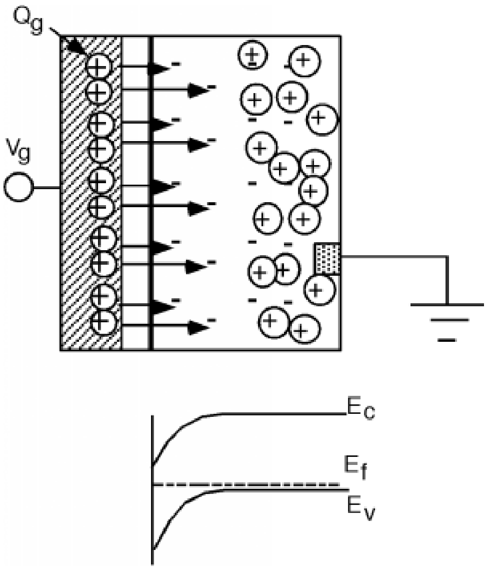


Figure 2.9: Threshold, E_f is getting close to E_c.

The electric field extends further into the semiconductor, as more negative charge is uncovered and the bands bend further down. But now we have to recall the electron density equation, which tells us how many electrons we have:

$$n = N_c e^{-\frac{E_c - E_f}{kT}} \tag{2.12}$$

A glance at Figure 2.9 reveals that with this much band bending, E_c the conduction band edge, and E_f the Fermi level are starting to get close to one another (at least compared to kT), which means that n, the electron concentration, should soon start to become significant. In the situation represented by Figure 2.9, we say we are at *threshold*, and the gate voltage at this point is called the *threshold voltage*, V_T.

Now, let's increase V_g above V_T. Here's the sketch in Figure 2.10. Even though we have increased V_g beyond the threshold voltage, V_T, and more positive charge appears on the gate, the depletion region no longer moves back into the substrate. Instead electrons start to appear under the gate region, and the additional electric field lines terminate on these new electrons, instead of on additional acceptors. We have created an *inversion layer* of electrons under the gate, and it is this layer of electrons which we can use to connect the two n-type regions in our initial device.

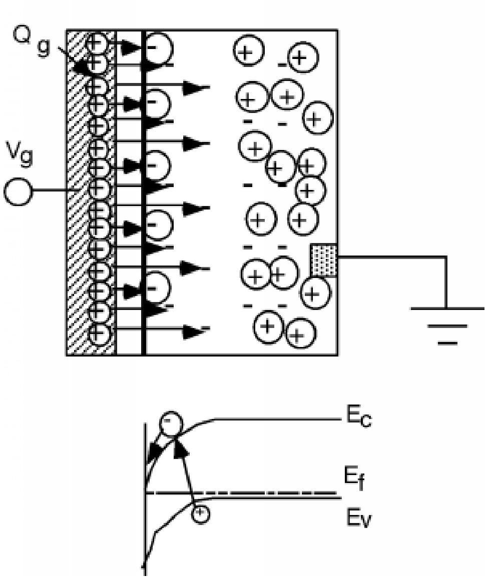


Figure 2.10: Inversion - electrons form under the gate.

Where did these electrons come from? We do not have any donors in this material, so they can not come from there. The only place from which electrons could be found would be through thermal generation. Remember, in a semiconductor, there are always a few electron hole pairs being generated by thermal excitation at any given time. Electrons that get created in the depletion region are caught by the electric field and are swept over to the edge by the gate. I have tried to suggest this with the electron generation event shown in the band diagram in the figure. In a real MOS device, we have the two n-regions, and it is easy for electrons from one or both to "fall" into the potential well under the gate, and create the inversion layer of electrons.

2.3 Introduction to the MOS Transistor and MOSFETs[3]

> NOTE: This module is adapted from the Connexions modules entitled *Introduction to MOSFETs* and *MOS Transistor* by Bill Wilson.

We now move on to another three terminal device - also called a transistor. This transistor, however, works on much different principles than does the bipolar junction transistor of the last chapter. We will now focus on a device called the *field effect transistor*, or *metal-oxide-semiconductor field effect transistor* or simply MOSFET.

In Figure 2.11 we have a block of silicon, doped p-type. Into it we have made two regions which are doped n-type. To each of those n-type regions we attach a wire, and connect a battery between them. If we try to get some current, *I*, to flow through this structure, nothing will happen, because the n-p junction on the RHS is reverse biased, i.e., the positive lead from the battery going to the n-side of the p-n junction. If we attempt to remedy this by turning the battery around, we will now have the LHS junction reverse biased, and again, no current will flow. If, for whatever reason, we want current to flow, we will need to come up

---

[3] This content is available online at <http://cnx.org/content/m33805/1.1/>.

with some way of forming a layer of n-type material between one n-region and the other. This will then connect them together, and we can run current in one terminal and out the other.

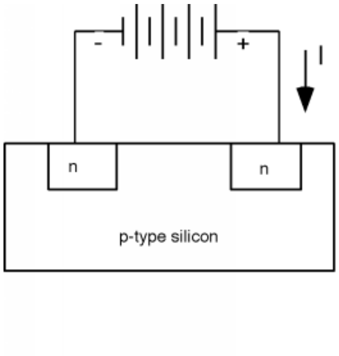


Figure 2.11: The start of a field effect transistor.

To see how we will do this, let's do two things. First we will grow a layer of SiO_2 (silicon dioxide or silica, but actually refered to as "oxide") on top of the silicon. To do this the wafer is placed in an oven under an oxygen atmosphere, and heated to 1100 °C. The result is a nice, high-quality insulating SiO_2 layer on top of the silicon). On top of the oxide layer we then deposit a conductor, which we call the gate. In the "old days" the gate would have been a layer of aluminum; hence the "metal-oxide-silicon" or MOS name. Today, it is much more likely that a heavily doped layer of polycrystalline silicon (polysilicon, or more often just "poly") would be deposited to form the gate structure. Polysilicon is made from the reduction of a gas, such as silane (SiH_4), (2.13).

$$\mathrm{SiH}_{4(g)} \rightarrow \mathrm{Si}_{(s)} + 2\,\mathrm{H}_{2(g)} \tag{2.13}$$

The silicon is polycrystalline (composed of lots of small silicon crystallites) because it is deposited on top of the oxide, which is amorphous, and so it does not provide a single crystal "matrix" which would allow the silicon to organize itself into one single crystal. If we had deposited the silicon on top of a single crystal silicon wafer, we would have formed a single crystal layer of silicon called an epitaxial layer. This is sometimes done to make structures for particular applications. For instance, growing a n-type epitaxial layer on top of a p-type substrate permits the fabrication of a very abrupt p-n junction.

> NOTE: Epitaxy, is a transliteration of two Greek words *epi*, meaning "upon", and *taxis*. meaning "ordered". Thus an epitaxial layer is one that follows the order of the substrate on which it is grown.

Now we can go back now to our initial structure, shown in Figure 2.11, only this time we will add an oxide layer, a gate structure, and another battery so that we can invert the region under the gate and connect the two n-regions together. Well also identify some names for parts of the structure, so we will know what we are talking about. For reasons which will be clear later, we call the n-region connected to the negative side of the battery the source, and the other one the drain. We will ground the source, and also the p-type substrate. We add two batteries, V_{gs} between the gate and the source, and V_{ds} between the drain and the source.

