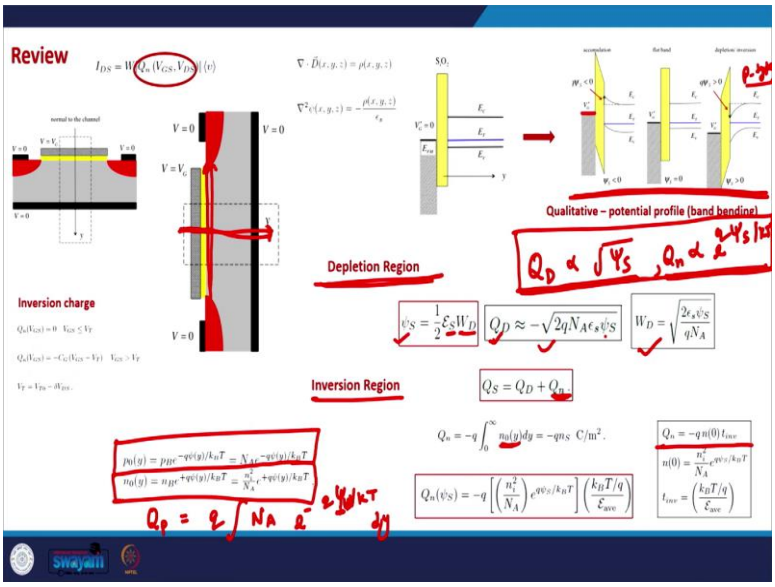
Physics of Nanoscale Devices
Prof. Vishvendra Singh Poonia
Department of Electronics and Communication Engineering
Indian Institute of Technology, Roorkee

Lecture - 52
MOSFET Electrostatics, Threshold Voltage

Hello everyone. Today we will continue the discussion on MOSFET Electrostatics and we will come across a new idea or we will formally discuss the idea of Threshold Voltage, which is I guess most of us know about threshold voltage, but this is not yet mathematically discussed. So, let us quickly review what we have seen.

(Refer Slide Time: 00:48)



In the MOSFET electrostatics, in generally in conventional MOSFETs we only do electrostatics in one direction which is the y direction along the direction from the gate to the body, gate to the substrate. Because in the long channel MOSFETs generally the electric field in along the direction of the channel is low and the electric field in the perpendicular direction which is the direction from the gate to the body is generally high.

So, generally that is how that is why this is the typical typically only 1D electrostatics suffice for conventional MOSFETs, but as you will see in coming classes that in the short MOSFETs the electric field in the other direction is not so low and that is why we also need to consider the electric field in the source-drain direction and we need to deal with the 2D electrostatics in the short MOSFETs.

But that is for the later on. And these are the key points of the electrostatics discussion and I hope these things are clear to all of you. This is the way the bands are visualized in when we apply various kind of gate voltages and when the gate voltage is a positive value it leads to the depletion in the P type semiconductor, it leads to the depletion of holes in the semiconductor.

And if this becomes even more positive, the gate voltage becomes even more positive in that case it leads to inversion. So, in the depletion region, which means that there is a positive gate voltage, but still the inversion is not there, the inversion charge is not there in that case, the this Ψ_S which is the surface potential of the semiconductor is related to the depletion width and the electric field at the surface. The charge is given by this expression, and the depletion width is given by this expression.

And in the last class we saw that, in the inversion region in addition to the depletion charge there is also this mobile charge or also what is known as the inversion charge, and the inversion charge is calculated by integrating the electron density in the semiconductor along y direction, the electron density comes from the fundamental carrier statistics of the semiconductor from this relationship.

And finally, what we see is that this inversion charge can be represented in this expression or it can in short be written like this. Now, there are there is an important observation that we need to make here. The observation is that the depletion charge is related to Ψ_S in this way. So, the Q_D charge is directly proportional to the square root of Ψ_S, but the inversion charge is exponentially related to the Ψ_S which is the surface potential.

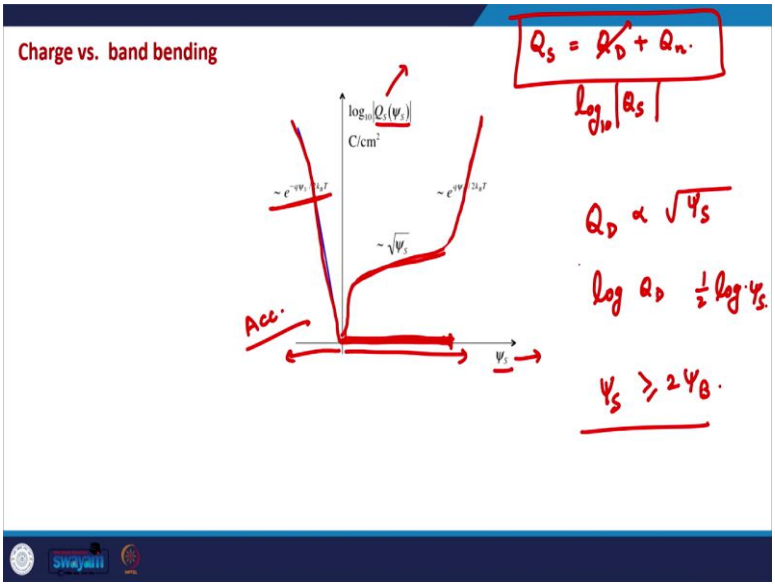
So, this Q_n is exponentially related to the surface potential and there is there are other factors as well here, and that is the key difference here. So, what it means is that, for the low positive gate voltages when the inversion has not yet happened and the only the depletion charge is there in that case the charge will vary according to this relationship.

But as soon as there is strong inversion in the semiconductor the this the inversion charge or the charge in the semiconductor increases exponentially. Now, you might think what would be the charge in the accumulation region and just to quickly remind you the accumulation region is the region, when a negative gate voltage is applied which means

that now the holes are accumulated in the semiconductor and essentially that is given by this relationship.

So, the accumulation charge which will be a positive charge is given by q times $\int N_A \exp(-q\Psi(y)/kT)dy$ and this Ψ is negative because a negative gate voltage is there which results in a negative surface potential. So, this $\Psi(y)$ will be negative and this integration needs to be done in y direction. So, even this accumulation charge is also exponentially related.

(Refer Slide Time: 05:41)



So, just to summarize all these various regimes, we can what we can do is we can plot a relationship between the semiconductor charge and the surface potential. So, the semiconductor charge Q_S is if you recall is the summation of depletion charge plus inversion charge, and in order to properly understand this we plot the log of the log 10 of Q_S take a mod as well we generally take a mod.

