

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

Lecture – 29
MOS-CV

(Refer Slide Time: 00:27)

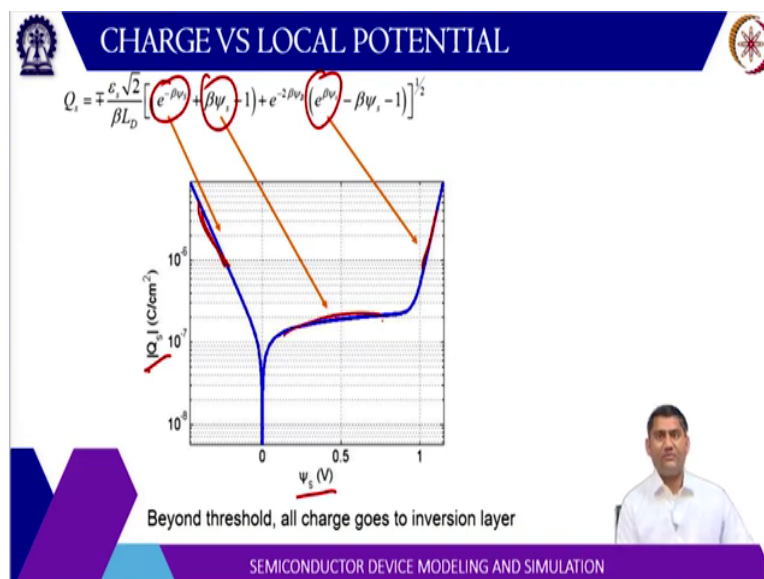
L29 MOS-CV

- THRESHOLD VOLTAGE
- INTERFACE CHARGES

SEMICONDUCTOR DEVICE MODELING AND SIMULATION

Hello, let us continue our discussion on MOS capacitor. So, again we are dealing with the ideal MOS so far and we have solved the Poisson equation.

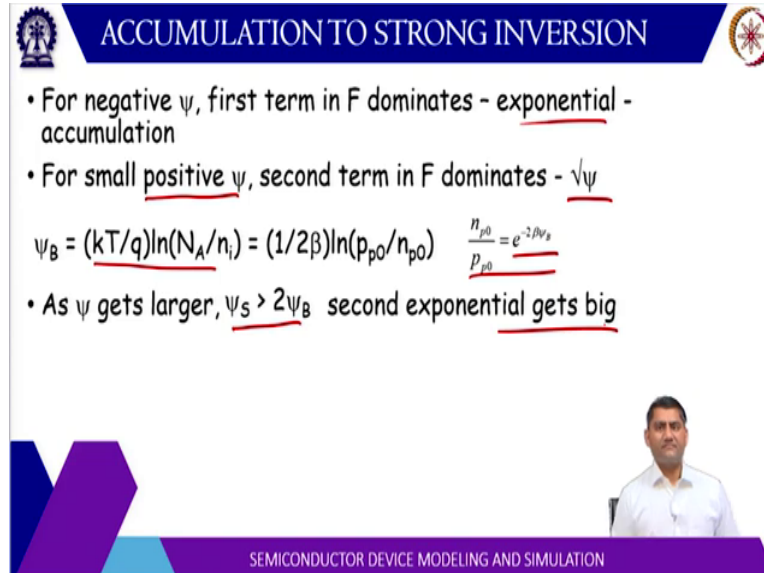
(Refer Slide Time: 00:38)



So now, let us define the other parameters that are related to MOSFET, so, this is a typical Q versus surface potential curve for a MOS structure and where we have identified the three

regions. This is accumulation, this is depletion region and this is your inversion region. And in inversion region and accumulation region this is exponential term in depletion region. It is inverse proportional to the square root of the voltage.

(Refer Slide Time: 01:06)



ACCUMULATION TO STRONG INVERSION

- For negative ψ , first term in F dominates - exponential - accumulation
- For small positive ψ , second term in F dominates - $\sqrt{\psi}$

$$\psi_B = \frac{(kT/q)\ln(N_A/n_i)}{(1/2\beta)\ln(p_{p0}/n_{p0})} \quad \frac{n_{p0}}{p_{p0}} = e^{-2\beta\psi_B}$$