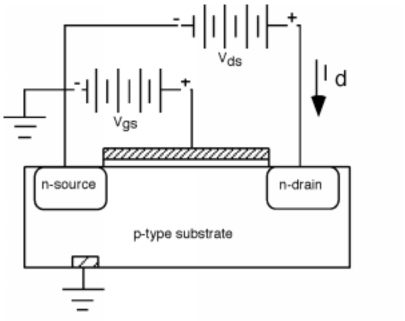


Figure 2.12: Biasing a MOSFET transistor

It will be helpful if we also make another sketch, which gives us a perspective view of the device. For this we strip off the gate and oxide, but we will imagine that we have applied a voltage greater than V_T to the gate, so there is a n-type region, called the channel which connects the two. We will assume that the channel region is L long and W wide, as shown in Figure 2.13.

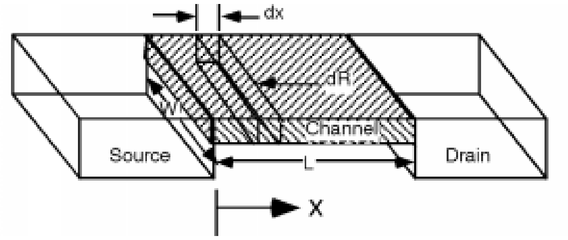


Figure 2.13: The inversion channel and its resistance.

Next we want to take a look at a little section of channel, and find its resistance $d(R)$, when the little section is $d(x)$ long, (2.14).

$$d(R) = \frac{\mathrm{dx}}{\sigma_s W} \tag{2.14}$$

We have introduced a slightly different form for our resistance formula here. Normally, we would have a simple σ in the denominator, and an area A, for the cross-sectional area of the channel. It turns out to be very hard to figure out what that cross sectional area of the channel is however. The electrons which form the inversion layer crowd into a very thin sheet of surface charge which really has little or no thickness, or penetration into the substrate.

If, on the other hand we consider a surface conductivity (units: simply mhos), σ_s, (2.15), then we will have an expression which we can evaluate. Here, μ_s is a surface mobility, with units of $\mathrm{cm}^2/\mathrm{V.sec}$, that is the quantity which represented the proportionality between the average carrier velocity and the electric field, (2.16) and (2.17).

$$\sigma_s = \mu_s Q_{\text{chan}} \tag{2.15}$$

$$\bar{v} = \mu E \tag{2.16}$$

$$\mu = \frac{q\tau}{m} \tag{2.17}$$

The surface mobility is a quantity which has to be measured for a given system, and is usually just a number which is given to you. Something around 300 $\text{cm}^2/\text{V.sec}$ is about right for silicon. Q_{chan} is called the surface charge density or channel charge density and it has units of $\text{Coulombs}/\text{cm}^2$. This is like a sheet of charge, which is different from the bulk charge density, which has units of $\text{Coulombs}/\text{cm}^2$. Note that:

$$\begin{array}{rcl} \frac{\text{cm}^2}{\text{Voltsec}} \frac{\text{Coulombs}}{\text{cm}^2} & = & \frac{\frac{\text{Coul}}{\text{sec}}}{\text{Volt}} \\ & = & \frac{I}{V} \\ & = & \text{mhos} \end{array} \tag{2.18}$$

It turns out that it is pretty simple to get an expression for Q_{chan}, the surface charge density in the channel. For any given gate voltage V_{gs}, we know that the charge density on the gate is given simply as:

$$Q_g = c_{\text{ox}} V_{\text{gs}} \tag{2.19}$$

However, until the gate voltage V_{gs} gets larger than V_T we are not creating any mobile electrons under the gate, we are just building up a depletion region. We'll define Q_T as the charge on the gate necessary to get to threshold. $Q_T = c_{\text{ox}} V_T$. Any charge added to the gate above Q_T is matched by charge Q_{chan} in the channel. Thus, it is easy to say: (2.20) or (2.21).

$$Q_{\text{channel}} = Q_g - Q_T \tag{2.20}$$

$$Q_{\text{chan}} = c_{\text{ox}} \left(V_g - V_T\right) \tag{2.21}$$

Thus, putting (2.20) and (2.15) into (2.14), we get:

$$d\left(R\right) = \frac{d\left(x\right)}{\mu_s c_{\text{ox}} \left(V_{\text{gs}} - V_T\right) W} \tag{2.22}$$

If you look back at Figure 2.12, you will see that we have defined a current I_d flowing into the drain. That current flows through the channel, and hence through our little incremental resistance $d\left(R\right)$, creating a voltage drop $d\left(V_c\right)$ across it, where V_c is the channel voltage, (2.23).

$$\begin{array}{rcl} d\left(V_c\left(x\right)\right) & = & I_d d\left(R\right) \\ & = & \frac{I_d d(x)}{\mu_s c_{\text{ox}}(V_{\text{gs}} - V_T) W} \end{array} \tag{2.23}$$

Let's move the denominator to the left, and integrate. We want to do our integral completely along the channel. The voltage on the channel $V_c\left(x\right)$ goes from 0 on the left to V_{ds} on the right. At the same time, x is going from 0 to L. Thus our limits of integration will be 0 and V_{ds} for the voltage integral $d\left(V_c\left(x\right)\right)$ and from 0 to L for the x integral $d\left(x\right)$.

$$\int_0^{V_{\text{ds}}} \mu_s c_{\text{ox}} \left(V_{\text{gs}} - V_T\right) W d\text{Vc} = \int_0^L I_d dx \tag{2.24}$$

Both integrals are pretty trivial. Let's swap the equation order, since we usually want I_d as a function of applied voltages.

$$I_d L = \mu_s c_{\mathrm{ox}} W \left(V_{\mathrm{gs}} - V_T\right) V_{\mathrm{ds}} \tag{2.25}$$

We now simply divide both sides by L, and we end up with an expression for the drain current I_d, in terms of the drain-source voltage, V_{ds}, the gate voltage V_{gs} and some physical attributes of the MOS transistor.

$$I_d = \left(\frac{\mu_s c_{\mathrm{ox}} W}{L} \left(V_{\mathrm{gs}} - V_T\right)\right) V_{\mathrm{ds}} \tag{2.26}$$

2.4 Light Emitting Diode[4]

> NOTE: This module is adapted from the Connexions module entitled *Light Emitting Diode* by Bill Wilson.

Let's talk about the recombining electrons for a minute. When the electron falls down from the conduction band and fills in a hole in the valence band, there is an obvious loss of energy. The question is; where does that energy go? In silicon, the answer is not very interesting. Silicon is what is known as an *indirect band-gap* material. What this means is that as an electron goes from the bottom of the conduction band to the top of the valence band, it must also undergo a significant change in momentum. This all comes about from the details of the band structure for the material, which we will not concern ourselves with here. As we all know, whenever something changes state, we must still conserve not only energy, but also momentum. In the case of an electron going from the conduction band to the valence band in silicon, both of these things can only be conserved if the transition also creates a quantized set of lattice vibrations, called *phonons*, or "heat". Phonons posses both energy and momentum, and their creation upon the recombination of an electron and hole allows for complete conservation of both energy and momentum. All of the energy which the electron gives up in going from the conduction band to the valence band (1.1 eV) ends up in phonons, which is another way of saying that the electron heats up the crystal.

In some other semiconductors, something else occurs. In a class of materials called *direct band-gap* semiconductors, the transition from conduction band to valence band involves essentially no change in momentum. Photons, it turns out, possess a fair amount of energy (several eV/photon in some cases) but they have very little momentum associated with them. Thus, for a direct band gap material, the excess energy of the electron-hole recombination can either be taken away as heat, or more likely, as a photon of light. This radiative transition then conserves energy and momentum by giving off light whenever an electron and hole recombine. This gives rise to the light emitting diode (LED). Emission of a photon in an LED is shown schematically in Figure 2.14.

