Microelectronics: Devices to Circuits

Professor Sudeb Dasgupta
Department of Electronics and Communication Engineering
Indian Institute of Technology Roorkee

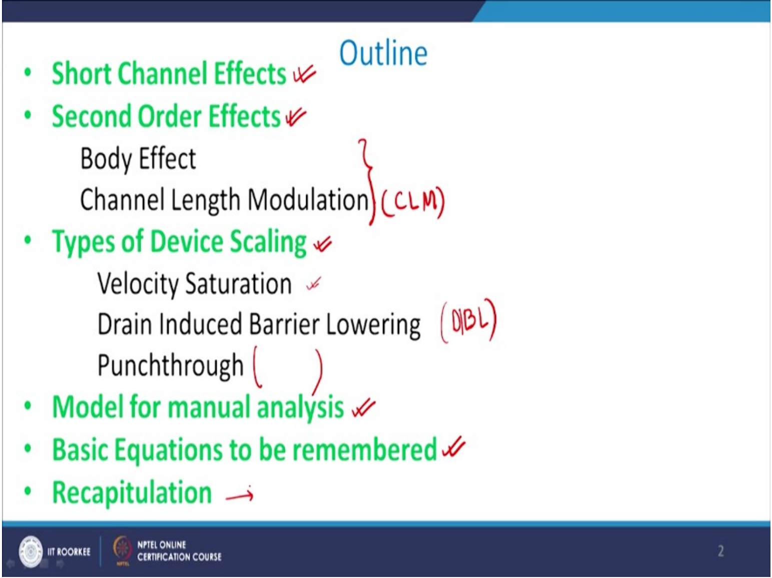
Lecture 16 - MOS Transistor Basics – III

Hello everybody and welcome again to the NPTEL online certification course on Microelectronics: Devices to Circuits. As we have, in our previous interactions, we have understood basic BJT which we have already done quite a lot and then we started with CMOS technology and to understand that, we understood what is the MOSFET technology and we understood therefore in the previous 2 slides or previous 2 interactions or modules, what is a MOSFET really, the structure of MOSFET and the functionality of MOSFET.

So first order, we saw that the MOSFET therefore can be the recapitulated as a current source when working in the saturation region, it can work as a voltage variable resistor when it is working in the linear region of operation. These two things we have already seen as a device. We have also seen that the MOSFET can be used as a switch when your gate voltage swings above and below the threshold voltage of the device. We have also seen that there are two types of devices, enhancement mode MOSFET and depletion mode MOSFET.

In one case, the threshold voltage is positive and in the other case, the threshold voltage is negative. Enhancement mode is also known as normally off-device and depletion mode is also referred to as normally on-device. What we will be looking subsequently is the following thing right and the topic of today is transistor basics. We will be going into the part three part.

(Refer Slide Time: 1:58)



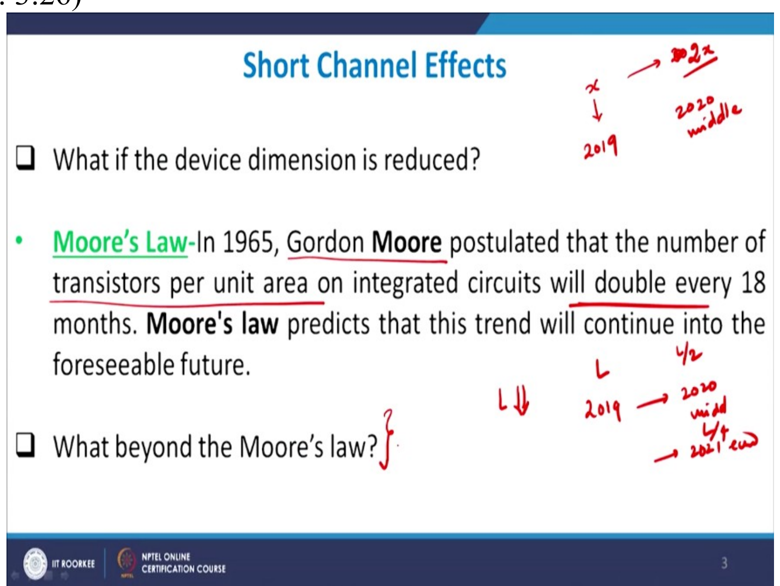
Within part 3, the outline of my this topic will be, we will be looking at various short channel effects, right. We will be looking at second order effects. Short channel effect primarily means that as you reduce the dimensions of the device, what second-order phenomena or effects comes into picture which is detrimental to the performance of the device, that is basically a short channel effect. We will be also looking at body effect as channel length modulation, this is CLM, right?

And this happens at extremely low dimensions and how does it influence the output electrical characteristics of the device, we will be looking into that. We will be also looking at the type of device scaling. Primarily we will be looking into two types of device scaling. One is electric field scaling, another is the voltage scaling. And then we will see its relative advantages and disadvantages. Within short channel effects, we will be looking into velocity saturation effects, right?

We will be actually looking into drain induced barrier lowering, also referred to as DIBL and we will be looking at punch-through which is basically a hot carrier effect phenomena. Then, once the MOSFET is, MOSFET as a device is clear to me, we will like always this MOSFET to be used in circuit simulation purposes. So can we have techniques of doing some amount of circuit simulation using, so can we have therefore some electrical equivalent models of MOSFET which can be used for circuit purposes?

We will also look at basic equation we should, we need to remember and then recapitulate the whole talk, right.

