

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

Lecture – 32
MOSFET

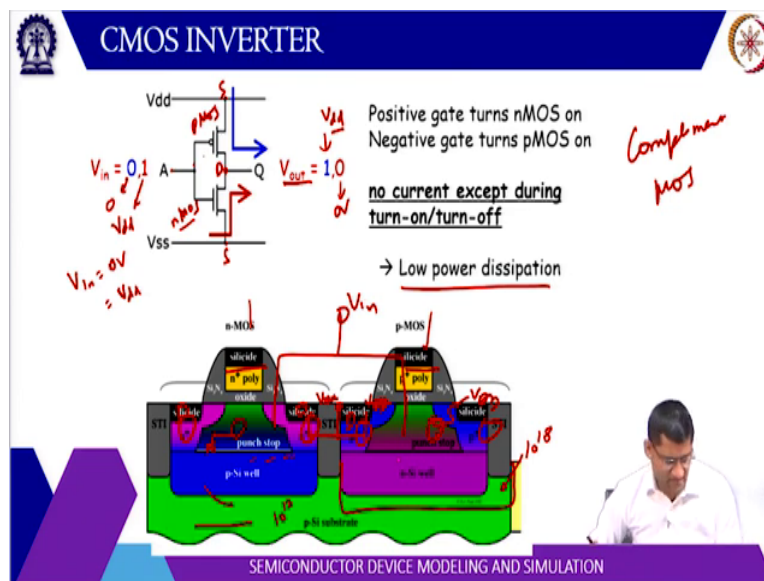
Hello, welcome to lecture number 32.

(Refer Slide Time: 00:28)



We will continue our discussion on CMOS, MOSFET structures. So, we will discuss two things, CMOS inverter and then non-idealities in the MOSFET device.

(Refer Slide Time: 00:37)



So, typical CMOS, if you see here CMOS refers to complementary metal oxide semiconductor devices. So, complementary means there is a nMOS here and there is a pMOS here. So, nMOS is connected between the output and ground. So, when this conduct nMOS conduct this output is connected to the ground pMOS is connected between the supply and the output. So, when pMOS conduct the output is connected to the supply.

And the gates of these two pMOS and nMOS are connected together. So, 0 refers to a state 0 so that 0 volt one is referred to the state one so that is Vdd voltage. So, when we apply a 0 volt then this nMOS will be off because the gate voltage for nMOS has to be you have to apply a positive voltage for n was to turn on. So, V_{gs} is 0 for 0 volt input. So, when V_{in} is 0 volt, this nMOS is off pMOS will be on because if you see with respect to source.

This is a drain, this is a source so, with respect to source, the gate is at lower potential. So, this pMOS channel will be on and then output will be connected to supply, so, your output will be Vdd. When input is Vdd then this gate of nMOS will turn the nMOS on and pMOS V_{gs} will be 0, so, this will be off. So, your output is connected to the ground, so, your output is 0 volt.

So, if you see here in either case whether the state is 0 that means output is 0 volt or the state is one that means output is Vdd. There is no current flow as such. There is no part between the Vdd and ground either nMOS is on then pMOS is off or if pMOS is on, nMOS is off. So, in steady state it does not draw any current only when it is transiting from 0 to 1 or 1 to 0 this will draw some power.

So, overall conclusion is that this structure actually draws very less power or low power dissipation. So, if we see it how do we fabricate this kind of CMOS structure? Let us say be a p-type substrate here. Then in p-type substrate we can easily follow the form, the nMOS structure, so, we have a source here, drain here, gate here, this oxide thickness. This is little exaggerated, the oxide actually quite very, very thin actually few tens of nanometre.

Then this is basically connected to the substrate so, this is called body contact. So, you can bias the body also and then this basically is your nMOS structure. Then on the same substrate, we have to make a pMOS structure but that requires a n well, so that because

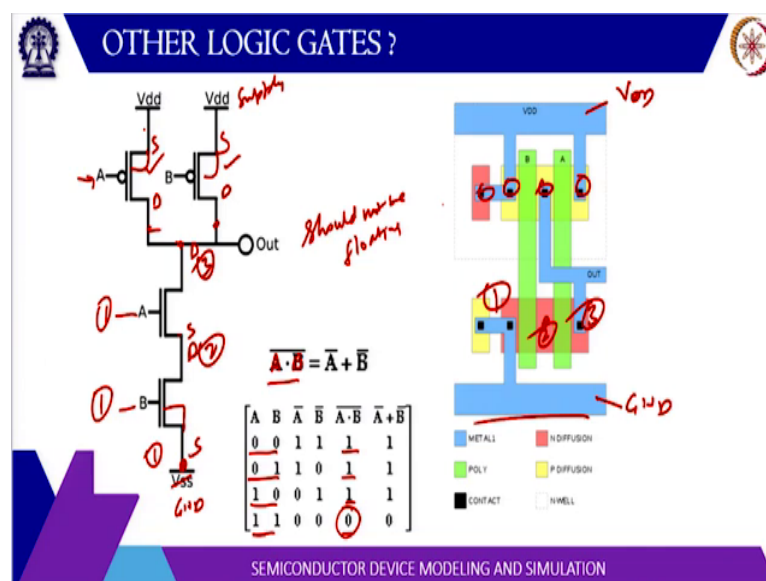
n-type of state. So, how do we create n-type substrate? What we do? In the p substrate we create a n well. So that means, if this is p substrate with let us say 10 is to power 17 dope in.

We will dope it with 10 is to power 18 n-type, so, it becomes n well because it will compensate distance of 17 and 10 is to power 18 extra donors will be there. So, this will become n-type substrate now. Now, here we can make p + source, p + drain and the body contact and the gate here so, this will be your pMOS. Now, these has to be connected so, if you see here, this gate and this gate they are connected together.

So, this is your input and then pMOS source is connected to Vdd and these two drains are connected together. Actually, there is no difference between, you know there is a symmetrical device, so, you can write this as a source. You can write this as a source this has a drain. So, the source is connected to Vdd and this drain is connected to the drain of nMOS and this is connected to ground. So, output is taken from here V out.

So, this is a structure or this is the planar showing here but actually there are lot of other things also there, if you see here. So, this is basically your source drain contact here it is a silly site then you have this n + poly instead of metal for nMOS then p + poly for pMOS and this is the n well. And the doping is actually not that uniform but follow certain restricted by the doping method process and then we have this kind of a structure.