---

[4] This content is available online at <http://cnx.org/content/m33801/1.1/>.

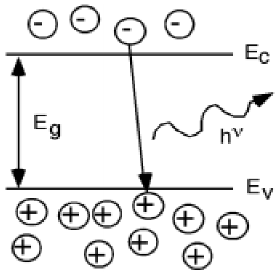


Figure 2.14: Radiative recombination in a direct band-gap semiconductor.

It was Planck who postulated that the energy of a photon was related to its frequency by a constant, which was later named after him. If the frequency of oscillation is given by the Greek letter "nu" (ν), then the energy of the photon is just given by, (2.27), where h is Planck's constant, which has a value of 4.14 x 10^{-15} eV.sec.

$$E = h\nu \tag{2.27}$$

When we talk about light it is conventional to specify its wavelength, λ, instead of its frequency. Visible light has a wavelength on the order of nanometers, e.g., red is about 600 nm, green about 500 nm and blue is in the 450 nm region. A handy "rule of thumb" can be derived from the fact that $c = \lambda\nu$, where c is the speed of light (3 x 10^3 m/sec or 3 x 10^{17} nm/sec, (2.28).

$$\begin{aligned} \lambda\,(\mathrm{nm}) &= \frac{hc}{E(\mathrm{eV})} \\ &= \frac{1242}{E(\mathrm{eV})} \end{aligned} \tag{2.28}$$

Thus, a semiconductor with a 2 eV band-gap should give off light at about 620 nm (in the red). A 3 eV band-gap material would emit at 414 nm, in the violet. The human eye, of course, is not equally responsive to all colors (Figure 2.15). The materials which are used for important light emitting diodes (LEDs) for each of the different spectral regions are also shown in Figure 2.15.

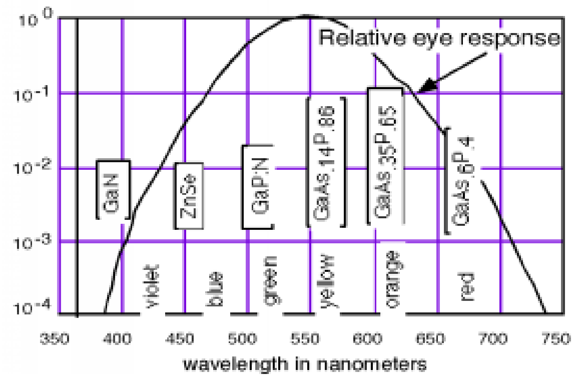


Figure 2.15: Relative response of the human eye to various colors.

It is worth noting that a number of the important LEDs are based on the GaAsP system. GaAs is a direct band-gap semiconductor with a band gap of 1.42 eV (in the infrared). GaP is an indirect band-gap material with a band gap of 2.26 eV (550 nm, or green). Both As and P are group V elements. (Hence the nomenclature of the materials as III-V (or 13-15) compound semiconductors.) We can replace some of the As with P in GaAs and make a mixed compound semiconductor $\mathrm{GaAs_{1\text{-}x}P_x}$. When the mole fraction of phosphorous is less than about 0.45 the band gap is direct, and so we can "engineer" the desired color of LED that we want by simply growing a crystal with the proper phosphorus concentration! The properties of the GaAsP system are shown in Figure 2.16. It turns out that for this system, there are actually two different band gaps, as shown in Figure 2.16. One is a direct gap (no change in momentum) and the other is indirect. In GaAs, the direct gap has lower energy than the indirect one (like in the inset) and so the transition is a radiative one. As we start adding phosphorous to the system, both the direct and indirect band gaps increase in energy. However, the direct gap energy increases faster with phosphorous fraction than does the indirect one. At a mole fraction x of about 0.45, the gap energies cross over and the material goes from being a direct gap semiconductor to an indirect gap semiconductor. At $x = 0.35$ the band gap is about 1.97 eV (630 nm), and so we would only expect to get light up to the red using the GaAsP system for making LED's. Fortunately, people discovered that you could add an impurity (nitrogen) to the GaAsP system, which introduced a new level in the system. An electron could go from the indirect conduction band (for a mixture with a mole fraction greater than 0.45) to the nitrogen site, changing its momentum, but not its energy. It could then make a direct transition to the valence band, and light with colors all the way to the green became possible. The use of a nitrogen recombination center is depicted in the Figure 2.17.

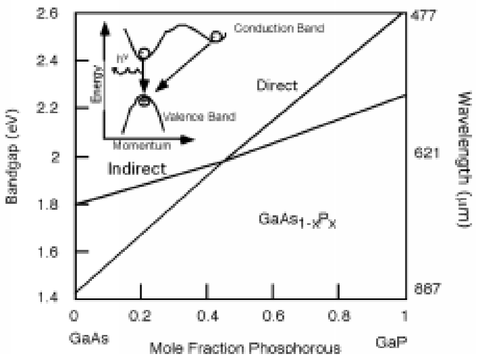


Figure 2.16: Band gap for the GaAsP system

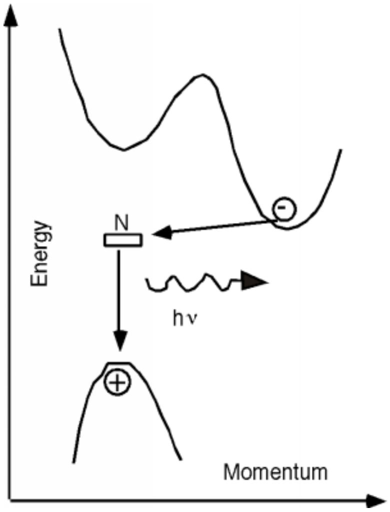


Figure 2.17: Addition of a nitrogen recombination center to indirect GaAsP.

If we want colors with wavelengths shorter than the green, we must abandon the GaAsP system and look for more suitable materials. A compound semiconductor made from the II-VI elements Zn and Se make up one promising system, and several research groups have successfully made blue and blue-green LEDs from ZnSe. SiC is another (weak) blue emitter which is commercially available on the market. Recently, workers at a tiny, unknown chemical company stunned the "display world" by announcing that they had successfully fabricated a blue LED using the II-V material GaN. A good blue LED was the "holy grail" of the display and

CD ROM research community for a number of years. Obviously, adding blue to the already working green and red LED's completes the set of 3 primary colors necessary for a full-color flat panel display. Furthermore, using a blue LED or laser in a CD ROM would more than quadruple its data capacity, as bit diameter scales as λ, and hence the area as λ^2.

2.5 Polymer Light Emitting Diodes[5]

This module was developed as part of a Rice University course CHEM496: *Chemistry of Electronic Materials*. This module was prepared with the assistance of Pui Yee Hung.

2.5.1 Introduction

In 1990, electroluminescent (EL) from conjugated polymers was first reported by Burroughes et al. of Cambridge University. A layer of poly(*para*-phenylenevinylene) (PPV) was sandwiched between layers of indium tin oxide (ITO) and aluminum. When this device is under a 14 V dc bias, the PPV emits a yellowish-green light with a quantum efficiency of 0.05%. This report attracted a lot of attention, because the potential that polymer light emitting diodes (LEDs) could be inexpensively mass produced into large area display area. The processing steps in making polymer LEDs are readily scaleable. The industrial coating techniques is well developed to mass produce polymer layers of 100 nm thickness, and the device could be patterned onto large surface area by pixellation of metal.