(Refer Slide Time: 3:26)



Let us look at the various short channel effects. Now why short channel effect came into picture? See, in 1965, Gordon Moore, he was a scientist at IBM and he predicted a very important trend, very famously known as Moore's Law which is very famous within the semiconductor industry as well as in academia. He stated that for every one and a half year growth of semiconductor industry, you would expect to see that the number of transistors per unit area will be doubling right which means that if the number of transistors per unit area is X today, then after so if it is at, this is 2019 let us suppose and then at middle of 2020, right, I would expect to see this the number of devices, active devices per unit area to be equals to 2X right.

And this was given by Gordon Moore who stated that the number of active devices will increase and why will it increase? To increase the functionality of the chip and to enhance the functionality of the chip. So those companies or those industries or those people who are able to sustain this Moore's Law have been able to increase the functionality of the chip and therefore the profitability is there when it follows Moore's Law. Easier said than done because when you want to reduce the, when you want to increase the number of charge carriers or number of active devices per unit area, the only option available to you is that you reduce the dimension of the device.

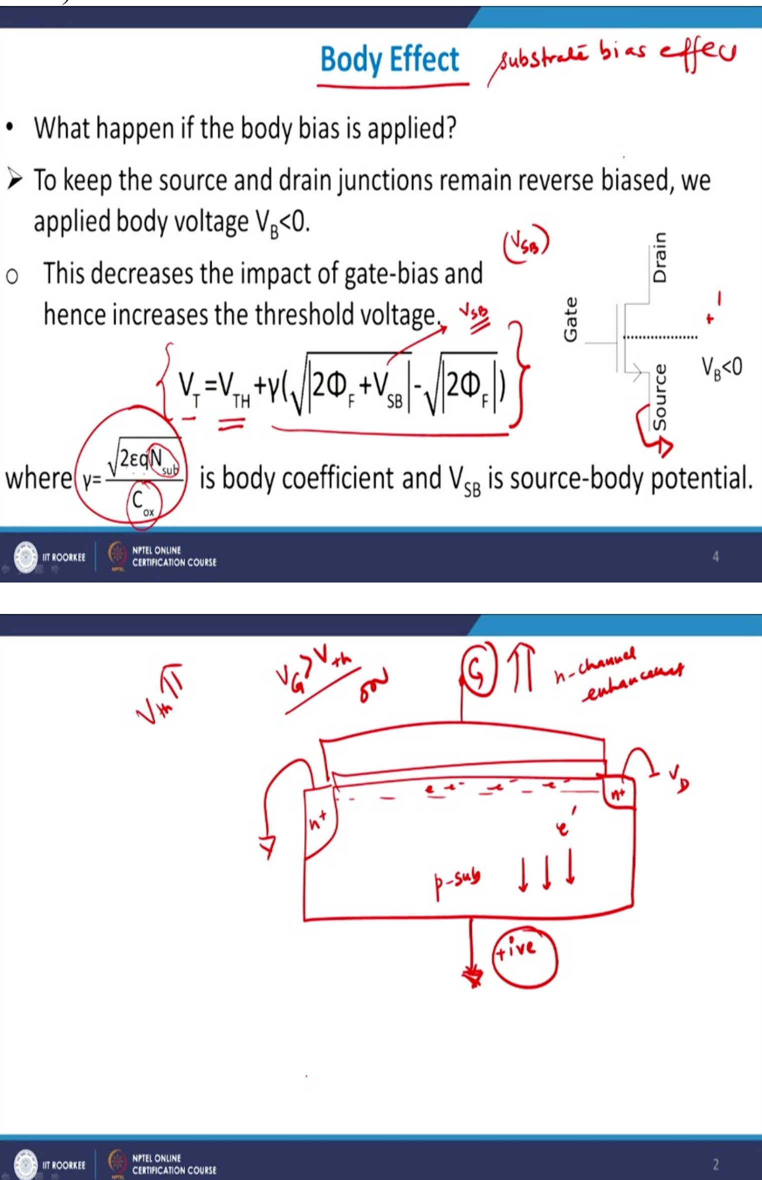
Because your silicon area is always fixed right and you cannot increase it drastically beyond a particular point because of fabrication limitations. So the only option available to you, if you want to follow Moore's law is that you want to reduce the dimension of the device right. So how do you reduce the dimension of the device? For example, in a MOSFET, how do you reduce dimension of the device? You simply channel length, you lower it. So if today your channel length in 2019 is L, then in the middle of 2020, 2020 middle I would expect to see this to be as L by 2.

Similarly, at 2021 end, 2021 end, this should be equals to L by 4 and so on and so forth. So you see for every, for example, 2 to 3 years of growth approximately, you are almost quartersizing your device, right? And so the idea was therefore that in case of a FET for example, when you are reducing the channel length, you are also making the source and drain come closer to each other right. Now theoretically speaking, therefore if you continue with Moore's Law, a time will come when that length and drain will touch each other or will be close to each other and they will be shorting.

Once the shorting is there, it will not work as a FET anymore. So therefore people are trying to find out alternatives to your bulk MOSFETs so that your performance, electrical performances do not get degraded at low dimensions. So that is the reason, I was saying what is beyond Moore's Law right and I would recommend that you can have a look at large number of research papers as well as large number of videos on the YouTube and a large number of open literature which actually gives you an idea what happens beyond Moore's Law.

Now therefore, if these dimensions are reducing, then there are certain things which are known as short channel effects which comes into picture. So that is what is we are referring to. That is once the dimensions are reduced because of Moore's Law, you have short channel effects coming into picture.