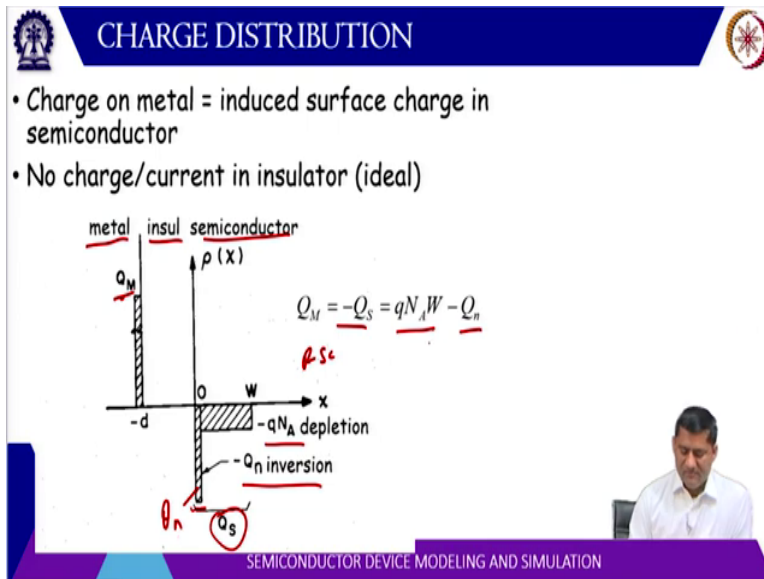
- As ψ gets larger, $\psi_s > 2\psi_B$ second exponential gets big

SEMICONDUCTOR DEVICE MODELING AND SIMULATION

So, for a negative potential first term, in the F. So, this is the first term in the F that dominates and then this is the exponential function and this accumulation. For a small positive side, the second term which is proportional to the square root of psi. The second term beta psi s that actually dominates. And then psi B is the bulk potential that is kT by $q \ln N_A$ by n_i . And so, n_{p0} by p_{p0} is a exponential – $2\beta\psi_B$.

That we already discussed and as size most more than $2\psi_B$. The second exponential starts become active because now it has overcome this term here – $2\beta\psi_B$. So, this becomes big and there is a inversion region basically.

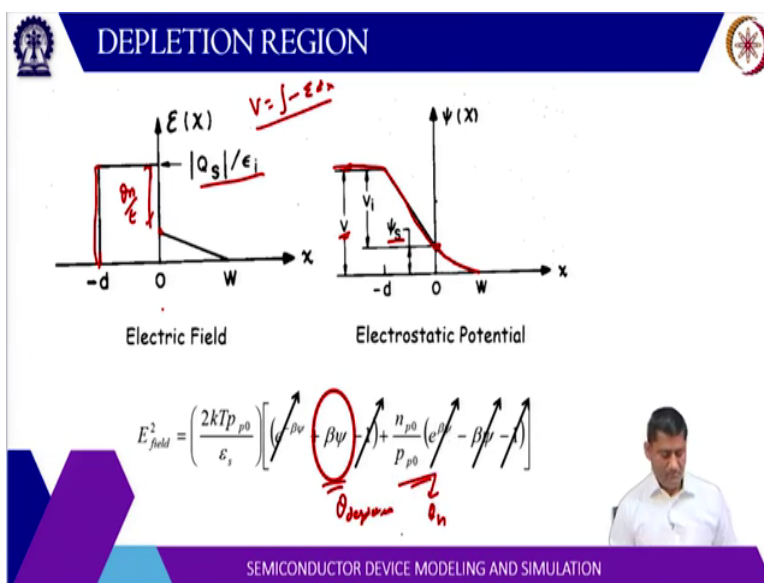
(Refer Slide Time: 02:10)



Now, if you look at the total picture, so, there is a metal here. There is an insulator oxide here and there is semiconductor here. So, in the semiconductor there is a depletion region and the charge density can be $-q$ times N_A assuming it is p-type semiconductor. Then, when there is a large positive bias here at the metal, it goes into the inversion and then this inversion charge is actually a very thin layer at the semiconductor and oxide interface.

And this charges, let us say, total charge is Q_S so, same charge has to be on the metal, so, $Q_M = Q_S$ what bit opposite polarity, so, Q_M is $-Q_S$ which is a depletion charge plus the inversion charge. So, depletion charge is $-qN_A$ and inverse charge we call it Q_n . So that is the expression here.

(Refer Slide Time: 03:13)



Now, from the charge profile we can also calculate the electric field. So, let us calculate the electric field. Here the electric field is 0 because total charge it is easier is $Q_M + Q_S$ that is 0. Now, you cross this boundary, the charge is Q_M . So, there will be field here that will be Q_M by ϵ that electric field will exist here. Now, here there is no charge so that this electric field will be constant here then there is a $-Q_N$ here.

