Semiconductor Device Modelling and Simulation
Prof. Vivek Dixit
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology – Kharagpur

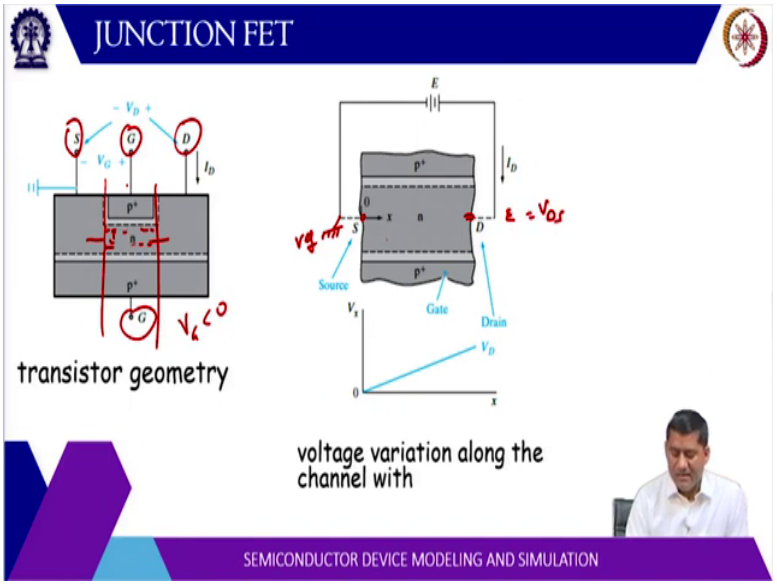
Lecture – 27
Field Effect Transistor

Hello, welcome to lecture number 27 and today we will discuss about the Field Effect Transistors.

(Refer Slide Time: 00:30)

Specifically, we will consider several configurations like junction field effect transistor then metal semiconductor field effect transistor and also high electron mobility transistor. So, we will briefly touch them but we will try to drive current voltage equations for first. Now, this phenomena of field effect is pretty old early 1920s actually, this idea was patented. But due to unavailability of the semiconductor pure semiconductor and the necessary insulator interface they were not realized.

(Refer Slide Time: 01:12)



But as the technology developed so then in after 1947 the transistor was invented then BJT came into the picture. Then after that you know this silicon, silicon oxide interface they were made pure enough. Then MESFET and all these they came into the effect and they actually, capture the market. So, to understand the phenomena field effect transistor, let us consider first configuration which is junction field effect transistor.

So, here you can see the typical geometry and this n-type channel this is basically the n-type region is a channel region. And there are two contacts on the either side. So, let us say one contact is source, other contact is drained. And the voltage is applied across these two. As long as this channel is open, the current will flow and that can be easily calculated V = ir and the resistance will basically be proportional to the or I would say, inversely proportional to the number of carrier, the carrier density and the geometry.

Then you see another terminal called the gate terminal. And then these two terminals are written as gate that means these two are internally sorted. So, the basically same voltage applied here. Now, what happens when you apply a voltage? You look at only this region. That is shown here on the right side. So, when you apply a voltage across these two region, if G is positive then this is forward wires.

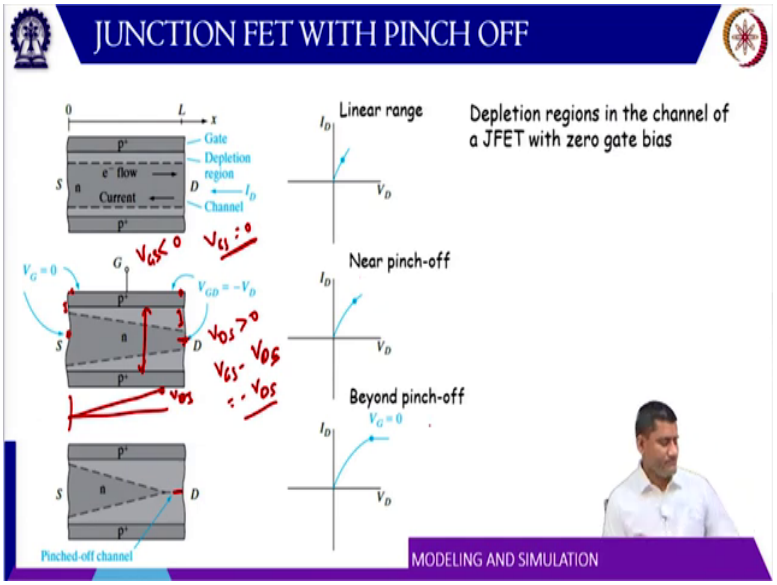
So, it is not really changing the chain of region but another current will flow through the gate. But generally, we do not want any current to flow through the gate, so, this is usually reverse biased and the voltage at the gate is usually less than 0. So that this is reverse bias. Now, if

this is reverse bias then the depletion region here will cover some of this channel region. And the channel will become basically narrow.