(Refer Slide Time: 06:00)



What about other logic? Can we implement or gate, nor gate or nand gate. So, here you see one example of a nand gate. Now, nand gate is $A \cdot B$ bar then you can draw that table basically,

so, $A \cdot B$ means, when both are 0 output will be 1. When any 1 of 1 another is 0 still output will be 1 when both are 1 then output is 0. So that means, when both are one then output should be pulled down.

So, you can easily figure out that this nMOS have to be in series. So, when both are 1 both are conducting then this will go to 0. Otherwise, it will not go to 0. So, any one of them is 0 then output is not connected to the ground. So, how to connect it to the Vdd? Because when it is not 0, it should only floating. So, idea is that output should not be floating, it should be connected either to the Vdd supply or to the ground.

So, this is the ground and this is the supply. So, when either A or B is 1 it should be connected to the supply. So, we connect this pMOS in parallel. So, if A is 1 then this transistor, this pMOS will be on and when B is 1 this pMOS will be on and when both are 1 then both are on basically. So and how do we implement this kind of logic gate? Now, we have how many gates, 2 pMOS and 2 nMOS.

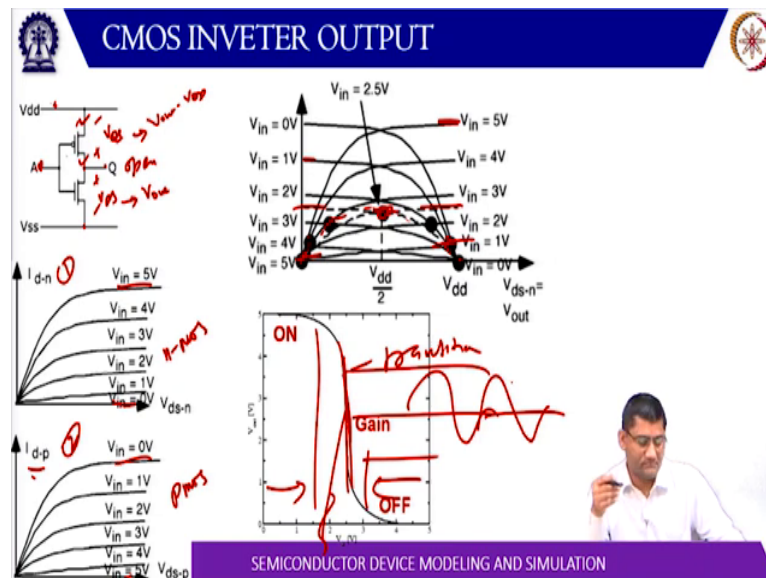
So, there is a technical layout, so, here you can see the layout basically so, here this is your Vdd this is a metal layer, this is your ground this is also a metal layer. Now, this metal is connected to the ground is connected to if you notice here. This ground is connected to this is the drain, this is the source, this is the drain, this is the source. So, it is connected to the source of one of the nMOS ground and this source is connected to the body.

So, you see here this body and source are connected together for this structure and then upper Vdd is connected to the source and drain. So, two sources is connected to the two sources, so, this is basically a structure with two sources. So, an A B are the two inputs, one to the nMOS one to the pMOS. So, this source and drain are connected together. So, this is the source centre connected together and this is the let us say let me type the point here.

This is 1, so, this is basically 1 here then this is 2 here, so, this is 2 here, this is 3 here so, this is 3 here, so, it is connected to the now these two are actually this point. And then of course, both the sources are connected to the Vdd and this body is connected to the Vdd. And you can see the green color is basically the two gates for nMOS and pMOS. The upper one are the pMOS lower one are the nMOS and these are the metal.

Now, you see there is some kind of intersection in these lines but in physical, design there is no intersection because they are at different layers. So, these are the whenever there is intersection they are a different layer. So, it does not really affect. So, this is just a preview that how the mass structure can be used to design the logic gates?

(Refer Slide Time: 10:17)



Now this is the CMOS inverter output analysis, so, you can yes see the I V characteristic for this is I d V ds this is for nMOS. This is I D V DS per pMOS. So, when input is 0, this is not conducting when input goes to 5 it is conducting. For pMOS when input is 5 volt that means V_{GS} is 0. So that it is not conducting and when input is 0 then V_{GS} is -5 volt. This is conducting. Now, what you can see? I d is same for both of them.

The drain current has to be same because this less A Q is open and the V DS. So, this is V DS for n-type this is V DS for p-type, so, this is drain, so, this plus minus this is plus minus. So, this total $V_{DS1} + V_{DS2} = V_{DD}$. So, you can say let us say this is your V DS which is equal to V_{out} , so, this V DS is basically $V_{out} - V_{DD}$. So, if you want to plot it, what we can do? We can plot this figure 1 and figure 2 together.

So, if you see the picture here so, figure one and figure two are plotted together. And the intersection of these two will tell you the operating point. So, when input is 5 volt, I d through the pMOS will be 0. So, this curve the bottom curve but I dn will be 5. So, you see here I dn is 5 this curve here and this is 0, so, this curve here. So, they both the curve actually intersect here. So, the when V in is 5 volt V out is 0 because pMOS is open nMOS is short.

If you increase, let us say both are 2.5, 2.5 then this is somewhere here and this is somewhere here they will intersect and this is the maximum. So, this is 2.5 and V_{out} is also 2.5 and this is the region where maximum current will flow because both are conducting because the situation is that both have the same current. And then of course, we go further you decrease the input voltage.

So, less input is 1 volt, so, this is the 1 curve and then for now this is for one for pMOS and one for nMOS they will intersect here. So, this is what actually you will get. So, as you increase the input or you change the input from 5 volt, 4 volt, 3 volt, 2 volt, 1 volt, 0 volt. The output is changing from 0 to 5 volt. So, thus you can write something like this. So, V_{in} is changing from 0 to 5. Your output is changing from 5 to 0.

So, this region up to this point it is on, up to this point it is off in this region this is a transition region. And this is also used for the gain purpose because slope is quite high here. So, if you have a small change in the input voltage that will give a large change in the output voltage. So, this region is also used for the gain purpose. So, it is used as digital device and it can also be used as an analog device.

So, for analog, application will be to get the gain here and digital application will be basically to store the states either 0 or 1.

