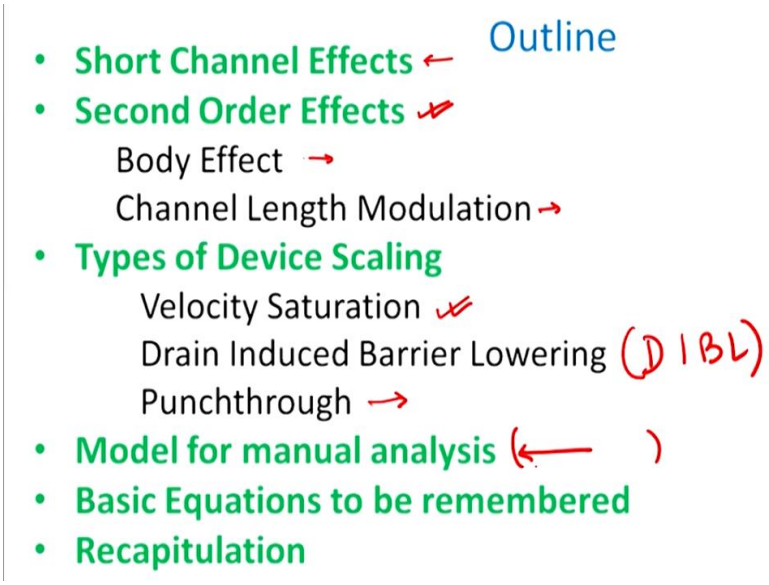
CMOS Digital VLSI Design
Prof. Sudeb Dasgupta
Department of Electronics and Communication Engineering
Indian Institute of Technology – Roorkee

Module No # 01
Lecture No # 03
MOS TRANSISTER BASICS-III

Hello everybody, welcome to the third lecture of NPTEL online certification course on CMOS Digital VLSI Design. In the first two lectures, we have actually dealt with the basics of MOS device, how a MOS device works? What are the various characteristics of MOS device both electrical and structural? We have also looked into the fact what is the meaning of sub threshold slope and the various governing equations for the MOSFET working both in the saturation as well as in the linear and non-linear region. What will be looking in this lecture is give you the idea of short channel effects which you in front of you.

(Refer Slide Time: 01:02)



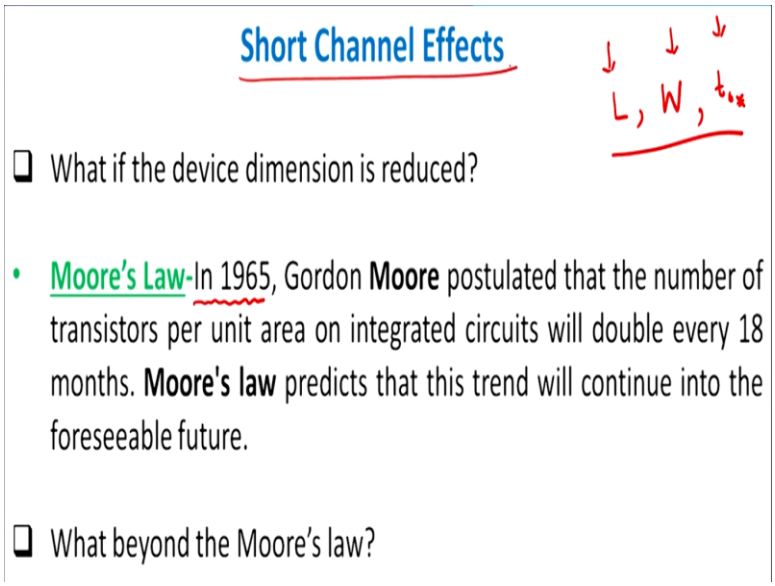
We look into second order effects we will have first short channel effects then second order effects which we will looking into the body effect and the channel length modulation effect. These two effects this body effect and channel length modulation is primarily because of reduced dimension of the device these are all of second order effects.

We will also look into the various types of device scaling and in which we will be referring into three important issues related to the device scaling and these are velocity saturation and drain induced barrier lowering DIBL also referred to as DIBL and we will also look into the punch through phenomena in a MOS device, right, which will be we will be also having a

look into what are the various model equation which means that the this part is primarily to do what is to give an idea but if I know the over underline characteristics of the device.

Can I have a set of mathematical equations, which will very easily be able to mimic the working conditions of a MOSFET. So that will be dealing with this case and then we will be recapitulating at the end of the whole talk.

(Refer Slide Time: 02:09)



What happened was that in 1965, Gordon Moore gave a very important law which is known as Moore's law. Those who are in VLSI domain or those who are working in this domain will be easily appreciate in this point. And he told that for every one and half year growth in industry, you would actually see doubling of the number of transistor per unit area of the chip. This was required, this was not a natural law but it was a law which he was able to predict by virtue of certain parameters in terms of profitability and turnaround time in the market.

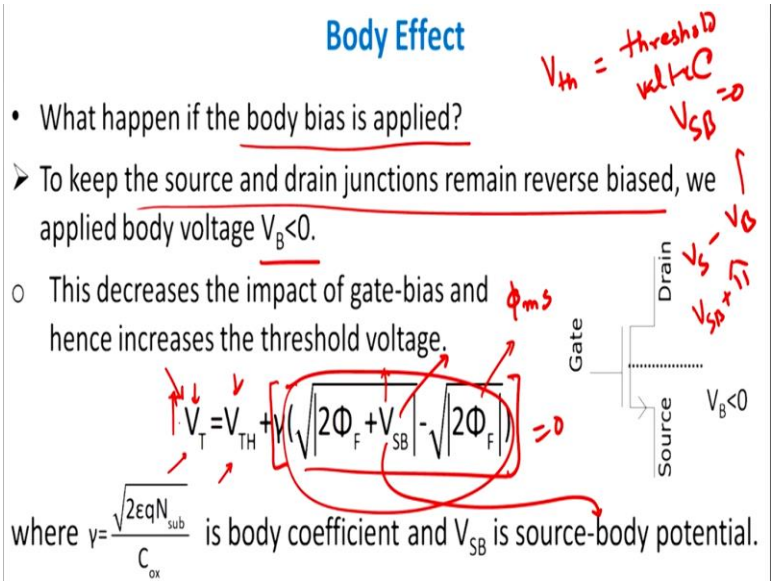
And he actually stated that every doubling has to be there. Now, if I want to double the number of transistors per unit area of the chip the only option available to me is to reduce the dimension of the device by half, right. So if you do not do anything else and you reduce the dimension of the system by half, you will be able to accommodate twice the names, number of things in the same amount of area. That what is very important that you have a Moore's law in, that the, if you can reduce the channel length W, if you can reduce the width W and channel length L and you reduce the width W.