When channel becomes narrow, it is current carrying capability will be reduced. So that means it will draw basically, less current for a given source 10 voltage. So, let us say initially we do not have any gate voltage that is open. So, if we apply a voltage across source centre then it will follow some linear relationship. We can assume that voltage is linearly increasing. So, let us say this is x direction.

So, this voltage will linearly increase. So, if this is a reference voltage, let us say this is a reference voltage. Then this voltage will be equal to E or we also call it V DS drain source voltage and it is legally increased.

(Refer Slide Time: 03:55)



Now, if we apply a gate bias, so, V G is p-terminal so, you said to less than 0. Now, you notice one more thing here this V DS is greater than 0. So, if this is the reference voltage, so, we can always express the voltage in terms of some reference voltage. Let us say this source is a reference. So, V G can be written as V GS. So, voltage of gate with respect to source so, if we have 0 voltage, so, let us say initially, we start with V GS = 0.

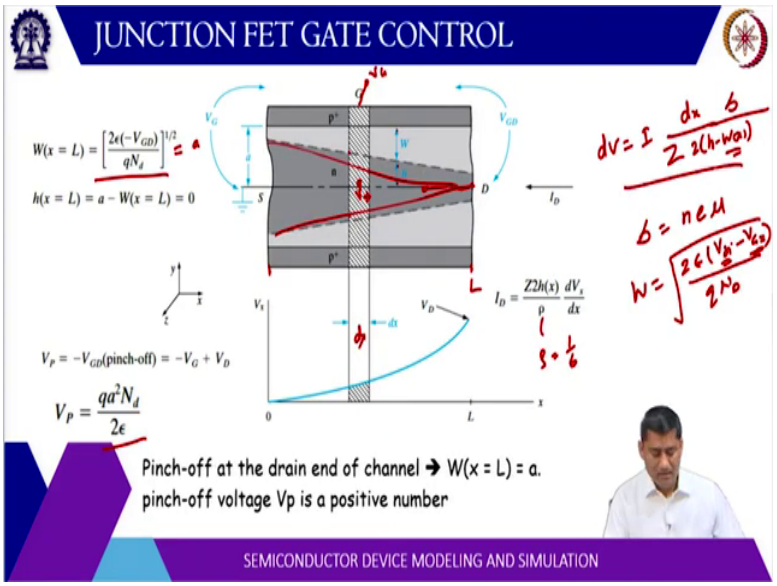
So that means the voltage seen here and here this is a p-terminal. So, everywhere is the same voltage. So, this voltage is 0 here. So, there will be certain depletion region width here that we have drive for p n junction. And as you approach towards the drain side because the

voltage is increasingly linearly. So, this voltage is 0 here and this voltage is V DS here. So, the difference between these two terminals, the voltage is V GS – V DS.

Now, you see V GS we have taken to be 0, so, the voltage here is actually minus V DS. So that means, if V DS is less a few volt. Then this will be negative, so, this means it is more reverse biased at this terminal than at the gate source terminal. So, the depletion width is bigger here and at certain voltage, when the depletion width is equal to this total width. So because this depletion width is approaching from both side, so, you can say half of this height.

So, when the depletion width is half of this height then we can say this is pinch-off and that condition is called pinch-off. So that means there is no more n-region here is how depletion region basically. And beyond pinch-off this current becomes constant. So, up to this pinch-off, what is happening? This current is increasing but when it approaches near pinch-off this current slows down basically the rate of increase is not that much and beyond pinch-off it becomes kind of constant.

(Refer Slide Time: 06:14)



Now, we can drive this relationship you can consider let us say this is at x = 0. This is x = L the length of this region. Now, at distance x you can consider this element dx. So, the voltage drop across this region let us say is dv is equal to let us say I is the current times the resistance of this region. So, resistance will be length that is dx divided by area. So, area can be written as the width into the paper.

