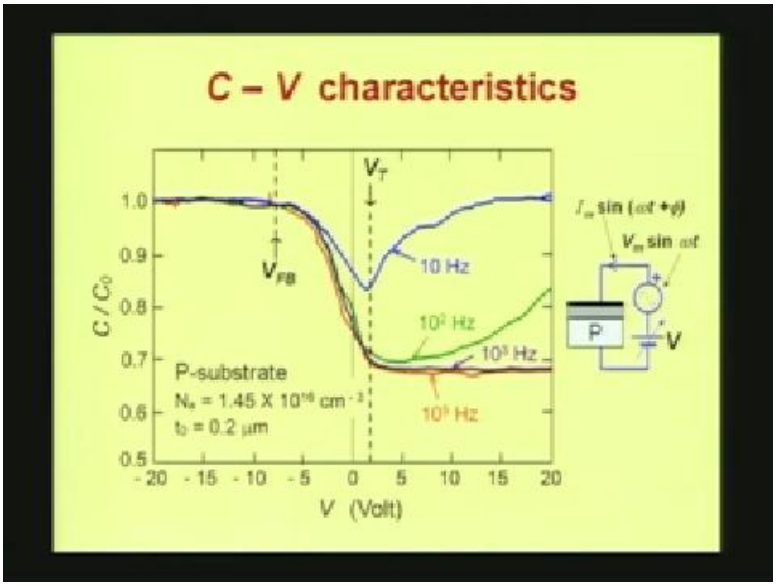
Solid State Devices
Dr. S. Karmalkar
Department of Electronics and Communication Engineering
Indian Institute of Technology, Madras
Lecture - 33
Metal-Oxide-Semiconductor (MOS) Junction

In the previous lecture we completed the discussion on Bipolar Junction Transistor. In this lecture we shall begin a new topic that is the Metal-Oxide-Semiconductor Junction. Now, for reasons that we will discuss this structure that is the Metal-Oxide-Semiconductor Junction, it is very often called the capacitor. Now let us see what we will try to address in the next few lectures. The most important thing we would like to explain for the MOS junction or the MOS capacitor is the C-V characteristics shown in the following slide. Let us look at the characteristics in detail.

(Refer Slide Time: 08:43)



On the X axis you have the voltage and on the Y axis you have the capacitance of the structure divided by C_0. We will see what C_0 means. Basically this is the normalized capacitance. So here you also have the biasing arrangement under which the capacitance characteristics are being measured so you have a DC bias applied and a small signal is superimposed on the DC bias. So this is a sinusoidal small signal voltage. Now the capacitor is a two terminal device and the structure is that there is a semiconductor, in this particular example it is p-type then on the semiconductor you have a silicon dioxide layer which is shown by this hatched portion and then on the top of this silicon dioxide layer you have a metal electrode.

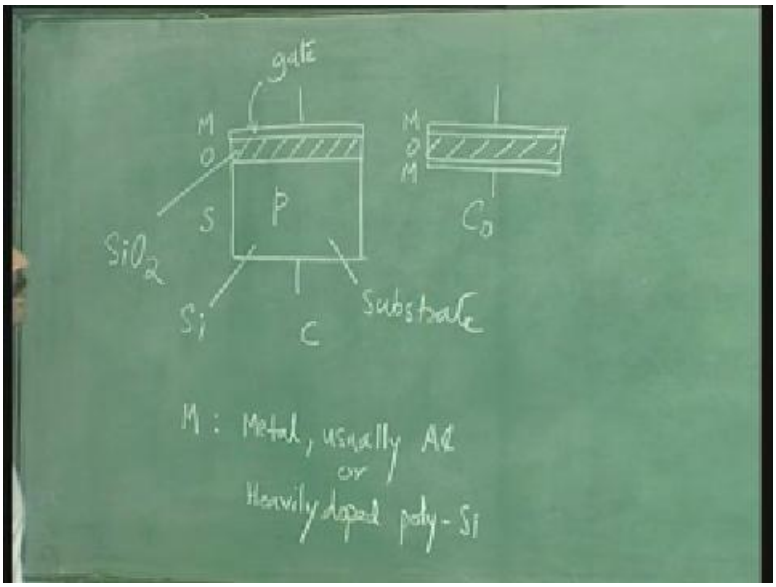
Now between the metal electrode and the p-type semiconductor there is a capacitance that is being measured. Now how do you know that there is capacitance here between these two in terms of the terminal characteristics? Well when you apply a small signal voltage there is a small signal current that results in the circuit. The relation between the small signal current or the sinusoidal current and the sinusoidal voltage tells you whether this structure is resistive or capacitive. If it is a purely capacitive structure then the phase difference between this current

and the voltage indicated by phi here would be 90 degrees though omega here indicates the frequency of the small signal. Now when you vary a DC voltage as shown in the X axis here you find that the capacitance of the device is constant over a certain range and then beyond a voltage indicated by this symbol V_{FB} or the Flat-Band voltage the capacitance starts falling.