(Refer Slide Time: 14:31)

BJT VS MOSFET

- RTL logic vs CMOS logic
- DC Input impedance of MOSFET (at gate end) is infinite
Thus, current output can drive many inputs → FANOUT
- CMOS static dissipation is low $\sim I_{OFF} V_{DD}$
- Normally BJTs have higher transconductance/current (faster!)

$I_C = (q n_i^2 D_n / W_B N_D) \exp(q V_{BE} / kT)$ $g_m = \partial I_C / \partial V_{BE} = I_C / (kT/q) \rightarrow \text{higher}$	$I_D = \mu C_{ox} W (V_G - V_T)^2 / L$ $g_m = \partial I_D / \partial V_G = I_D / [(V_G - V_T)/2]$
---	--

- Today's MOSFET $I_D \gg I_C$ due to near ballistic operation

SEMICONDUCTOR DEVICE MODELING AND SIMULATION

We can also compare the MOSFET with the bipolar junction transistor. The logic that is used with BJT we call it RTL or TTL resistor transistor logic or transistor, transistor logic. In

CMOS we call it CMOS logic. So, these things you will learn in more detail in case in your digital electronics. So now, in case of MOSFET, it consumes 0 current because input is connected to the gate. Gate is drawing no current.

So, you can connect MOSFET output to any number of MOSFET because they are not drawing any current. So, number of inputs that output can drive is called FANOUT. So, they will have high FANOUT. And of course, the static power distribution is also 0 almost 0 because in steady state, state 0 or state 1 there is no path between the BCC or supply and the grounds current is 0 or almost 0 there, except some leakage current.

So, in case of BJT, if you recall the expression for the collector current, it is some coefficient times exponential qV_{BE} by kT and if you differentiate with respect to V_{BE} , you will get I_C divided by kT by q which is a thermal voltage $V_{thermal}$, so, these are trans conductance. In case of MOSFET the trans conductance is I_D divided by $V_G - V_T$ by 2. So, this is usually small because this V_T is actually thermal is quite small, so, this again will be quite large.

But if you want to increase the conductance of MOSFET, this I_D has to be increased and I_D actually can be increased. If you make the device very small because in that small device, the conduction mechanism change instead of diffusive transport, it becomes ballistic transport. So and that is the ballistic operation you can have high I_D and this new generation MOSFET they are actually quite fast they are comparable to BJT.

(Refer Slide Time: 16:44)

NON IDEAL MOSFET

- If work function differences and oxide charges are present, threshold voltage is shifted just like for MOS capacitor:

$$V_T = V_{FB} + 2\psi_B + \frac{\sqrt{2\epsilon_s q N_A (2\psi_B)}}{C_i}$$

$$= \left(\phi_{ms} - \frac{Q_f}{C_i} \right) + 2\psi_B + \frac{\sqrt{2\epsilon_s q N_A (2\psi_B)}}{C_i}$$
- If the substrate is biased wrt the Source (V_{BS}) the threshold voltage is also shifted

$$V_T = V_{FB} + 2\psi_B + \frac{\sqrt{2\epsilon_s q N_A (2\psi_B - V_{BS})}}{C_i}$$

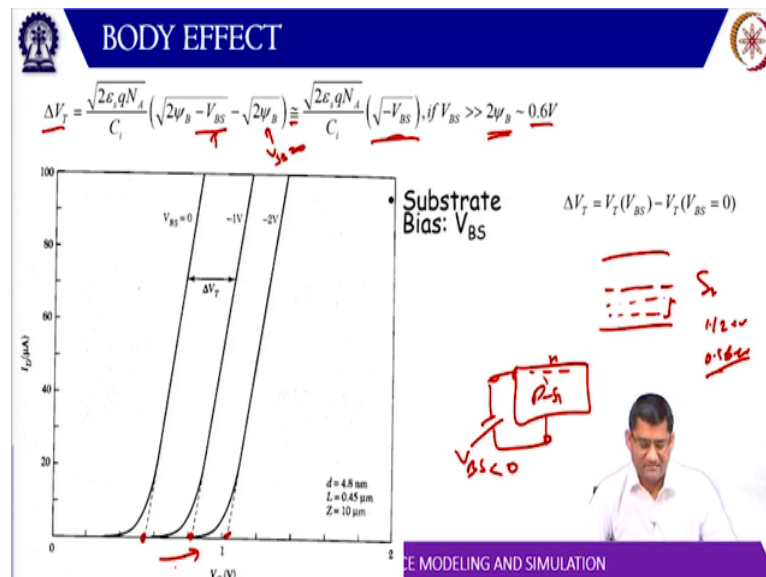
SEMICONDUCTOR DEVICE MODELING AND SIMULATION

Now, we will discuss now the non-idealities in the MOSFET. We have discussed the threshold voltage so, threshold voltage of course, the flat band voltage plus 2 psi B. That is the band bending inside of semiconductor and psi V is the potential difference between the or the energy difference divided by Q between the Fermi level and the intrinsic level. And this is the term due to the charge inside the depletion region so, this is $-Q_D / C_{oxide}$.

And of course, flat band voltage is $\phi_{ms} - Q_{effective} / C_{oxide}$ inside the oxide divided by C oxide. Now, this is 2 psi B if the body and the source they are connected together, if the body and source have different voltage then this will modify how it will modify? This is the band bending here and this is your Fermi level here. If here they are not connected together means the Fermi level is not same there is some voltage here.

Let us say this has some voltage V_r here at the source, so, the Fermi level will go down for positive voltage. So now, we have this 2 psi B difference now it requires this also psi B. So, the difference will be 2 psi B plus this voltage here and that is voltage is voltage of the source with respect to the body or you can write V_{SB} or if you write in terms of V_B with respect to S then it will be $2\psi_B - V_{SB}$. So, this expression actually get modified to this form. So, this is the exact expression.

(Refer Slide Time: 18:48)



Now, we can use this expression to understand the body effect and further calculate the or use in the gradual channel expression to get the modified expression for the drain current. So, when we apply this body bias, the change in the threshold voltage will be let us say this is V

SB and this is for $V_{SB} = 0$. So, the difference of 2 multiplied by $2q \epsilon N_A$ by C_i that is the difference in the threshold voltage due to the body bias.


So, this is called body bias or body effect. So, if you apply a higher body bias voltage then typically, this V_{SB} is on or off 0.6 volt. Because if you see here this is the silicon, the band gap is around 1.12 electron volt, so, half is around 0.56 electron, volt and then Fermi level somewhere here. So, it changes up to here. So, this range is typically 0.3 volt, so, the 2 side is 0.6 volt. Now, your body bias can be much larger than this thing.

