Introduction to Semiconductor Devices

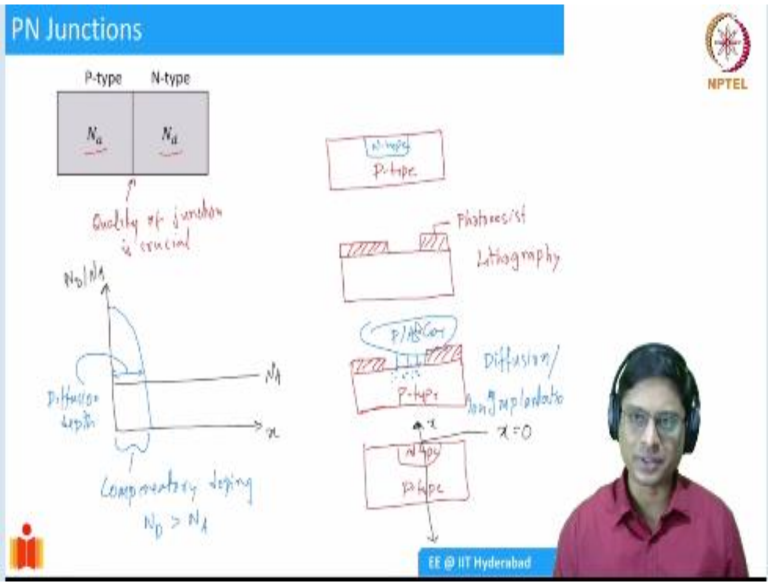
Dr Naresh Kumar Emani
Department of Electrical Engineering
Indian Institute of Technology – Hyderabad

Lecture – 4.1

P-N Junctions - An introduction

Hello everyone, welcome back to Introduction to Semiconductor Devices. This week, we will start discussing one of the most fundamental semiconductor devices, which is a PN junction.

(Refer Slide Time: 00:21)



PN Junctions

So, a PN junction is simply you know, can be thought of as something that is shown here. So, we can take a piece of P type semiconductor and a N type semiconductor. So, P type semiconductor is doped with acceptor ions and N type semiconductor is doped with donor ions and then if they form a junction, then you can have a PN junction so, simply as a name implies. But it is not that simple.

The reason for that is this particular junction, you know, it is very, very crucial; quality of this junction is crucial. If you simply take a piece of N type semiconductor and a piece of P type semiconductor and join them together, we are going to have a lot of defects on the interface. Because we saw that, you know, the lattice matters a lot. And whenever the lattice breaks at the edge of a semiconductor, you have a lot of what to call is dangling bonds so, unsatisfied covalent bonds.

When you have these dangling bonds that can result in trap states within the band gap. And that can deteriorate the quality of electronic devices. So, you cannot simply take a piece of N type and P type and put them together to form a PN junction. So, what can we do? Well, we have to follow a series of processing steps to realise that PN junction. So, what we can do is in its simplest form, we can take a wafer.

Let us take a piece of wafer, which is, you know, let us assume that it is doped P type, uniformly doped with P type semiconductor, P type dopants. So, P type semiconductor will take and then I will convert some region of it into N type. The way I can do that is by doing a series of processing steps. So, if you look at the latest technology right now, if you want to make a chip, the process steps involve something like 2000 or more steps.

It is a very, very complicated process. But you know, we can simplify it, we will say that I want to have a small region in the centre here; I want to make it N type. This is what I want. So, how would I make it N Type? So, what I can do is, I will take this P type wafer and then I will do a processing step which is called as lithographic. So, essentially, what I will do is, I will cover this wafer with what is called as a photo-resist.

The way to do that, again, it is involved; we do not need to get into it. I will cover it. So, essentially, what this photo-resist does is; it will block some type of atoms. So, I will do this lithographic step. And then I will do a step which is called as diffusion. So, what we can do is we will take this wafer covered with photo-resist and then put it in a chamber containing blueprints.