Since the initial discovery, and increasing amount of researches has been performed, and significant progress has been made. In 1990 the polymer LED only emitted yellowish green color, now the emission color ranged from deep blue to near infra red. The efficiency of the multi-layer polymer LED even reached a quantum efficiency of >4% and the operating voltage has been reduced significantly. In term of efficiency, color selection and operating voltage, polymer LEDs have attained adequate levels for commercialization. But there are reliability problems that are symptomatic of any organic devices.

2.5.2 Device physics and materials science of polymer LEDs

A schematic diagram of a polymer LED is shown in Figure 2.18. A polymer LED can be divided into three different components:

A. **Anode**: the hole supplier, made of metal of high working function. Examples of the common anode are indium tin oxide (ITO), gold etc. The anode is usually transparent so that light can be emitted through.
B. **Cathode**: the electron supplier, made of metal of low working function. Examples of the common cathode are aluminum or calcium.
C. **Polymer**: made of conjugated polymer film with thickness of 100 nm.

---

[5] This content is available online at <http://cnx.org/content/m25670/1.2/>.

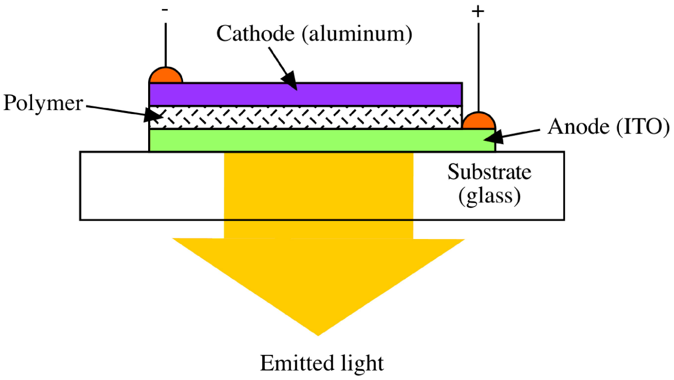


Figure 2.18: Schematic set-up of polymer LED.

When a polymer LED is under a direct current (dc) bias, holes are injected from the anode (ITO) and electrons are injected from the cathode (aluminum). Under the influences of the electrical field, the electrons and holes will migrate toward each other. When they recombine in the conjugated polymer layer, a bound excited states (excitons) will be formed. Some of the excitons (singlets) then decays in the conjugated polymer layer to emit light through the transparent substrates (glass). The emission color will be depended on the energy gap of the polymers. There is energy gap in a conjugated polymer because the π electron are not completely delocalized over the entire polymer chain. Instead there are alternate region in the polymer chain that has a higher electron density (Figure 2.19a). The chain length of this region is about 15-20 multiple bonds. The emission color can be controlled by tuning this energy band gap (Figure 2.19b). It shows that bond alternation limits the extent of delocalization. Table 2.1 summarizes the structure and emission color of some common conjugated polymers.

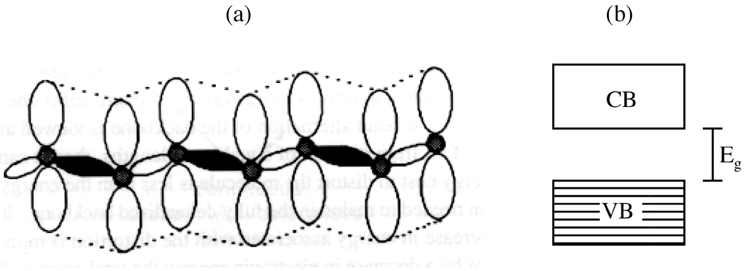


Figure 2.19: Alternation of bond lengths along a conjugated polymer chain (a) results in a material with properties of a large band gap semiconductor (b), where CB is the conductive band gap, and VB is the valence band, and E_g is the band gap.

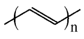
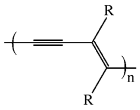
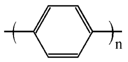
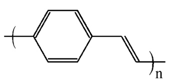
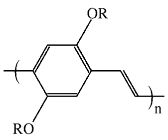
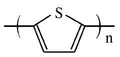
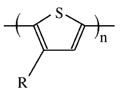
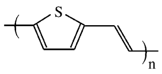
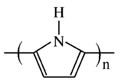
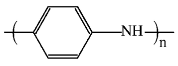
Polymer	**Chemical name**	**Structure**	**π-π* energy gap (eV)**	**Emission peak (nm)**
PA	*trans*-polyacetylene		1.5	600
PDA	polydiacetylene		1.7	
PPP	poly(*para*-phenylene)		3.0 (red)	465
PPV	Poly(*para*-phenylenevinylene)		2.5 (green)	565
RO-PPV	poly(2,5-dialkoxy-p-phenlyenevinylen)		2.2 (blue)	~580
PT	polythiophene		2.0 (red)	
P3AT	Poly(3-alkythiophene)		2.0 (red)	690
PTV	Poly(2,5-thiophenevinylene)		1.8	
PPy	Polypyrrole		3.1	
PAni	Polyaniline			3.2

Table 2.1: Example of common conjugated polymers.

2.5.3 Approaches to improve the efficiency

Efficiency for any LED is defined:

$n_{\mathrm{ext}} = n_{\mathrm{esc}} * n_{\mathrm{int}}$

where $n_{\mathrm{ext}}$is the external quantum efficiency, n_{int} is the internal efficiency (represents the fraction of injected carrier, usually electron, that is converted to photon), and n_{esc} is the escape efficiency (represent fraction of photons that can reach to the outside).

The most common way to improve the internal efficiency is to balance the number of electrons and holes which arrives at the polymer layer. Originally, there are more holes than electron that arrive of the polymer layer because conjugated polymers have a higher electron affinity, and as a consequence will favor the transport of hole than electron. There are two ways to maintains the balance:

1) Match the work function of electrode with the electron affinity and ionization potential of the polymer.
2) Tune the polymer's electron affinity and ionization potential to match the work function of the electrode.

The escape efficiency is also important because a polymer LED is made up of layers of materials that have different refractive index, and some of the photon generated from the excition may be reflected at the boundary and trapped inside the device.

2.5.3.1 Improvement in internal quantum efficiency using low working function cathode

Conjugated polymer is electron rich, the mobility for hole is higher than electron, and more holes will arrive in the polymer layer than electrons. One way to increase the population of the electron is to use a lower working function metal as cathode. Braun and Heeger have replaced the aluminum cathode with calcium results in improved internal efficiency by a factor of ten, to 0.1%. This approach is direct and fast but low working function electrode like calcium will be oxidized easily and shorten the devices' life.

2.5.3.2 Improvement in internal quantum efficiency using multiple polymer layers

A layer of poly[2,5-di(hexyloxy)cyanoterephthalylidene] (CN-PPV, Figure 2.20) is coated on top of PPV to improve the transport and recombination of electron and holes (Figure 2.21).

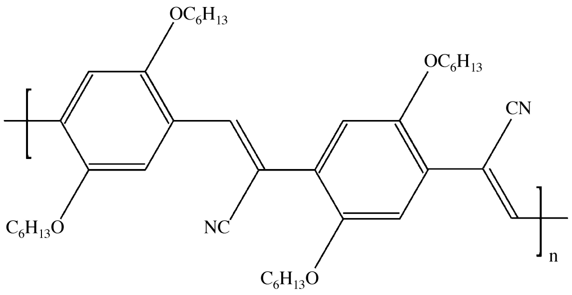


Figure 2.20: Structure of CN-PPV.