(Refer Slide Time: 6:59)



The first effect which I wanted to stress in front of you is basically the body effect right? Or also known as substrate bias effect, substrate bias effect or also known as body effect. We will come to this as we discuss. Let me explain to you in a very simple terms and then we can move forward, right? So if I have a p type substrate right and we have discussed this earlier also, we have p type substrate and I have got N $^+$ source, N $^+$ drain and I have got silicon dioxide metal and then you have gate, source is grounded right.

And then the drain is, you are applying a voltage V_D. Now in the substrate side, we generally ground it. For most practical purposes, we ground it. But let us suppose we do not ground it and

we start giving let us suppose a positive bias. A positive bias if you give and if it is an N channel enhancement model MOSFET right, then we can explain to you, as we have discussed this point V_G should be greater than equals to V_{TH} for the device to be on, which means that there will be large number of charge carriers here.

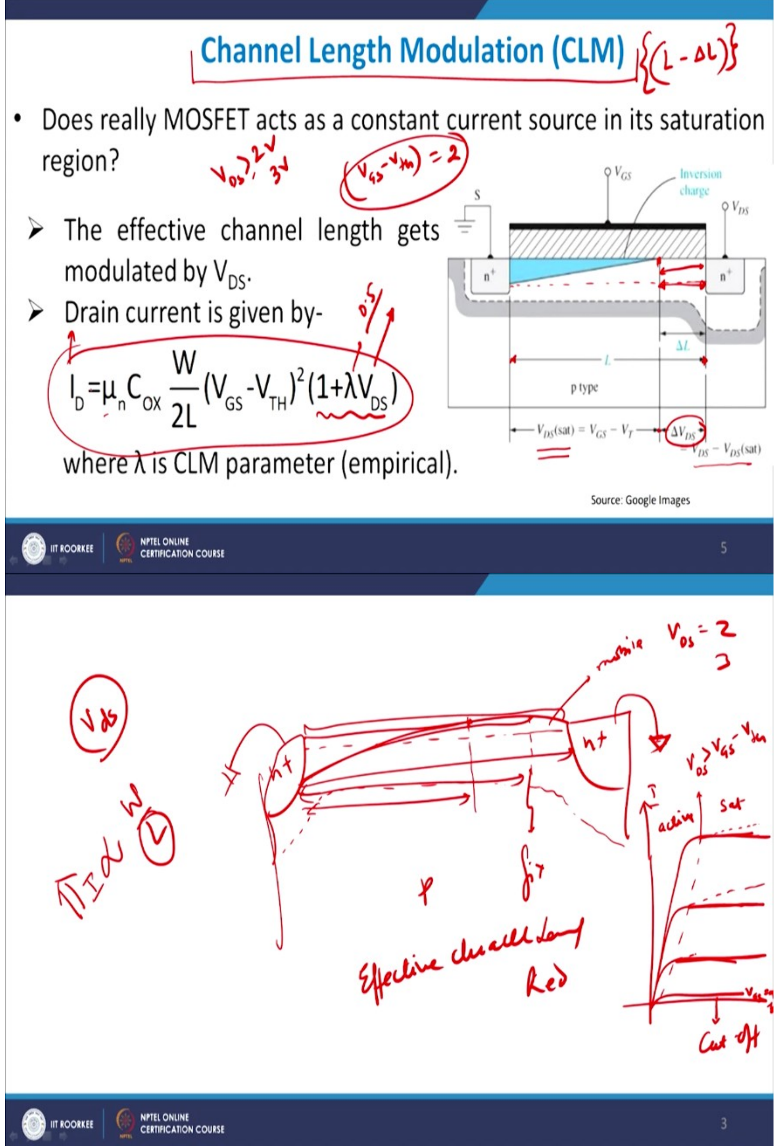
What are the charge carriers? Electrons, free electrons right, will be here provided V_G is greater than V_{TH}, threshold voltage. But now, you apply the positive bias on the substrate which means that there is this positive bias will be pulling these electrons towards itself. These electrons will be pulled here. As a result, your gate voltage has to further increase in order to form the channel. So what happens to threshold voltage? The threshold voltage rises. This is known as the substrate bias effect.

That if the substrate which was ordinarily grounded right, you apply a positive bias in case of an N channel MOSFET, you end up having a threshold voltage increased. Not a good idea but that is what is happening. Now so the definition is something like this that when your V_{TH}, V_{TH} is the value of threshold voltage without body bias effect and this is with body bias. And this is with V_{SB} is basically my substrate source, source to bulk potential right.

And you have to give the source to bulk potential, we do not have to worry about its sign, we do not have to worry about sign. If it is source to bulk, source is generally grounded right. Bulk will be given, so if we give bulk positive dimension, positive, then source to bulk will be positive. If it is negative, source to bulk will be negative in dimensions or in nature. γ is referred to as a body coefficient parameter given by 2ε Q into N_{sub}. N_{sub} is the substrate concentration of the doping concentration and C_{ox} is basically the oxide capacitance per unit area.

And Phi_F is basically my Fermi potential, right. So we have understood therefore that this V_{SB}, substrate voltage will have an adverse impact on the threshold voltage. So you actually started with a low threshold voltage and you might end up having a large threshold voltage which means that you require a larger gate voltage to switch it on, which means that you have a larger power dissipation for a particular device, which means that if there is a short channel effect, because of body bias, n channel enhancement MOSFET you give a substrate bias greater than 0, you might end up having a potential or having a threshold voltage which is larger than expected. This is the first thing.

(Refer Slide Time: 10:22)



The second short channel effect is basically by channel length modulation effect, right. And it is pretty simple and easy to explain but we will still nonetheless explain each one of them individually before moving forward. So we have understood what is basically a body bias or a back gate bias. Let me therefore give you an idea about what is channel length modulation effect. See, in our previous discussion, when we were discussing devices we came to the fact that whenever you have a channel, channel basically in our case means that it is basically having free charge carriers connecting source to drain in at least as much it is n type as it was initially p type.