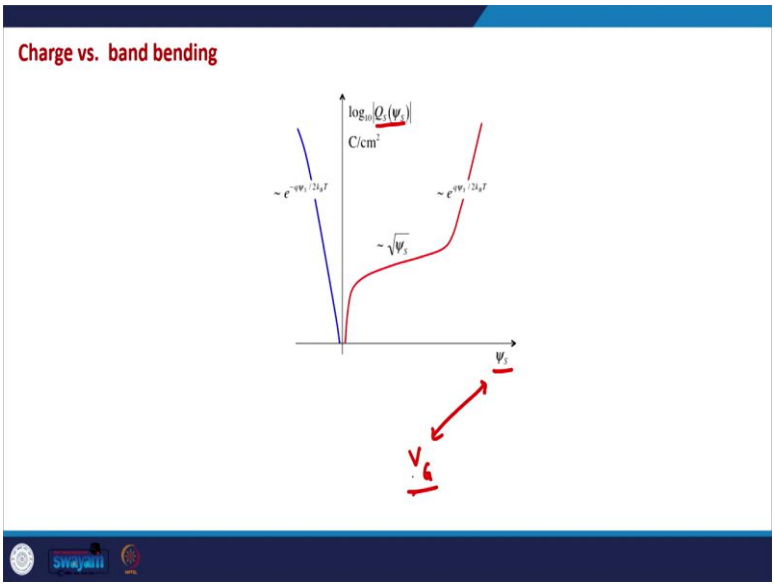
Because this Q_S is negative sometimes, but to try to understand how the semiconductor charge depends on the surface potential, we generally take we generally plot the log of mod of Q_S versus Ψ_S and this region when Ψ_S is negative is the accumulation region because the negative gate voltage is there which is resulting in the negative surface potential.

And in that region this Q_S will be only this depletion charge will be 0 and the only charge that will be there will be the mobile charge, and that will be exponentially related to the surface potential. So, that is why we see as exponential dependence of Q_S in the accumulation. In this in the positive half when Ψ_S is positive initially there is the depletion of holes and the inversion charge is very less and in that case this Q_n is negligible as compared to Q_D and the only charge that is there is just the Q_D charge.

And the Q_D charge as all of us know, the Q_D charge is related to Ψ_S in this way ok. So, which means that this log of Q_D will be 1/2 logΨ_S. So, that is why there is this slow dependence of the semiconductor charge as a function on Ψ_S in the for the low values of Ψ_S, but as soon as the inversion starts which means that now the Ψ_S is greater than or equal to 2 Ψ_B, in that case, this charge which is the mobile charge, this Q_n charge starts dominating and it starts becoming more than the Q_D charge.

So, this Q_D becomes negligible as compared to Q_n and again this Q_n has exponential dependence on Ψ_S. So, again this plot of Q_S versus Ψ_S follows the exponential path. So, this is I would say, in MOSFET electrostatics this is an important observation; how the MOSFET charge or how the semiconductor charge which is essentially the charge in the channel region of the semiconductor varies as a function of Ψ_S.

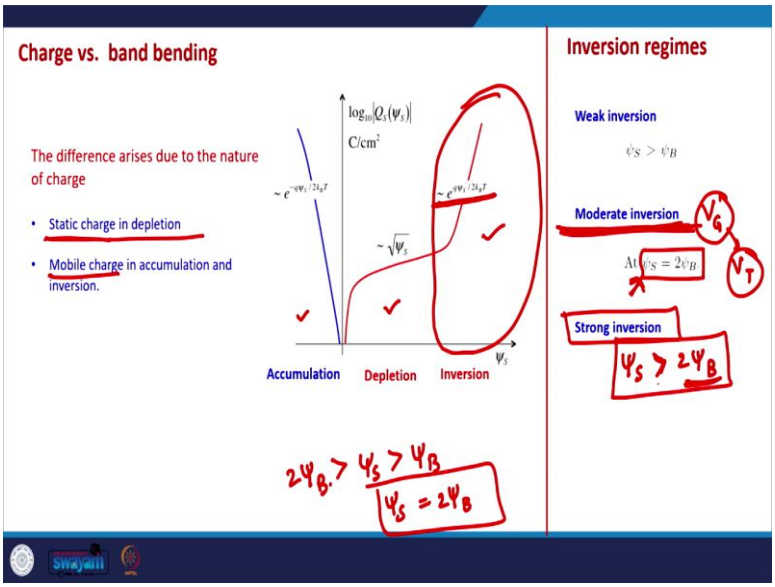
(Refer Slide Time: 09:04)



So, now we have seen the dependence between Q_S and Ψ_S, but actually the free variable in our hands is V_G. V_G means, the gate voltage that is what we can change externally. So,

we would also like to see a relationship between V_G and Ψ_S because that way we will properly understand the relationship between Q_S and V_G. So, that will be our next discussion, and from there this idea of threshold voltage will come about.

(Refer Slide Time: 09:37)



But before going into that discussion, let us have a quick sort of review or let us quickly recall what happens in various regimes. So, these are the various regimes and so, this is the accumulation, this is depletion and this is the inversion regime. And as we know that in the depletion region, the charge the major charge is the static charge, in the accumulation and the inversion region we have the mobile charge dominating the semiconductor charge.

And so, this is an important point here, because even in there is a difference in the nature of charges various charges and if we just focus on the inversion regime, because that is the most important regime for operation of the MOSFETs. There is this weak inversion, weak inversion is when the semiconductor is slightly n type at the interface which means that Ψ_S is greater than Ψ_B, but it is less than 2 Ψ_B.

So, in this regime, even in the weak regime, the mobile charges is very less and the total charge in the semiconductor is dominated by the depletion charge or the static charge. The moderate inversion is the inversion, or it happens when Ψ_S is equal to twice of Ψ_B. So, this is known as the moderate inversion and this is also the defining point for the

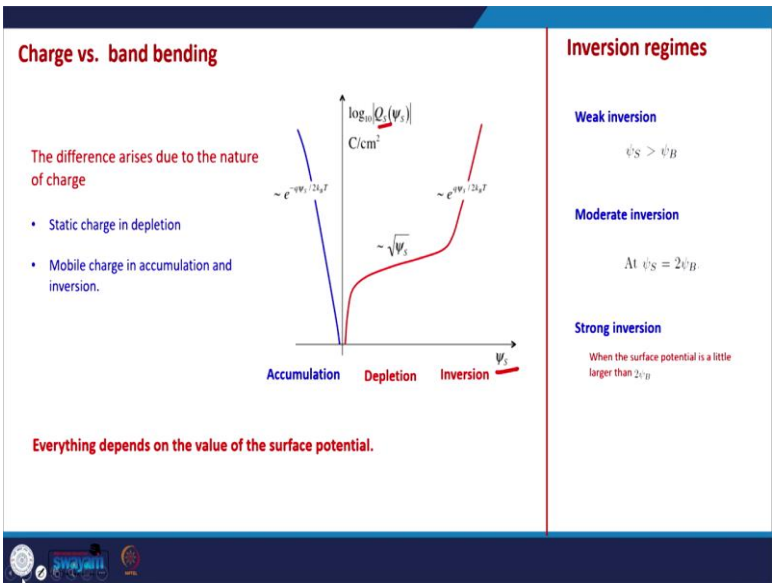
threshold voltage because the voltage at which Ψ_S becomes 2 Ψ_B is defined as the threshold voltage.