For example, if we want to make it N type, we need to put it in a chamber containing phosphorus or arsenic atoms; you can have a gas of phosphorus and arsenic, heat it up so that the phosphorus and arsenic atoms get a lot of energy. And then those that energetic ions will actually get into; will diffuse into the wafer. So, originally, this was a P type wafer and you could imagine basically a gas of everywhere.

You create a chamber to put it in a chamber containing; I will just show gaseous cloud here essentially. So, basically this will be phosphorus or arsenic gas, which is heated. So, when it is heated, some of these atoms will actually hit this surface and then go in; diffuse into the substrate. So, that will actually cost the doping to change. This sort of process, I will call as diffusion. So, I will do a diffusion step.

So, simply, you know, if you take a room and then you open an alcohol bottle let us say, the alcohol vapour will come out and then you can smell it. So, it is alcohol atoms, vapours are diffusing in the room. Similarly, here, the gaseous atoms will diffuse into the lattice. Of course, because this is a lattice, it is a solid; the rate of diffusion is very, very low. It is not easy for the gas to actually go through the lattice.

So, that is why we have to increase the temperature. Typically, this is done at 700, 800 degrees centigrade or so, high temperature, so that the diffusion happens faster. So, we can do this process, or this was originally in the 60s and 70s ways to do this, but nowadays, more of it is done by what is known as implantation, I will also mention that implantation. So, what is ion implantation?

So, essentially, what we do in this process is as I said, diffusion is slow. So, we do not depend on diffusion, but instead we accelerate these ions, phosphorus, and arsenic ions, you can accelerate. We can pass it to a particle accelerator and then they acquire a higher energy, kinetic energy and then you bombard them into the lattice. So, you can make the doping process faster.

And there are some advantages of course, you know, and there are disadvantages as well. So, anyway, that is the topic of a processing or VLSI technology course. For us essentially, it is enough to know that we get a diffusion of these ions. So, essentially you have some arsenic or phosphorous atoms penetrating into the P type substrate and then we will simply remove the photoresist.

And then what you will be left with, something which is looking like this. So, P type and then I have a region of N type. That is how we will realise a PN junction. So, if you are paying attention, you might recognise that we did what is called as compensated doping. Remember, compensated doping we mentioned in the last week. So, essentially if you try to, let us say plot, I will take this axis and I will label it as x.

If you plot, you know, the doping concentration along x and I will take this to be the junction, x = 0, I will take it. We will also call it metallurgical junction, essentially the junction between the P and the N regions that the point where the type of the doping, so we will plot it that way. Let we simplify it. We will start using this notation in the next part.

Right now, let us assume this is x = 0; the top of the vapour is x = 0. I did this for a reason anyway. So, let us plot out the doping as a function of x, this x, I should take x is going downwards. So, this is doping. Let us, this will be N_D or NA. So, when I do this, essentially, you can originally see the fitting, the original wafer, that was a P type of wafer. So, I will have a doping concentration which was uniform.