That is what is basically known as a strong inversion phenomena. And the strong inversion therefore, you will have direct connectivity path available to you. Now what you do is that, on

the drain bias if you do not give any voltage then the substrate, then the depletion region on the substrate or the depletion region on the channel, near the channel this is N^+, N^+ and this is P. So if you do not give any bias, this is grounded, this is also grounded. Then the depletion thickness will look something like this, it will go something like this and it will be something like this.

It will be symmetrical in dimensions right. Now if you go on increasing it, a time will come when you will have inverted layer here. So this will be mobile charge carrier, this will be mobile and you will have fixed charge carriers here, fixed here. Again we have still grounded our drain. Now what I do is, I do something else and that is that I increase the value of your V_{DS}. So I go on increasing the value of V_{DS} right. And let us see how does it influence the overall current carrying capability of the MOSFET.

If you remember yesterday's talk or the previous lecture's talk, you will see that when we were discussing saturation, we saw that this is parabolic in nature and this was happening at V_{DS} greater than V_{GS} - V_{TH}, this was saturation right. This was active right. And this was cut off. What was cut off? Whenever my V_{GS} equals to 0, I define this to be as cut off or equal to threshold voltage, we define the lower value to be cut off. So V_{GS} equals to V_{TH} with cut-off.

Now you see, at saturation also referred to as active in some of the textbooks. If you look very carefully, this is almost a straight line which primarily meaning that the current is constant independent of the voltage right. And it starts to behave like constant current source. But in reality, we will see that the channel modulation effect does not allow you to do that. Let us see why. So as you make your drain voltage higher and higher, the depletion thickness here becomes larger and larger and it starts to eat away into the channel.

And as it starts to eat away into the channel, you see the inversion charge is only restricted up to this much point. However, if you did not have applied any bias and your drain voltage was 0, you are inversion layer would have been something like this. Right? All filled up with free electrons but now since we have apply a drain bias which is quite large, the depletion region itself eats away into the inversion layer and your effective inversion charge carriers are shifted from this region.

So the channel length was which was initially till this to this point, from this to this point has actually reduced by a factor ΔL. Therefore the effective channel length is $L - \Delta L$. Right. And therefore your, if you define channel length as that length at which the inversion charge is available, then by that definition, you will automatically have a reduced channel length which is given by $L - \Delta L$ at this particular point. Right. Which means that therefore, if I give a potential $V_{DS\ sat}$, $V_{DS\ sat}$ is the saturation drain to source voltage, then $V_{GS} - V_{TH}$ is the voltage drop till this much point. Right?

And the extra V_{DSsat} which makes it saturation falls across this region. So thisΔV_{DS} which is $V_{DS} - V_{DS\ sat}$, is the voltage which is required here. For example let us suppose, to form the, this is 2 volts. So $V_{GS} - V_{TH}$ to require to form is 2 volts, right. So I require that V_{DS} should be greater than 2 volts for saturation. Fine. That is okay. But then if I give 3 volts V_{DS}, then out of the 3 volts, 2 volt actually goes to form this channel right, this channel.

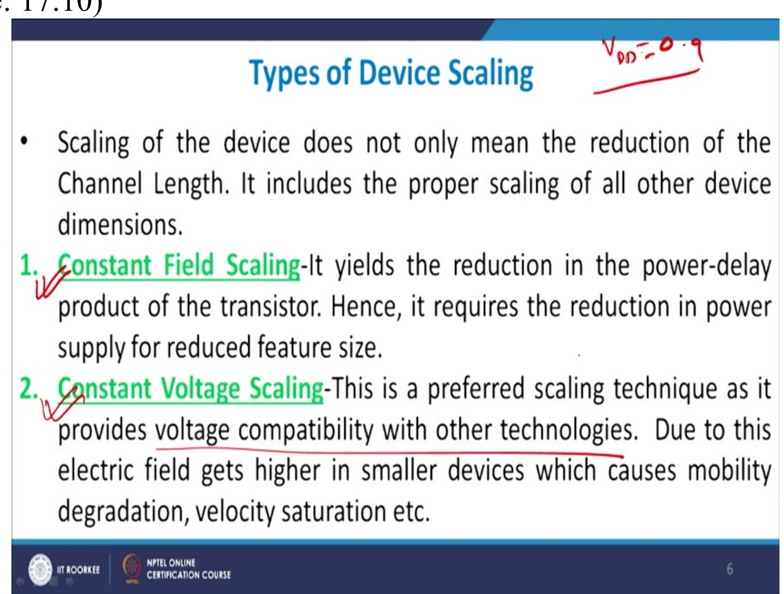
And the rest 1 volt appears across this particular region. Fair. And that is known as ΔV_{DS} or which is given by $V_{DS} - V_{DS\ sat}$ right? As a result, if you look very carefully therefore that therefore what I will see here is, so I had, so now my depletion, so my inversion layer was somewhere here when we apply V_{DS} say equals to 2 volts. You increase it to 3 volts, the inversion layer further comes down. It was initially here, now it is here, which means that the effective channel length is actually reducing.

It was initially, initially it was this much right. Now it became this much right. And now further increasing V_{DS} became this much. Fine, which means therefore that current which is proportional to W by L right, so L is discussing with increasing V_{DS} which means that the current is rising and therefore assuming that current is constant in the saturation region is a wrong assumption and therefore the current will actually show an increase. So I will show an increase.

