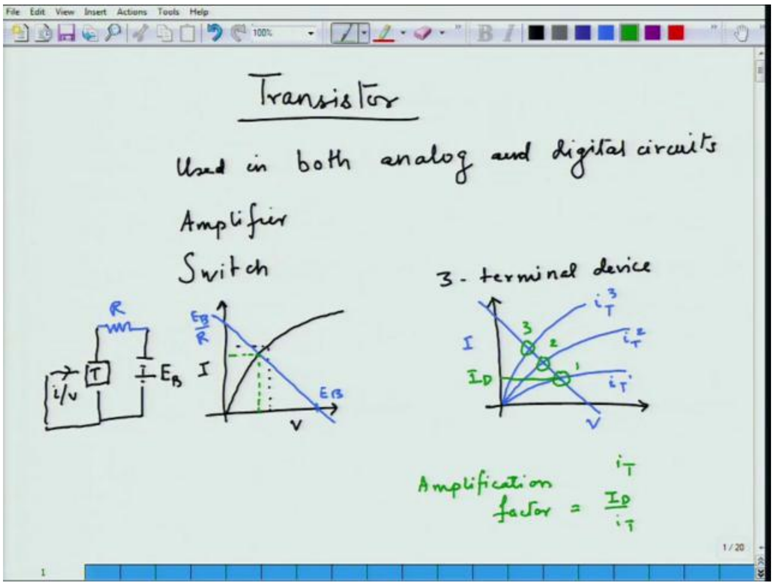
Optoelectronic Materials and Devices
Prof. Dr. Monica Katiyar
Department of Materials Science and Engineering
Indian Institute of Technology, Kanpur

Module - 4
Optoelectronic Device Physics
Lecture - 37
Transistor

Today, we are going to discuss a very important device, transistor.

(Refer Slide Time: 00:26)



Transistor

Used in both analog and digital circuits

Amplifier

Switch

3-terminal device

Amplification factor = $\frac{I_D}{i_T}$

As you know that transistors are very important, they are used in both analog and digital circuits. They can be used for and they are used for basically for two functions; as a amplifier and as a switch. So, if we look at this very important device. Let us first try to understand what kind of a electronic device we are trying to generate. So, if you before we get in to the details of how this device is made. How we get the characteristics from semiconducting device transistor, characteristics from that?

Let us just spend some time to understand the applications or what kind of characteristics we want in a transistor. So, if we have a device for example, in a circuit, which is a two terminal device. I apply voltage across this. Then, normally I am expecting I V characteristics from this two terminal device, which I am going to call it, make any arbitrary shape, may be of this nature. So, depending on what is the voltage that I will apply, I will get a current.

I have no other control on this particular device. Now, if I changed this device into a three terminal device and transistor is a three terminal device. I am going to provide a control. So, this is a transistor. I am going to provide a control through a third terminal to this transistor. This control could either be a current or it could be a voltage. I am trying to see if I can change different type of device curves in this case. So now, I have a device in which I have provided a third terminal to my transistor.

This third terminal current or voltage applied can has a capability of changing the output of the device. So, this could be the third terminal current. It could either be a voltage also. One value, at a different value will have a different I V characteristics and the third value. So, this is the importance of a three terminal device, where the third terminal signal is being used to control the I V characteristics of the device. Now, how is it used with an amplifier or a switch?

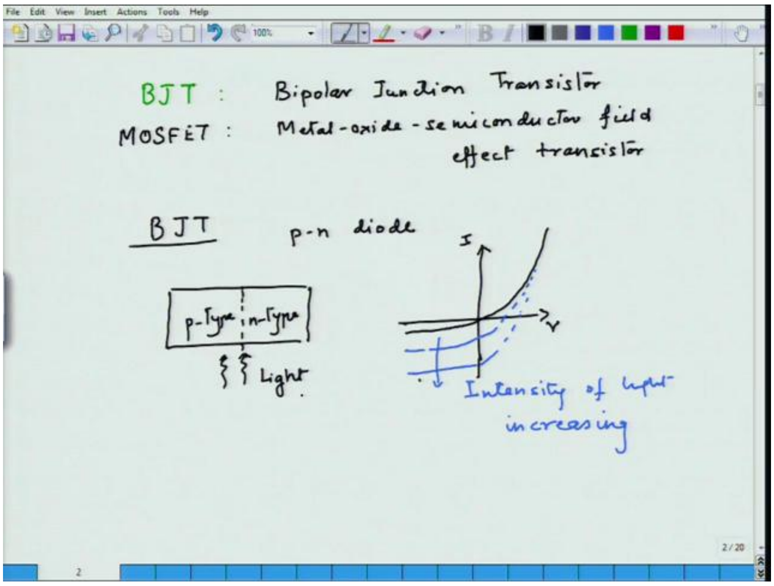
To understand that, let us put a load in this circuit. If I put a load on this circuit and I want to see how the circuit operates across this load. Then, I need to in order to find the operating condition. I should make the load line, which is going to be like this; this would be the value, if for the resistor that I have put here. For the load resistor is this; is when there is no current flowing. This is the value for the current in the maximum value for the current in the resistor. This particular device then will operate at this point and at this voltage and current.