So, the gate voltage at which Ψ_S becomes equal to 2 Ψ_B is known as the threshold voltage. But even at this point, the mobile charge or the inversion charge is still not too much, it is still it is just comparable to the depletion charge at the interface per unit area. So, that is why it is known as the moderate the inversion, because even at the at this point the depletion charge is quite significant in the semiconductor and at the interface the mobile charge is just equal to the depletion charge per unit area.

And the strong inversion happens when Ψ_S is slightly more than 2 Ψ_B and at this point if we if you recall this inversion charge increases exponentially, and so it suddenly dominates the semiconductor charge. The semiconductor charge can be assumed to be entirely the mobile charge in the strong inversion regime.

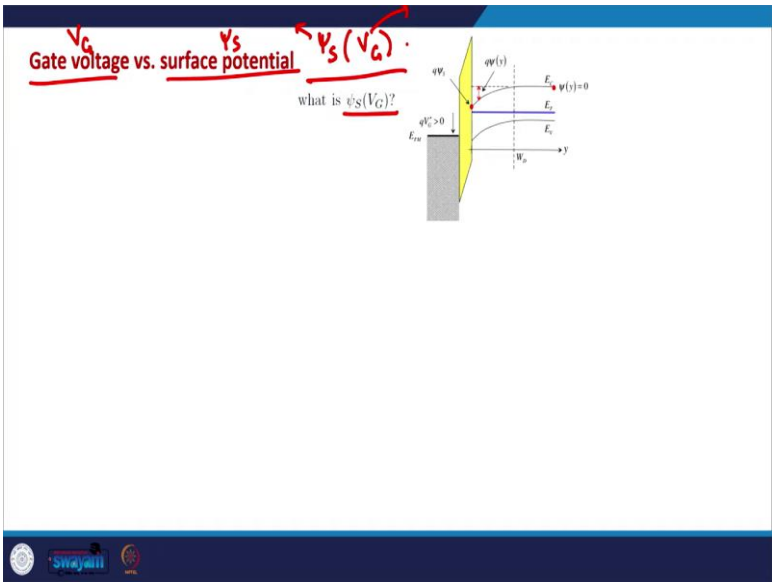
So, generally the MOSFETs are operated not exactly at Ψ_S is equal to 2 Ψ_B, they are operated slightly above a few kT values above the $2\Psi_B$ surface potential and this is known as the strong inversion regime. So, please try to understand these intricacies of the inversion regimes as well. We will not spend too much of a time here, because this is all part of the conventional electrostatics theory of MOSFETs.

(Refer Slide Time: 13:08)



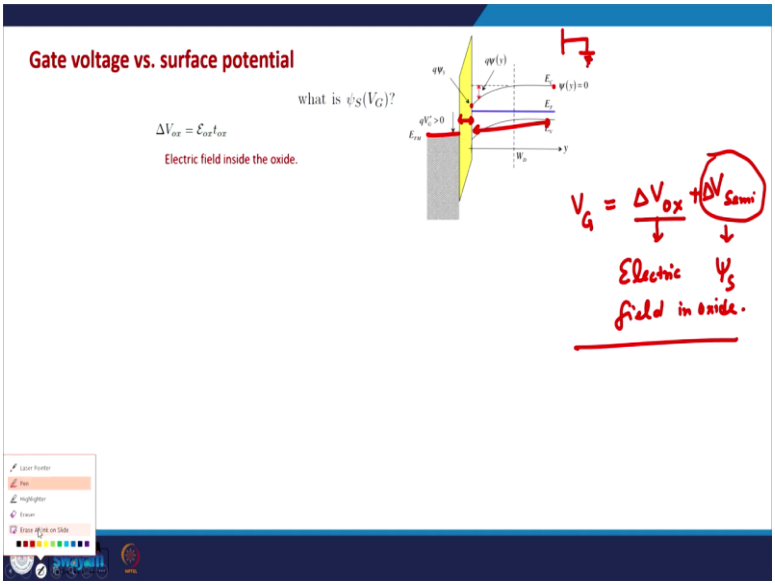
So, now we understand the relationship between Q_S and Ψ_S.

(Refer Slide Time: 13:24)



Now onwards we will try to see the relationship between Ψ_S the surface potential and V_G because V_G is the voltage that is in our hands. So, to start with what is that we are trying to find out? We are trying to find out a relationship between Ψ_S and V_G or we are trying to see how Ψ_S is a function of V_G, because V_G is the independent variable, psi S is the dependent variable; Ψ_S depends on the applied V_G.

(Refer Slide Time: 13:58)

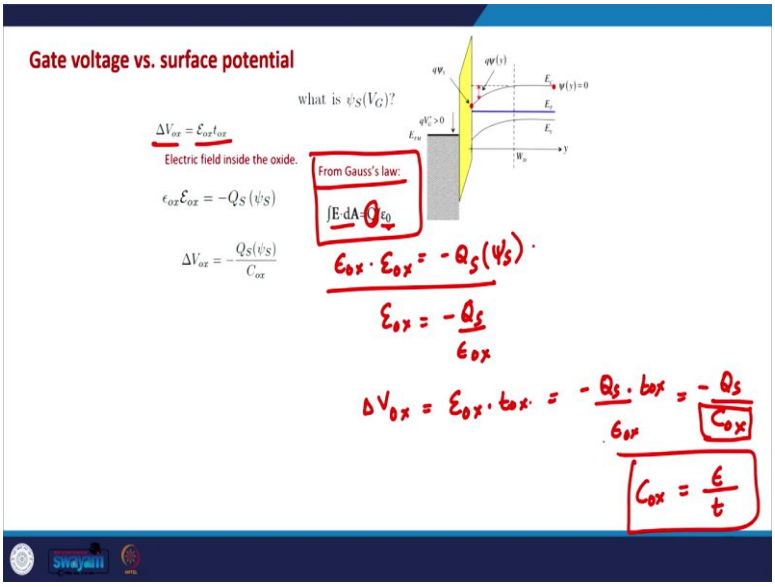


So, the one way to understand this is that, the voltage that we apply on the gate terminal is dropped in the oxide and in the semiconductor ok. So, the gate; because deep inside

the semiconductor or at the other terminal, then we generally ground the body contact. So, the gate voltage can be written as the voltage drop inside semiconductor and the voltage drop in the oxide. What is the voltage drop in the semiconductor?