This increase is primarily because of an increase in V_{DS} and reduction in the effective channel length. That is what is shown here and therefore the drain current is approximated as this value. Remember, if you remember the previous talk, it is $\mu_n C_{ox}$ W by 2L into $(V_{GS} - V_{TH})^2$. Even extra term here which is $1 + \lambda V_{DS}$, right? λ is defined as the CLM parameter or channel length modulation parameter. It is an empirical constant varying from 0 to 1 and V_{DS} is the applied voltage which you see.

So if λ is equal to say 0.5, then for every increase in V_{DS}, I would see that 50 percent of V_{DS} will be responsible for increase in the current at higher values of drain voltage. And therefore, from the IV characteristics of the MOSFET, what thing becomes clear is that the current will increase with increasing voltage. Right? The current will go on increasing with increase in voltage; otherwise we are not allowing to do that because we are making it fixed value of voltage, right? So that is what is known as channel length modulation effect.

(Refer Slide Time: 17:10)



Now coming back to therefore to the basic concept of device scaling, we have two types of device scaling. The scaling will be one is known as the constant field scaling or electric field scaling and we have constant voltage scaling which is available with us. Now scaling does not only mean scaling the channel length. When you scale, when you do a structural scaling, when you scale up or scale down, you scale up everything. So you scale down channel length, you also scale down the oxide thickness, you scale down the metal contact, you scale down the base, the source contact and the drain contact.

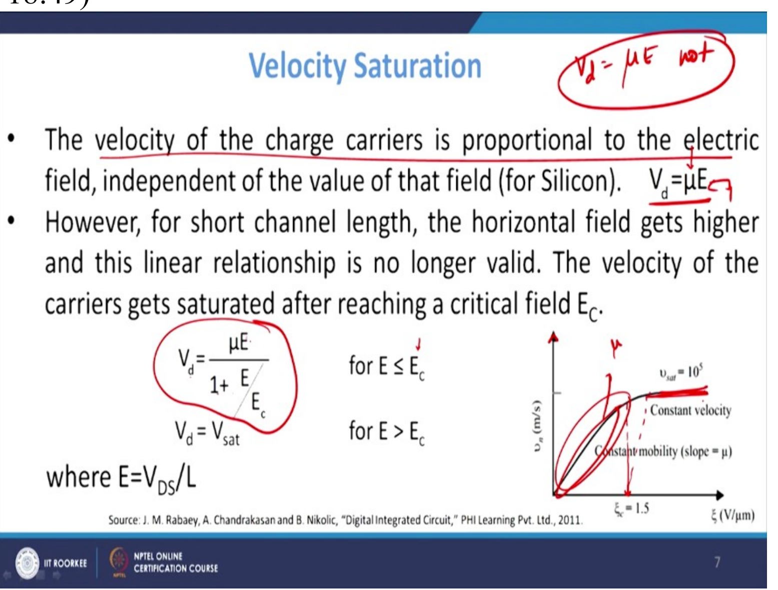
So everything has been scaled down or scaled up, right. So it is not just this channel length which you are doing it together. Now the constant field scaling primarily means that in this case we assume, what we do is that as we scale down our dimensions, we also take into consideration the electric field remains constant. Right? To do that, you have to scale down your voltages in a

straightaway fashion so that V by L is constant. Right? V by channel length or V by dimensions is almost constant. What is constant voltage scaling?

Well, this is a preferred scaling technique as it provides voltage compatibility with other technologies. Due to this the cost you pay for it is that you start, in the 1st case whereas when you lower your channel, you also lower your V_{DD}. In this case, you do not do that. You keep your voltage constant right. So when you keep your voltage constant then the electric field enhancement is there, mobility reduction is there and so on and so forth. But then, it is compatible with technology.

So if you want to work with a 90 nanometer technology, if I use a V_{DD} of 0.9 volt, when I do not have to do any scaling and it will work, give me an optimized result available to me. Right? So for a single amount of scaling, this gives me a very well, a well-defined result as far as this device is concerned.

(Refer Slide Time: 18:49)



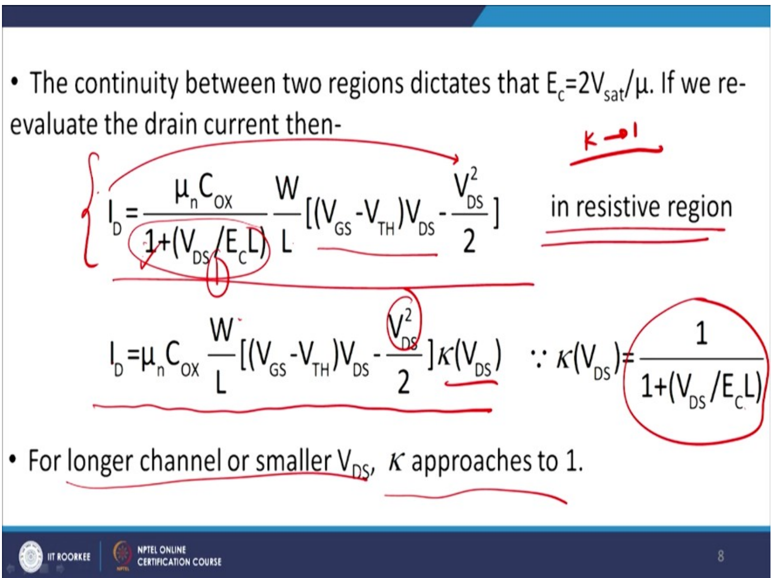
We come to the velocity saturation but as we have discussed earlier, that the velocity of the charge carriers is proportional to the electric field. So V_{Ddrift}, this V_D is known as the drift velocity, is equals to μ, mobility of the charge carriers multiplied by the electric field. Right? However for a short channel device, the electric field in the longitudinal direction and the horizontal direction and transverse direction are both equally high. Right. And therefore this linear relationship of V_D equals to μ times E is not valid anymore, is not valid.