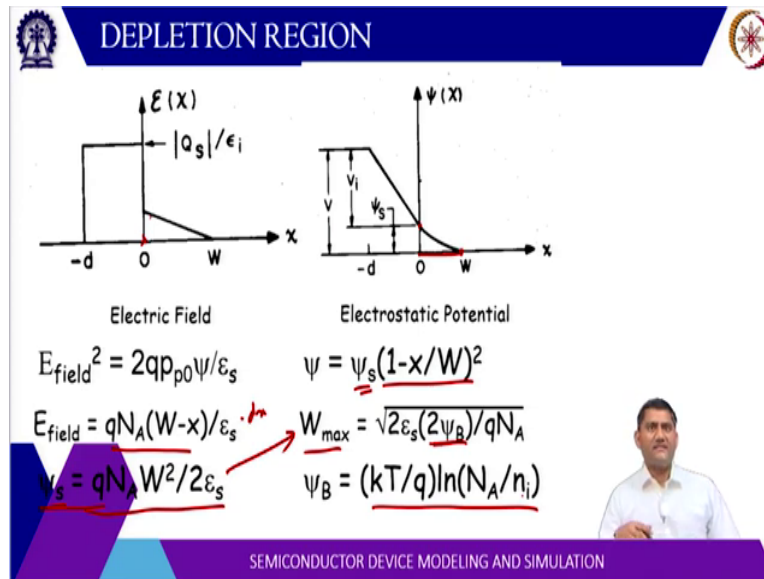
So, this will create $-Q_N$ by ϵ electric field, so, there will be sharp jump here. Whenever you encounter a sheet of charge there will be a sudden change in the electric field. So, this electric field will reduce now. Then you have this depletion charge which is uniformly distributed. So that means, the electric field will be if you recall $\nabla \cdot \mathbf{E} = \rho$ by ϵ . So because ρ is constant, so, E will be linear.

So, this electric field will linearly vary here and then beyond this point, this electric field will be 0 so that you can see here. So, this is a metal charge here. So, after this point, electric field is constant that is Q_S by ϵ or Q_M by ϵ . Then here is Q_N by ϵ so, this gap is Q_N by ϵ and then this is linear. So, this is electric field curve and then, of course, if you integrate electric field that $V = \int -E dx$.

You will get a potential here so, for linear it will be parabolic so, from W depletion, width to $x = 0$ this will be parabolic. Then there is a jump here in the electric field that will not change of potential because potential cannot change instantaneously. So, this potential will be same but now a field is more here. So, the potential will change linearly because field is constant, so, this will be linear and beyond.

This potential is constant because electric field is 0. So, here is the potential ψ_s and here is the ψ that is applied voltage V . Now, here is a depletion region. So, this term is highlighted that controls the depletion region here and this term controls the inversion charge, Q_N and that controls the Q depletion region.

(Refer Slide Time: 05:59)



Now we can write some expressions also here. So, $E^2 = q^2 N_A \psi / \epsilon_s$ and if you take the square root, if you write in terms of the potential width, so, E field can be written as $qN_A(W-x)/\epsilon_s$ as a function of position. At surface this x is 0 here. So, the electric field, with $qN_A W$ by ϵ_s and the potential will be area under the electric field. So that will be half height into width.

So, this is the electric field as a surface $qN_A W$ by ϵ_s times half W that will be $qN_A W^2$ by $2\epsilon_s$. That will be the potential at the semiconductor oxide surface and in the region of semiconductor. This is the field so, if you integrate dx , you will get the potential and then, if you separate out the ψ_s from here, you will get $1 - x/W$ whole square. That is how the potential is varying.

So, at $x = W$ this potential is 0 and at $x = 0$ this potential is ψ_s . And then, of course, you can find out the W_{max} . That is a depletion width, so that can be obtained from here that will be $2\epsilon_s \psi_B$ by qN_A . You can substitute $\psi_s = 2\psi_B$ so, ψ_s is equal to so, if you substitute here so, from this equation, you can get the W_{max} so, $2\epsilon_s \psi_s$ by qN_A and ψ_s is $2\psi_B$ here.

So, beyond $2\psi_B$ that is means stronger version depletion width does not increase further. So, all the charge that is coming after a strong inversion that is in the inversion layer only. So, we can assume that W_{max} will occur when the surface potential is $2\psi_B$. And ψ_B is of course, we know $kT/q \ln(N_A/n_i)$. So, this is a handy expression that you can easily

evaluate and use them to visualize the field and the potential profile inside MOS metal oxide semiconductor structure.