Here, also able to reduce the t oxide thickness. Then we define these to be, as a scaling. So the direct influence of Moore's law was that scaling came into picture. But I need to know how I will reduce the dimensions of the device, so that, it does not affect overall functionality of the whole system and the Moore's law is still getting validated. So that was the major region why short channel effect comes into picture.

What happened was, but if you reduce the channel length which are discussed in the previous two lectures and you reduce your widths certain phenomena comes up or gets more shown in the IV characteristics. These are all known as short channel effects. As the name suggests short channel why because channel length has been reduced.

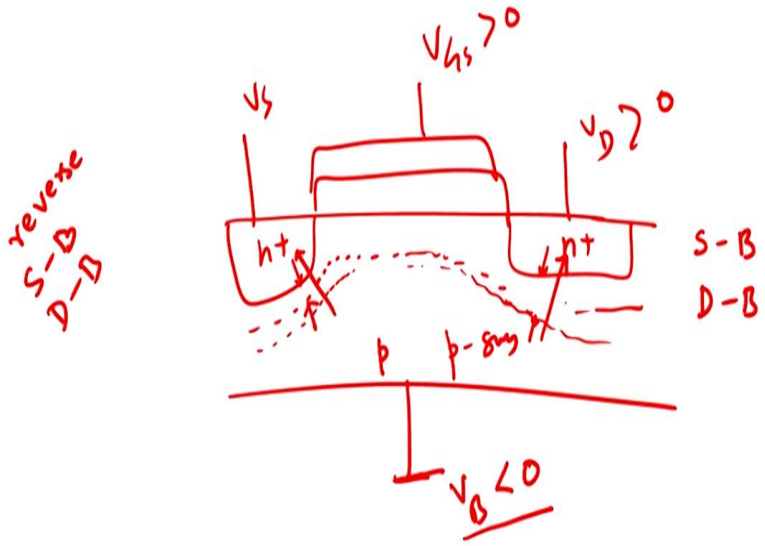
W has been reduced and certain phenomena are more heavily visible to you in an IV characteristics. What we see so first what will be looking into is basically the body effect. Please remember, in our previous discussions that I had assumed that a my body is a fourth terminal and my body should be kept at the negative most bias for an NMOS, right, if you remember correctly in a previous discussion this was what we had taken care of.

(Refer Slide Time: 04:38)



Let us see what happens if you apply a body bias right, if you apply a body bias, let us see what happens.

(Refer Slide Time: 04:44)



Now, if you look back, if you draw the MOSFET diagram again enhancement mode MOSFET only. So, I will when I mean MOSFET I mean enhancement mode. I have n + region I have p type substrate, right, substrate here. This is my body VB and this is the gate and this is my gate here VGS, VD and this is my VS, right. They will be depletion region formed here and this depletion region will go like this and will go like this if provided VGS greater than 0 and V D also greater than 0 under such a criteria this will happen, right.

Now if you apply a bias here, VB which is also negative, please understand. This is p type, this is n +, n +. So if you give VB much larger than but less than 0, you are actually increasing the depletion thickness here, as well as here, right. So what will happen this depletion thickness which some become something like this will increase, right. So you are therefore still maintaining the reverse bias condition between source and bulk and drain and bulk body.

So the condition was that I have to sustain or have to have reverse bias between source and bulk and drain and bulk and this will be maintained if you simply give VB less than 0. If you make VB greater than 0, then you are in a problem that then this will forward bias this junction as well as this junction. And there will be heavy current flow through the device, it will burn out the device, right. So just to give you an idea about what us talking about therefore to keep the source and drain junction always in reverse bias, we make VB less than 0.

Now, let us see what happens? The we will look at the expression first and then we go to the physical discussion. VT is given as VTH + gamma times 2 times Phi F + VSB – 2 Phi F. 2

Phi F is a constant which is given again by the previous discussion the KT by 2 times K T by Q is the value of Phi F which you see Fermi level between. So this is basically you are, sorry, Phi ms metal semiconductor function difference which you see.

$$V_T = V_{TH} + \gamma\left(\sqrt{\left|2\Phi_F + V_{SB}\right|} - \sqrt{\left|2\Phi_F\right|}\right)$$

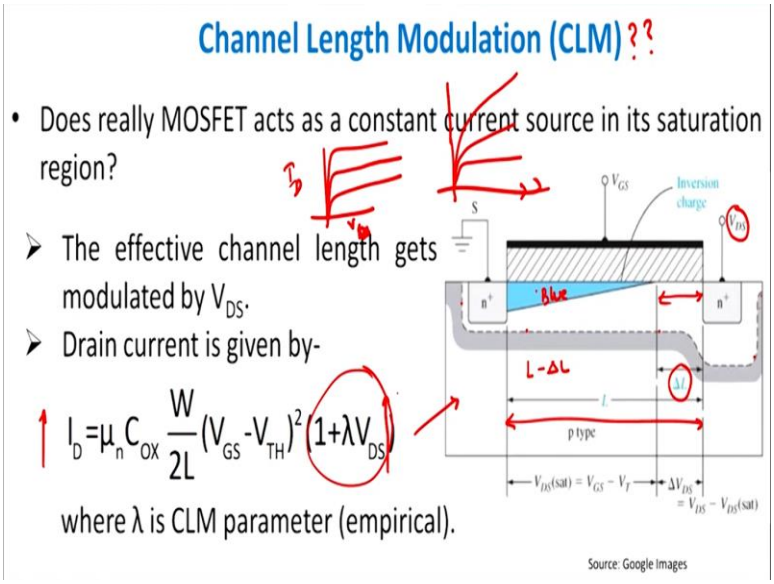
If you look very carefully this gamma is defined as the body coefficient factor and VSB is the source to body voltage. Now, if you, as I discussed with you in the previous just now discussion that if I make it therefore VSB if I make it more and more negative, right. Then this depletion region becomes larger and larger, agreed? As the depletion thickness becomes larger and larger on this side, I have to apply a larger gate voltage in order to break this channel.