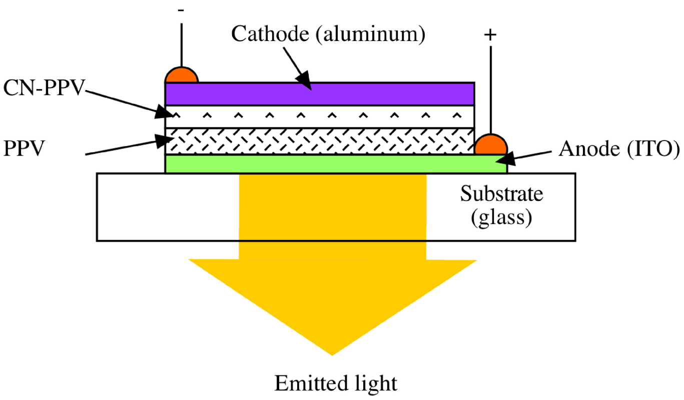


Figure 2.21: Schematic representation of a CN-PPV and PPV multi-layer LED.

The nitrile group in the CN-PPV has two effect on the polymer.

1) It increases the electron affinity so electrons can travel more efficient from the aluminum to the polymer layer. And metal of relative high working function like aluminum and gold can be now be used as cathode instead of calcium.
2) It increases the binding energy of the occupied π and unoccupied π* state but maintain a similar π-π* gap. So when the PPV and CN-PPV is placed together, holes and electron will be confined at the heterojunction.

The resulting energy levels are shown in Figure 2.22.

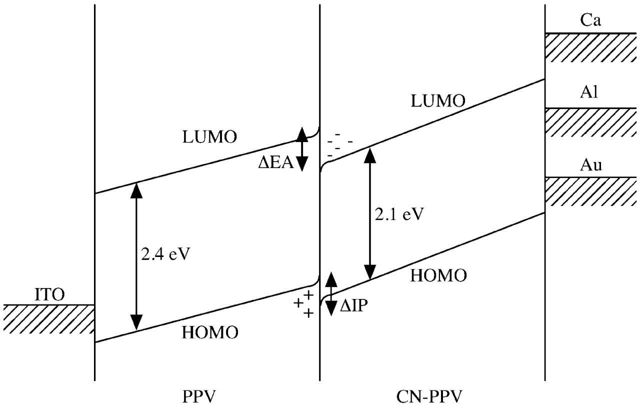


Figure 2.22: Schematic energy-level diagram for a PPV and CN-PPV under foreword bias. Adapted from N. C. Greenham, S. C. Maratti, D. D. C. Bradley, R. H. Friend, and A. B. Holmes, *Nature*,1993, **365**, 62.

The absolute energies of levels are not known accurately, but the diagram show the relative position of the HOMO and LUMO levels in the polymers, and the Fermi levels of the various possible metal contacts, the differences in electron affinity (ΔEA) and ionization potential (ΔIP) between PPV and CN-PPV are also shown (Figure 2.22).

At the polymers interface there is a sizable offset in the energies of HOMO and LUMO of PPV and CN-PPV, the holes transported from the ITO and the electrons transport from the aluminum will be confined in the heterojunction. The local charge density will be sufficiently high to ensure the holes and electrons will pass within a collision capture radius. This set-up increases the chance for an electrons to combine with holes to form an excition. In addition, the emission will be close to the junction, far away from the electrode junction which will quench the singlet excitions. The result is that a multi-layers LED has an internal quantum efficiency of 10% and external quantum efficiency (for light emitted in foreword direction) of 25%.

Based on this approach, a couple of polymers have been developed or modified to produce the desirable emission color and processing property. The drawback of this method is that desirable properties may not be commentary to each other. For example, in MEH-PPV an alkoxy side group (RO) is introduced to PPV so that it can be dissolved in organic solvent. But the undesirable effect is that MEH-PPV is less thermally stable. Moreover in multiple layers LEDs, different polymer layers have different refractive indices and a fraction of the photons will undergo total internal reflection at the refractive boundaries and cannot escape as light. This problem can be overcome by Febry-Pert microcavity structure.

2.5.3.3 Improvement in external quantum efficiency using microcavity

Fabry-Perot resonant structures are also used in inorganic LED, and are is based on Fermi's golden rule:

$K_r \sim | <M> | r_{(v)}$

where M (the matrix element of the perturbation between final and initial states) depends on the nature of the material, and $r_{(v)}$ can be altered by changing the density of various density states, e.g. using a luminescent thin films to select certain value of V.

In building a microcavity for a polymer LED, the polymer is placed between two mirrors. (Figure 2.23), in which one of the mirrors is made up of aluminum, the other mirror (a Bragg Mirror) is form by epitaxial multilayer stacks of Si_xN_y and SiO_2.

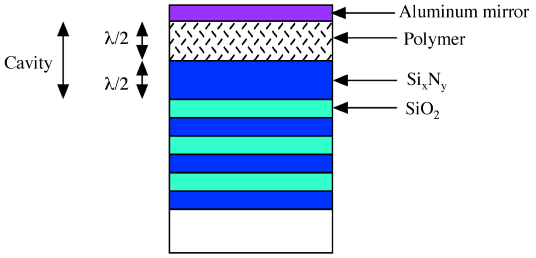


Figure 2.23: Schematic set-up of micro-cavity.

2.5.3.4 Improvement in internal quantum efficiency: doping of polymer

Doping is a process that creates carrier by purposely introducing impurities and is very popular method in the semiconductor industry. However, this technique was not used in polymer LED until 1995, when a co-polymer polystyrene-poly(3-hexylthiphene) (PS-P3HT) was doped with $FeCl_{3.}$ Doping of MEH-PPV with iodine has improved the efficiency by 200% and the polymer LED can be operated under both forward and reverse bias (Table 2.2). The doping is accomplsihed by mixing 1 wt% MEH-PPV with 0.2 wt% I_2. The molar ratio of MEH-PPV to I_2 is 5:1. That is a huge "doping " ratio when you compare the doping concentration in the semiconductor.

	Un-doped	**Doped**
Turn on voltage (V)	10	foreword 5, reversed 12
External efficiency (%)	4×10^{-4}	8×10^{-3}

Table 2.2: Results of iodine doping of an Al/MEH-PPV/ITO-based LED.

2.5.4 Polymer LEDs on a silicon substrate: an application advantage over inorganic LEDs

In the initial research polymer LEDs were in direct competition with the inorganic LEDs and tried to achieve the existing LED standard. This is a difficult task as polymer LEDs have a lower long term stability. However, there are some applications in which polymer LEDs have a clear advantage over their more traditional inorganic analogs. One of these is to incorporate LEDs with the silicon integrated circuits for inter-chip communication.

It is difficult to build inorganic LEDs on a silicon substrate, because of the thermal stress developing between the inorganic LED (usually a III-V based device) and the silicon interface. But polymer LEDs offer a solution, since polymers can be easily spin-coated on the silicon. The operating voltage of polymer LED is less than 4 V, and the turn on voltage can be as low as 2 V. Together with a switching time of less than 50 ns, make polymer LED a perfect candidate.