So, there is a width this Z the Z here so, width into the paper. And this height here which is let us say, h is the height here and W is a width of w is a region. So, this is two times h – W, W is a function of x, h is a constant times L by area times resistivity or sigma L by a. So, this is multiplied by the conductivity. And conductivity here we know we can write in terms of the carrier consultation and the mobility.

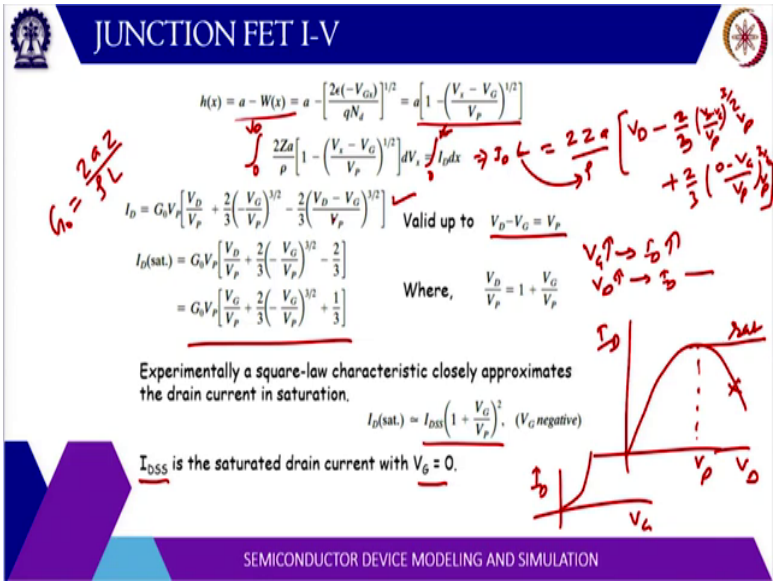
So, we can write it sigma, so, sigma = ne mu. So, n is the **(()) (07:53)** here, e is the charge on electron, mu is the mobility of the electron. So, the expression for depletion width we can recall the relationship between the p n junction and we can write W is equal to two times epsilon V bi – V Gx at position x V Gx with respect to position x. So, this is, let us say, V G let us say this voltage is V x.

So, the difference is V Gx divided by q times the doping consultation then square root. Now, sometimes this is approximated, as we generally ignore V bi we just write V Gx. So, at pinch-off this V Gx is V GD. So, for at x = L this expression can be written as 2 epsilon – V GD by qN d square root. And of course, a pinch-off this is equal to a. And that voltage is pinch-off so, when we equate it to a, we can get the value of pinch-off. So, this V GD value when the W is equal to a.

We call it pinch-off so that of course, you can find out that V = a square times qN d by 2 epsilon. So, a square times, qN d by 2 epsilon, so that is a pinch-off voltage. Beyond this voltage what happens? If you increase further voltage then this pinch-off will move slightly to this side and most of the voltage actually, will drop here means high field will exist. So, there are no carriers as such but the carrier coming from this region they will quickly sweep through this region.

So, current remains constant it does not increase the beyond this pinch-off point. And of course, when you substitute here then we can separate the two sides. So, we can write, now sigma and rho this rho here is one over sigma. So, if you integrate it here then we can take this x to the other side. So, this is one over this thing and we can integrate from, let us say, 0 to V DS and x from 0 to L.

(Refer Slide Time: 10:19)



So, what we will get? So, this expression here is rewritten here in this form, so, 2Za by rho. Now, this seems to epsilon V GD by qN d that is related to V pinch-off, so, we can replace this region by V p. So, we can write V GX as V x – V G divided by V p. So, this h a – W simplifies to this form and we can substitute here. So, this now is in terms of dV x. So, now, we can integrate it 0 to V D and this is 0 to L.

So, because this, drain current is constant throughout the length because there is no other way to for the current to flow. So, this will be simply I D times the length. And left side will be if you integrate it, so, you will get basically 2Za by rho times V D. So, what is done here? Because this, if you integrate here, I 0 L = 2Za by rho times V D minus now, this is square root, so, it is integral will be 3 by 2 this power to the power 3 by 2.