The voltage drop in the semiconductor is essentially Ψ_S. Because Ψ_S is the electrical potential at the surface and deep inside it is 0. So, the total voltage drop is Ψ_S. But what is the voltage drop in the oxide? So, that will be given by the electric field in the oxide ok. So, that is an important point again here.

(Refer Slide Time: 15:18)



That in order to understand the relationship between V_G and Ψ_S we need to understand what is the voltage drop in the oxide and that depends on the electric field in the oxide. And in order to find out the electric field in the oxide we use the Gauss's Law. And what does the Gauss law says?

The Gauss law says that inside a closed surface, the surface integral of the electric field is given by the total charge enclosed in that surface divided by the permittivity of that medium essentially. So, with this, what we can say that or what we can say is that ϵ_{OX} times electric field in the semi in the oxide is $-Q_S$, because where Q_S is the charge in the semiconductor.

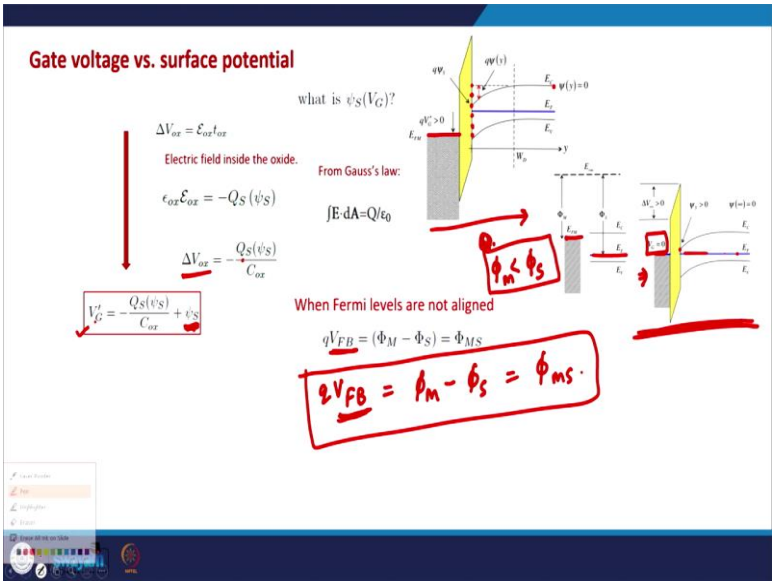
Because if we and most of the charge in the semiconductor we assume that it drops very close to the interface, and by using the Gauss's law we can write down this. We are

assuming that there is no interface charge in the semiconductor. So, that is why this is what we can write down. This Q_S is the charge per unit area and it depends on the surface potential as well.

So, E_{OX} is essentially $-Q_S/€_{OX}$, where $€_{OX}$ is the dielectric constant of the oxide or the permittivity of the oxide, and that is why the voltage drop in the oxide is which is essentially E_{OX} times t_{OX} is $-Q_S$ divided by or Q_S times $t_{OX}/€_{OX}$ or it can be written as $-Q_S/C_{OX}$ where C_{OX} is the capacitance of the oxide.

And the capacitance per unit area of any material is given by this, total capacitance is just multiplication of area to this constant, but if we want to find out the capacitance per unit area it is just €/t ok. So, this ΔV_{OX} finally, is $-Q_S/C_{OX}$.

(Refer Slide Time: 17:43)



And so, that is why the gate voltage the total gate voltage is ΔV_{OX} which is $-Q_S/C_{OX} + \Psi_S$ which is the voltage that drops inside the semiconductor. And please note it down here that we have written a prime on V'_G. What does the prime means here?

The prime means that this is this may not be the actual gate voltage, this is a modified gate voltage because in most of the cases the Fermi level of the metal and the Fermi level of the semiconductor they may not be aligned with each other in 0 bias condition. So, there might be a work function difference between the metal and the semiconductor, also

there might be some surface charges, there might be some static charge at the interface between the oxide and the semiconductor.

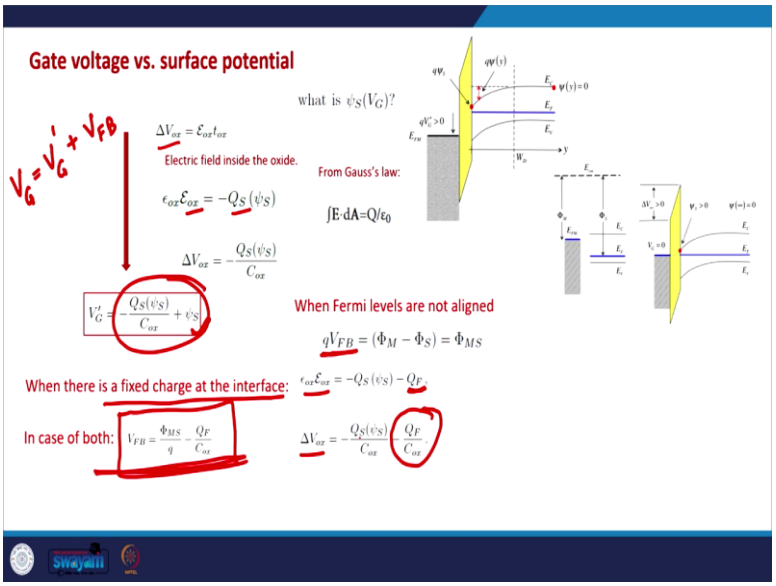
So, these charges come because of the fabrication, because of the imperfect fabrication and these are in most of the cases unavoidable. So, this is in a way this V_G' prime is the modified gate voltage which essentially takes into account the effect of the flat band or I would say the work function difference and the surface charges.

Let me sort of quickly do that here as well. So, when in no bias condition when there is a work function difference between the metal between the gate and the semiconductor which means. So, in this example we have shown that phi those work function of the metal is less than the work function of the semiconductor. So, in that case what happens is when this MOS structure is formed in that case, there will be a band bending even at 0 gate voltage. So, this will be the case when these will be joined.