2.5.5 Reliability and degradation of polymer LEDs

In terms of the efficiency, color selection, and driving voltage, polymer LED have attained adequate level for commercialization. However, the device lifetime is still far from satisfactory. Research into understanding the reliability and degredation mechanisms of polymer LEDs has generally been divided into two area:

1) Photo-degradation of polymer.
2) Interface degradation.

2.5.5.1 Polymer photo degradation

Photoluminescece (PL) studies of the photo-oxidation of PPV have been undertaken, since it is believes that EL is closely related with PL.

It was found that there is a rapid decay in emission when PPV is exposed to oxygen. Using time resolved FTIR spectroscopy an increase in the carbonyl signal and a decrease in C=C signal with time (Figure 2.24). It was suggested that the carbonyl group has a strong electron affinity level to charge transfer between molecules segment in the polymer, thereby dissociating the excition and quenching the PL.

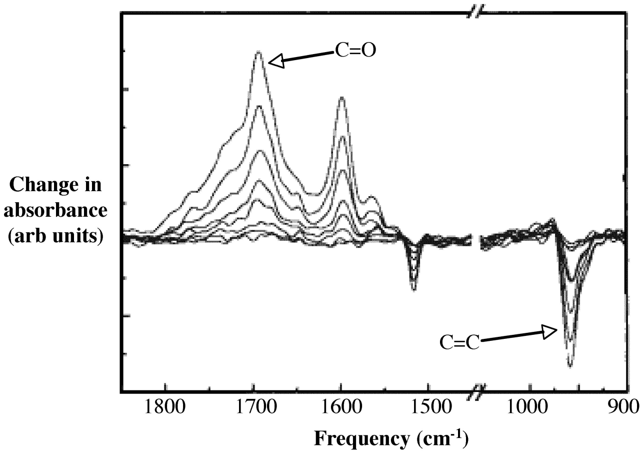


Figure 2.24: FTIR as a function of photo-oxidation of PPV. Adapted from M. Yan, L. J. Rothberg, F. Papadimitrakopoulos, M. E. Galvin and T. M. Miller, *Phys. Rev. Lett.*, 1994, **73**, 744.

Similar research was performed by Cumpston and Jensen using BCHA-PPV and P3OT (Figure 2.25) and exposing them to dry air in UV irradiation. In BCHA-PPV, there is an increase in carbonyl signal with time, while the P3OT remain intact. A mechanism proposed for the degradation of BCHA-PPV involves the transfer of energy from the excited triplet state of the PPV to oxygen to from singlet oxygen which attack the vinyl double bond in the PPV backbone. And P3OT dose not has vinyl bond so it can resist the oxidation .

Figure 2.25: Structure of (a) BCHA-PPV and (b) P3OT.

The research described above was all performed on polymer thin films deposited on an inert surface. The presence of cathode and anode may also affect the oxidation mechanism. Scott et al. have taken IR spectra from a MEH-PPV LED in the absence of oxygen. They obtained similar result as in Yan et al., however, a decrease in ITO's oxygen signal was noticed suggesting that the ITO anode acts like a oxygen reservoir and supplies the oxygen for the degradation process.

2.5.5.2 Polymer LED interface degradation

There are few interface degredation studies in polymer LEDs. One of them by Scott et al. took SEM image of the cathode from a failed polymer LED. The polymer LED used ITO as the anode, MEH-PPV as the polymer layer, and an aluminum calcium alloy as cathode. SEM images showed "craters" formed in the cathode. The craters are formed when the cathode metal is melted and pull away from the polymer layer. It was suggested that a high current density will generate heat and result in local hot spot. The temperature in the hot spot is high enough to melt the cathode. And when it melt, it will pull away from the polymer. This process will decrease the effective cathode area, and reduce the luminescence gradually.

2.5.6 Bibliography

- D. R. Baigent, N. C. Greenham, J. Gruner, R. N. Marks, R. H. Friends, S. C. Moratti, and A. B. Holmes, *Synth.Met.*, 1994, **67**, 3.
- B. H. Cumpston and K. F. Jensen, *Synth. Met.*, 1995, **73**, 195.
- J. H. Burroughes, D. D. C. Bradley, A. R. Brown, R. N. Marks, K. Mackay, R. H. Friend, P. L. Burns, and A. B. Holmes, *Nature*, 1990, **347**, 539.
- N. C. Greenham, S. C.Maratti, D. D. C. Bradley, R. H. Friend, and A. B. Holmes, *Nature*, 1993, **365**, 628.
- J. Gruner, F. Cacialli, I. D. W. Samuel, R. H. Friend, *Synth. Met*, 1996, **76,** 197.
- M. Herold, J. Gmeiner, W. Riess, and M. Schwoerer, *Synth. Met.*, 1996, **76**, 109.
- R. H. Jordan, A. Dodabalapur, L. J. Rothberg, and R. E. Slusher, *Proceeding of SPIE*, 1997, **3002**, 92.

- I. D. Parker and H. H. Kim, *Appl. Phys. Lett.*, 1994, **64**, 1774.
- J. C. Scott, J. Kaufman, P. J. Brock, R. DiPietro, J. Salem, and J. A. Goitia, *J. Appl. Phys.*, 1996, **79**, 2745.
- M. S. Weaver, D. G. Lidzaey, T. A. Fisher, M. A. Pate, D. O'Brien, A. Bleyer, A. Tajbakhsh, D. D. C. Bradley, M. S. Skolnick, and G. Hill, *Thin solid Films,* 1996, **273**, 39.
- M. Yan, L. J.Rothberg, F. Papadimitrakopoulos, M. E. Galvin, and T. M. Miller, *Phys. Rev. Lett.*, 1994, **73**, 744.

2.6 Laser[6]

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

What is the difference between an LED and a solid state laser? There are some differences, but both devices operate on the same principle of having excess electrons in the conduction band of a semiconductor, and arranging it so that the electrons recombine with holes in a radiative fashion, giving off light in the process. What is different about a laser? In an LED, the electrons recombine in a random and unorganized manner. They give off light by what is known as *spontaneous emission*, which simply means that the exact time and place where a photon comes out of the device is up to each individual electron, and things happen in a random way.

There is another way in which an excited electron can emit a photon however. If a field of light (or a set of photons) happens to be passing by an electron in a high energy state, that light field can induce the electron to emit an additional photon through a process called stimulated emission. The photon field *stimulates* the electron to emit its energy as an additional photon, which comes out *in phase with the stimulating field.* This is the big difference between *incoherent light* (what comes from an LED or a flashlight) and *coherent light* which comes from a laser. With coherent light, all of the electric fields associated with each phonon are all exactly in phase. This coherence is what enables us to keep a laser beam in tight focus, and to allow it to travel a large distance without much divergence or spreading out.