(Refer Slide Time: 08:38)

THRESHOLD VOLTAGE

- Total voltage across MOS structure = voltage across dielectric + ψ_s

$$V_T = V_i + \psi_s = -\frac{Q_s}{C_i} + 2\psi_B$$

$$|Q_s| = qN_A W_{\max} = qN_A \sqrt{\frac{2\epsilon_s \psi_s (inv)}{qN_A}} = \sqrt{2\epsilon_s q N_A (2\psi_B)}$$

$$\Rightarrow V_T = \frac{\sqrt{2\epsilon_s q N_A (2\psi_B)}}{C_i} + 2\psi_B = \frac{-Q_d}{C_i} + 2\psi_B$$

Handwritten notes in red ink:

- $Q = CV$
- $C_i: C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$
- $\psi_s \approx 2\psi_B$

SEMICONDUCTOR DEVICE MODELING AND SIMULATION

Now, the total applied voltage V_T drops across oxide, so, we can write V_T as voltage across insulator or you can write V_T as voltage across oxide + ψ_s is the potential drop in the semiconductor. Now, inside the oxide it basically acts like there is a charge here on the metal layer. The inversion charge here is a depletion charge here this is Q_n , this is $Q_{depletion}$. So, this oxide is basically like a capacitor here.

So, the potential can be written as $Q = CV$. So, V which is basically Q by C . So, Q on either side, let us say $-Q_s$ by C_i or where C_i is same as C_{ox} is in some text. It is C_{ox} it is oxide capacitance or insulated capacitance that is ϵ_{ox} by the width, so that is t_{ox} . So, t_{ox} is a thickness of this oxide region. So, $-Q_s$ by $C_i + 2\psi_B$. This is at the strong inversion region.

Now, actually if you go beyond strong inversion region, the change in potential will be very small because the corresponding change in the charge is exponential. If you look at this curve, Q versus ψ_s for accumulation, it increases exponentially, for depletion it stays for some time, for certain value then beyond 0.7 it again goes exponential. So, if you see here, the change in charge concentration grows exponentially.

So, for a small change in ψ_s there is a large change in the charge. So, your Q_s is $qN_A W_{\max}$. There is a maximum depletion width so that expression we gave already used. So, if

you multiply W_{max} by qN_A W_{max} we have obtained here, if you multiply this by qN_A then you will get $qN_A \psi_B$. And then of course, you can express this total voltage which is actually we call it threshold voltage.

So, this is drop across capacitors Q_s by C_{ox} so, you substitute Q_s here. So, this is a voltage across oxide and this is a voltage across semiconductor. So, this is just at the threshold, so, threshold is by definition the voltage at which $\psi_s = 2\psi_B$. So, this is a threshold voltage and this threshold voltage is for ideal MOS structure where we have assumed that ϕ_{ms} is 0 and we have not considered any charges inside the oxide.

(Refer Slide Time: 11:42)

BOUNDARY CONDITION

Recall, $V_T = V_{ox} + \psi_s = -\frac{Q_s}{C_{ox}} + 2\psi_B$

$\epsilon_{ox} V_{ox}/t_{ox} = \epsilon_s \psi_s/(W/2)$ Before Inversion

After inversion there is a discontinuity in D_{ox} due to surface Q_{inv}

$V_{ox}(\text{at threshold}) = \frac{\epsilon_s(2\psi_B)}{(W_{max}/2)C_i}$

$D = \epsilon E$
 $\epsilon_{ox} E_{ox} = \epsilon_s E_s$

SEMICONDUCTOR DEVICE MODELING AND SIMULATION

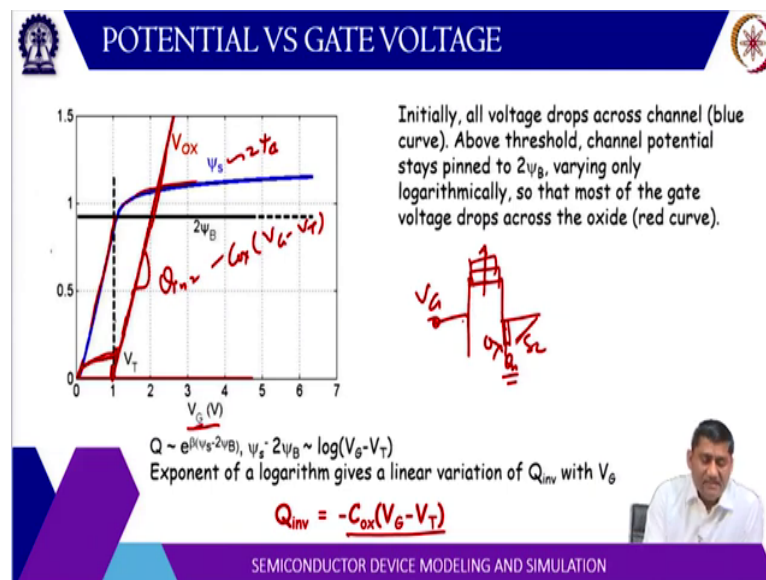
Now, let us recall that threshold voltage is Q_s by $C_{ox} + 2\psi_B$ and at the interface this is oxide here. So, there is a electric field here and then there is a jump in the electric field here and then this goes like this linear. So, since at the interface if you recall that there are two regions, one is oxide, one is semiconductor, the electric field is not continuous but the flux density D the displacement flux density is constant.