Now, for our two terminal device, you have only one operating position for the load in this particular case. But if I have the same load with a three terminal device, then I have possibly three operating conditions. So, this would be the first operating condition second and third. What is a switch here? The value of the third terminal signal in this case, I am showing current. It could also be voltage; will decide whether the device is going to operate at point 1, 2 or 3. So, that is how a transistor which has a third terminal to control its response; can give you a switching affect.

In the same manner, if I look at this signal; if I have a third terminal, which has a value of i T. Third terminal signal is i T in the circuit. I am getting a signal which I will call I D. Then, I am getting an amplification factor in this three terminal device of value I D over i T. Of course, when I say amplification, I need to make a device such that I D is going to be many times the value of i T. That is possible to make in these devices.

So, you see how a three terminal transistor device can be used either as a amplifier or as a switch. These functions are very important for many electronic devices. In order to know more about this, one can refer to some courses in the electric engineering department. Now we talk about more on the device aspect only. What kind of devices can I make which will give me transistor characteristics?

(Refer Slide Time: 06:32)



BJT : Bipolar Junction Transistor

MOSFET : Metal-oxide-semiconductor field effect transistor

BJT

p-n diode

So, there are two kind of solid state transistors two basic kinds of solid state transistors; one is called BJT. This is bipolar junction transistor. The other one that we will discuss is MOSFET which from earlier lecture we have had discussion metal oxide semiconductor field effect transistor. By far, MOSFETs are more commonly used compared to the BJTs. They have definite advantages over BJT's. But, for the discussion in this course, let us discuss the main operating principles of both of these transistors.

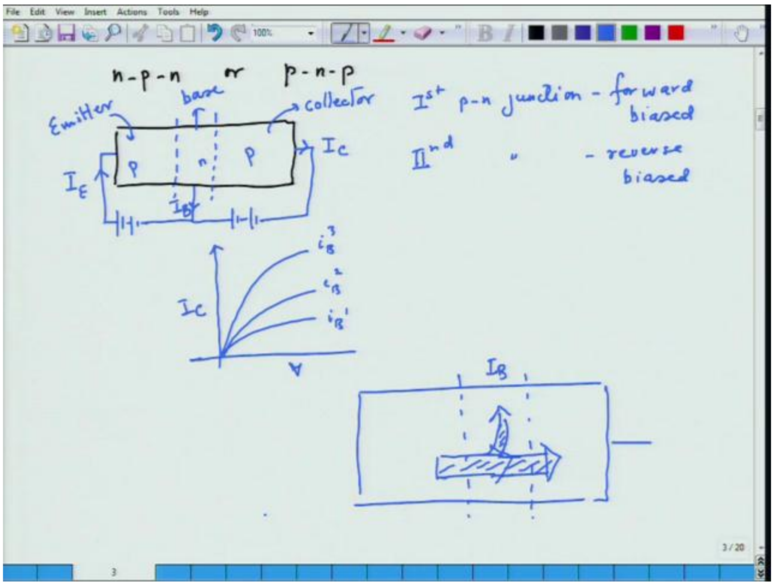
So, if I first let us talk about BJT's. The basic principle of a BJT is based on the p n diode. In our earlier lectures, we have seen that if I have a p n diode, the p n junction where the material is p type here and n type here. Then, if I connect it to a circuit and apply proper bias, I will get a behavior, a diode behavior, which this equation we have derived earlier.

Now, this particular diode, if it is working in the reverse bias; at that point, if I am going to apply let us say take a simple example. If I shine light on it; then, we know that this is a situation where this diode is been used like a solar cell or as a photo detector. What happens in that case is as we shine more light, we create more carriers in the depletion region of this p n diode.

The reverse current increases in this manner; this is the direction in which the intensity of light is increasing. So, if we now look at this curve, we will see that by just taking a p n junction and applying light to it, we have created a transistor. You may not see it in the same way that was drawn earlier. But, if I reverse this basically, if I take the negative access to the positive access; this is basically a junction behavior. I have a device whose response can be changed by applying how much light intensity I shine on it.