Then this can be approximated as a square root of $-V_{BS}$ now, V_{BS} is usually negative because if you see here in case of nMOS that is p-type silicon. This is your body here and there is a source of channel here. Then V_{BS} is basically like this B is positive. So, this value has to be negative because you have a n channel here. So, if you apply a positive voltage here, this will become forward bias. We do not want it to be forward bias.


So, this V_{BS} is actually negative, so, this p-type silicon is this P n junction is reverse biased, so, V_{BS} is always negative for n channel device and as you increase the V_{BS} magnitude, this ΔV_T will actually increase. So, if you see here for $V_{BS} = 0$ that is this is a threshold voltage, then if you apply a 1 volt, this is the threshold voltage if you only 2 volt this is a threshold voltage. So, this threshold voltage actually increases with the body bias.

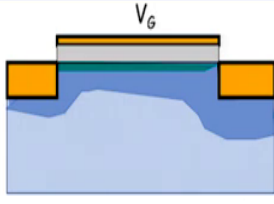
Now, this technique can be used, in situations where you do not want some device to accidentally turn on, while operating with other MOSFET devices. Because in case of integrated circuit, all these devices are sitting on the same substrate and the situation is pretty complex. So, this is one technique what device fact is used to control the unwanted turning on off a device?

(Refer Slide Time: 21:51)



BODY EFFECT





The threshold voltage is increased due to the depletion region that grows at the drain end because the inversion layer shrinks there and can't screen it any more. ($W_d > W_{dn}$)

$$Q_{inv} = -C_{ox}[V_G - V_T(y)], \quad I = -\mu_{eff} W Q_{inv} dV(y)/dy$$

$$V_T(y) = V_{FB} + \psi + \sqrt{2\epsilon_s q N_A \psi / C_{ox}}$$

$$\psi = 2\psi_B + V(y)$$

$$I_L = \int \mu_{eff} W C_{ox} [V_G - V_{FB} - (2\psi_B + V) - \sqrt{2\epsilon_s q N_A (2\psi_B + V) / C_{ox}}] dV$$

$$I = (W \mu_{eff} C_{ox} / L) [(V_G - V_{FB} - 2\psi_B) V_D - V_D^2 / 2 - 2\sqrt{2\epsilon_s q N_A} ((2\psi_B + V_D)^{3/2} - (2\psi_B)^{3/2}) / 3 C_{ox}]$$

$\frac{\partial \psi}{\partial V}$
 similar to previous
 body effect

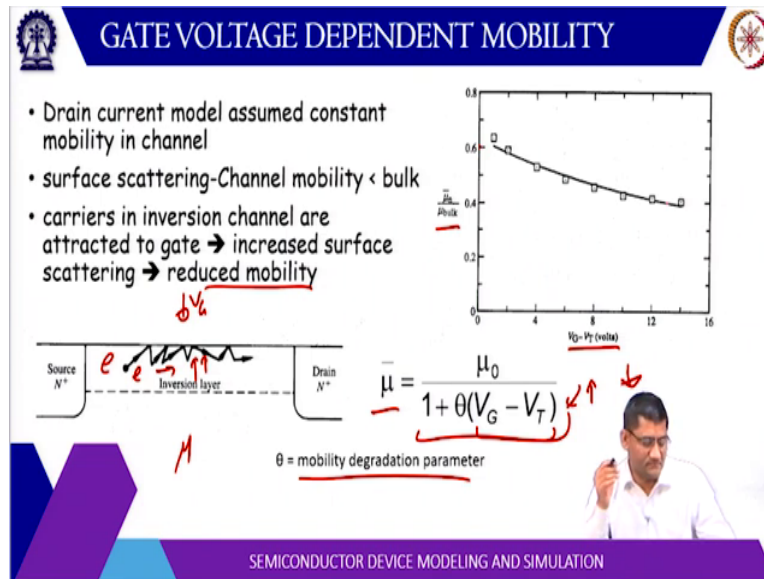
SEMICONDUCTOR DEVICE MODELING AND SIMULATION

Then if you recall that equation for the gradual channel approximation, so, if we include this term due to the body bias. So, this is your the depletion charge Q_D so, $-Q_D$ by C_{ox} . Now, it has a voltage here this is the channel voltage. So, this is a voltage, the difference between the channel voltage and the body. So, let us say body bias is, certain voltage and then let us say with respect to the source and the channel voltage V .

And V is changing from V_{source} to V_{drain} , so, if you integrate this part also then the rest of the term is same as we obtained, so, this is similar to previous lecture. The extra term is basically integrating this square root term. Then you get this power to the power 3 by 2 and this is the effect of body bias, so, it actually controls the current. So, there is a some negative coefficient is here due to this body bias.

Because body bias actually increases the threshold voltage and therefore, it will reduce the current.

(Refer Slide Time: 23:17)

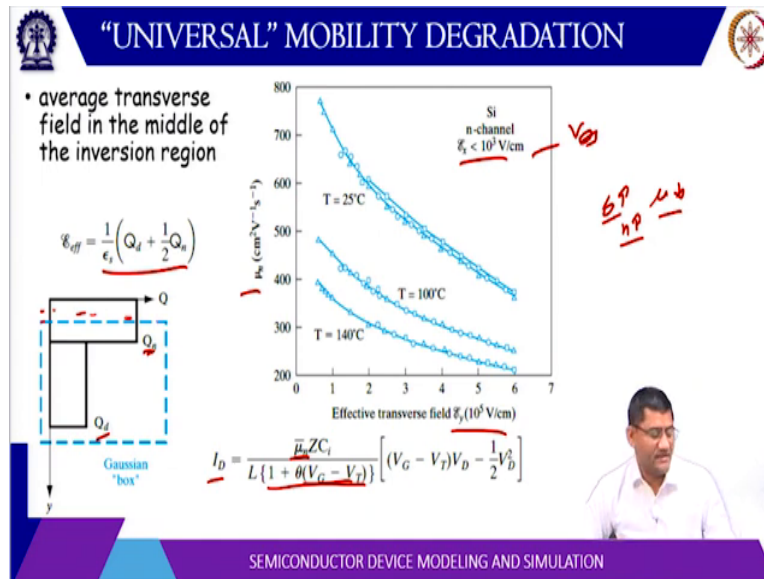


Now, another mechanism in case of MOSFET device. What happens? The mobility of this channel which we assume to be μ is not constant. It depends on the gate voltage. Now, how does it depend on the gate voltage? As if there are electrons here carriers they are moving through this inversion layer from source to drain. Now, from the surface, there is a scattering here. When you increase the gate voltage then the field vertical field will be more.