So, this is 2 by 3 divided by 2 by 3 by 2 so, power 2 by 3 this is V x – V G so, you can substitute one at 0, one at V D. So, let us some say top one, so, V D – V G by V p. Then there is 1 by V p will appear here if you differentiate so because we are integrating so, it has to be multiplied by V p. So, what we can do? We can take that V p out then for 0 it will become + 2 by 3 0 – V G by V p to the power 3 by 2 times V p.

And if you take V p out so then this will be divided by V p it will simplify to this expression and here G naught is basically 2aZ by rho times L because this L will come here. Now, this expression we have derived assuming there is no pinch-off. So, this means this expression is valid till pinch-off that means, when V D – V G = V p.

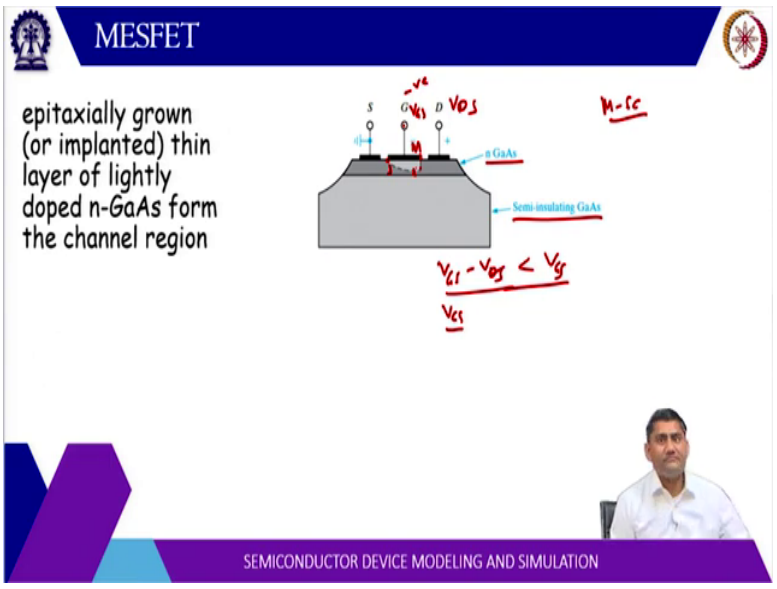
And of course, beyond saturation this will be fixed at the value V D – V G = V p so, we can substitute here V D – V G = V p. So, because V D – V G is V p so, this will become 1, so, this will be – 2 by 3 and this is V G by V p to the power 3. So, if you further increase the drain voltage, this current is not going to increase. So, if you increase the gate voltage, so, if you increase the V G then current will increase.

But if you increase the V D then current is not increasing, it is constant. So that is what is happening beyond pinch-off. And of course, you can write a Matlab code to plot this curve, so, what you will basically get it will somehow increase this and then go down. So, this point is pinch-off. So, beyond this you will not follow this equation, so, this equation will not be valid.

So, beyond this, you will just write, replace it with a constant. So, you will not follow this curve but it will be constant. So, this will be your I D, V D curve. Now, this is the linear region and this is called saturation region. So, if you plot in the saturation region then you can plot I D versus V G. And that will follow some kind of curve and which is I D assistance 1 + V G by V p whole square.

And where I DSS is the saturation current at V G = 0. So, if you substitute V G = 0 this I D sat = I DSS. So, this is the saturation current. So, you can say I sat here.

(Refer Slide Time: 14:44)



Now, similar concept is used in the metal semiconductor field effect transistor. So, in JFET we use two junctions, p n junction. In case of MESFET what we use? We directly put metal

on the semiconductor. So, of course, we cannot use very thick substrate because then you cannot increase this depletion region and semiconductor region to a larger extent. So, because you know this, it will look at very high voltage.