So, in this particular case, in these particular examples, the light emission is my third terminal to that I need. But when I am making an electronic device generally, I am not going to use light as a third terminal. So, in order to then make a electronic device, what we do is we use another p n junction to inject carriers in a in a reverse bias a diode. Hence, the bipolar junction transistor is a either n p n combination.

(Refer Slide Time: 10:40)



Ist p-n junction - forward biased

IInd " - reverse biased

There are two p n junctions here; n p n p n or p n p. So, in this particular case, let me first take the example of p n p here. The other one can be worked out in a similar manner to this discussion. So, what I am going to use here is I am going to have two junctions. The first junction is the p n junction. This is p type, this is n type. The second junction is very close to the first junction which is n p junction. So, I am going to now bias these junctions such that that the first p n junction is biased forward, forward biased. The second one is reverse biased.

So, in this situation, if we look at what will happen in this device, we will find that if I am going to bias it. The bias would be such that it is forward biased. For the second one, the bias would be such that it is reversed biased. Then, I am going to have a current flowing through this one, which we call the emitter current. This particular part is of the bipolar junction transistor is called emitter. There will be a current, which we call coming out from the middle portion, which we called base current.

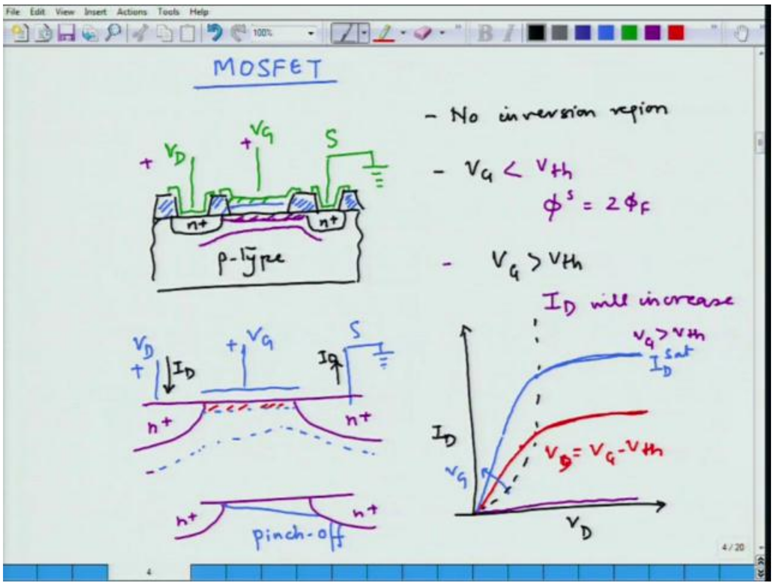
This part is called the base. Then, there is a current which is coming out from here, which we call collector current. This part of the bipolar junction transistor is called the collector. Now, in this situation, the first junction, which is forward biased, is going to have inject lot of holes in the base region. The base region is this. The base region is taken such that that these injected carriers then do not get enough time to recombine.

Then, they get into the depletion region of the second p n junction and are and go through the circuit because of it being reversed bias. So, whatever current you are going to flow inject carriers, you are going to inject in the first junction are basically taken up by the second junction. They flow through the device. So, in this particular bipolar junction, we get a behavior in which the third terminal is the base. It is the base current which is the signal.

The behavior for the collector current versus the voltage for different base currents is going to be of this kind. Here, we have a transistor. So, using a single diode and using two diodes in combination, we get a bipolar junction transistor. We use the base current as a third terminal to control the current of the device in this bipolar junction transistor. The important part in this particular discussion device is that whatever in the base whatever current is or carriers are injected in the base, they must not recombine in the base.

So, in this particular case, if the holes are injected in the base, from this forward bias junction, most of them should continue and get collected by the collector in the outer circuit. Only a very small part goes to the base current. So, the total injected carriers are collected. For that reason the device structure should be made correctly so that we get a good transistor behavior. So, this is a bipolar junction transistor. We can have n p n also and get the transistor behavior from this particular device. Now, let us discuss the other important transistor which is MOSFET.

(Refer Slide Time: 15:37)



MOSFET

- No inversion region
- $V_G < V_{th}$

  $\phi^s = 2\phi_F$
- $V_G > V_{th}$

  I_D will increase

Now, for this discussion, we are going to base our discussion on the knowledge of the earlier lecture where we discussed metal oxides semiconductor capacitor. So, we are going to see how we use that metal oxide semiconductor capacitor, first to make a transistor. So, what did we do there? We had a semiconductor let us say, this particular case. Also, I take a p type semiconductor. On top of that semiconductor, I have oxide. I am making contact through the metal on top. So, this is making the metal contact. Here, I was applying a voltage.