And when the vertical field is more then this electron will be pushed more towards the interface. And when they are near to the interface, there will be greater scattering due to the surface effect. So, this increased scattering due to the surface will reduce the mobility and that effective mobility is given by a relation μ_0 divided by $1 + \theta(V_G - V_T)$. So, as $V_G - V_T$ increases.

This is the denominator term it is increasing, so, your μ is actually decreasing. Then this θ is called mobility degradation parameter. So, this is the μ by μ_{bulk} for different gate voltages you see it is 0.6 to going to 0.4. So, there is a significant change in this effective mobility and that is due to the surface scattering arising from the vertical electric field due to the gate. So, they are more close to the surface and there is a more scattering.

(Refer Slide Time: 24:55)



So, based on this there is a universal mobility degradation curve, so, you can use this μ_n by $1 + \theta(V_G - V_T)$ instead of μ and rest of the terms remain same. So that is the expression for the current, including the mobility degradation. Now, how this is calculated? So, this $V_G - V_T$ what we do? We consider the middle of this channel region so, middle of this inversion Q and inversion charge region.

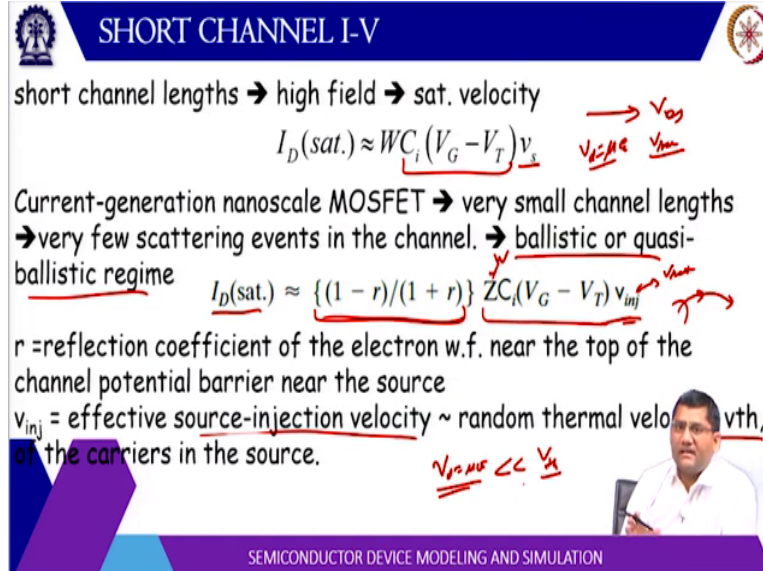
And there we calculate the field so that field will be the charge enclosed divided by epsilon. So, this is a depletion charge plus half of the inversion charge. So that divided by epsilon will be the effective electric field. And we plot this mobility versus effective electric field and that comes out to be a constant basically at a given temperature. So, as long as this, the field in the x direction that is due to V_{DS} is limited, is small.

Then we have uniform universal mobility degradation curve. So, regardless of the structure or material means n-type, p-type or the doping level we get same curves **(()) (26:25)** so, the doping level we get the same curve. Then of course it changes with the temperature that is because of the mobility is a function of temperature and as the temperature increases, a mobility actually decreases due to enhanced scattering.

So, although the conductivity increases but the mobility decreases because conductivity depends on the N because nI increases when you increase the temperature. But mobility always decrease with the temperature because at high temperature there is a higher scattering. Now, all it difference is that for electron we take easily in the middle and for whole we take at around one third of this region.

So, the reason is not done but that follows this universal mobility, regression curve or both electron and all both follow this curve.

(Refer Slide Time: 27:14)



SHORT CHANNEL I-V

short channel lengths \rightarrow high field \rightarrow sat. velocity $\rightarrow V_{\text{sat}}$

$$I_D(\text{sat.}) \approx WC_i(V_G - V_T)v_s$$

Current-generation nanoscale MOSFET \rightarrow very small channel lengths \rightarrow very few scattering events in the channel. \rightarrow ballistic or quasi-ballistic regime

$$I_D(\text{sat.}) \approx \left\{ \frac{(1-r)}{(1+r)} \right\} ZC_i(V_G - V_T)v_{\text{inj}}$$

r = reflection coefficient of the electron w.f. near the top of the channel potential barrier near the source

v_{inj} = effective source-injection velocity \sim random thermal velocity of the carriers in the source. $v_{\text{inj}} \ll v_s$

SEMICONDUCTOR DEVICE MODELING AND SIMULATION

Now, there is another effect called short channel effect, so, short channel effect basically, when there is a high field in the x direction or along V_{DS} . The mobility they cannot increase the drift velocity beyond saturation velocity. So, instead of μE we can write V_{sat} because the field is so, high that it is saturation velocity. You cannot express use expression μE because that will give you that will overestimate the velocity.

So, we replace this drift velocity by the saturation velocity. So, your current is basically $C v$ that is a q times W Times v_s . So, this occurs at very small channel lengths and in this for a small channel length, there are few scattering event and then we call it ballistic or quasi-ballistic transport region. And here the wave function of the electrons they play a role. So, if you have studied some quantum mechanics, you can write this transmission and reflection functions.

So, $1 - r$ by $1 + r$ so that is coming from the wave nature and then of course rest is W times C_i this say W is the same W times C_i $C_{ox} V_G - V_T$ times injection velocity. So, this with injection velocity which is used instead of saturation velocity, so, with injection velocity, they actually move. So, it is like a projectile so, from the source these are injected and they directly land here at the time because there is insignificant amount of scattering.

So, they just go like a projectile and land in the drain region. So, this kind of expression is used for ballistic region. Here r is a reflection coefficient of electron wave function near the top of the channel potential barrier at the source. And this is the source injection velocity and which is similar to the thermal velocity at the temperature of operation. Because in general, if you see that when we calculate $V_d = \mu \times E$ this is pretty small compared to the thermal velocity.

So, there are some velocity is quite large but when the length is small, they are randomly moving with the thermal velocity, so, in source should **(0)** **(30:03)** also they are moving with same thermal velocity. But here length is so, small there is no scattering, so, they directly land at the drain and contribute to the current flow. So that is what happens at the short channel MOSFET and then we get the soft channel IV characteristic.