Because depletion thickness is large, in order to make it smaller by gate voltage has to become larger and larger, this is threshold voltage will become higher and higher. With this basic knowledge what I am going to show tell you therefore is, if you make a therefore VSB large, high VSB. This expression will give you a larger value and therefore your threshold voltage will increase.

VTH 0, this is VTH which is threshold voltage, old voltage, right, at VSB = 0. When you V SB = 0, this 0 becomes like this, this equation vanishes and VT goes to VTH. So at VSB = 0, the whole quantity is = 0 and these two quantities become equal, as VFB become more and so VSB is what? VS – VB, right. So if you are bulk, you are making it negatively large, you are making VSB more positive, right.

And therefore as you make it more and more positive, this quantity which is VSB starts to increase and therefore threshold voltage increases. This is what is known as a body effect, right. So in body effect, in an n channel enhancement mode MOSFET, if you go on rising the value of the, if you go and making it more and more negative, you actually end up having a larger threshold voltage available to you at the end of the day this is the first.

(Refer Slide Time: 08:44)



The second one is basically your, channel length modulation effect. What is channel length modulation effect? See, till now, we were assuming that channel was defined as that distance between source and drain where the number of electrons is almost equal to the number of volts actually present at the particular time. This was the discussion we had till now in an inversion region.

With this basic idea, let us see what happens. In front of you, this is a structure, this edge is basically a depletion region, this dotted curve which you see is basically a depletion region, right. Now, what happens is, as the VDS drain to source voltage goes on increasing, the depletion thickness between drain and the bulk goes on becoming larger and larger and it starts eating into the channel in a very layman language, right. So a channel is a blue one is a channel.

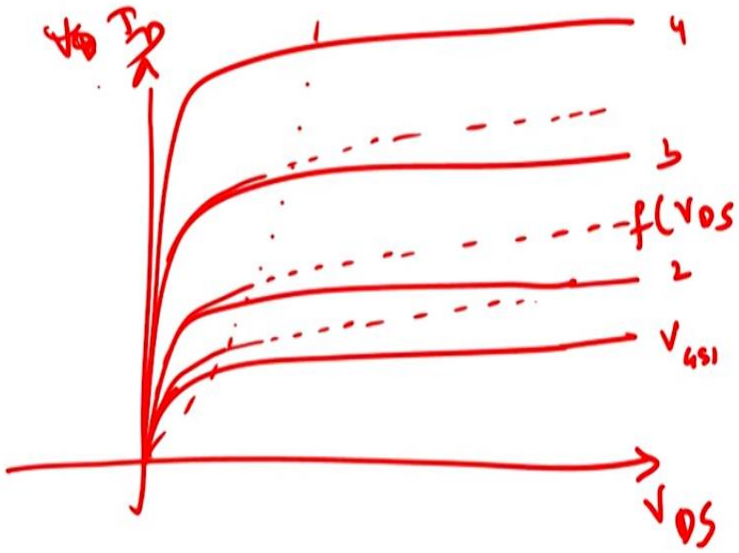
This blue one is your channel and as you increase the value of VDS. This eats away into the channel and this is the depletion region which is available to me. So what has the channel become you and you had initial channel L this much but you ended up having n channel equals to L - delta L. When delta L is the depletion thickness on the drain side, right. As a result, your effective L has actually reduced.

It is become in fact smaller than capital L by how much factor delta L. The delta L is nothing but the depletion thickness on the drain side drain to bulk side. As a result, what you will see is very important property that this results in an enhanced on current or the on current which I assumed in an earlier discussion to be constant which means the if I plot ID versus VDS, right V DS.

It was almost like this constant which you will see for various values of V GS now it cannot become constant and the reason is something like this. It will become something like this. right. Why this being like this? Because, now at this age or at this stage**,** since the channel in this shortn down, right channel in the shortn down.

The VDS value drain to source voltage also starts to play a role in determining the value of ID. So initially are ID was constant independent of VDS and we say that we enter into saturation region. Now, there is an added term here, which is (1 + lambda VDS) because of which current actually shows an increase as the VDS actually increases. However, it might be very large increase, but it might be very small increase. But however there will be a increase nonetheless.

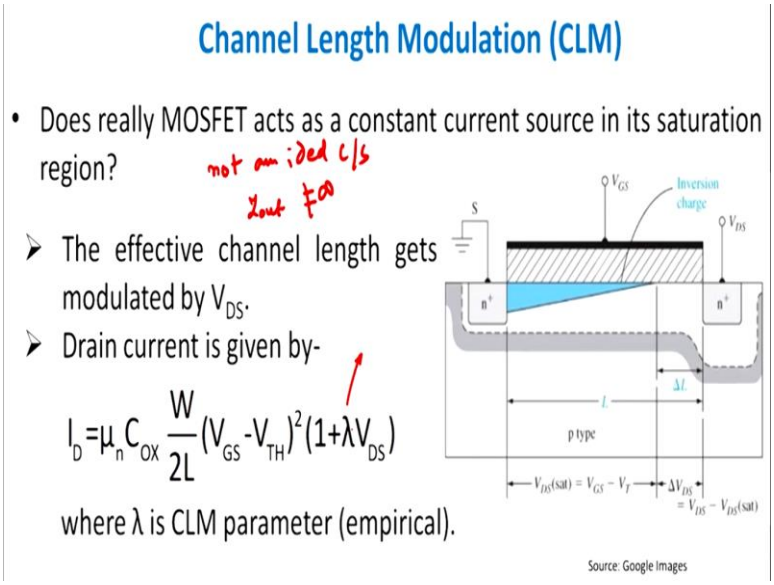
(Refer Slide Time: 11:20)



So, what we do? We have this. So, this was your without CLM without channel length modulation this is what you get for various values of VGS, right. This is VGS1, 2, 3, 4. Now, which CLM this is the onset of saturation as I discussed with you.Here saturation **is** is getting on. Now and I will do is if I plot it I should I ideally get like this. But now I start to get something like this. So, I get like this and then I get like this.