So how do we restructure an LED so that the light is generated by stimulated emission rather than spontaneous emission? Firstly, we build what is called a heterostructure. All this means is that we build up a sandwich of somewhat different materials, with different characteristics. In this case, we put two wide band-gap regions around a region with a narrower band gap. The most important system where this is done is the AlGaAs/GaAs system. A band diagram for such a set up is shown in Figure 2.26. AlGaAs (pronounced "Al-Gas") has a larger band-gap then does GaAs. The potential "well" formed by the GaAs means that the electrons and holes will be confined there, and all of the recombination will occur in a very narrow strip. This greatly increases the chances that the carriers can interact, but we still need some way for the photons to behave in the proper manner. Figure 2.27 is a diagram of what a typical diode might look like. We have the active GaAs layer sandwich in-between the two heterostructure confinement layers, with a contact on top and bottom. On either end of the device, the crystal has been "cleaved" or broken along a crystal lattice plane. This results in a shiny "mirror-like" surface, which will reflect photons. The back surface (which we can not see here) is also cleaved to make a mirror surface. The other surfaces are purposely roughened so that they do not reflect light. Now let us look at the device from the side, and draw just the band diagram for the GaAs region (Figure 2.28). We start things off with an electron and hole recombining spontaneously. This emits a photon which heads towards one of the mirrors. As the photon goes by other electrons, however, it may cause one of them to decay by stimulated emission. The two (in phase) photons hit the mirror and are reflected and start back the other way . As they pass additional electrons, they stimulate them into a transition as well, and the optical field within the laser starts to build up. After a bit, the photons get down to the other end of the cavity. The cleaved facet, while it acts like a mirror, is not a perfect one. Some light is not reflected, but rather "leaks"; though, and so becomes the output beam from the laser. The details of finding what the ratio of reflected to transmitted light is will

---

[6] This content is available online at <http://cnx.org/content/m33802/1.1/>.

have to wait until later in the course when we talk about dielectric interfaces. The rest of the photons are reflected back into the cavity and continue to stimulate emission from the electrons which continue to enter the gain region because of the forward bias on the diode.

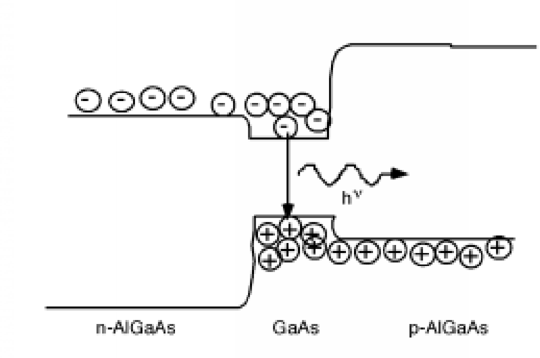


Figure 2.26: The band diagram for a double heterostructure GaAs/AlGaAs laser.

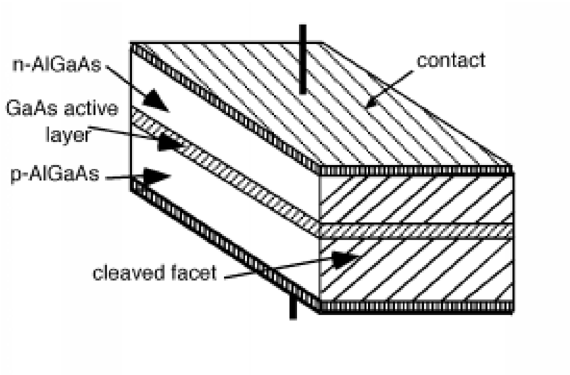


Figure 2.27: A schematic diagram of a typical laser diode.

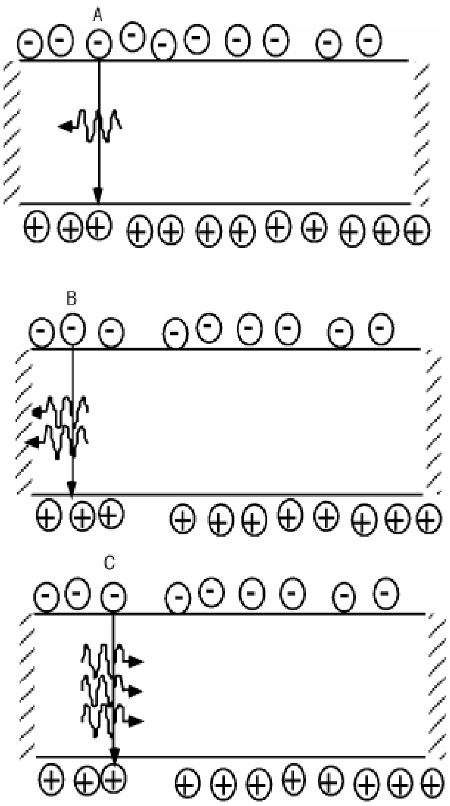


Figure 2.28: Build up of a photon field in a laser diode.

In reality, the photons do not move back and forth in a big "clump" as we have described here, rather they are distributed uniformly along the gain region (Figure 2.29). The field within the cavity will build up to the point where the loss of energy by light leaking out of the mirrors just equals the rate at which energy is replaced by the recombining electrons.

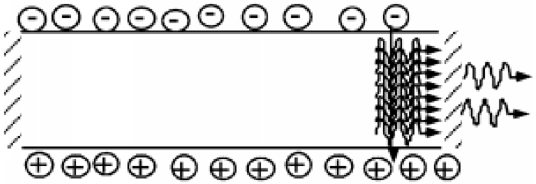


Figure 2.29: Output coupling in a diode laser.

2.7 Solar Cells[7]

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

Now let us look at the opposite process of light generation for a moment. Consider the following situation where we have just a plain old normal p-n junction, only now, instead of applying an external voltage, we imagine that the junction is being illuminated with light whose photon energy is greater than the band-gap (Figure 2.30a). In this situation, instead of recombination, we will get photo-generation of electron hole pairs. The photons simply excite electrons from the full states in the valence band, and "kick" them up into the conduction band, leaving a hole behind. This is similiar to the thermal excitation process. As can be seen from Figure 2.30b, this creates excess electrons in the conduction band in the p-side of the diode, and excess holes in the valence band of the n-side. These carriers can diffuse over to the junction, where they will be swept across by the built-in electric field in the depletion region. If we were to connect the two sides of the diode together with a wire, a current would flow through that wire as a result of the electrons and holes which move across the junction.

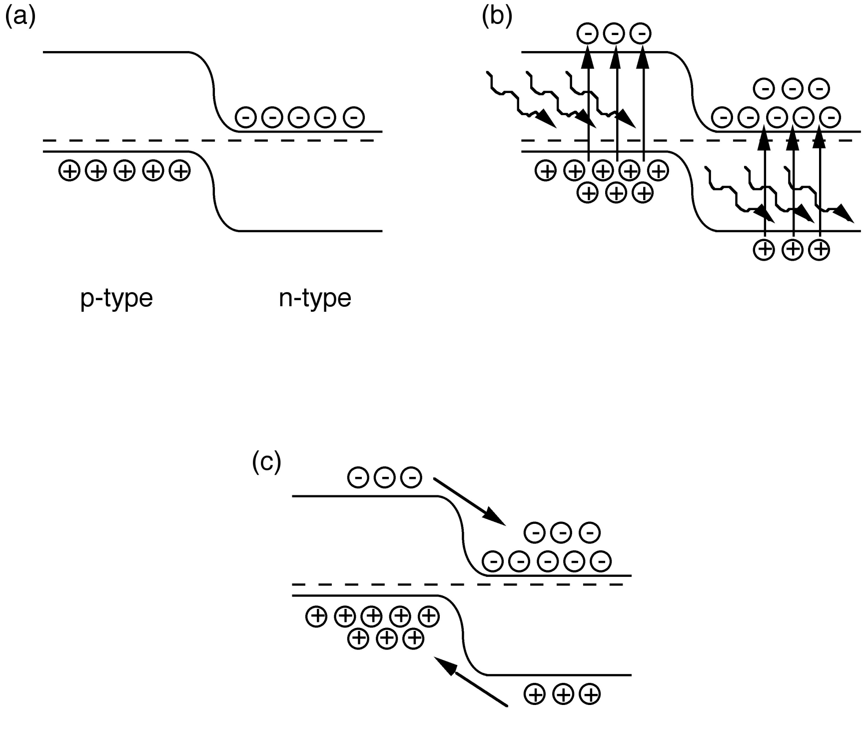


Figure 2.30: A schematic representation of a p-n diode under illlumination.