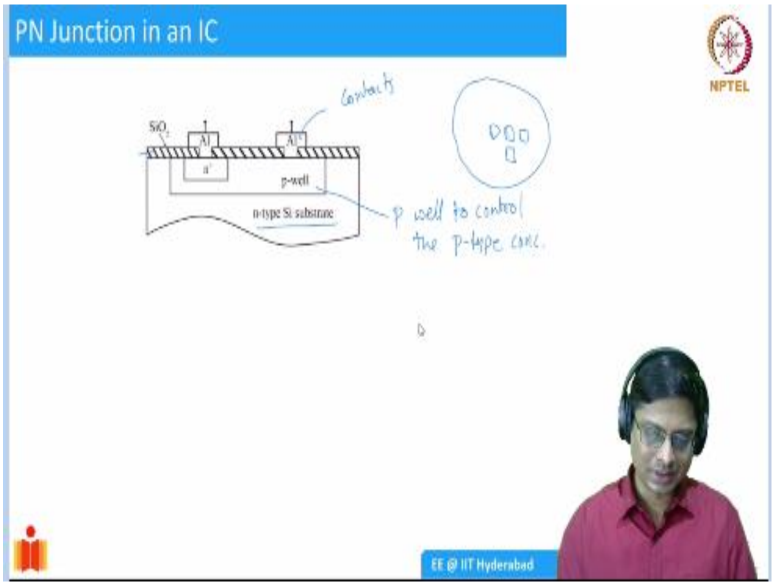
So, I will have N_A uniformly doped region and then I introduced N type dopants by diffusion process, but this diffusion process is not going to be, you know, not going to produce uniform doping. So, typically, diffusion process produces a shape which is looking like this. So, you have more of the dopants on the top surface and as you go into the wafer, the concentration reduces.

So, essentially, this is going to be the depth of diffusion, diffusion depth you could say, so, this can be controlled. Let us say for example, I take higher temperature, I can increase the diffusion depth. If I do it for a longer time, I can increase the diffusion depth, but there has some limitations. Within some limitations, we can treat these parameters and you get some profile like this.

If you are having implantation, the profile will be slightly different, but for practical purposes, we can assume the same for as far as P-N junctions are concerned. So, if you get this, so what is happening? Essentially, in this region, you have compensatory doping. So, if you really want to make it P type, we need to make sure that sorry, if you really want to make it N type, you need to make sure that N_D that you are doping that is greater than N_A, only then you get the N type doping.

If not, you do not get the N type doping. So, this is a serious limitation for us. But so, we will tell you how to overcome this in a moment. So, this is how you fabricate a PN junction.

(Refer Slide Time: 10:12)



PN Junction in an IC

So, now let me try to show you a slightly more complicated process, this is actually the cross section of an actual PN junction that might be fabricated in an IC. You see that you know, it is a little bit more involved. We said that you know if you have a fixed type of dopant, P type. If you take a P type substrate, then you do not have a whole lot of flexibility, because everywhere all across a wafer, your devices have to have the same type of doping, same amount of doping that is not advisable or that is not convenient for us.

We would like to have independent doping, you know. Remember, if you have a wafer, which is like let us say 12 inch wafer, you will be making I do not know, 1000s and 1000s of diodes, let us say on it. We are going to make like multiple diodes on the same wafer, you are not going to use an entire wafer for one diode. If you use the entire P type substrate as you know as one of the P regions, then you do not have that much flexibility in changing the doping concentrations.

So, we introduce what is known as, we start with an N type wafer again, this is one type of process where you could always have different, different types of processes depending on what is convenient for us. So, we start with an N type wafer. P well, essentially it is introduced to control the P type concentration. So, I can make 2 adjacent diodes and both of them can have a different P type concentration.

So, I can have different P wells depending on you know, I can adjust the dose in my implantation step. If you use diffusion, it might not be possible but ion implantation has an advantage that you can actually adjust the dose locally and you can have different P wells for different types of P type dopants. And then of course, introduce N type well. Well, that again is you know, again a lithographic step has to be done; even for the P well, there has to be a lithographic step.

So, you create a you know, photo-resist mask. Mask is essentially something that will block the diffusion or implantation process. So, you open a window where you want a P well implant or diffuse, then again do a photolithography, open a smaller window, implant or diffuse the N type wafer N type sorry, N type dopant. So, that way, you can get a PN junction, but you also want to have connections.

Always, we want connections to the external world you cannot simply leave it like that. So, we have to make what are known as contacts. Remember, if you are talking about bulk size you know, a macroscopic devices, it is very easy for us; it is easy to visualise, but you have to do it on the micrometre scale or even; right now P-N junctions are made in the nanometer scale.