In case of capacitor here, the voltage applied, we call the gate voltage with respect to a reference which we will define little later. The p type semiconductor is also contacted where we apply the bias with respect to the device here and see what happens. So, just to review what we did last time was that depending on what is my electronic structure is, what is the work function difference in other things are and what is a applied value of VG? I am going to have different states of carrier accumulation or depletion or inversion in this portion of the semiconductor. That particular behavior, I am going to use in a MOSFET device.

So, let me make the structure of a MOSFET device around this particular MOS capacitor. So, this particular oxide is grown and then metal is contacted. Now, I need to make use this channel. So, I need to make contacts to that channel. So, I am going to make these two portions of my semiconductor n plus. In order to make these two portions, I must have had some amount of oxide here. This is for the MOS portion. This is also the oxide. I am making contacts to the gate metal; that is the contact to the gate metal. I am going to make contacts through this n plus portions also to connect to the channel.

This I am going to call drain and source. Since, this is a symmetric device, it does not matter which one I call drain and which one I call source. So, there is going to be I am going to call this the source. I am going to ground it. This is going to be voltage applied here which is V drain. So, here I have a device now. What I have done here is taken the MOS capacitor and built around it a structure to access the channel which is made by the MOS capacitor.

Now, what I need in order for this device to conduct. If I am going to apply a voltage between source and drain let us say V G is not there, third terminal is not there. If I am going to apply a voltage between source and drain, these are n plus. This is p type and there is going to be a diode created here. One of these diodes whether going from n to p or p to n is; one is going to be forward bias. The other one might be reverse bias. In that case, I have not going to have any current flowing through here.

So, when there is no gate voltage, when there is no accumulation of charges, when no inversion region then, even if I apply a source drains voltage; that is going to be no conduction through here. Then, I apply V G. When V G is increased, first I will get a depletion region. Depletion is going to be created in this manner because if I am applying a V G source is 0 and V D is at some positive value. Let me take this to be a positive value.

Hence, there is a less, more depletion in the V D side and less in the source side. So, there will be a depletion region if I continue to increase V G. When V G is less than the threshold voltage, if you recall the threshold voltage; in case of a MOS capacitor was the value at which the inversion sets in. This we have discussed earlier at which point phi of S was equal to 2 phi of F for a MOS capacitor. So, when V G is less than V threshold, we only had depletion.

When I have the value of V G which increases from threshold, it means inversion layer starts. Then, I am going to get a layer here which is going to have in a p type semiconductor electrons. This layer is then going to connect the source to the drain. I am going to have a current flowing through the device. I D will increase. So, let me summarize what I have just said. In this particular device, where we have a MOS capacitor in the center if there is no voltage applied and there are no access carriers in the channel here.

Then, when I am applying a drain voltage, there will be no current flowing because this is a n p p n combination here. When I increase the gate voltage, then a current will flow because basically by when I have a channel, here I am reducing the resistance here. Initially, the resistance is too high. When the resistance is reduced there is going to be a flow in the current and I am going to increase the flow of current. So, if I show this behavior in a more simpler schematic, I can plot only the behavior of source and drain and the enlarge it.

So, let us say this is n plus the depletion region is in this manner. The channel is created close to the interface with the gate when I am going to apply V G. Here, I have the source which is grounded normally and V D which is positive. So, this is the situation in which I am looking at the whole behavior and rather increase the V G. The current is going to change. So, let us look at the I D. If I plot for this particular device, for different values of V G, the drain voltage V D refers to the drain voltage and the drain current. The drain current is normally shown in this direction positive I D and in this direction I D here.

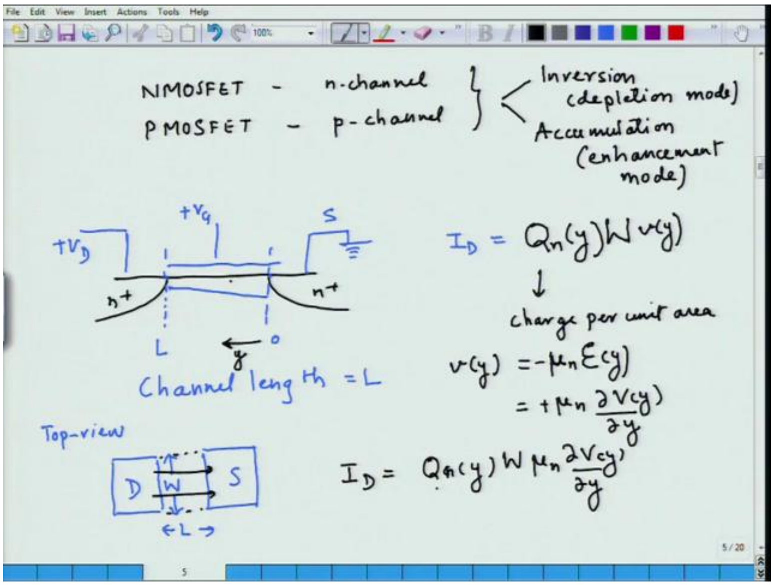
So, if I am plotting I D, V D when there is no voltage applied, what I am going to see? I am going to see absolutely no I D because the resistance is too high between the source and drain. There is going to be no current flowing. When V G is greater than V threshold for the MOS capacitor then, I am going to start seeing increase in the current. As the V G is increasing then my charge accumulation here increases. If I am going to see more charge and accumulation when the V G is increasing; so, there will be less resistance.