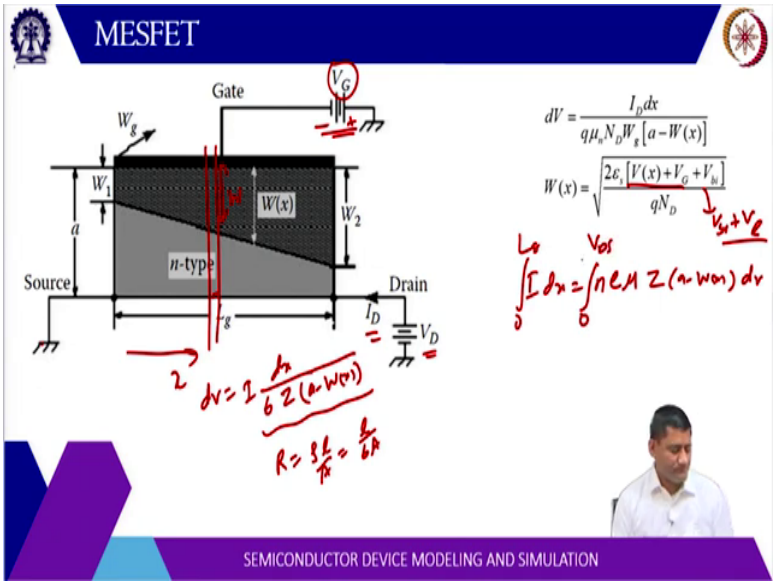
So, what we do? We use semi-insulating substrate here. So that means it is not doped. It has enough resistivity. Then on top of this, we deposit a layer of n-type gallium arsenide which is highly doped. Then and this is usually quite thin order of let us say some micron few 10, 100 nanometre like that. Then it there is a metal layer that is deposit on top of it. So, this metal is in direct contact with entire semiconductor.

So, this contact we already discussed and we call it short key contact or metal semiconductor contact. So, this junction also follows a same trend. The depletion region forms on the semiconductor side. So, if this is n-type and this is a metal layer, so, when we apply a negative voltage on the gate side, the depletion will extend to the gallium arsenide side. If we apply the positive voltage to the gate then there will be current through the gate and n-type gallium arsenide because then this will be forward biased.

So, in both the cases, junction FET and the MESFET, this junction is kept reverse biased. So that no current flows through the gate then there is a voltage here V DS and let us say this is V GS. So, the drain voltage is more than the source voltage. So, the reverse bias at the gate drain side. So, this is the side is more because this voltage difference is V GS – V DS. So, it is even more negative than on the source side.

So, the source side it is V GS only. Now, V DS is positive so, this voltage is less than V GS so, it is having greater reverse bias here. So, the depletion width is more here, so that means channel is getting shrink here. So, for this also, we can write a similar I b equation.

(Refer Slide Time: 17:28)

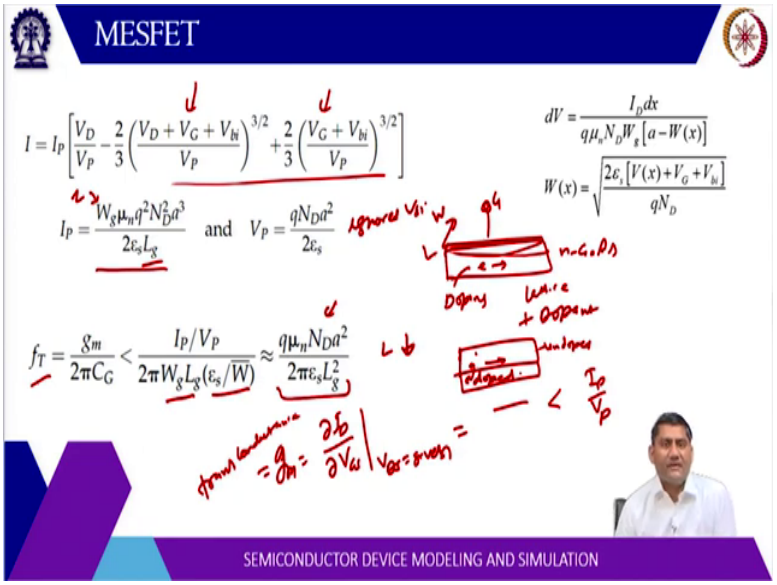


Let us consider this region this L is the length of this semiconductor. And we have this gate connected here V G now, note here the polarity this is plus, this is minus. So, this V G is actually, negative. Then on the source side, this is ground, so, this voltage is V G only and there is some depletion width here. And on the drain side there is some higher voltage here, V D so, this depletion width is even more.

So, here also as a let us say, distance x we can choose this region, so, this again we write dv = I dx by conductivity times area. So, area will be Z the width into the paper Z times a – W. So, this is let us say W, so, a – W as a function of x. This is basically the resistance term R = rho times l by A or l times sigma by A. Now, W of course you can recall again, 2 epsilon by qN d times the voltage across the gate and the channel at position x.