Which way would the current flow? A quick look at Figure 2.30c shows that holes (positive charge carriers) generated on the n-side will float up to the p-side as they go across the junction. Hence positive current must be coming out of the anode, or p-side of the junction. Likewise, electrons generated on the p-side will fall down the junction potential, and come out the n-side, but since they have negative charge, this flow represents current going into the cathode. We have constructed a *photovoltaic diode*, or *solar cell*.

---

[7] This content is available online at <http://cnx.org/content/m33803/1.4/>.

Figure 2.31 is a picture of what this would look like schematically. We might like to consider the possibility of using this device as a source of energy, but the way we have things set up now, since the voltage across the diode is zero, and since power equals current times voltage, we see that we are getting nada from the cell. What we need, obviously, is a load resistor, so let's put one in. It should be clear from Figure 2.32 that the photo current flowing through the load resistor will develop a voltage which it biases the diode in the forward direction, which, of course will cause current to flow back into the anode. This complicates things, it seems we have current coming out of the diode and current going into the diode all at the same time! How are we going to figure out what is going on?

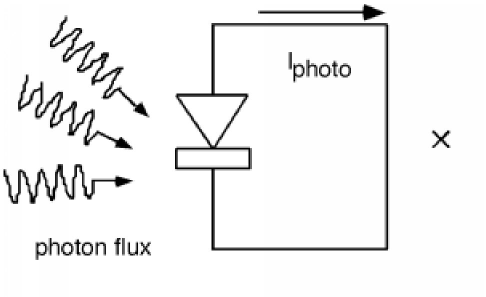


Figure 2.31: Schematic representation of a photovoltaic cell.

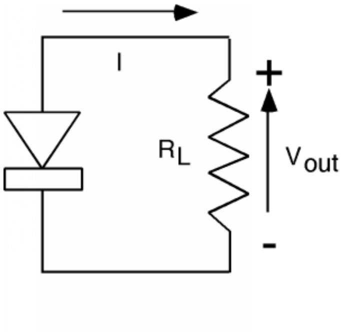


Figure 2.32: Photovoltaic cell with a load resistor.

The answer is to make a model. The current which arises due to the photon flux can be conveniently represented as a current source. We can leave the diode as a diode, and we have the circuit shown in Figure 2.33. Even though we show I_{out} coming out of the device, we know by the usual polarity convention that when we define V_{out} as being positive at the top, then we should show the current for the photovoltaic, I_{pv} as current going into the top, which is what was done in Figure 2.33. Note that $I_{pv} = I_{diode} - I_{photo}$, so all we need to do is to subtract the two currents; we do this graphically in Figure 2.34. Note that we have numbered the four quadrants in the I-V plot of the total PV current. In quadrant I and III, the product of I and V is a positive number, meaning that power is being dissipated in the cell. For quadrant II and IV, the product of I and V is negative, and so we are getting power from the device. Clearly we want to operate in quadrant IV. In fact, without the addition of an external battery or current source, the circuit, will only run in the IV'th quadrant. Consider adjusting R_L, the load resistor from 0 (a short) to ∞ (an open). With R_L, we would be at point A on Figure 2.34. As R_L starts to increase from zero, the voltage across both the diode and the resistor will start to increase also, and we will move to point B, say. As R_L gets bigger and bigger, we keep moving along the curve until, at point C, where R_L is an open and we have the maximum voltage across the device, but, of course, no current coming out!

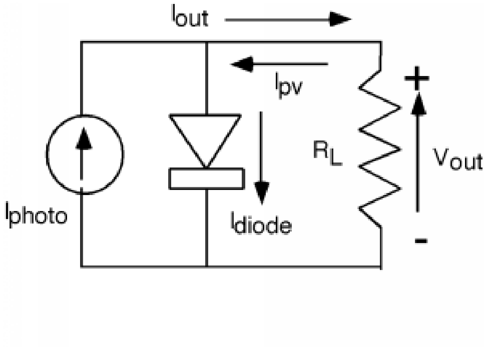


Figure 2.33: A model of a PV cell.

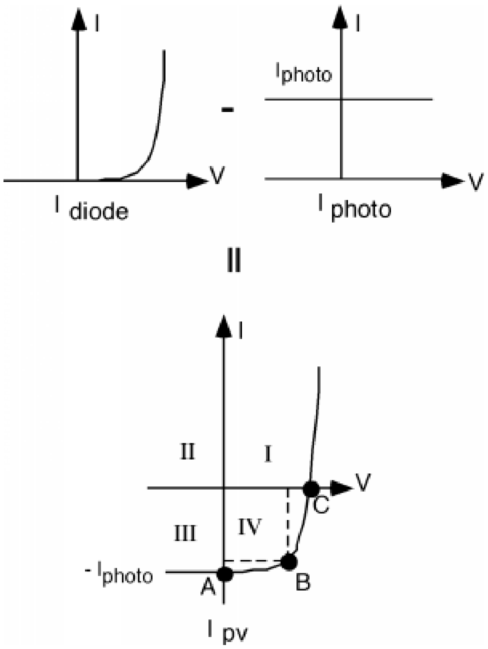


Figure 2.34: Combining the diode and the current source.

Power is VI so at B for instance, the power coming out would be represented by the area enclosed by the two dotted lines and the coordinate axes. Someplace about where I have point B would be where we would be getting the most power out of out solar cell.

Figure 2.35 shows you what a real solar cell would look like. They are usually made from a complete wafer of silicon, to maximize the usable area. A shallow (0.25 μm) junction is made on the top, and top contacts are applied as stripes of metal conductor as shown. An anti-reflection (AR) coating is applied on top of that, which accounts for the bluish color which a typical solar cell has (Figure 2.36).

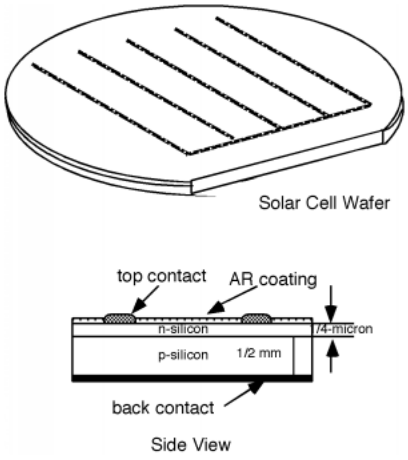


Figure 2.35: A schematic diagram of a real solar cell.

Figure 2.36: A solar cell showing the blue tint due to the AR coating.

The solar power flux on the earth's surface is (conveniently) about 1 kW/m^2 or 100 mW/cm^2. So if we made a solar cell from a 4 inch diameter wafer (typical) it would have an area of about 81cm^2 and so would be receiving a flux of about 8.1 Watts. Typical cell efficiencies run from about 10% to maybe 15% unless special (and costly) tricks are made. This means that we will get about 1.2 Watts out from a single wafer. Looking at B on 2.59 we could guess that V_{out} will be about 0.5 to 0.6 volts, thus we could expect to get maybe around 2.5 amps from a 4 inch wafer at 0.5 volts with 15% efficiency under the illumination of one sun.