So, post saturation I should actually get a constant value of current but no now I am getting a current which is a function of VDS. This is VDS versus ID. And this is an important property which makes our life difficult and makes it very difficult to overall see the things.

(Refer Slide Time: 12:12)



Which therefore means that I cannot assume therefore MOSFET to behave like an ideal current source in saturation. So, this is the only important property and therefore it will be not an ideal current source and therefore it Zout therefore will not be equal to infinity but will never finite value.

And this lambda is refer to as a channel length parameter generally less than one it is much smaller than one but this is what is known as the CLM parameter and it is empirically derived. It is devised after large amount of experimentations and it have been derived quite a lot.

We come to the, therefore this is the direct consequence of Moore's law but if I want to have larger number of devices per unit area of the chip. I want that the devices should be scaled properly, right. It should be scaled such that I have larger number of devices being available per unit area of the chip. There are generally two types of scaling which is available, scaling means not only channel length scaling. Please keep in mind this very very important.

(Refer Slide Time: 13:15)

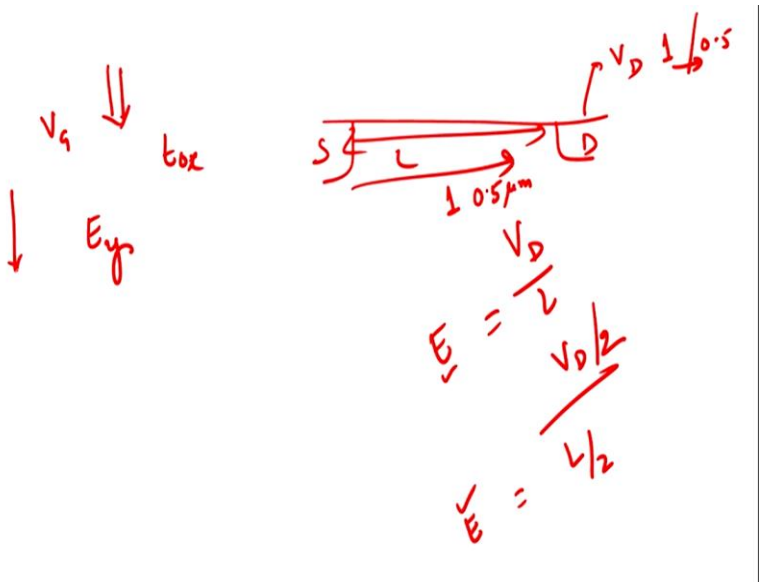
Types of Device Scaling

- Scaling of the device does not only mean the reduction of the Channel Length. It includes the proper scaling of all other device dimensions.

1. Constant Field Scaling-It yields the reduction in the power-delay product of the transistor. Hence, it requires the reduction in power supply for reduced feature size.
2. Constant Voltage Scaling-This is a preferred scaling technique as it provides voltage compatibility with other technologies. Due to this electric field gets higher in smaller devices which causes mobility degradation, velocity saturation etc.

This is not only the reduction of the channel length but also includes the overall dimension reduction. So you have to reduce your W, you have to reduce your Tox, you have to also reduce your voltages, current so and so forth. With this idea in mind, we generally have two types of scaling theory available to us. One is known as constant electric field scaling theory another is known as the constant voltage field scaling theory. What is the constant field theory means it is very simple and straight forward. I will give you an idea.

(Refer Slide Time: 13:47)



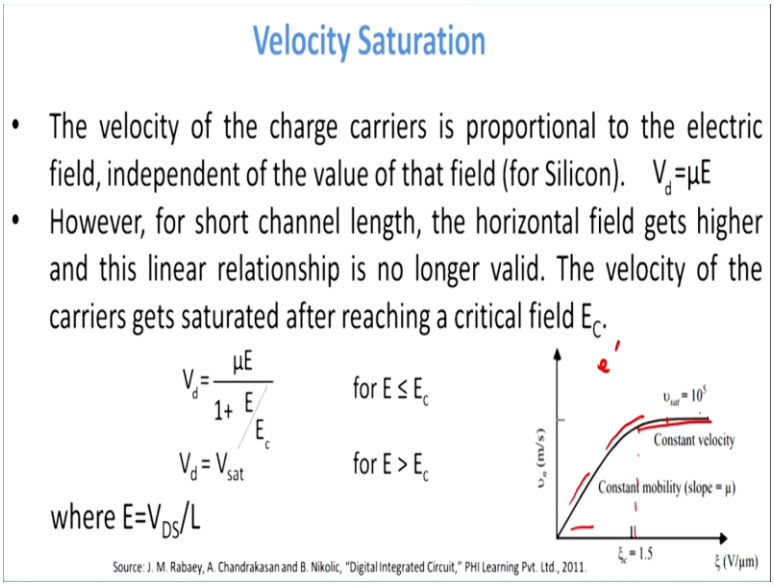
Constant electric field primarily means that if I have a channel and I have a source and drain here, right and this is the channel length L. I apply a voltage V here then VD/L is the electric field in this direction, right. If you now reduce the channel length by half, so I make it L / 2. I also need to reduce the value of voltage VD by half in order to have the constant electric field which you see in front of you.

So I need to reduce this by V.D / 2. So If I was initially having 1 volt and channel length 1 micron if I reduce it by 0.5 micron, right, I need to reduce it by 0.5 volts to 0.5volts. This is known as constant electric field scaling that means the electric field along longitudinal as well as in transverse direction will always be scaled down.

Similarly, VG was the applied gate voltage on the gate side, I have an oxide thickness available to me, right. I can find the value of electric field in Y direction which is the transverse direction and therefore if I scale down the value of tox. I also need to scale down the value of gate voltage. This is known as constant electric field module, right. So, the constant electric field module tells me therefore that it actually does what it actually keeps the electric field constant. It does not let the electric field move larger than its own value.

The second aspect or the second idea is constant voltage scaling. In which we do not try to manipulate anything else but we try to keep the value of voltages constant, right. So I have constant electric field which is which means that my as I reduce the dimensions voltages has to be also reduced. In constant voltage scaling module, we always keep the voltage constant available to me, right. And this is scaling technique which is preferred over all the techniques.