Because at 0 V_G value, the Fermi levels need to be aligned and in order to align the Fermi levels, the conduction band and the valence band extremas will bend. So, this is the case with no applied voltage on the gate terminal when there is a work function difference. And in order to neutralize this band bending we need to apply V_{FB} voltage on the gate terminal and this V_{FB} is given by, V_{FB} is 1/q times ($\Phi_M - \Phi_S$), or Φ_{MS}.

So, when we apply this voltage on the gate terminal, this bend bending is neutralized and in that case the flat bands are achieved. So, that is why this gate voltage is known as the flat band voltage.

(Refer Slide Time: 21:12)

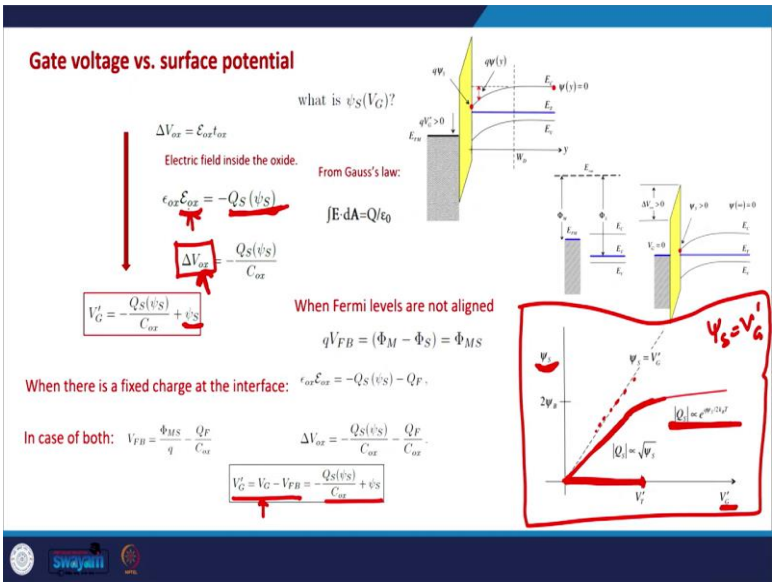


So, in all those materials in which the metal Fermi level and the semiconductor Fermi levels are not aligned in that case we in addition to these voltages the voltage drop in the semiconductor and in the oxide, we also need to account for the flat band voltage. Also if there is a fixed charge at the interface because of the fabrication because of various conditions, we also need to account for Q_F while calculating the electric field in the oxide and not only Q_S we also need to account for Q_F in this calculation.

So, that is why this ΔV_{OX} becomes $-Q_S/C_{OX} - Q_F/C_{OX}$. So, in total we need to generally this quantity is absorbed in the flat band voltage and this flat band voltage is written as $\Phi_{MS}/q - Q_F/C_{OX}$. So, this is the additional voltage that we need to apply on the gate terminal in order to achieve the perfect or ideal conditions that we have been discussing so far.

So, apart from this voltage we need to account for the flat band voltage in the semiconductor and that will. So, V_G the actual gate voltage will be $V_G' + V_{FB}$ ok.

(Refer Slide Time: 22:44)



So, or V_G' is V_G - V_{FB} and this is given by this value. So, now, this is the relationship between Ψ_S and V_G'. So, this is a complex relationship I would say this V_G' and Ψ_S they are related to each other. Because this Q_S which includes the static charge and the depletion charge and the inversion charge this is a complex quantity, and then we have this Ψ_S term as well.

So, analytically it is not straight forward to do this calculation to exactly figure out a relationship between Ψ_S and V_G', but we can qualitatively understand this relationship from this plot. And what this plot says is that, when the gate voltage is below the threshold voltage or in other words the MOSFET is operating in the depletion region.

So, the most of the charge in the material or most of the charges because of the depletion charge, in that case the Ψ_S the surface potential or the semiconductor potential almost follows the gate voltage. So, which means that so, this dotted line if you see is the line of Ψ_S is equal to V_G'.

And in the sub threshold regime when V_G' is less than V_T' in that case, the actual relationship between Ψ_S and V_G' is quite close to this line. So, which means that the Ψ_S value this the semiconductor potential follows closely the gate voltage.

And the reason for that is that the semiconductor charge in this region is less because it is only because of the depletion charge, and that is why this electric field in the oxide is

less because it depends on the semiconductor charge and that is why the voltage drop in the oxide is also less. So, which means that, most of the voltage drops in the semiconductor. So, below threshold most of the voltage drops in the semiconductor and Ψ_S is closely following the V_G'.

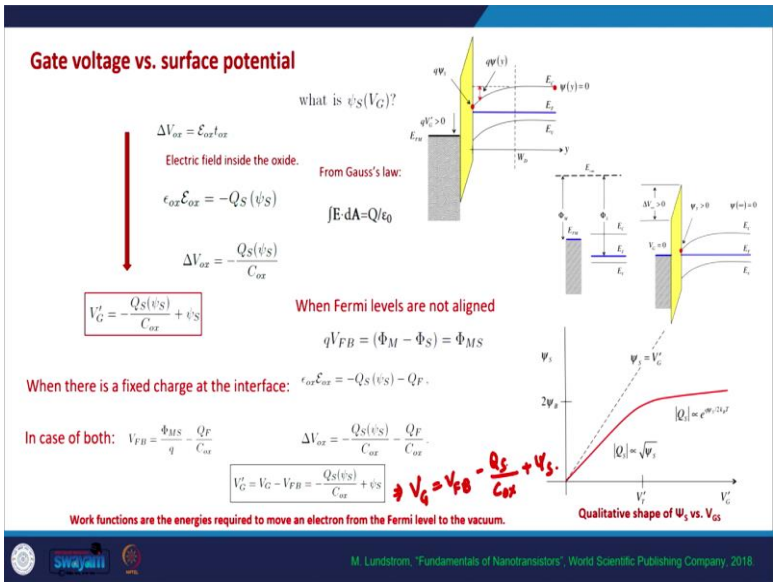
But as soon as the inversion starts which is exactly at this point when V_G' is equal to V_T', in that case the semiconductor charge is exponentially dependent on Ψ_S. Which means that the semiconductor charge increases abruptly and that also increases the electric field in the oxide also the voltage drop in the oxide. So, after the inversion has started after the strong inversion, there is a huge charge buildup in the semiconductor that leads to a huge voltage drop in the oxide.

And this Ψ_S actually deviates this goes away from the this plot this line, which means that after the inversion Ψ_S does not follow the gate voltage it departs away from the gate voltage. Because even a slight increase in the gate voltage results in a huge increase in the semiconductor charge and as a consequence a huge increase in the voltage drop in the oxide.