So, this is V G increasing in this direction. So, V G is greater than V threshold. But then, it continues to grow and finally, it saturates out. So what is happening at higher V D? So, at a particular given V G, if I am applying the drain voltage for this value, there it works as resistors. But, at one particular point what we see is that the V G becomes equal to; the drain voltage here becomes equal to V G minus V threshold. At this particular point, I have basically the voltage drop here between the gate and the capacitor is 0.

As a result, this particular channel is pinched off. So, if I plot only this particular portion the channel at this value is going to look the channel is stop getting pinched off right at the edge of the drain. As a result now, there is increase in resistance and the current gets reduced after this pinch off point. Then eventually, it saturates out. So, if I plot for different V G same behavior occurs. When I increase the V G, there is more carrier accumulation. Resistance is less. Eventually again, there will be a point where V D will become equal to V G minus V threshold.

At that point again, there is pinch off of the channel. This is known as pinch off of the channel. Beyond that point, the I D saturates. This is called I D saturation. One can plot the line for the pinch off for this behavior. So, you can see how a MOS capacitor is changed into a transistor. This is a behavior that we are looking for. We are looking for a transistor behavior in which the third terminals in this case. This is not the current. But, the voltage is used to get a different I D V D behavior for different values of V G. Now, this particular MOSFET what we just discussed was made in p type semiconductor. The channel was n type. Hence, we talk two different kind of MOSFET.

(Refer Slide Time: 27:54)



NMOSFET - n-channel
PMOSFET - p-channel

Inversion (depletion mode)
Accumulation (enhancement mode)

$I_D = Q_n(y) W v(y)$

↓

charge per unit area

$v(y) = -\mu_n \mathcal{E}(y)$

$= +\mu_n \frac{\partial V(y)}{\partial y}$

$I_D = Q_n(y) W \mu_n \frac{\partial V(y)}{\partial y}$

One is called n MOSFET which is n channel. The channel which is created due to inversion one is called n type. The other one is p MOSFET where the channel created it is a p channel. Further, there are more variations in these MOSFET. Each of them can have, can be either created. The channels can be created by inversion as we discussed. That is known as a depletion mode. It can also be created in accumulation. This is the term that comes from the MOS capacitor when the majority carriers are getting accumulated near the semiconductor oxide interface. This particular transistor is then called enhancement mode.

So, it is clear that the similar behavior of the transistor can be obtained in different type of materials. You can obtain it by different mechanisms. Now, let us look at what if we can figure out how much is this value of the I D that I am going to get in a MOSFET. In order to try to analyze that, let us look at the channel itself. If we continue our example this is a n type n MOSFET. A channel that we created and I am going to enlarge the figures to explain things is over here.

I am making this as if it is same. But, it is not going to be the same because we apply a V G here. This obviously is at this n it is 0. The source is 0. I am going to apply V D source here. It means that at this particular point the voltage difference at this point is the gate bias minus the 0. But, a voltage difference at this point is a gate bias minus the drain bias. So, the voltage is not the same as we grow go in this particular channel from source to drain.

So, if I take define this point as 0 for the channel and this point L. So, channel length for my MOSFET is L. In that case, the voltage applied here is more compared to here. This is always going to be in this particular nature. Now, we want to calculate what kind of current what would be the I D that will be flowing through this channel that has been created. So, the current that will flow through this channel is going to be equal to whatever is the charge here multiplied by the speed at which that charge is moving.