If you want to have that sort of precision, think about what needs to be done. You cannot do it like what you do in the regular welding kind of thing you can do. It is a much, much smaller dimension. So, the whole field of nanotechnology right now whatever is known as nanotechnology, most of it is developed for the sake of advancing the semiconductor devices. I mean these tools and techniques, whatever I mentioned diffusion, lithographic and so on implantation, all these were developed, so that we can achieve nanoscale devices.

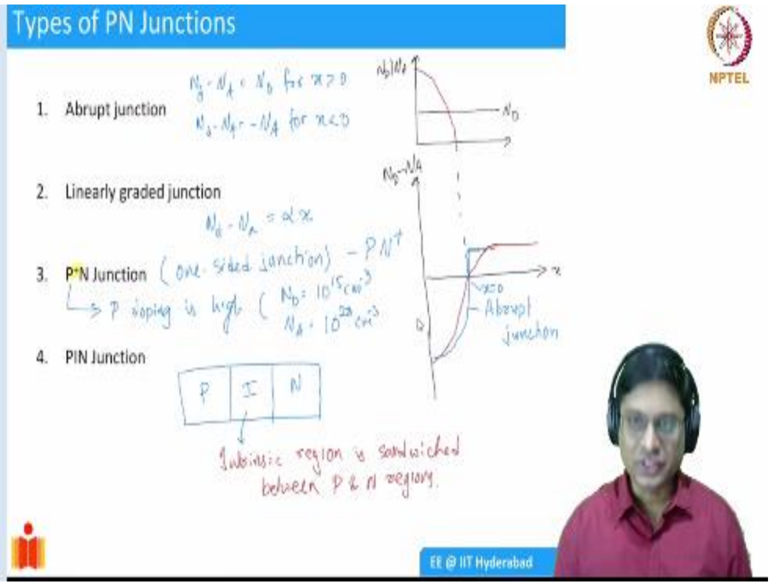
So, I would say that nanotechnology whatever we know today is a by-product of semiconductor revolution. So, we can make this P type; PN junctions and then we have to make a contact. How will you make a contact? Again lithographic step. You do lithography step, put a photo-resist or you know, in this case, we are actually; photo-resist is actually; it can be removed easily. It is an organic material.

So, if you want to actually protect and you want to insulate the PN junction, you will put in silicon dioxide and then open a small window in that. This is insulation silicon dioxide. And the way to open it into a window is basically to put a silicon dioxide and then try to etch it in a small window. Again, it is a little bit more complex, but I will make a connection to aluminium.

In this case, I have deposited aluminium, showing that it is an aluminium. So, then you can make a connection. So imagine you know, this is just the lowest part of the most basic level. If you have let us say 10 diodes on your wafer, how will you connect it? Imagine you know, you are actually making these diodes in, let us say, you know, at the spacing of you know, 100 nanometers each, then you want to connect them in some way, we have to think of how to metal; I mean, you have to run a metal line.

You might have seen a PCB in a computer or you know, in a TV. If you open a PCB, you will see a lot of these copper wires going all around. So, those are all connections between devices. We do something similar in IC technology, wherein we run various levels of metals and then we make these interconnections. So, that is how you make a PN junction in the realistic sense.

(Refer Slide Time: 15:21)



Types of PN Junctions

1. Abrupt junction $N_D - N_A = N_D$ for $x > 0$; $N_D - N_A = -N_A$ for $x < 0$
2. Linearly graded junction $N_D - N_A = \alpha x$
3. P^+N Junction (one sided junction) – PN^+
   → P doping is high ($N_D = 10^{15}\,cm^{-3}$, $N_A = 10^{20}\,cm^{-3}$
4. PIN Junction

Intrinsic region is sandwiched between P & N regions.

Now, let me tell you the various types of P-N junctions. So, we briefly mentioned this, you know, this sort of a distribution of doping concentration, wherein, initially we had a, let us say, substrate with let us call it right now, I will take it as, for convenience, I will take it as N_D. That means I am having a N type substrate and N_D or N_A will be the doping concentration.