So, $D = \epsilon E$ so, ϵ on the oxide side and E on the oxide side is equal to ϵ on semiconductor side times E on electric field on semiconductor side. So, this is basically $\epsilon_{ox} V_{ox}/t_{ox} = \epsilon_s \psi_s/(W/2)$ that is ϵE on the oxide side on semiconductor side it is $\epsilon \psi_s/(W/2)$ that is before inversion. So, after inversion there will be discontinued in D also. That is because of this qN charge here.

So, whenever there is a surface charge density D will become discontinuous now. As long as there is no infinite sheet of charge or last sheet of charge D will be continuous, so, upload depletion region, this device continuous and then, we have to take this Q universe into account there. So, B oxide the voltage drop of across oxide is a threshold is you can say if you write here if you saw rearrange this equation.

So, V_{ox} can be calculated as ϵ_s by ϵ_{oxide} . So, this is ϵ_s times ψ_s divided by W by 2 divided by t_{ox} by ϵ_{oxide} that is B_{oxide} . So, ψ_s is actually $2\psi_B$ and W is W_{max} by 2 and this is C_{oxide} or C_i ϵ_s by T . So, this should be divided by C_i basically. So, this is 1 by C , this is 1 by C . So, the voltage across oxide can be estimated at the threshold using this expression.

(Refer Slide Time: 14:35)



Now, this is a plot again with respect to the gate voltage. So, gate is basically this metal, oxide and semiconductor so, this is the gate voltage. So, a threshold voltage so, this is the threshold voltage here. You can see up to the threshold voltage, the voltage across oxide is almost constant but the voltage across semiconductor actually increases. Because depletion which keep on increasing here.

So, this is a voltage across offside. This is the voltage across semiconductor. But beyond threshold what happens? Here Q_n comes into the picture, so, this field actually increases. So now, extra this charge actually appears here, so, this field actually keep on increasing in the oxide. So, beyond threshold, the voltage in the oxide region increases but the surface potential remains more or less constant to is equal to $2\psi_B$.

So that variation is fairly small. So, we can say that Q inversion is actually C_{ox} times $V_G - V_T$. So, this is V_T , here this is a linear line, so, this is basically you can write. This Q inversion is equal to $-C_{ox}$ times, $V_G - V_T$, this value is V_T , here and this value is V_G here. So, $V_G - V_T$ this region is linear.

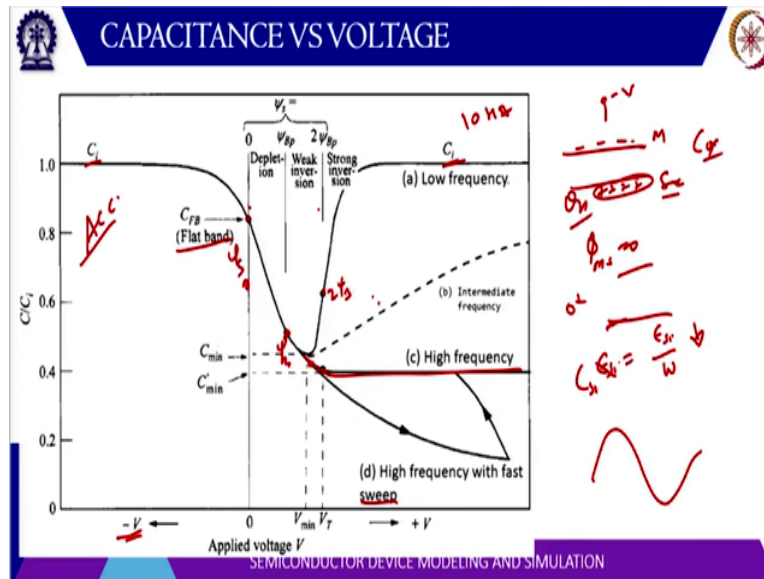
(Refer Slide Time: 16:02)