So that is V x + V G because of this polarity here because V G is now this voltage source V G. So, because this is negative here. So, this is simply V x + V G but if we say which is a gate voltage then it will be V x – V G so, V x + V G + V bi. So, this basically, you can recall it is V bi plus reverse bias, voltage V r. So, V r is actually, here V x + V G. Then of course, when we can integrate it like this I dx = sigma, sigma is n e mu times Z times a – W dv. And then we integrate 0 to L written here and this is dv is 0 to V DS.

(Refer Slide Time: 20:17)



So, when we integrate it, we will get very similar expression that we got for the junction FET. So, this is I p and I p is the width. So, this is same as the Z here the width inside the paper times mu n q square N square D a cube by 2 epsilon L is the length. And V P is again this expression has ignored V bi. Now, so when you plot this I versus V D we will get very similar characteristic as we got for the junction FET.

The only difference here is that the amount of charge stored is only on the n-type gallium arsenide side or on the semiconductor side. On metal side charge can be very, very easily moved around. So, generally the short key diodes are fast as far as switching characteristic are concerned. Another thing you can notice here this is your n-type, gallium arsenide so, this is your gate here.

Now, what is happening here? We are modulating this channel only. So, it has a doping and these electrons are also moving in this channel only. So, these electrons encounter is scattering from lattice plus the dopant atoms, so, both the scatterings are there. So, what people actually, do? They make a multilayer structure here and we they separate out let us say this is a doped region and this is undoped region here.

And if this band gap is small then these electrons in the doped from the doped region will spill over to the undoped region so, here the channel will form. So, in this reason they will encounter less scattering because now, it will only have lattice scattering here and the dope ends are here. And that is the basis of high electron mobility transistor, so that we will also discuss. And of course, it is frequency characteristic on the obtained using g m by 2 pi C G.

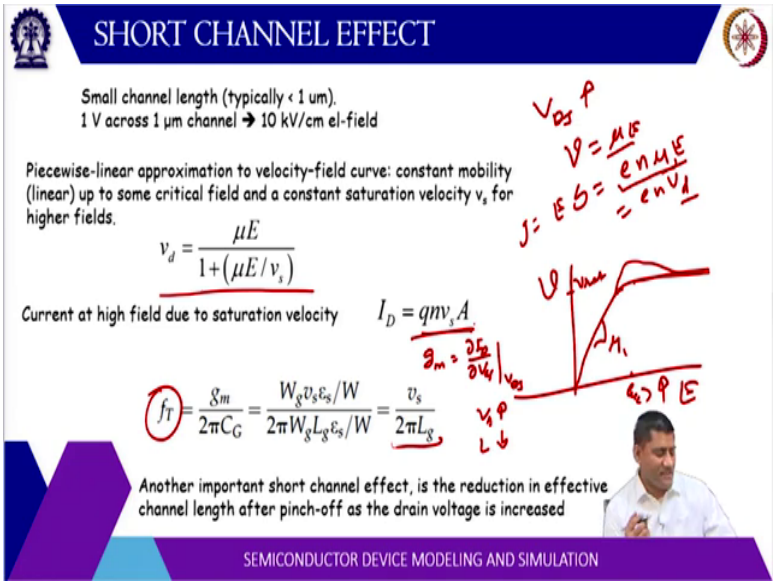
So that is that tells you about the unity gain frequency and g m is defined as del I D by del V GS for a given V DS. And if you differentiate this one, you will get this around it will be, I p by V p minus differentiate this term, so, I p is not dependent voltage. So because V G appearing this is 0 basically because this is V D. So, only difference with respect to this V G and with respect to this V G.

And if you select certain V DS so, at V D = 0. They both are equal and they will probably they will cancel out, basically so, at higher V D this difference will exist and you will get some finite trans conductance g m is trans conductance. But the value will get really, will be basically, less than I p by V p. Because this negative term will come here then divide by 2 pi C G. Now, C G is the gate capacitance.

So, if you look here, this is the gate area, so, the gate will have a length L and the width, let us say W or Z. And then the gap stands will be this area that is W times L times epsilon by depletion width so, this W is the depletion width. So, we substitute the value of I p and V p this is the basically expression that we will get so, q mu n N D times a square by 2 epsilon L g square.