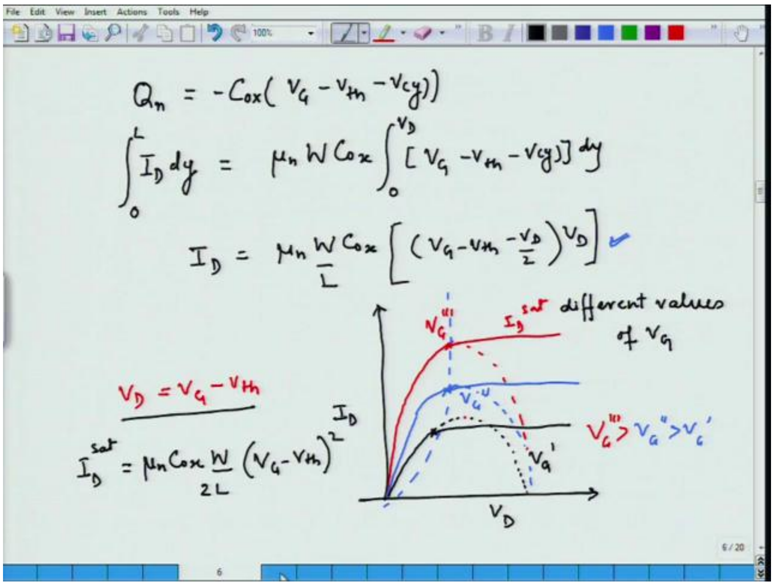
So, if I take for this channel then this channel also has a width. If I am going to plot the top view of the MOSFET then, basically the contact the drain contact and the source contact. This is a drain. This is a source contact. Below, we have the gate. This is the width of the channel. This is the length of the channel that we have enlarged over here. So, the current that I am trying to calculate between source and drain; the I D current is going to be equal to in this particular channel that has been created. How many charges are flowing per centimeter square per second from drain to the source.

So, the charges that are flowing is going to be equal to the I D times total amount of charge in the inversion layer the n channel layer. That as I just mentioned is going to be a function of y. So, I am going to call this direction as y and multiplied by this. Q is the charge inversion charge per unit area. So, any point y on in the channel will be the charge per unit area multiplied by the width of the channel multiplied by the velocity at which the carriers are moving at point y.

That will depend on what is the voltage at different points. So, it is going to be a function of y. That is why it comes into this equation. Now, if we analyze this further, we can find out what v of y is going to be. So, the velocity of carriers at particular point y we can write that from our earlier discussion on drift. It will depend on the mobility multiplied by and it is a n channel; so, for electrons multiplied by the electric field at point y. The charge for electron is negative.

So, it is going to be in the negative direction. Electric field, if we take the assumption that the voltage between source and drain is V D, is going to be gradient of V at point y. Again a negative sign, that takes it out. So, we have velocity of a carriers at point y is going to be given by this. So, the total current I D as I want to figure out is going to be equal to Q of n at point y multiplied by the width of the channel and mobility of electrons in the channel and the gradient of the voltage in the channel. Now, I need to estimate what this charge is going to be. For that we will go back again to our MOS capacitor discussion. So, what is a charge in the channel?

(Refer Slide Time: 34:38)



$$Q_n = -C_{ox}(V_G - V_{th} - V(y))$$

$$\int_0^L I_D\,dy = \mu_n W C_{ox} \int_0^{V_D} [V_G - V_{th} - V(y)]\,dy$$

$$I_D = \mu_n \frac{W}{L} C_{ox} \left[\left(V_G - V_{th} - \frac{V_D}{2}\right) V_D \right] \checkmark$$

$$\underline{V_D = V_G - V_{th}}$$

$$I_D^{sat} = \mu_n C_{ox} \frac{W}{2L} (V_G - V_{th})^2$$

Charge in the channel is nothing but due to inversion. We can go back to our capacitor’s value for the oxides C of ox, the gate voltage applied minus the threshold voltage minus the voltage at point y. So, this is the charge in the channel. If I now put it back into I D, the current at any particular point I D, total current has to be obtained by integrating it over the whole channel length. The total charge that is moving mu of n W C ox, integration for 0 to the drift voltage drain voltage minus V threshold minus V of y, d of y.