Right? Why? Because now the electric field is not only in this direction very strong, it is also very strong in this direction. So under the influence of these 2, you will have a new value of V_D and that value of V_D is given by this formula which you see in front of you, given as μ into E upon 1 - E less than E_C, where E_C is the critical electric field. So critical electric field is defined as that electric field after which the mobility is constant or after which the velocity is constant. So if you plot electric field velocity versus electric field for silicon, it looks something like this.

So this much point, since V versus E is always constant or maybe this much point right, if this is almost linear profile, so mobility is constant. Beyond this, the mobility starts to decrease and beyond a particular point, the mobility is actually very low or very small. And V_{sat} is typically very large. V_{sat} is the saturation velocity which you see. Now therefore, you see that your V_D here is equals to V_{sat} right? And that is understandably so that the applied drain voltage is almost near the saturation voltage and we get a very large value and the increase in velocity in case of, so rather than the pinch off of the depletion region, I would actually get a pinch off in the velocity of the charge carriers.

That will result in almost a saturated drain current for a MOS device for short channel effects because the electric fields are very very high right.

(Refer Slide Time: 20:55)

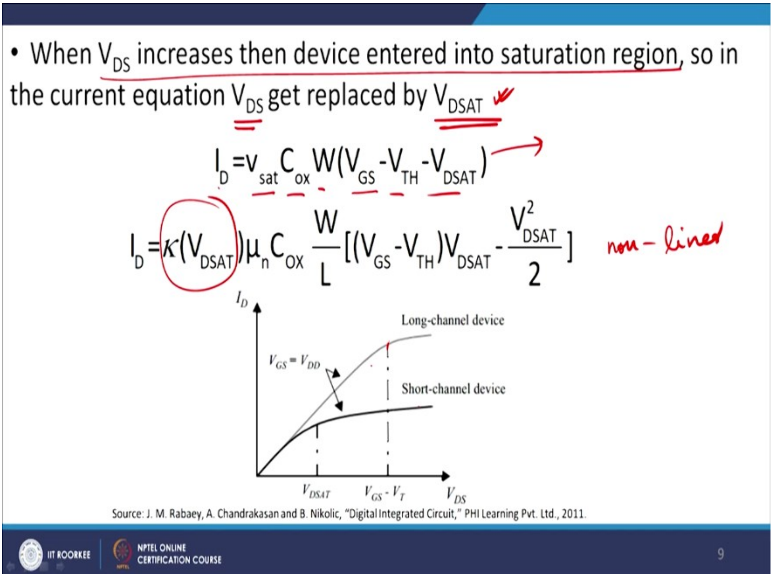


Now if you plot it, then we get, if we evaluate it, we get something like this that the drain current I_D is given by this basic formula, if you remember yesterday's talk, it was basically 1 + V_{DS} by E_CL and this will give you the value, W by L is coming here and V_{GS} - V_{TH} into V_{DS} and V_{DS}^2. So this is for the resistive region. And therefore you see, I_D has a non-linear dependency on V_{DS} because it is a square term which you see in front of you. Similarly, if you look at this curve, again, I_D has got a non-linear dependency on V_{DS} right? Where K V_{DS} is given by this one formula.

So this formula is nothing but this whole thing. Right? So I define K to be equals to 1 by 1 plus this thing and 1 + V_{DS} by E_C into L. E_C is a critical electric field, L is the channel length right. Now for, actually what is happening is for long channel lengths, K will approach to 1. So 1 plus 1 is equal to 2 and therefore you reach the conventional equation which we have already studied earlier right?

That μ_n C oxide by 2 WL into $(V_{GS} - V_{TH})^2$, right. So that is what we reach if provided K approaches to 1 when you have a longer V_{DS}, when you have a longer channel length or a smaller V_{DS}. And in those cases, we reach to this value in this case.

(Refer Slide Time: 22:16)

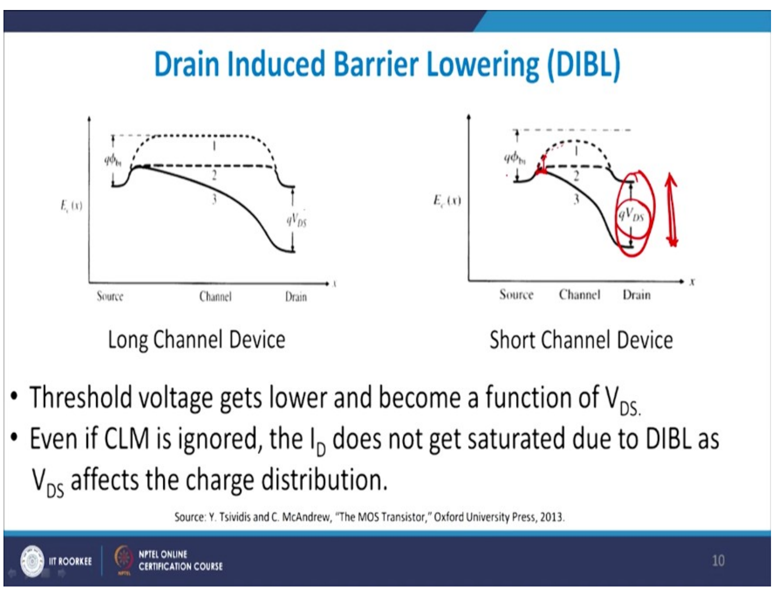


When V_{DS} increases therefore as you can see the device enters into the saturation region. We have already discussed this point. So what we do is that we replace V_{DS} by $V_{DS\ sat}$. Right? So whenever you are sure that the device has entered into saturation region, then the drain to source voltage which you apply is rather than writing it as V_{DS}, we write it as $V_{D\ sat}$. Right? And that is $V_{D\ sat}$. $V_{D\ sat}$ means saturation velocity drain voltage.