The slide is titled "CAPACITANCE" and features a diagram of a metal-semiconductor-oxide structure. The structure consists of a metal gate on top of an insulator layer, which is on top of a semiconductor layer. The semiconductor layer has an ohmic contact at the bottom. The diagram labels the metal, insulator, semiconductor, and ohmic contact. The thickness of the insulator is labeled d , and the thickness of the semiconductor depletion region is labeled x_d . To the left of the diagram, there is a circuit diagram showing two capacitors in series: C_i (insulator) and C_D (depletion). The equivalent circuit is shown as $\frac{1}{C} = \frac{1}{C_i} + \frac{1}{C_D}$ or $C = \frac{C_i C_D}{C_i + C_D}$. Below this, the insulator capacitance is given as $C_i = \frac{\epsilon_i}{d} = \frac{\epsilon_{ox}}{t_{ox}}$. The slide also includes a small video inset of a person in the bottom right corner and a logo in the top left corner.

Now, if you look at the metal semiconductor of oxide structure, there are two regions here. One is the oxide here insulator one is the semiconductor. So, inside the semiconductor there is a depletion region. So, there are two capacitors series. This is the insulator oxide capacitance C_{ox} this is the $C_{depletion}$. So, when the two capacitances are series, so $\frac{1}{C}$ is $\frac{1}{C_1} + \frac{1}{C_2}$. So, if C can return as $\frac{C_1 C_2}{C_1 + C_2}$.

And we know that oxide capacitors ϵ_{ox} by D or we write ϵ_{ox} by oxide thickness where D is the oxide thickness or t_{ox} .

(Refer Slide Time: 16:50)



Now let us, look at this picture first so, this is voltage is negative, so that means this is the accumulation region. So, in accumulation region this is metal, this is semiconductor. So, all the holes actually accumulate here and there is a negative charge at the metal surface. So, all the potential falls across the oxide, so, this capacitance is simply C_{oxide} or C_i insulator capacitance. Then, so, this is for negative voltage.

Now, let us come down to 0 volt, so, 0 volt there will be some potential across the metal, some potential across the semiconductor. Now, if you look at exactly 0 volt, we have assumed a flat band condition. We have assumed that ϕ_{ms} is 0. So that is, the bands will be flat. But if you go to $0 +$ then these bands will bend actually. So, when we say that capacitance at 0 voltage means plus minus variation at this point will give some variation in the depletion charge in the semiconductor.

So, there will be capacitance in the semiconductor that will come into the picture and we call this capacitance a flat band capacitance because your bands are flat and the capacitance is the combination of capacitance of the oxide in series with capacitance in the inside the semiconductor. As the depletion width increases the capacitance inside the silicon is $\epsilon_{silicon}$ by W .

So, W increases so, this capacitance of the silicon or the depletion region actually decreases. So, when two capacitances are in series, the overall capacitance will also decrease, so, it actually decreases. And it reaches maximum when it is in the W_{max} . So that is around 2 psi

B but at 2 psi B there is already Q n is there, the inversion layer is already there. So, slightly before that it will peak to the minimum.

And then of course again it will increase and it will be close to the oxide capacitance. Because now, all the charge is getting added here but in the form of electrons because now it is inverted. Now, this capacitance curve actually depends on the frequency because in case of MOS capacitor, this extra charge, when you apply a voltage, a large positive voltage, this electron have to come from somewhere and that takes time.

So, if the frequency is low, let that means around 10 hertz or something then these electron will have submission time to appear there and then at low frequency you will get this curve. But if you got high frequency then this charge carrier will not have time to appear. Because by that time they appear this is again gone to negative side like that. So, at high frequency this will be fixed. So, this will be $C_{ox} + C_D$ in series.

And then of course, if you do high frequency with fast sweep then you can get even lower capacitors. So, these are the MOS CV curve for different frequencies and different region. So, beyond flat band capacitance is the accumulation region. Then there is a depletion region. Then there is a weak inversion. So, psi s is between C_0 and 2 psi B. So, this is psi S = 0. Here this is psi B here and this is 2 psi B here and so on.