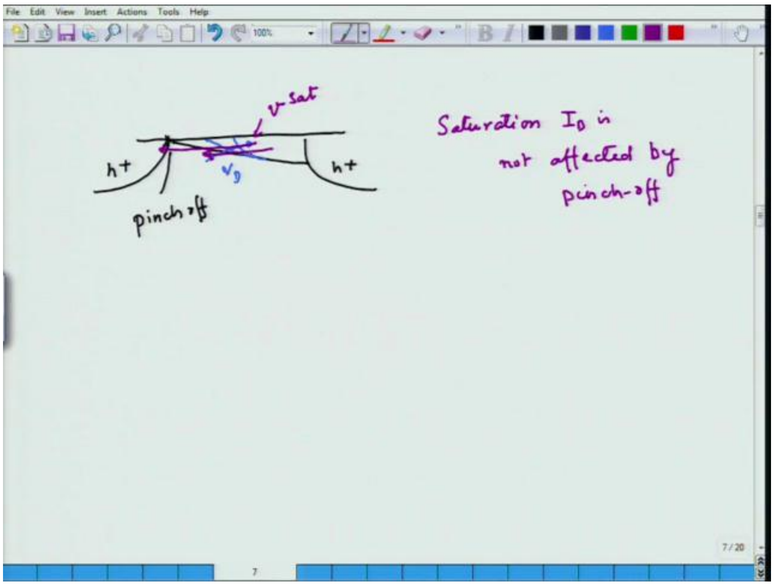
Now, this is, if I, one integrates this one, gets a expression for I of D. The drain current is given by once the inversions in is sets I; it is given by C of ox divided by the channel length V G minus V threshold minus V D by 2 times V D. So, this is the expression for drain current as a function of V D and V G. If we try to plot this expression, you can see how we will get the MOSFET behavior here. So, if I now try to plot V D verses I D. Once again, for different values of V G, I am going to get curves which for a particular value of V G 1, second curve will V at a higher V G, V G double prime.

If I go to even higher value so, this is for a higher V G. V G double prime is greater than V G prime. V G even higher value is going to be even higher than this. Now, this looks does not look like a transistor, right? It does not here; the initial linear behavior is correct. But, beyond that it seems to be getting into regime, where you have actually a negative resistance. Then, coming down to a 0 current; now, that is not what is expected. The reason for that is the equation that we used apply only at the beginning of the inversion layer at the point where we have discussed the pinch off takes place.

That equation does not apply. So, at the pinch off point when the applied drain voltage becomes equal to the V G minus V threshold. So, for each value of V G it occurs at a different point. So, this would be the pinch off point for V G prime. Pinch off point for V G double prime will be here. For V G 3, prime is up there. So, the pinch off starts from here. At pinch off after that, when we will continue to increase the gate voltage, the current does not increase. So, this equation does not apply for pinch off. So, at that point the current saturates. That is why we see a behavior like this. This is I D saturation.

So, in order to get the value of I D saturation; if we replace V D is equal to V G minus V threshold, we can calculate the value of I D saturation. It comes out to be equal to mu n C OX. That comes by substituting this in equation, in the earlier equation w divided by two of two by l v g minus v threshold square. So, you can see this behavior is right. The saturation value does not depend on the V D. It only depends on what gate voltage I am at. That would decide my saturation value. Now, this may puzzle you why does the pinch off occurs and why do things saturates. We have not explained that part. So, far so what we all we said was that.

(Refer Slide Time: 40:08)



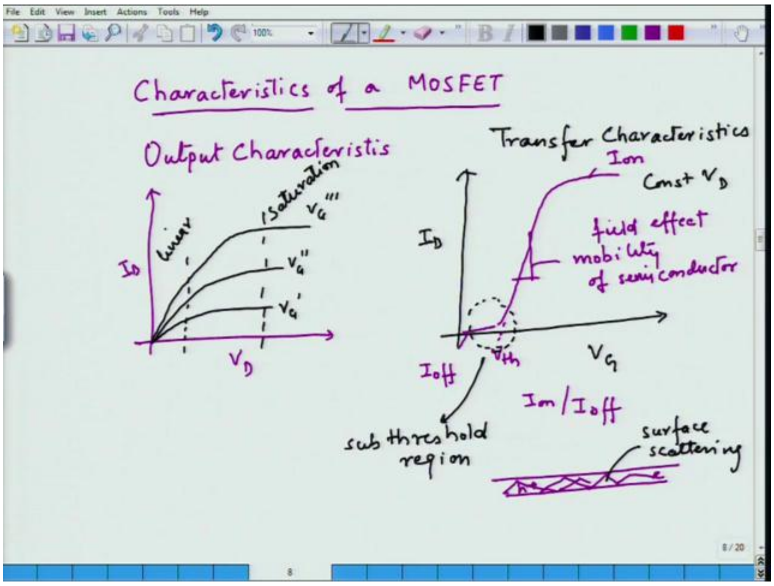
Saturation I_D is not affected by pinch-off

When I create the channel at some value of the V G, the channel gets pinched off at the drain. So, there is pinch off the channel. If I am going to continue to increase the V D, this pinch off is going to increase. So, this is going to move more in this direction with increasing V D. So, why is it that the current is getting saturated at this point? Now, in order to understand that, it would seem like that a channel, which had a low resistance.

Once, it is getting pinched off, it is in resistance is increasing. But, if with increasing resistance, one would expect, if one, if the earlier equation was correct, that the current will go down. But, that has not happened because the reason for that is that whatever carriers are injected here, they once they get start getting accelerated towards the drain. Even if there is a pinch off, they are collected by the source here. So, in the pinch off region, the carrier acceleration or the velocities are so high that the pinch off does not affect the current due to pinch off.