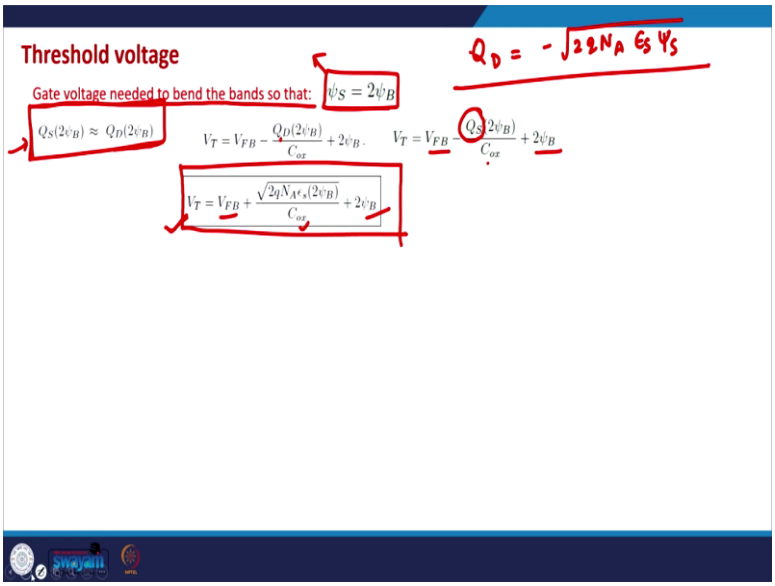
So, that is the qualitative I would say picture of the relationship between Ψ_S and V_G'. So, in the beginning of this lecture, we have seen a relationship between Q_S and Ψ_S, here we have seen a relationship between Ψ_S and V_G' and V_G' is related to V_G by this relationship V_G' is $V_G - V_{FB}$.

So, essentially the same kind of relationship will be there between Ψ_S and V_G. So, before inversion and after inversion this changes is noticeable change and this needs to be kept in mind while understanding the surface potential in the semiconductors ok.

(Refer Slide Time: 27:17)



(Refer Slide Time: 27:23)



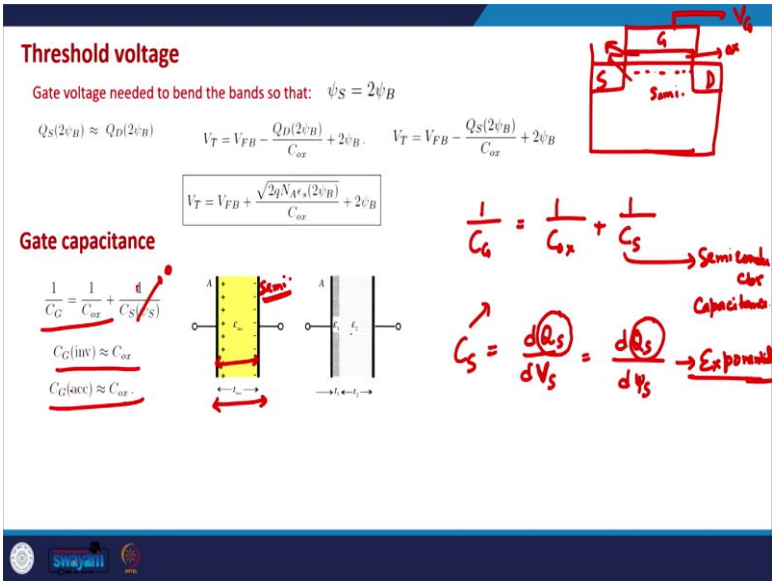
So, now let us move on to see, what happens or what is the threshold voltage. So, as we know that the threshold voltage is the gate voltage needed to bend the bands so that Ψ_S is equal to $2\Psi_B$. This definition now we have seen multiple times, we just need a mathematical form to this definition. And from this expression on the last slide this expression which is essentially V_G is equal to V_{FB} - Q_S/C_{OX} + Ψ_S, using this definition we will replace V_G by V_T when Ψ_S is 2 Ψ_B.

So, in this case what happens is V_T is equal to V_{FB} -Q_S(@ $\Psi^S=2\Psi^B$)/C_{OX} + $2\Psi_B$. So, we have replaced Ψ_S by $2\Psi_B$. And this is as I told you that this is the moderate inversion and even in this inversion, what we can assume is that the semiconductor charge is almost equal to the depletion charge the inversion charge is still not. So, significant.

So, this Q_S formula can be replaced by Q_D and this Q_D is related to Ψ_S if you recall Q_D is $\sqrt{(2qN_A€_S\Psi_S)}$. So, in this expression we need to replace Ψ_S by $2\Psi_B$. So, that way we will find out the expression for the threshold voltage. So, the threshold voltage is the flat band voltage plus $\sqrt{(2qN_A€_S.2\Psi B)}$)/C_{OX} + $2\Psi_B$.

So, this is the, this is one of the most frequently used expressions in the MOSFET electrostatics because this gives us the threshold voltage of a given MOSFET.

(Refer Slide Time: 29:37)



Let us quickly try to understand another important topic that is the gate capacitance. The gate capacitance is, essentially the capacitance of as seen from the gate terminal, and as we know that a MOSFET is also has a capacitive element. So, this is the gate terminal, this is the oxide and this is the semiconductor.

In the semiconductor what we have seen is that the charge changes as a function of the applied voltage. So, which means that the semiconductor also has some capacitance in it because whenever the charge is a function of the voltage, then it shows a capacitive

behavior and in that case. So, as seen from the gate terminal we will see the oxide capacitance and the semiconductor capacitance.

So, these are the 2 capacitive elements I would say, one is because of the this oxide which is an insulator and this is like a parallel plate capacitor, and then we have a semiconductor capacitance, but this is a function of the voltage. This capacitance is constant because this is like a parallel plate capacitor and the capacitance just depends on the thickness here.