(Refer Slide Time: 15:36)



We come to another important point, please understand and very important issue here and that is known as velocity saturation. Velocity saturation means that if you very well see, if you look at this graph velocity versus electric field, this is for an electron. You will see that as the electric field increases, the velocity increases in a linear fashion initially becomes the non-linear function of the electric field and then suddenly after certain point which is known

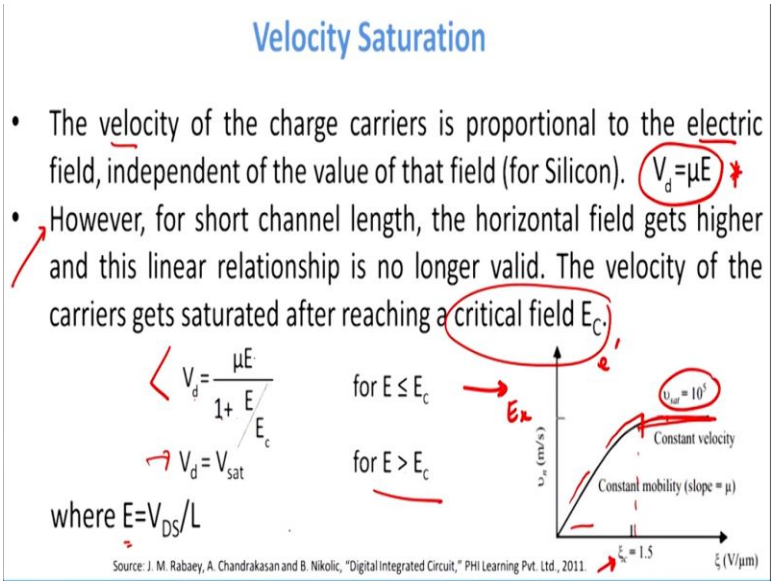
as the critical field and this case shown here at the velocity becomes constant independent of the electric field.

This is known as velocity saturation that means by velocity is actually saturated the exactly the same principle which you see when a rain drop falls from the sky. It first increases its velocity, then beyond the particular point it velocity becomes constant independent of the distance travelled.

And this is the reason why the reason is, as the velocity increases so does the scattering phenomena that means it heats larger number of charge carriers that it moves by. So if it, say it was moving with 1 meter per second. It was only hitting 5 or 6 ions in the path, right. And therefore it was scattering less. As the velocity increases, the number of scattering will also increase and therefore what will happen is the velocity will not be able to increase in that much heavy fashion because there it will always getting scattered, right.

And as a result, the velocity will become a constant phenomenon irrespective to electric field. So there will be two phenomena. One is the enhanced electric field which will like to drag the electrons from the channel towards source to drain and there will be another force which will be opposing it, that is scattering. When these two become exactly equal, my velocity becomes independent to the electric field, right. And this only for silicon for III-V it is something else.

(Refer Slide Time: 17:28)



So the velocity of the charge carrier as you can see is directly proportional to the electric field and therefore we define VD to be equals to Miu into E which is mobility into electric field. However that's very very important, the second point, the second bulletin is quite important.

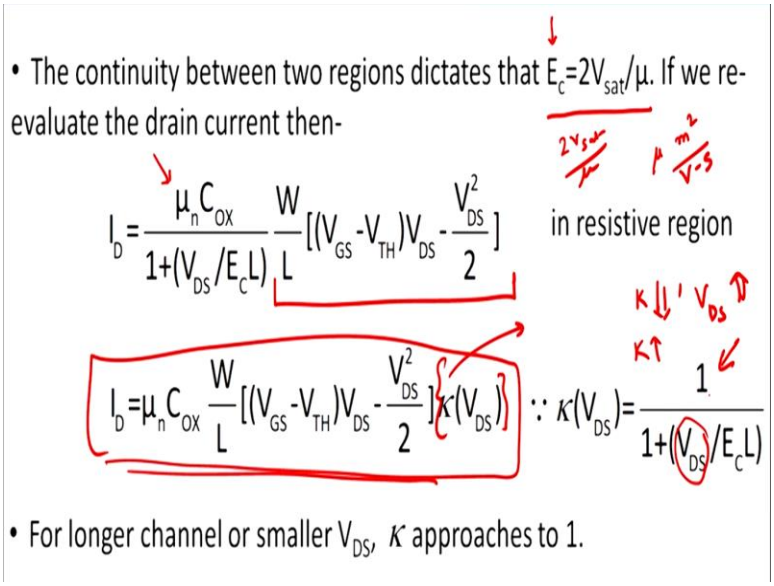
However, for short channel effects horizontal electric field which is basically this E x also starts to become higher and higher, right.

It starts to becomes higher and higher and beyond a critical point which is known as a critical electric field the velocity starts to saturate. So what we do is we define in two terms, right. The first term is that when E less than EC, EC is this point here. When E less than EC I will get velocity drip velocity VD to be = (Miu E) upon 1 + E / capital E C.

$$V_d = \frac{\mu E}{1 + \frac{E}{E_c}}$$

Beyond E equals to EC, I will get VD = VD sat that is what I am saying. Beyond this point, VD becomes VD sat and where electric field E is given as VDS / L, V DS is nothing but the voltage from source to drain difference divide by capital L, L is the effective channel length seen by the device, right. So this is what you get electric field at this particular point.

(Refer Slide Time: 18:37)



With this if you want to do it if you do a proper derivation of the previous result. I get that EC = 2 V sat / Miu 2 V sat is the velocity saturation value of voltage by Miu I will get critical electric field, right.

$$Ec = 2Vsat / \mu$$

So it is 2 V sat by Miu is the critical electric field which you see, Miu is Miu dimension is meter square per volt second, right. By if you want to evaluate again by the previous discussion if you look back, if you see this, this part is exactly the same we have been touched.