(Refer Slide Time: 20:51)

FLAT BAND CAPACITANCE

- Negative voltage = accumulation - $C \sim C_i$ OR C_{ox}
- Zero voltage - Flat Band

$V = 0 \Rightarrow \psi = 0 \Rightarrow C = C_{FB}$

$\psi_s \sim 1 + \psi$
 $\psi_s \sim 0$

$$\frac{1}{C_{FB}} = \frac{1}{C_i} + \frac{1}{C_D} = \frac{1}{\left(\frac{\epsilon_i}{d}\right)} + \frac{1}{\left(\frac{\epsilon_s}{L_D}\right)} \Rightarrow C_{FB} = \frac{\epsilon_i}{d + \frac{\epsilon_i}{\epsilon_s} L_D}$$

$\frac{d^2 \psi}{dx^2} = -\frac{q}{\epsilon_s} \left(\frac{N_A (e^{-\beta \psi} - 1)}{\beta} \right)$

Derivation:

$$\frac{d^2 \psi}{dx^2} = -\frac{q}{\epsilon_s} (N_A (e^{-\beta \psi} - 1)) \cong \frac{q N_A}{\epsilon_s} (\beta \psi) \rightarrow \psi = \psi_s e^{-x/L_D}, L_D = \sqrt{\frac{\epsilon_s}{q N_A \beta}}$$

$$C_{s,FB} = \frac{dQ_s}{d\psi_s} = \frac{d\left(\frac{\psi_s}{L_D}\right)}{d\psi_s} = \frac{\epsilon_s}{L_D}$$

$$\frac{d^2 \psi}{dx^2} = -\frac{q}{\epsilon_s} \left(\frac{N_A (e^{-\beta \psi} - 1)}{\beta} \right)$$

SEMICONDUCTOR DEVICE MODELING AND SIMULATION

Now, for negative voltage accumulation at flat band voltage, so that we can calculate by solving the Schrodinger equation. So, we know that at psi s = psi B. Sorry at psi s = 0. It is

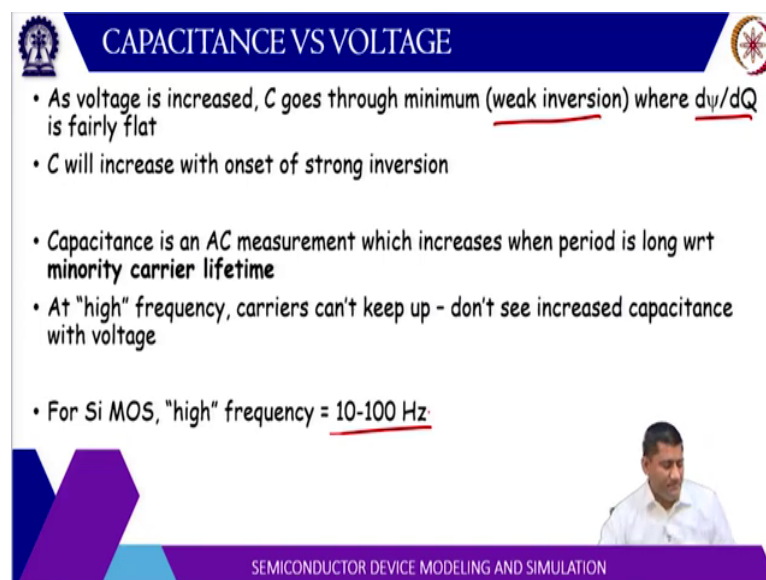
the basically accumulation term that is dominating so, other term we can ignore only we keep the holes from there. So, if you recall, that $D \frac{d^2 \psi}{dx^2} = -Q$ by epsilon silicon times p p naught e to the power $-\beta \psi - 1 - n$ p naught e to the power $\beta \psi - 1$.

So, this term will be insignificant, so only the first term will play all here, so that is, we have kept it here. So that is p p naught is $N_A \exp(-\beta \psi - 1)$ and then of course because size is close to 0. So, we can assume the size small so, e to the power ψ can we expect expressed as $1 + \psi$, so, this will be e to the power $\beta \psi$ can be $1 + \beta \psi$. So, this can be written as $\beta \psi$ only times qN_A by epsilon times $\beta \psi$.

Now, if you look at this equation, it is some coefficient times I so, the solution can be expressed as $\psi \propto \sqrt{X}$ by L_D , L_D is a square root of this coefficient epsilon by $qN_A \beta$ and of course, if you calculate C that is dQ by dV so, dQ by $D \psi$ s. So, Q is epsilon times ψ s by L_D . So that is the electric field, epsilon E and then that comes out to epsilon around by L_D .