So I get I_D equals to V $_{sat}$ into C oxide W into V_{GS} - V_{TH} - $I_{D\ sat}$. So this V $_{sat}$ is again broken down into K $V_{D\ sat}$ and therefore I get this into V_{DS}. So this is for the non-linear region right, for the non-linear region and this is for linear region of the operation of the device. Right. Which comes to the final expression therefore that if we have a long channel device and if you have a short tunnel device, automatically the short channel, so long channel device saturation will take place at a much higher $V_{D\ sat}$ whereas for a short channel device, it will take at a much lower V_{Dsat}.

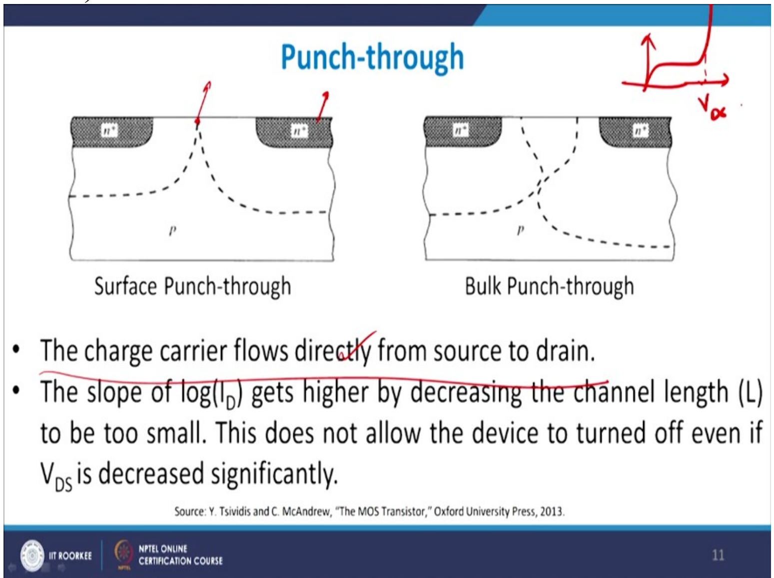
Because short channel, you require lesser amount of drain voltage to pinch off the carriers and therefore it is much smaller as compared to the first case. We come to another important topic and that is basically drain induced barrier lowering. At this stage, we will not go any further than saying that initially when drain and source are very far from each other, drain has got no influence over the source but what is happening is, as you go on bringing drain and source closer to each other, then the change in the drain voltage will also start influencing the source side.

(Refer Slide Time: 23:58)



So as you can see here that as you make your drain voltage higher, the channel length, the conduction band diagram, Q into V_{DS} there is a drop of voltage, right? Because into electron, you multiply. So there is a drop in the voltage, which means that you actually saw a larger current but now you actually see a dip. So there is always a decrease in the hump. As a result what happens is that that it affects and lowers the threshold voltage. The reason being, your built-in voltage is reduced. Right? Because of DIBL or DIBL effect. As a result, your charge carriers are more and more and current becomes more and more right and it affects the charge distribution model for this case.

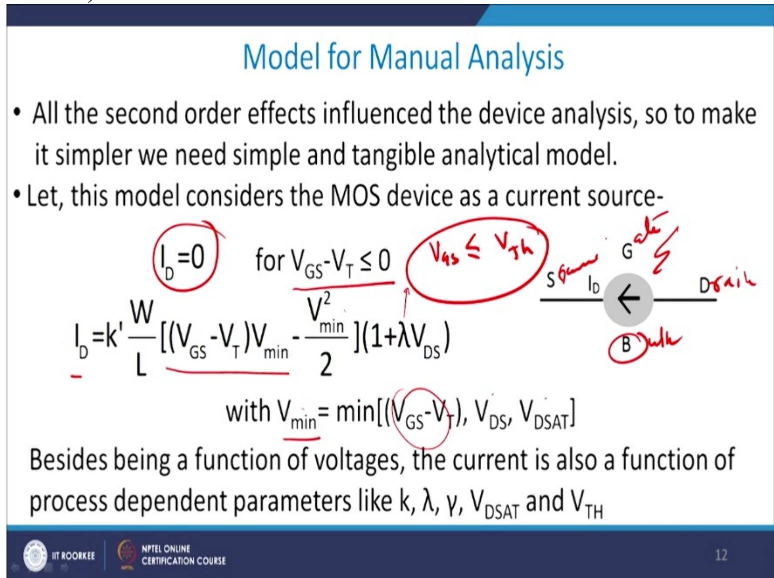
(Refer Slide Time: 24:46)



We come to the last part that is basically your punch through. What happens in a punch through is that if you go on increasing, as you can see, as you go on increasing the drain voltage, this depletion region touches each other, right? And this condition is known as punch through right. So charge carriers directly flow from source to drain. So you do not require the channel right? The drain voltage is so high that any electron can be pulled away from the source and it can pass through the depletion thickness and reach the channel and that is known as the punch through phenomena.