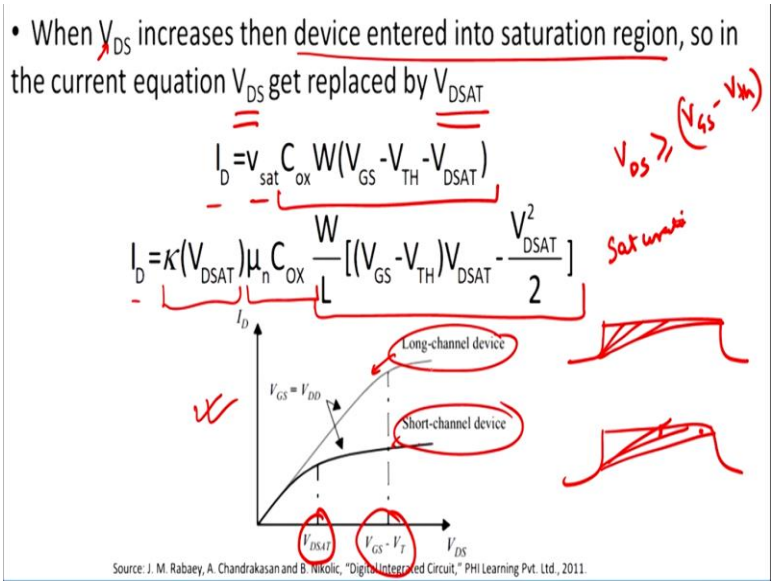
Only part is that we have replaced Miu / Miu N upon 1 + VDS / EC L by previous discussion and I get I D.

$$\mu \text{ by } \frac{\mu_n}{1+\frac{V_{DS}}{E_c}}$$

From here, I get I D to be equals to MuN C oxide and this whole term with this factor K right. This factor K will generally approach to 1 for very low values of VDS drain to source voltage, this quantity will becoming very large and it will becoming less than so K will be less than 1. For large values of VDS but K will be approaching 1 as VDS starts to go down and down, right.

And therefore there will be a maintain saturation taking place here quite interesting, if you just forget this part, right. This is exactly the same non-linear sorry, in the linear region the drain current which we are discussed in the previous two lectures. Only addition is the factor here K times VDS given by this formula which actually changes the drain current drastically, right. This is one important point which is mentioned and for henceforth onward we will be only referring to saturation velocity and V DSAT for all practical purposes, right.

(Refer Slide Time: 20:22)



So therefore as VDS increases, as VDS becomes larger and larger the device is force to enter into saturation region. Why? Please remember that VDS was greater than equals to V GS – V TH is the onset of saturation. So when a VDS is low, we are still in the linear region and as the VDS becomes larger and larger you go into non-linear region at the particular VDS value equals to V GS – V TH going to saturation region, right.

This saturation region will occur when the velocity of the charge carriers of becoming constant and is become independent of the electric field in this direction which means that the become independent of the value of the voltage applied between drain and source, right. Please understand, this is totally different from CLM, channel length modulation.

Here, we are assuming that there is no CLM there is no channel n modulation and therefore it is replaced by V DSAT. So, what we do is we replace VDS. What is known as V DSAT? V DSAT is nothing but the saturation drain voltage because, please understand VDSAT is the point at which the drain current starts to lose control over charge carriers and the velocity of the charge carriers have reach a fixed value and therefore current is fixed. So I D equals to V sat multiplied by this whole quantity and therefore I D will be given by K V DSAT into Mu n and C oxide into W by L into this whole quantity, right.

$$I_D = \kappa(V_{DSAT})\mu_n C_{OX} \frac{W}{L}[(V_{GS}-V_{TH})V_{DSAT} - \frac{V_{DSAT}^2}{2}]$$

And this is the only difference which we have when we consider velocity saturation as compare to other phenomena. So we therefore understand, we have two phenomena's to understand with one is a pinch of which we discussed in the first two slide another is your velocity saturation that means a pinch of so when the channel becomes very short this is what happens in long channel.

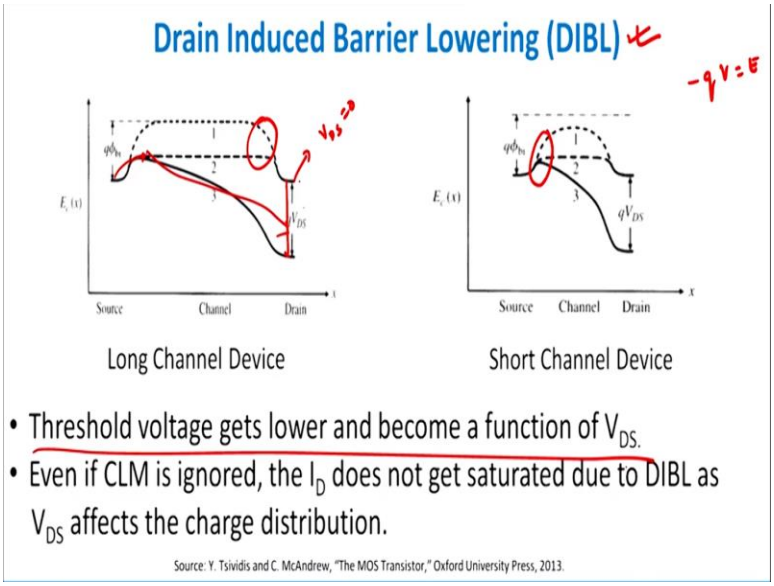
So as you increase the voltage, this pinch of here pinch of will go on increasing and move to this side. As you go to short channel devices when the channels is very short, then it is still remain like this as you go on increasing the voltage but at this point, it will be velocity saturated. It still decrease of course, decrease but the phenomena will be moreover velocity saturation, right. Because you have lowered your L therefore V / L becomes very high and if you can reach the critical electric field very fast and therefore output current, it becomes independent of the applied voltage.

So if you look at this figure here which u see in front of you, this figure right, then as you can see if you vary VD with ID I get IDVG, IDVD curve. This is for the short channel device and this is for the long channel device. This is for the long channel device and short channel device. And in both the cases, I we used VGS = VDD some value of VDS. As you can see

here, that short channel device my VDSAT is the reason why is it going to saturation which is much lower here.

Some long channel device it is V GS - VT which is relatively high. And therefore, the pinch of point for long channel device is large as compare to that in the short channel device. So these two things we should be very careful about that when you do a pinch of it is by virtue of velocity saturation in one case and pinching of by depletion region in other. The value of saturated voltage is slower lower in case of short channel device as compare to a long channel device.