(Refer Slide Time: 30:20)

SUB-THRESHOLD CHARACTERISTICS

- For gate voltage less than the threshold - weak inversion
- Diffusion is dominant current mechanism (not drift)

$$n(0) = n_i e^{\frac{q(\psi_s - \psi_B)/kT}{}} \quad \text{and} \quad n(L) = n_i e^{\frac{q(\psi_s - \psi_B - V_D)/kT}{}}$$

Source

$$I_D = J_D A = -qAD \frac{\partial n}{\partial x} \approx -qAD \frac{n(0) - n(L)}{L}$$

We can approximate ψ_s with $V_g - V_T$ below threshold since all voltage drops across depletion region

$$I_D = \frac{qAD n_i e^{-q\psi_B/kT}}{L} (1 - e^{-qV_D/kT}) e^{q(V_g - V_T)/kT} \approx \frac{qAD n_i e^{-q\psi_B/kT}}{L} (1 - e^{-qV_D/kT}) e^{q(V_g - V_T)/kT}$$

Sub-threshold current is exponential function of applied gate voltage
Sub-threshold current gets larger for smaller gates (L)

SEMICONDUCTOR DEVICE MODELING AND SIMULATION

Now, so far, we have assumed that below the threshold voltage there is no current. But if you see here this is the band bending here. So, when $\psi_s = \psi_B$ it is depletion but beyond ψ_B and $2\psi_B$ there is some inversion, not a strong inversion some small concentration is there. And that concentration of carrier will actually diffuse. So, if you look at the source and drain here, this is a source, this is a drain here.

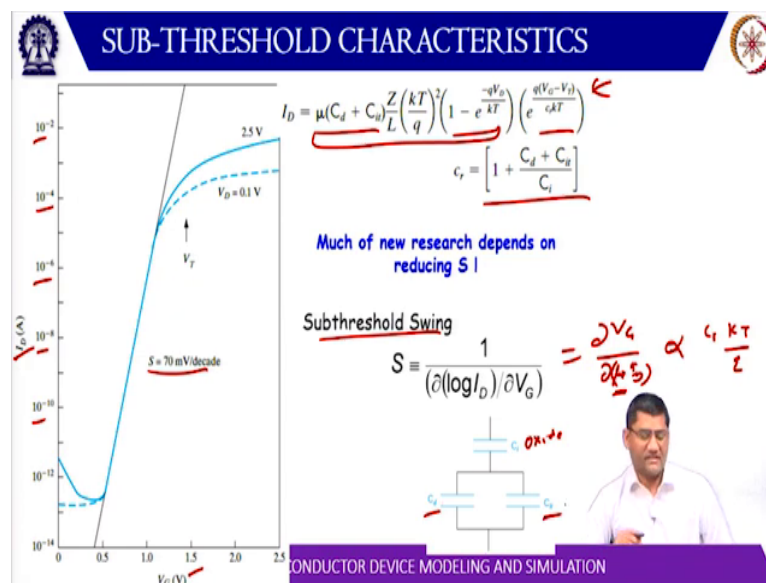
So, at source let us say it may be somewhere here, at the drain let us say it may be somewhere here, so, this is for the source this for the drain. So, there is a potential gradient. So, there is a concentration gradient. And this concentration gradient will call flow of these carriers due to

the diffusion. So, you can write at the source this is our source $n = n_i \exp(q\psi_s - \psi_B)$ by kT ψ_s is a surface potential, ψ_B is $e\phi$ and $e\phi$ difference divided by q .

And at the drain side so, length this is 0 to L so, $\psi_s - \psi_B = V_D$ by kT and because this concentration is small, we can assume a linear variation. And we can write the expression for the diffusion current qAD times d by derivative of the carrier concentration that can be n at 0 – n at L divided by L . Then if you further simplify, you will get some expression like this. So, here, $V_G - V_T$ is outside and there is a term here $1 - \exp(-qAD)$ by kT .

So, this is basically dependent on the V_{DS} the drain source voltage and that controls. So, this value is actually quite a small because V_D is small, so, this is close, so, the close to one but it is non-zero. This is non-zero. So, there is some amount of current flow for the some threshold region also. And when we design the MOSFET, we try to keep it as small as possible.

(Refer Slide Time: 32:32)



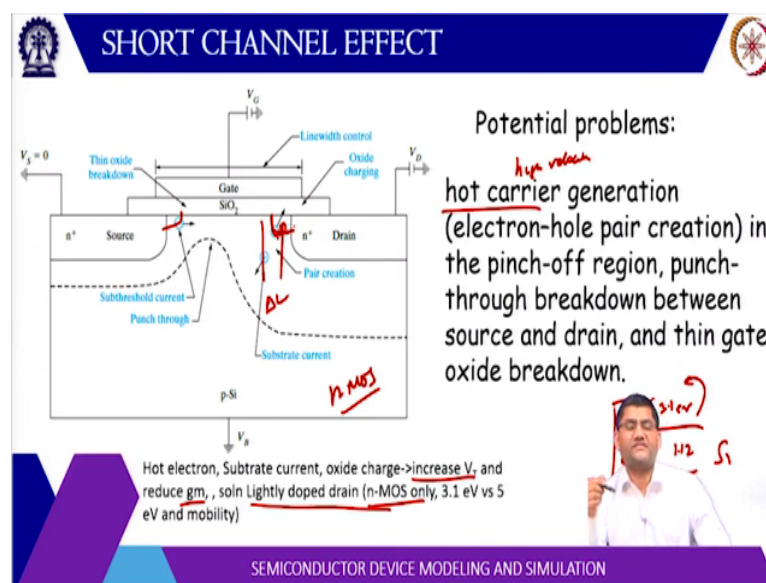
So, this is the I_D versus V_G for some threshold region. So, when V_G is less than V_T and this is logarithmic scale basically. So, you can write that $I_D = \mu \times C_d + C_i$. C_i is basically this capacitance due to the these interface traps then $1 - \exp(-qV_D)$ by V_T then $-qV_G - V_T$ by some coefficient times kT . And that coefficient c_r is basically $1 + C_d + C_i$ by C_i . Now, we can calculate this sub threshold showing here.