For most practical purposes, we do not use punch through phenomena in any case right. So what we do is that if you increase the value of V_{DS} significantly then this will happen. But in most of the cases, so therefore if you plot I_D versus V_D then it will be something like this and then suddenly it becomes large for V_{DS} somewhere here, it becomes large. We do not operate generally in this one but this is known as punch through which you see.

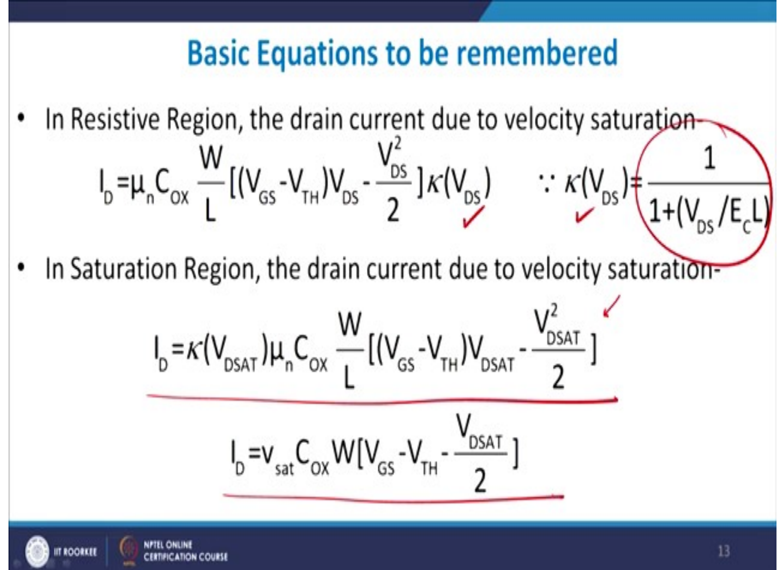
(Refer Slide Time: 25:33)



Now therefore the model which you apply for manual analysis which means that the model which you will be or second-order effects can be taken care of by a simple tangible analytical model and the model is that when V_{GS} - V_{TH} is less than 0 or V_{GS} is less than equals to V_{TH} for an n channel enhancement mode MOSFET, the current through the MOSFET is 0 and whenever my input voltage is larger than threshold voltage, then I get I_D to be equals to K' W by L V_{GS} - V_{TH} V_{min} - V_{min}^2 by 2 into 1 + λV_{DS} right.

V_{min} is basically V_{DS}, minimum value of V_{DS} and λ is the CLM parameter which you see and V_{DS} is the drain to source voltage which you see. So V min is basically minimum of the overdrive plus V_{DS} and $V_{D\,sat}$ right? Out of the 3, whichever is the really the smallest one, we assume that to be equals to value of V_{min} but typically, it is V_{DS} which comes into picture. And it is replaced by a current source as you can see with a current I_D and the source, drain, bulk and a gate. So this is bulk right, this is drain, and this is gate and this is source and this is the drain current which is flowing through the device.

(Refer Slide Time: 26:49)



So let me see what basic equations you need to remember as far as this lecture is concerned. The basic equation you need to remember is this one where K V_{DS} is given by this. This is also known as velocity saturation equation. Right? In the saturation region, as I discussed with you, the drain current is due to velocity saturation, which is therefore given by $V_{DS\,sat}$ and I_D equals to V_{sat} into C oxide and W by L V_{GS} -V_{TH} - $V_{DS\,sat}$ by 2. So once you know the value of V_{DSsat}, you can find the value of I_D in a much more detailed manner, right?

And that gives you an idea about the various basic equations you need to follow. So equations are exactly the same as you follow for a bulk MOSFET, small changes but when you do a short channel effect it is primarily because of velocity saturation. Whereas for long channel effect, it is primarily to do with pinch off at the drain side, right. And therefore these 2 are totally different issues which you need to find out. Let me therefore recapitulate today's lecture in detailed manner which we have collected.

(Refer Slide Time: 27:47)

Recapitulation

- The body bias plays a significant role in device analysis.
- The miniaturization leads to a several second order effects which increases the complexity of device analysis.
- Due to carrier velocity saturation, the drain current becomes the linear function of V_{GS}.
- What beyond the Moore's law? This is a vivid industry problem.

IIT ROORKEE | NPTEL ONLINE CERTIFICATION COURSE 14

We have understood what is basically a MOSFET technology, we have also understood that the body bias, which is the third terminal body bias, fourth terminal body bias plays an important role in determining the change in the threshold voltage of the device. Now, we have also understood that one of the important effects of miniaturization is the drain and source coming closer to each other and therefore you will have drain induced barrier lowering which means that the voltage change on the drain side will influence or will reduce the barrier on the source side because source side if you remember is basically a depletion width is there.

Now if I am able to somehow or other reduce the depletion width by giving an external bias and this case happens to be the bias given to the drain side and the drain is very close to the source side, so any variation in the drain side is replicated on the source aside. Then we are able to reduce the dimensions or we reduce the barrier on the drain side and thereby changing the threshold voltage of the device. Now the third point is due to carrier velocity saturation, I get the drain current is almost a linear function of V_{GS} right.

So it is $(V_{GS} - V_{TH})^2$ and therefore it is a linear function of V_{GS} right? Well. Beyond Moore's law, we do not know what will happen. There are many alternatives which are coming out but these alternatives need to be tested in terms of its performance analysis and more importantly, these structures should be able to be, can be fabricated in the current foundries which are available to us, right. So this is the basic idea or the basic concept or the MOS device second-order effects,

right. So we have dealt with the basic concept and we have understood the basic idea of this phenomenon. With these, let me thank you for your patient hearing and I will be open for any questioning. Thank you!