(Refer Slide Time: 23:40)



We come to the next issue of the second order effects and important problem area and that is what is known as a drain induced barrier lowering, a DIBL effect. This is basically shows channel and drain and what of plotted is basically the conduction band diagram from source channel and length. As you remember, if on the drain side, if you give apply positive voltages then – q into V is the energy of the electron.

And therefore you will see that the on the drain side, the conversion band will move lower and lower as you increase the drain voltage and that is what you see. So it was this was initially so this was the case when your VDS was equals to 0. We applied a drain to source voltage what it did it just drop down the value here.

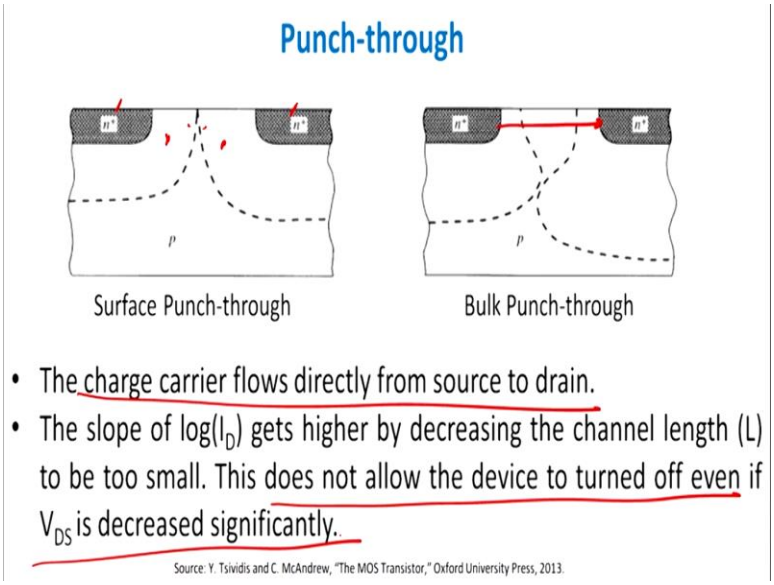
So now you see, the hump it was visible here, at this point has actually this no hump here and you can therefore if you can pick an electron and throw it at this point, it can easily slight down in energies domain and reach the value of drain side very easily, right. So this is one of

the short channel effects that in this case the drain starts to lower the barrier, right where at the source end.

At the source end, the drain starts to lower the barrier why because the drain is coming closer to the source. As drain is coming closer to the source and it is getting lower the value of the potential is lowered. I required a lower gate voltage to switch on the device. Why? Because since this side the height has been lowered I can afford to have a lower gate voltage to insert charge carriers or make it forward bias source to bulk forward bias and therefore threshold voltage go to lower. So what are the important things we have we take out of it we important things is the DIBL or DIBL we lower the threshold voltage very easily.

The saturation voltage the previous slide the saturation voltage will be much lower in short channel as compare to long channel device. And if you take CLM into the consideration in channel length modulation effect, the current will not that MOSFET will not behave like a ideal current source. It is start being like a non-ideal current source which means that currents becomes a strong function of the applied VDS and your output impedance is not equals to infinity. With this basic idea it makes explain to you what is punch through. Another phenomena when channel and the shortn as you lower the, see this.

(Refer Slide Time: 25:57)

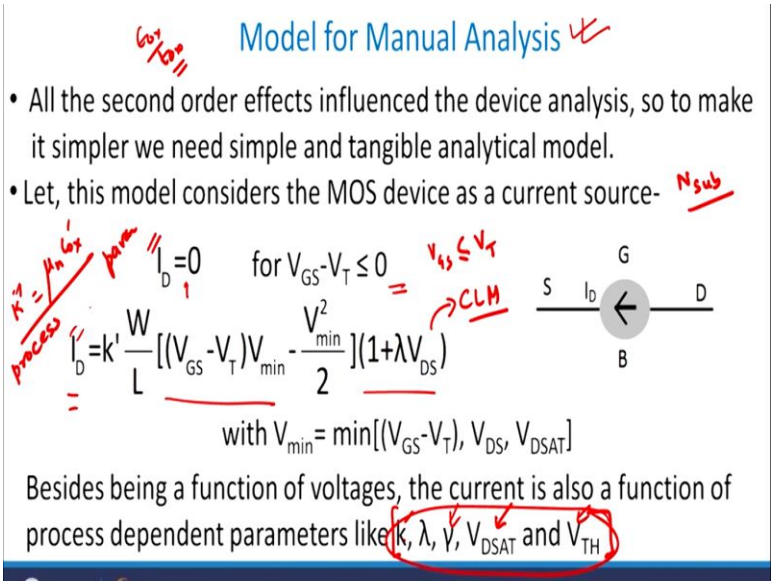


So if you look this is n+, n+ region have a p type substrate. This is the depletion region, right. If you bring it closer and closer chances are these depletion region will touch each other. And therefore, there will be a direct path between source and voltage. This is known as punch through. And therefore the charge carriers directly flow from source to drain and therefore

this does not allow you to turn off even if the VDS is decreased significantly why even if you decrease your VDS right, very very low.

You still have some current flow between source and drain why because this always a path available between source and drain for all practical purposes. This is known as a punch through phenomena of a short channel effect. So, the short channel effect has got an important phenomena that there will be always a problem of punch through, right. With this, we come to the last part of our talk here. And the last part is basically the model for manual analysis.

(Refer Slide Time: 26:49)



This is what we able be referring to which means that use a pen and paper and using simple mathematics as derived till now. Can we design a IDVD for us or a may be a drain current equation for us? As we discussed, if you assume all second order effects, right. We can assume, if you remove all second order effects. I can assume the MOSFET to be as a current source in saturation as a resistor in the linear region as per my previous discussion.

And current will be equals to 0 where when VGS is less than equals to VT again my a previous discussion, right. And current will be equals to K prime W by K prime is basically my Miu n C oxide. This is also known as process trans-conductance parameter. Also refer to as process parameter infact sorry, a refer to as process parameter here. And this is known as process parameter, right.