This is why this saturation is because the I D is not affected by the pinch off. But, there is no increase in the I D. There is also another explanation for this pinch off. It says that at the point of pinch off your v the velocity of the carriers is already close to the saturation of the carrier velocity. Hence, once we increase the V D beyond the pinch off, the carrier velocities cannot increase any further. That is why the saturation takes place. So, this is the reason for pinch off and the saturation in the MOSFET transistors.

(Refer Slide Time: 42:18)



Characteristics of a MOSFET

Now, in terms of characterizing the transistor; if tomorrow I need to go and buy a MOSFET transistor what kind of measurements or specifications I should have? There are the two. The one thing that we discussed so far is what is called output characteristics. Output characteristics are plotting the drain current verses the drain voltage for different values of V D. This is for different values of increasing the gate voltage. It is plotted between I D and V D.

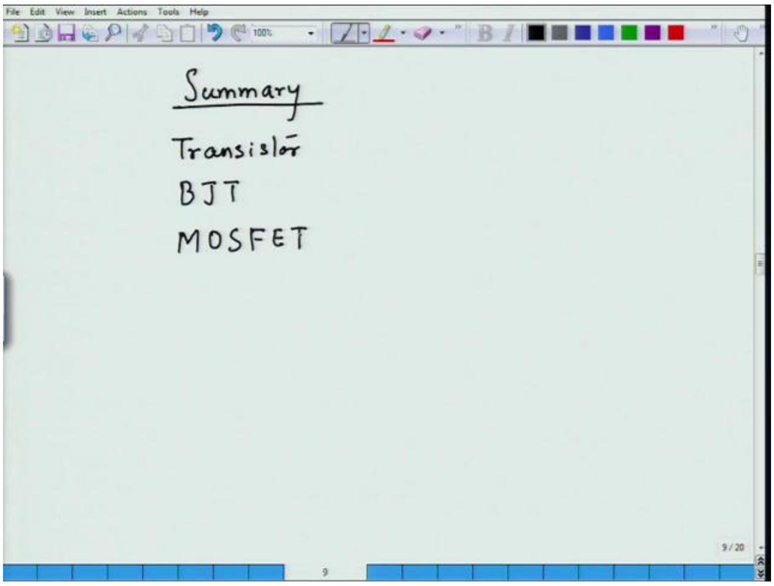
There is another characterization that we use often in that we used in MOSFETs is transfer characteristics. In this particular measurement, we are going to plot the drain current verses the gate voltage and how does that curve look. That curve is we can use the equation that we derived earlier for a particular values of for constant V D. We can plot, we can take a V D, which is at the in the linear region. This is a linear region of the MOSFET curve. This is a saturation region of the MOSFET curve.

This is linear region. This is the saturation region. We can plot the I D, V G. What we find is that this curve initially, when V G is very threshold, there is no increase in the current. At the threshold value, the current will increase and saturate for a constant V D. This particular point is what we it is a threshold for the MOSFET. We also define a current called I off. This is I on. So, we define a ratio of I on to I off. The slope of this particular transfer characteristic gives us the mobility of the semiconductor.

Now, generally this mobility is not the same mobility that we have discussed in module three. This one is called field effect mobility. The reason why it is different from the earlier bulk mobility that we have discussed for semiconductors in general because, in this particular case we have the interface. The carriers are only the electrons or holes are only moving in this particular channel. As a result, there is substantial amount of surface scattering.

Hence, typically the field effect mobility of semiconductors which is measured in MOSFET is less than the bulk mobility of the semiconductor. Finally, there is a this portion of the transfer characteristics, which is below the threshold where there is some amount of current is seen. That is known as the sub threshold region. It, one can think of this as a leakage current. A small value of that is probably more desirable for most applications. It relates to the defects at the interface and in the in the materials itself. So, this is also an important part of the characteristics of the MOSFET. If we summarize our discussion; so far what we have covered is summary of this lecture.

(Refer Slide Time: 47:07)



Summary

Transistor

BJT

MOSFET

We have covered a very important device transistor, how it is used as a switch or as an amplifier. Then, we have looked at what different kind of electronic devices can be used as a transistor. We have looked at bipolar junction transistor. We have looked at the functioning of a MOSFET. How to, what are the characteristics of a MOSFET device that we can calculate, and also use those characteristics to understand about the materials interfaces as well. With this, I will like to close this lecture. People who are further interested in understanding how the transistors are used in electronic circuits are requested to take up the lectures in the electrical courses in that area.