That means unity gain frequency increases if you reduce the length. So that physically, we can understand because now the time to transit for these electrons is less and therefore, speed will be more as you decrease the length. And if you increase the doping then large current will flow basically because now there are large number of carriers and this frequency will be more UNICOM frequency. And same thing, can we consider with respect to other parameters.

(Refer Slide Time: 25:39)



Now, when you go to higher voltages, let us say you go on increasing the V DS. Then we have use this expression that we have written that V = mu times E. So, we have used this one if you recall, we use sigma = e times n times mu. So, this actually, assumes because j is equal to the current densities is sigma times e so, mu E. So, this is e n times drift velocity. So, we have assumed that drift velocity is mu time C.

But that is true only below certain value of the electric field. So, if you plot this velocity versus electric field, it increases linearly and that slope is actually mu but after certain value it actually, saturates. In case of gallium arsenide is little more complex it goes like this and saturate. But anyway, there is some saturation velocity here. So, up to this electric field, where V is proportional to e this concept of b = mu E or defining sigma is as e and mu is valid.

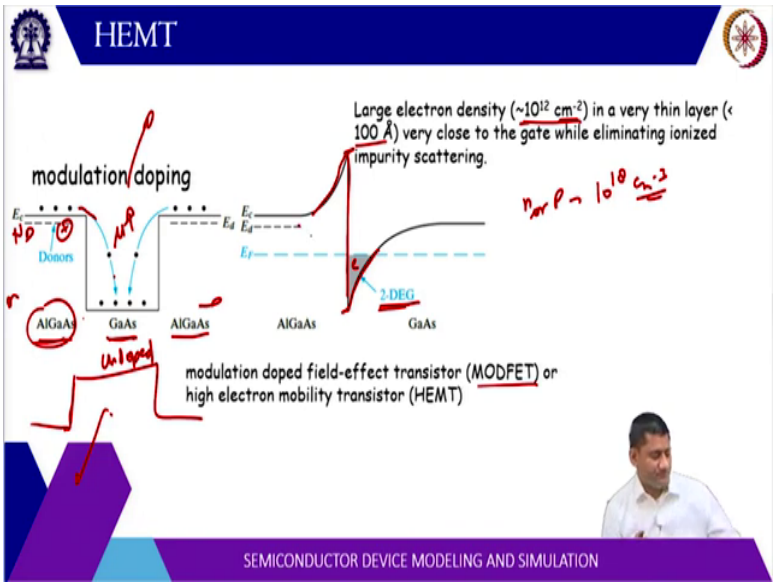
But beyond this electric field, this will not be valid and this is some critical electric field or at which the velocity of this carriers, saturate. So, more general expression is sometimes used that drift velocity is mu E divided by 1 + mu E by v s where v s is the saturation velocity. Once you operate in this region, so, at a higher voltage where field is more than this E critical.

Now, you cannot use this expression because the velocity is not changing with the electric field. Velocity is constant. So, what we will use? I D = qnv s times A where v is the saturation velocity. And again, if we calculate g m here, g m is del I D by del V GS for a given V DS. So that will be W g times epsilon V by W and divide by 2 pi C G. So that is v s by 2 pi L g.

So that means unity gain frequency increases if you have higher saturation velocity or you have a smaller length of the device. So, small devices are useful to be used at higher frequencies or you can choose a material where saturation velocity large. So, those materials can also be used for high frequency operation. So, same aspect if the length is small with respect to the applied voltage.

We will have to use this expression with the saturation velocity and if you are operating in the linear region then we can go ahead with the previous expression.

(Refer Slide Time: 28:42)



Now, in HEMT I just mentioned that if we separate out these electrons from their donor atoms. So, here you are seeing this picture aluminum gallium arsenide and the gallium arsenide and aluminum gallium arsenide. So, this aluminum gallium arsenide is doped with certain dopant, let us say N D but the band gap of aluminum gallium arsenide is more than the gallium arsenide.

So, if you put these two materials side by side, there will be some potential value or electrons as well as holes. So, these electrons will actually spill out here and this gallium arsenide is undoped. So, when you apply electric field here in the gallium arsenide, this electron will move if you use this as a channel. But these donors are coming from aluminum gallium arsenide.