So, basically, I can have a diffusion process, which will create my P region. So, if you do that, how will the doping profile look like? That is what we want to identify. So, to do that, what I will do is, I will make a; I will draw plot of $(N_D – N_A)$ versus x. So, you see that, you know, in this particular region, you have both N_D and N_A, so, $(N_D – N_A)$ will be lower.

So, if you are large x, we do not have a flat N_D like this. And as you come closer to this, let us this point, then your doping concentration changes, and it will become something like this. So, essentially, I am doing both in the same plot, so I am just showing (N_D – N_A). So, this is an important feature by actually controlling you know, the way the profile of you know, diffusion profile looks like, I can control the way your junction looks like this sort of behaviour.

Now, ideally, what you would want is you know, a profile like this very, there is a very sharp cut or sharp change between N type and P type agents. So, this, I would call as abrupt junction. So, wherein my profile basically is that $N_d - N_A = N_D$ for x greater than, let us call it the junction to be 0; right now, I am taking it to 0, x =0; I am taking this, change the notation method, please just note that.

In the previous graph, I wanted to show you from the top of the wafer, so I took x = 0 in the top, right now I am taking the junction x = 0. So,

$$N_d - N_A = N_D \ for\, x > 0$$
$$N_d - N_A = -N_A \ for\, x < 0$$

So, I have a junction which is very, very sharp that is known as an abrupt junction. But you see that it is clearly not going to be that you know, easy to get an abrupt junction, you can get an approximation to an abrupt junction, but most often are not, it is going to be something slightly different.

So, most of it, maybe it might be something like this, wherein your $N_D - N_A$ at the junction is actually proportional to x. There is a slope. So, you can also have junctions which are having doping concentration of the form.

$$N_d - N_a = \alpha x$$

So, what I am trying to say is the doping concentration is going to be proportional to x.

So, this is called as a linearly graded junction. So, I mean, as the name implies, the doping concentration is linearly changing with x. You can also have more complicated doping profiles, second profiles and all that, but it is anyway; it is a better approximation to the real situation that is why people study them, but for this course, you know, we will just solve a couple of problems with linearly graded profile.

Otherwise, most often, we need with abrupt junction that is a simplest form for us to analyse. You can also have what are known as P^+N junction. I do not know if you notice in one of the previous cases I have, there was a mention of P^+ or N^+ region. This P^+ means the P doping is high. So, basically related to the N doping, P region is much more highly doped.

So, it could be something like N, let us assume if N_D was, let us say 10^{15} cm^{-1} ,N_A let us say is 10^{19} or 10^{20} cm^{-1}. So, basically the P is highly doped related to N that is why we call it as P^+ junction. So, sometimes, these are also called as one-sided junction. It will become clearer in the next couple of lectures why it is one sided; one sided junction we call them.

You could also have you know, PN^+ junction, N can be highly doped related to P that is also possible, it is not going to make any difference, but anyway P^+N junction or PN^+ junction. So, basically one of the materials is highly doped, other materials moderately doped, so, we will call this. And you can also have another interesting sort of a semiconductor device which is known as PIN P ion diode, PIN junction.

So, here what we do is, we have the same material let us say, but now I have P region, N region and intrinsic region in between basically, here intrinsic region is sandwiched between P and N regions. So, I leave some part of it, let us say it is perfectly compensated so, you have effectively intrinsic region. This has a lot of applications in especially optoelectronic devices.

We will see that in the last couple of weeks of the course, but anyway, in the next week, we will see. In this week or next week, we will see how to analyse a PIN junction. It becomes very interesting. We understand the electrostatics of the PIN junction. And then optoelectronics part, we will see in the end. So, these are different types of junctions. So, in the next module, we will talk about the electrostatics. So, I will meet you there. Thank you so much.