So, at flat band capacitance you have oxide capacitance which is epsilon by D or $t_{\text{oxide}} + 1$ over epsilon by L_D . That is the capacitors from the depletion region. So, it is an expression for flat band capacitance so, this is slightly less than the oxide capacitance.

(Refer Slide Time: 23:14)



CAPACITANCE VS VOLTAGE

- As voltage is increased, C goes through minimum (weak inversion) where $d\psi/dQ$ is fairly flat
- C will increase with onset of strong inversion
- Capacitance is an AC measurement which increases when period is long wrt minority carrier lifetime
- At "high" frequency, carriers can't keep up - don't see increased capacitance with voltage
- For Si MOS, "high" frequency = 10-100 Hz

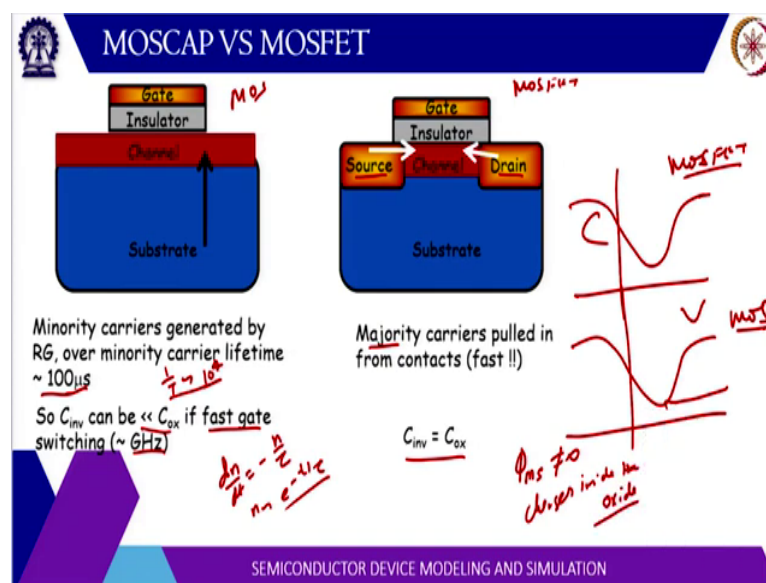
SEMICONDUCTOR DEVICE MODELING AND SIMULATION

Now so, as the voltage is increased C goes through a minimum during the weak inversion, where $d\psi$ by dQ is fairly flat. Then of course at the onset of strong inversion C again increases the capacitance again increases because the contribution from the semiconductor

side is now not that significant. And of course, when you consider frequency based measurement.

Then if the frequency is much smaller that is comparable to the inverse of minority carrier lifetime. Then, in case of inversion, a strong inversion, the capacitance will be same as the oxide capacitance. But at high frequencies these carriers cannot follow the voltage signal and then that will lead to increased capacitance with the voltage. So, of course, we should remember that for most high frequency is roughly 10 to 100 hertz because that is related to the inverse of the carrier lifetime. So that is quite large basically.

(Refer Slide Time: 24:28)



Now, you can compare these two structures one is the mass structure, so, this is metal oxide semiconductor other is MOS field effect transistor. So, here what we have done, we have made the two contacts. There is a source here there is a drain here. So, now, if you compare their CV characteristic, regardless of the frequency the MOSFET will give this curve. Why? Because now, this inverted carriers are they do not need to be generated.

They can be drawn from the source and drain side. So, they can quickly come so, for MOSFET the CV characteristic will be like this but in case of MOS capacitor the curve will be frequency dependent because in case of MOS these carriers have to be generated. So, therefore, because the carry lap time is order of hundred micro second. So that means your $1/T$ is basically 10^4 raised to power 4.

So that is kilohertz so, if you see their lifetime that is quite fast and that means once you go to order of kilowatts, they cannot keep up with the with the carrier consultation. So that means carrier lifetime is basically rated like this $\frac{dn}{dt} = -\frac{n}{\tau}$. So, τ is the carrier lifetime, so, there is a time it takes to recombine or generate so, n can be written as e to the power $-t/\tau$.

So, if τ is 100 micro second that is it takes 100 micro second to generate a percentage of carriers. So, your C inverse will be much smaller than the C oxide. If there is a fast switching there order of few megahertz or gigahertz in case of MOSFET this measured carriers are pulled from the source and drain region. So, your inversion capacitors will be same as the oxide capacitance because now, it can easily get the charge from these two sources.

So, thank you very much next class we will discuss about the real diode with difference in the work functions, ϕ_{ms} will be non-zero and we will also consider the effect of charges inside the oxide region or the insulator region. Thank you.