So, this is basically $\frac{dV_G}{dI_D}$ on log scale, so that can be calculated as if you see here with respect to V_G these terms are not depend on V_G only this term is dependent on V_G . So, this will be \log of I_D will be qV_G by $C_r kT$. So, this will be $C_r kT$ by q and C_r is

this term kT by q . And this log is basically natural log so, if you take log 10 then they will factor of 2.3 will come here.

So, this is proportional to C_r times kT by q . So, this is a subthreshold swing and there is a limit to this one around 70 millivolt per decade but the current is actually pretty small here and below the threshold. So, at threshold it basically, suits up. And how do we just get this expression? Because you can see this is the oxide capacitance here that is in series with this copper capacitance existing at the silicon dioxide and silicon interface. So, one is due to the depletion region other due to the interface traps, so, they are in parallel.

(Refer Slide Time: 35:04)



Now, short channel effect if you look further now, the potential problem here is basically the hot carrier a generation. Now, hot carrier are basically those carriers with high velocity. So, this velocity is much larger comparable to the thermal velocity and they actually acquire this velocity for high field. So, especially in the pinch-off region in this region, ΔL where the field is quite high, this electron occur high velocity.

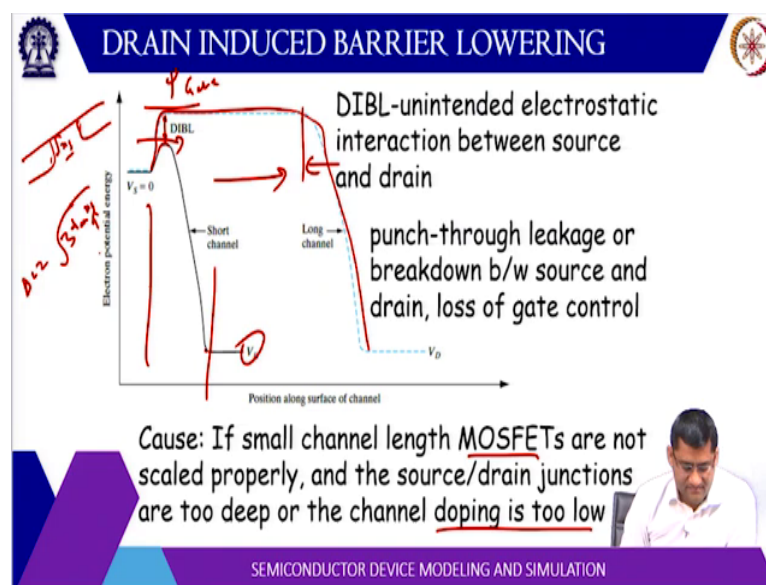
And then they basically punch through breakdown, may take place between the source and drain. And they may hit this silicon dioxide also so, what they do basically? These hot electrons, they induce some charges, they increase the threshold voltage and if the threshold voltage will reduce, if threshold voltage will increase then the trans conductance will decrease. So, what is done here? The solution is having lightly doped drain.

So, this is the drain here then we have little n-type region which is less doped. Now, this problem is more severe in case of nMOS. Now, why nMOS? If you look at the band diagram, this is silicon, this is silicon dioxide. So, the silicon band gap is around 1.12 and silicon dioxide band gap is around 9 electron volt. And the barrier between the conduction band and this and barrier and conduction balance band.

And this is around this barrier is 3.1 electron volt and this barrier is around 5 electron volt. So, this is larger actually virtually it is not solved. So, this is 5 electrons. So, this is 1 so, $3 + 5 + 8 + 1$ so, 9 electron volt. So, the barrier for electron is small, so, where these electrons are energetic, they can overcome this barrier. So because this region gets thin here because this oxide is thin here but in case of holes, this barrier is 5 electron volt.

So, even if they have high electric field, the barrier is pretty high so, easily it is not a problem. So, generally in case of nMOS we have this lightly doped region here near the drain because it is the region here where there is a maximum electric field. So, in around the trend we have a lightly doped n-type region called LDD to overcome this hot electron effect or short channel effect.

(Refer Slide Time: 37:50)

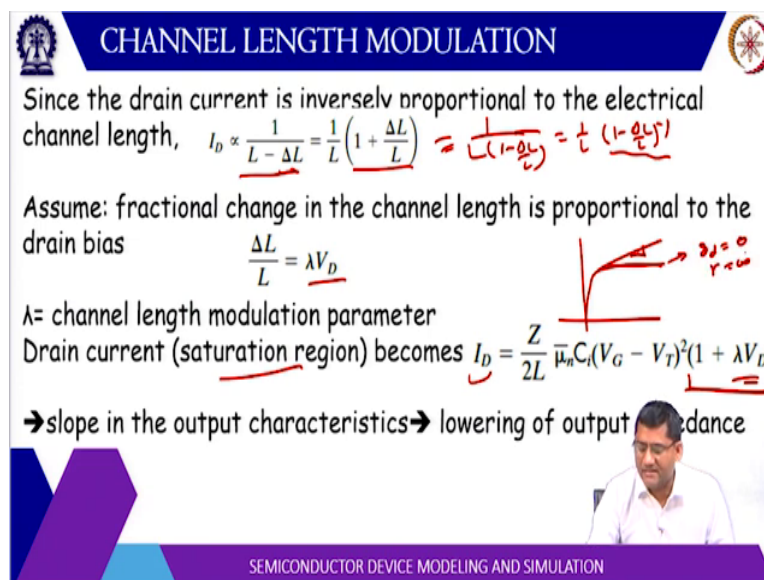


Then there is another phenomena called drain induced barrier lowering, so that is seen in case of short channel devices. What happens? The distance between source and drain is small and when you apply a drain voltage, especially at high drain voltage, this barrier is pulled down. Now, for long channel if you see here, this barrier is intact. So, even if there is some change here, it is away from the source but in case of a small length, this drain is near to the source.

So, when it pulls down the barrier, it can actually pull down the barrier on the source side also. So, when you pull down the barrier on the source side, the control of the gate becomes less significant. So, there may flow very high current there and that we call punch through leakage of breakdown source and drain, loss of gate control. So, these are the effect of drain induced barrier lowering.

So, if these MOSFETs are not scaled properly then or if the source drain junctions are too deep. Then or the channel dropping is too low then this kind of situation may arise. So, to counter this usually channel doping is tailored or we design the source and drain regions with small junction depth. So, you recall the delta s expression, delta L was 3 times $t_{ox} X_j$ so, this X_j is capital X.