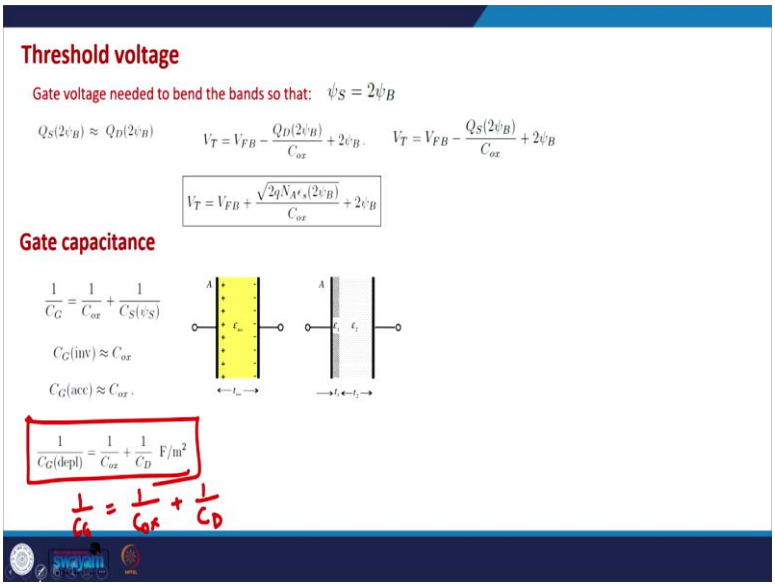
So, this is the picture here and the total capacitance or this becomes a series of 2 capacitors and the total gate capacitance will be $1/C_{OX} + 1/C_S$ which is the semiconductor capacitance ok. So, just to give you couple of pointers here in the inversion region, and this what is this C_S? C_S is dQ/dV. So, always this is how the capacitance is defined the change in charge as a function of the voltage.

And C_S will be the change in semiconductor charge as a function of the voltage applied which is the semiconductor voltage. So, it instead of dV it should be $d\Psi_S$ or dV let us just write is dV_S the voltage across the semiconductor which is also the Ψ_S voltage. And in the accumulation and the inversion regimes as all of us know that this Q_S is an exponential quantity as a function of Ψ_S.

So, this is also dQ_S divided by $d\Psi_S$, and Q_S is exponentially related to Ψ_S. So, this is also an exponential C_S is also exponential quantity which means that for a large enough accumulation and in strong inversion this is a huge quantity C_S the semiconductor capacitance is a huge capacitance and when we have two series capacitance is the total capacitance is actually limited by the smaller capacitance.

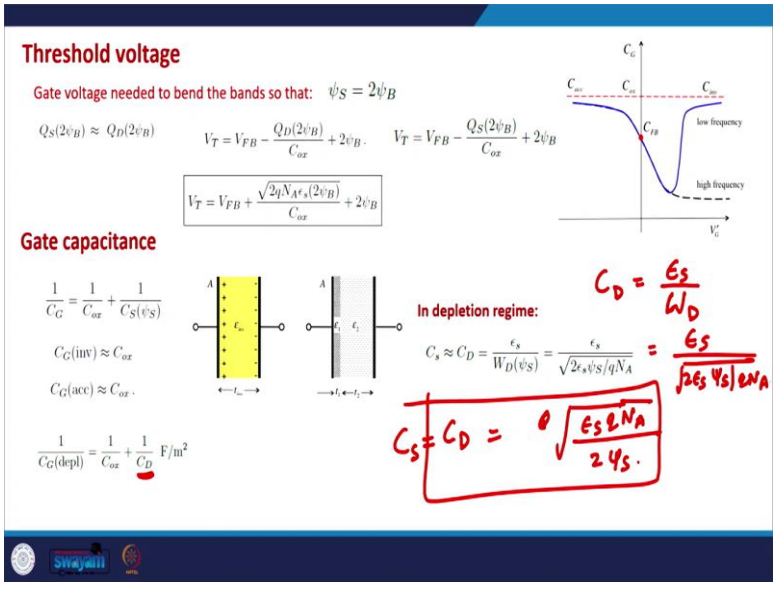
Because if C_S is extremely large then, $1/C_S$ will be a small number can be all approximately 0 and in that case C_G will just be equal to C_{OX}. So, what we can say is that in accumulation and in strong inversion the total gate capacitance is almost equal to the oxide capacitance. However, in the depletion region the semiconductor capacitance is not very large and in that case we need to account for this.

(Refer Slide Time: 33:48)



So, for that we need to understand that, that in the depletion region the gate capacitance will be derived from this formula, because still we have the two capacitors in series, and the gate capacitance will be dependent on the depletion capacitance of the semiconductor.

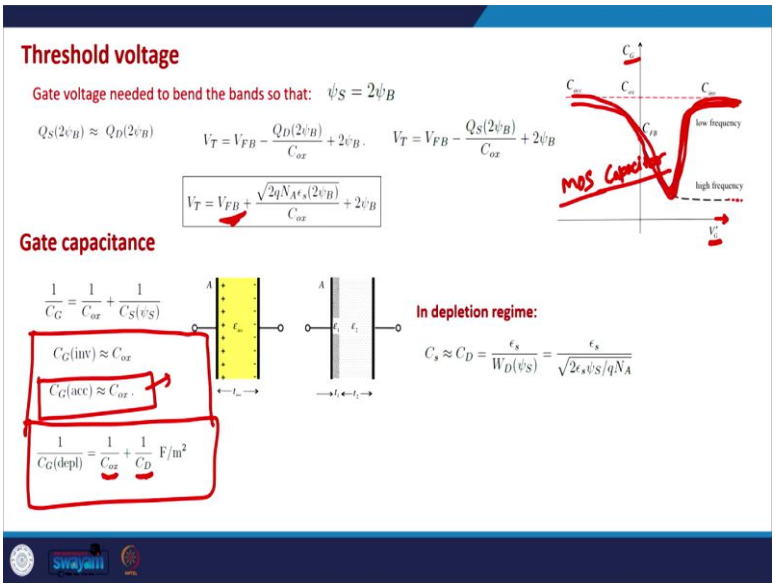
(Refer Slide Time: 34:06)



So, in the depletion region, generally the width of the depletion. So, in the depletion region what we can says that, the capacitance of the semiconductor will just be ϵ_S / W_D the depletion width, and the depletion width as a function of Ψ_S is $\sqrt{(2\epsilon_S\Psi_S/ qN_A)}$.

So, which means that the depletion capacitance will be $\sqrt{(\epsilon_S q N_A/2\Psi_S)}$. And in the depletion regime, the mobile charge or the inversion charge is very less. So, the only capacitance that the semiconductor has is the depletion capacitance.

(Refer Slide Time: 35:17)



So, the total in that case the total capacitance or the gate capacitance will be just a series combination of the depletion and the oxide capacitance. So, that way we can calculate this C_G in the depletion region. So, in this way we have seen what is the gate capacitance in various operational regimes of the MOSFET.

Now, in order to understand the relationship of C_G the gate capacitance on the gate voltage, V_G' please remember that this is V_G' because generally this effect of flat band voltage is not considered is not being considered at this point. So, there is a difference in this relationship the relationship between C_G and V_G at low frequency and at high frequency.