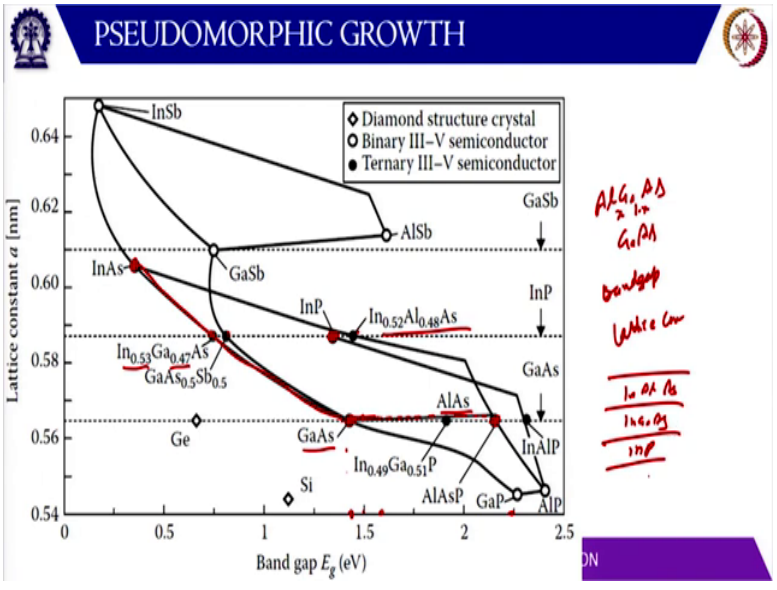
That means now, when they move through this region they will not encounter the scattering due to this donor ions. So, their scattering is reduced, so, their mobility will be actually, more. Now, when you apply a voltage bias here these bends are not that flat rather, they show some kind of curve like this. So, this goes like this, this discontinuity is constant regardless of the applied voltage.

So, these electrons basically spill out here fill this region and this we call 2-dimensional electron gas because they are confined in one direction but they are free to move in 2D. So, this 2-dimensional electron gas is basically the fundamental way and the high electron mobility turns stripe. They are also called modulation doped field effect transistor because there are this doping is modulated so that the donors are separated from the carriers.

So, in this 2-DEG the electron density is pretty large around 10 is to power 12 per square centimetre. Please note the dimension here generally, we say that n the electron or hole concentration, n or p is around 10 is to power 17 or 18 per cubic centimetre. So that is the bulk concentration. In 2-DEG this is basically per unit area, so, it is not transition 12 per square centimetre in a very thin layer.

So, the size is typically 100 Armstrong that is 10 nanometre and is close to the gate and it is basically away from the ionized impurity scattering.

(Refer Slide Time: 31:18)



Now, how do we choose two materials? So that you can understand of this picture, so, if you see here in the previous slide, we use aluminum gallium arsenide and gallium arsenide. So,

this multilayer has to be there and they are band gap has to be different. So, band gap has to be different but their lattice constant should be same. So, if you see here, where the gallium arsenide it is here. This is the gallium arsenide.

Where is the aluminum arsenide, aluminum arsenide is here. So, we cannot put aluminum arsenide or gallium arsenide. If you see their band gap gallium arsenide band gap is around 1.4 aluminum arsenide is around more than 2 basically. So, the band gap is large here. But if you take a compound so, typically it will follow some kind of linear year. So, you take a small friction, so, it is aluminum x gallium 1 – x some mole friction is there.

So, it is kept you know up to 10, 20 percent, like that not very large. So, if you see here, so, somewhere here, so, the band gap is now more here. Let us say, 1.4 to 1.6 here but the lattice constant is very much same gallium arsenide or aluminum gallium arsenide. So, throughout the range the lattice constant is same. So that is one reason we can choose this system of material. Similarly, other material that is popular.

We grow indium phosphide as a host material or a substrate on top of that we grow indium gallium arsenide or indium aluminum arsenide. So, indium gallium arsenide is they see this is indium arsenide, this gallium arsenide, so, indium gallium arsenide will fall on this line. So, particular percentage which is lattice mesh to indium phosphide is 53 percent indium, 47 percent gallium, so that is lattice mesh.

So, this is another combination that is used so on an indium phosphide layer you can have this indium gallium arsenide with this composition or you can have this indium aluminum arsenide with this composition. So, these three materials have different band gap but same lattice constant. So, they can also be used. So, some other different combinations of material that are used. So, thank you very much in next class. We will consider one example related to MOSFET.