(Refer Slide Time: 39:38)



CHANNEL LENGTH MODULATION

Since the drain current is inversely proportional to the electrical channel length, $I_D \propto \frac{1}{L - \Delta L} = \frac{1}{L} \left(1 + \frac{\Delta L}{L} \right) \approx \frac{1}{L} \left(1 + \lambda V_D \right)$

Assume: fractional change in the channel length is proportional to the drain bias

$$\frac{\Delta L}{L} = \lambda V_D$$

λ = channel length modulation parameter

Drain current (saturation region) becomes $I_D = \frac{Z}{2L} \mu_n C_i (V_G - V_T)^2 (1 + \lambda V_D)$

→ slope in the output characteristics → lowering of output resistance

SEMICONDUCTOR DEVICE MODELING AND SIMULATION

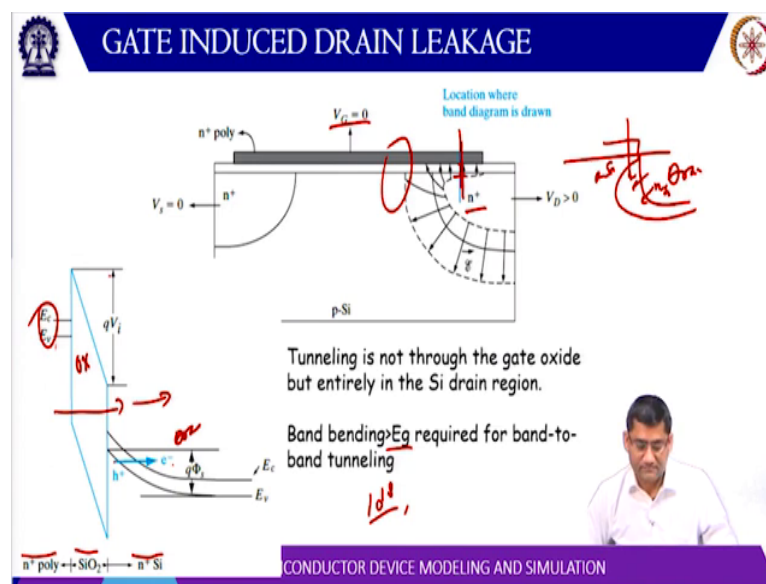
Then there is a phenomenal channel length modulation, so, here effective length of the channel actually reduces. So, your I_D scales as 1 over $L - \Delta L$ so because ΔL is small. So, we can write it as 1 over L times $1 - \Delta L$ by L . So, this is basically 1 over times $1 - \Delta L$ by L power -1 . So, this can be approximated as $1 + \Delta L$ by L . And then this is proportional to the drain voltage, so, we call it with some parameter lambda this lambda V_D .

So, this your drain current in case of saturation region get modified with one extra term, $1 + \lambda V_D$. So, this is very similar to the case we discussed in BJT, where we at early voltage. So, the I_C was some expression for $I_C + 1 + V$ by V_a or so, same phenomena is

there in case of MOSFET but the name is different here it is called channel length, modulation in BJT at base width modulation.

So, this give arise to the slope in the saturation region. So, instead of having this, it becomes sloppy and that is slope is related to the lambda here. So, this basically so, when it is flat, the G out was 0 that means r was infinite. Now, this will give r although large but relatively finite. So, it affects the output impedance it lowers down the output impedance.

(Refer Slide Time: 41:26)



Then another phenomena is gate induced drain leakage. So, at the gate tunnel there is some overlap with respect to the drain. So, this region oxide which is quite thin and if you look along this line so, this is along this line. So, this is your oxide and this is your drain region. So, from oxide and the drain region, these electrons here in the metal so, this is the voltage dope across oxide.

And this metal is beta poly silicon, silicon oxide, n-type silicon. So, this will be here, so, this electron can make a transition here through the turn length because this oxide actually quite thin it is shown bigger but it is quite thin actually. So, if you notice here, this is silicon region here this is draining. So, when drain is doped here you can thin like this. This is the drain. Drain is doped here.

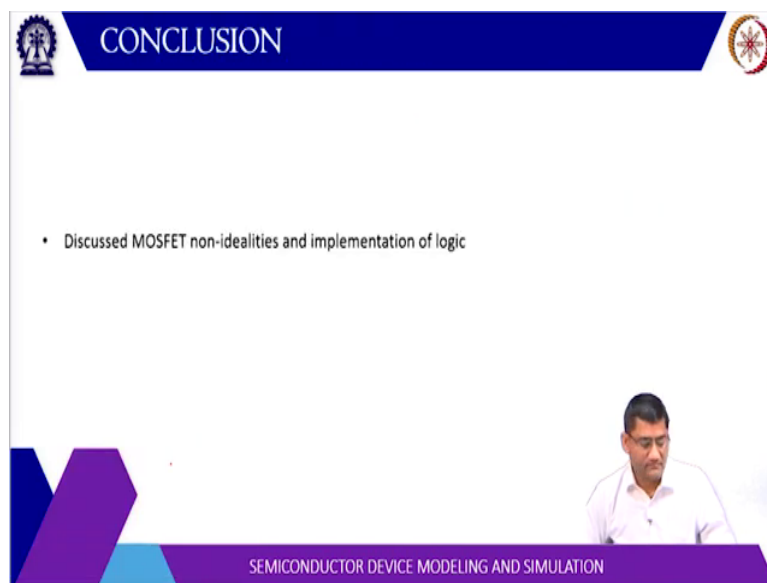
Now, these diffuse basically, so, they diffuse like this. So, this is the gate here. So, what you have here? You have silicon p-type silicon and this is n-type silicon. So, here, this tunnelling may take place into the oxide, so, the band bending that is required. It should be greater than

the band gap region. So, this is p-type poly silicon. So, you have E_c and E_v here and then this is filled basically.

So, this is not through the gate oxide so, this is your gate oxide. It is entirely into the drain region. And the band gap of course the band bending should be more than the band gap because then this balance band is overlapping here and it can easily make a transition to the region where there are state for these electrons. Now, this typically occurs at moderate doping of 10^{10} is to power 18 because when there is low doping the band bending is small.

When there is high doping then of course the band bending may be too much and it may concern some higher energy level. So, where this may not be feasible for this electron to tunnel through so, at moderate doping around 10^{10} is to power 18 this gate induced drain leakage is usually seen.

(Refer Slide Time: 44:54)



So, in this lecture we have discussed the non-idealities and the some logic based on complementary MOS structure. Thank you very much.