And what is the difference? At low frequency when we do we apply an ac signal of low frequency what happens is that, in the accumulation nothing changes because if the gate voltage is an ac signal and the semiconductor charge is the holes the majority carriers they can respond to any frequency of the gate voltage, but in the inversion region what happens is that, the gate voltage is a negative voltage.

So, there is a large negative dc voltage and along with this there is a small ac signal as well and in that case what happens is, that if the frequency of the gate voltage is extremely high then the semiconductor charge cannot respond to that frequency. Because the semiconductor charge is made of the minority charge carriers and the minority charge carriers or minority charges the rate of generation is limited by the recombination generation rate.

So, if the frequency is above a few megahertz in that case this inversion charge cannot respond to that frequency and the semiconductor will have only the depletion capacitance and at high frequency as you can see the gate capacitance will be equal to the capacitance in the depletion region.

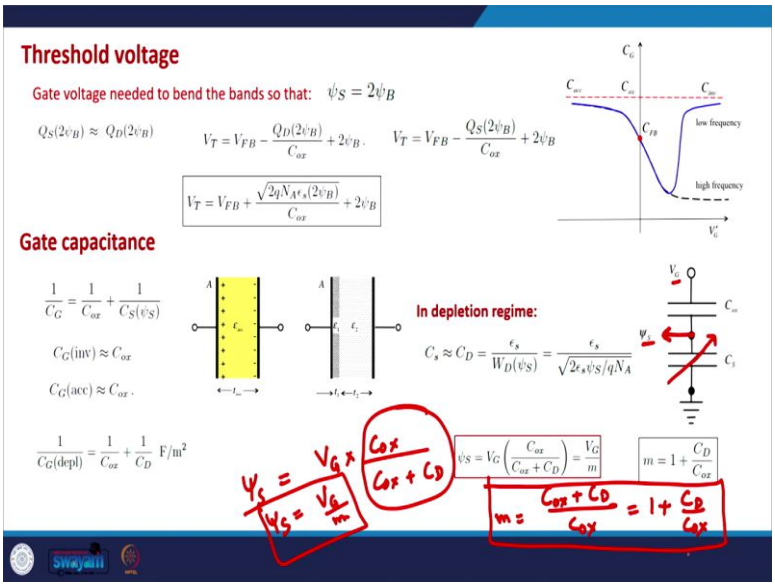
So, this is the capacitance in the depletion region and this is a cumulative effect of C_{OX} and C_D, in the accumulation region the gate capacitance is quite close to the oxide capacitance, but at low frequency when the gate voltage is applied we apply a gate voltage of low frequency in that case the inversion layer can respond to that low frequency and the total gate capacitance is almost again equal to the oxide capacitance.

So, this is a key point here in the gate capacitance, and this needs to be kept in mind while designing various circuits of MOSFETs that at high frequency this charge inversion charge may not respond to the gate voltage. But if we consider the total MOSFET in that case this charge this is the true this is true for MOS device MOS capacitor.

But if we consider the MOSFET in that case what happens is that, this inversion layer charge which is essentially the electrons that can come from the source and the drain region and even at high frequency in the MOSFET we will see this behavior this low frequency behavior will be seen ok.

Because now the electrons need not come from recombination generation process, they can come from the contacts as well they can because they are the majority carriers there. So, this is the frequency this will be the frequency or this will be the CV relationship for all frequencies ok.

(Refer Slide Time: 39:29)



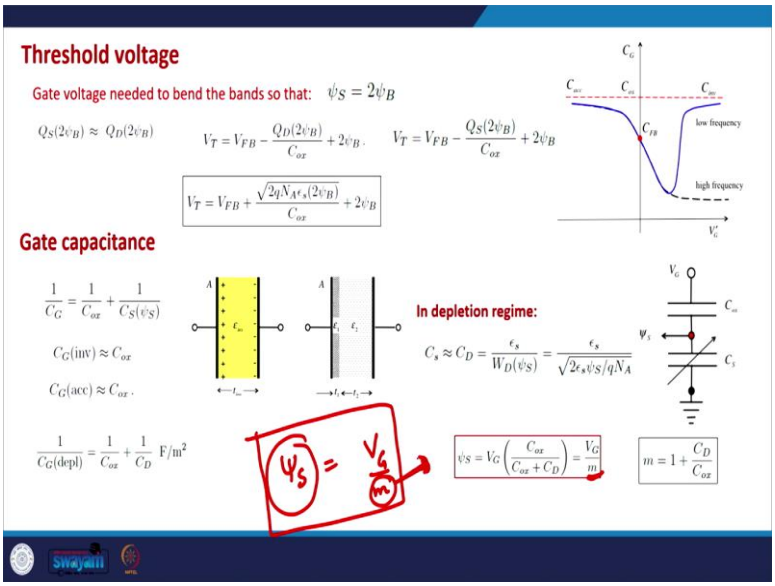
Now, the last point in today's discussion is the relationship between V_G and Ψ_S in terms of various capacitances. We have already seen a relationship between V_G and Ψ_S in terms of charge, but a more I would say intuitive picture is the relationship between Ψ_S and V_G as a function of various capacitances.

So, as you can see that this entire MOSFET can be sort of summarized in this way where this is the gate terminal then we have an oxide this is the interface between the oxide and the semiconductor and then we have the variable semiconductor capacitance and the voltage at the semiconductor surface is Ψ_S ok.

So, if we this is a series of capacitances, and if we try to find this Ψ_S in terms of V_G we just need to apply the voltage division rule in a series of capacitors and that rule says, what does that rule say is that Ψ_S is V_G times $C_{OX}/(C_{OX} + C_D)$ or $C_S \sim C_D$ is the case we are considering the depletion regime, and in total what we can do is we can define this to be V_G divided by m.

So, Ψ_S is V_G/m, where m is $(C_{OX} + C_D)/ C_{OX}$ or $1 + C_D/C_{OX}$. So, this m is known as the body coefficient and this is always a number larger than 1, but now we have a very intuitive expression between Ψ_S and V_G.

(Refer Slide Time: 41:27)



And that is Ψ_S is V_G divided by m. So, this m actually governs how much of the gate voltage drops across the semiconductor because if this m is close to 1 in that case Ψ_S will be almost equal to V_G. So, what we can say is that almost all the voltage drops across the semiconductor.

So, in an in our ideal or in order to make the MOSFET better and better what we try to do is, we try to make m as close to 1 as possible ok. So, we will stop here. I will recommend all of you to think more about the body coefficient this coefficient m, and we will start from this point, and we will see various ways in which m is brought closer to 1.

So, thank you and see you in the next class.