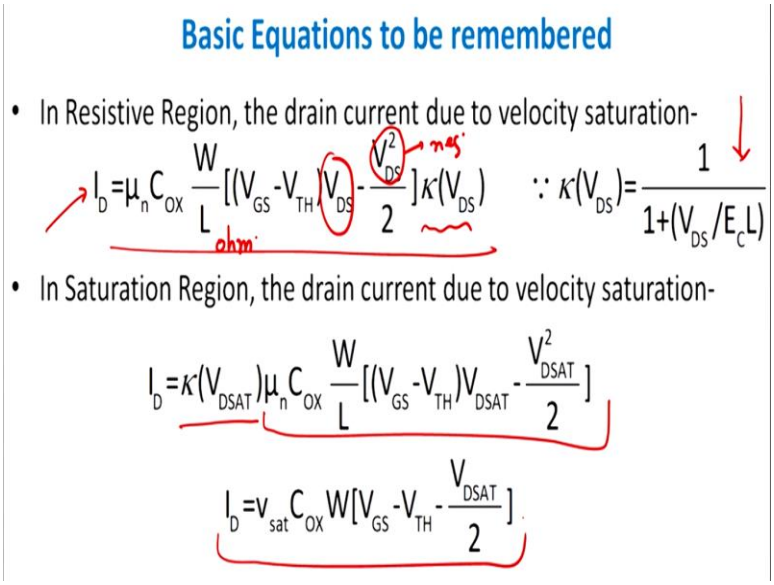
$$I_D=k'\frac{W}{L}[(V_{GS}-V_T)V_{min}-\frac{V_{min}^2}{2}](1+\lambda V_{DS})$$

And what we get from here is basically that ID will be equals to 0 for gate voltage lower than threshold voltage. And ID will have a value equals to K prime W / L VGS – VTH mean, a + 1 + lambda VDS this is assuming CLM into consideration, right. So therefore, if you want to calculate the current of a MOSFET apart from its 3 terminal voltages, you just require these 5 parameters and these are all process dependent parameters. So also process has been fixed, you are able to fabricate the MOSFET.

If you know the value of K which is nothing but Miu n into C oxide, Miu n is given to you by the user, C oxide is nothing but Epsilon ox by Tox, so if you know T ox, you know C ox and therefore you know K prime. You know lambda by large amount of experimentation, you know body coefficient factor again from body coefficient factor again here, you can easily determine it because it depends upon the value of substrate bias conditions again known to you.

V DSAT is again known to you when it calculate the IDVD characteristics. Threshold voltage again known to you. If all these 5 parameters are known to you. If it easily calculate the value of current using a manual analysis, that is the reason we are showing this for the manual analysis.

(Refer Slide Time: 29:08)

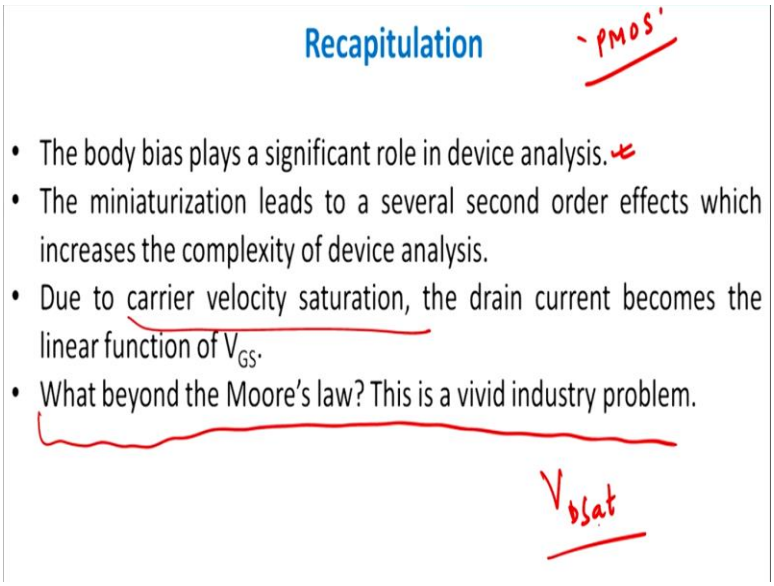


So let us see to recapitulate what important equations do you need to unit to be remembering these equations. And these equations are the first equation is basically the resistive region, the drain current should be of this form. Please understand, this is square term so of a lower value of VDS this can be negligible and therefore it only depends primarily on the value of VDS,

Ohm's law is followed at low values of VDS. If you take velocity saturation into consideration, you will have KVDS here, this term will be coming into picture.

In saturation region, you will have K times VDS and the same equation which you have derived in the previous slide and this will be the value of ID which you will get for your a basic equation. So for drain current equation you get the something like this, right.

(Refer Slide Time: 29:49)



So to recapitulate the whole idea the whole process flow what you have learnt from whereas the body plays a very significant role in device analysis. So therefore to recapitulate again or to put appreciate this point that you have bulk voltage should be much much lower as compare to the negative value of the system whole system which means that if you do not want any bias phenomena or body effect phenomena, try to keep the bulk voltage to the most negative value when you are doing a N MOSFET, for P MOSFET just the reverse.

Please, do it as a home work for p MOS. So I shown it for n MOS, please show it for PMOS yourself please do it yourself. We look into the fact that if we do a miniaturization by virtue of Moore's law, you will have many second order effects coming into picture some of them are CLM, we have velocity saturation and so and so forth. And then what we have seen is that due to carrier velocity saturation, drain current is becoming though independent of VDS it becomes a linear function of VGS for most practical purposes.

And what is beyond Moore's law is obviously a research topic and can be looked into by its a still a valid industry problem and can be looked into by many researchers over the years have I have been looking into it for over the years. So, just conclude this third lecture here we I

will just say the to important take away from all these three lectures as I can use MOSFET as a current source provided this no CLM. In such CLM, it does not behave as a ideal current. I can use MOSFET as a resistor in the linear region and that advantage of it that by changing the gate voltage I can change the on resistance and therefore the on current.

The third important point is that it the if I assume that the channel length has been shorten it becomes velocity saturated and the value of the pinch of point actually reduces and the we define therefore that point to be as the VDSAT. V DSAT is nothing but the point that drain to source voltage at which the saturation takes place in case of a MOSFET, right. With this the idea I will just stop at this stage and thank you for your listening. Thank you very much.