Now in this region what you find is that the capacitance is independent of the frequency of this signal. But beyond this range as you move to the right you find the capacitance for different frequencies is slightly different. Until you come to this point here which is called the threshold voltage denoted by the symbol V_T the capacitance falls and at high frequencies the capacitance beyond this V_T point or threshold voltage point saturates. At very low frequencies of about 10 hertz or lower however the capacitance raises again to value that you had for very high negative voltages.

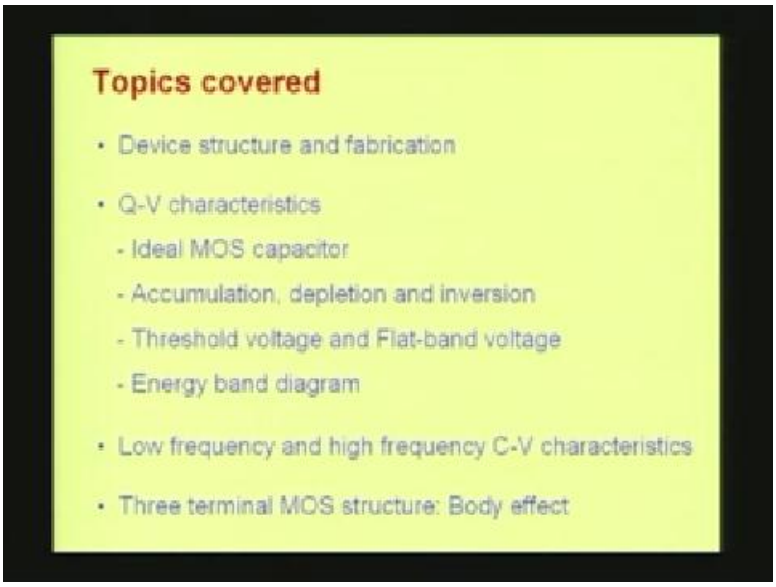
Now these are the characteristics assuming a p-type substrate. So we would like to explain this variation in the capacitance as a function of voltage and also as a function of frequency. And then we would like to derive formulae for V_{FB} that is the Flat-Band voltage and V_T the thermal voltage. Now let us explain what is meant by this C_0 capacitance which is used for normalizing the capacitance of the device. Now this C_0 can be explained as follows:

(Refer Slide Time: 14:36)



This is your structure. So this is the metal, this is the silicon dioxide and this is the semiconductor which is silicon. So O stands for oxide silicon dioxide. Now C_0 indicates the capacitance of a structure which has metal electrodes on either side. So this is metal, this is oxide and this is again the same metal. This is a normal parallel plate capacitor containing the oxide layer as the dielectric. Now the thickness of the oxide layer is the same as the thickness of the oxide layer in the actual MOS capacitor. So capacitance of this particular structure is C_0 [08:07] between these two terminals whereas the capacitance of this MOS structure is C. So C is normalized with respect to C_0 so what is this curve? In the slide it shows that the maximum value of the capacitance of the MOS structure is equal to C_0. Now we shall discuss the capacitance voltage characteristics in the following manner. First we will discuss the device structure and how it is fabricated.

(Refer Slide Time: 11:11)



Then we shall discuss the charge voltage characteristics, the Q here stands for charge. In the charge voltage characteristics first we will discuss the idea of an ideal MOS capacitor. Please note that we are using the words junction and capacitor interchangeably. This is because for the MOS junction there are no characteristics other than the capacitance characteristics. For a PN junction you have the current voltage characteristic that is the DC current voltage characteristics. For the capacitance however you do not have any DC current voltage characteristics because there is an oxide layer that insulates the metal from the semiconductor and when you apply voltage to the structure there is no DC current. However, when you apply an AC voltage there is an AC current.

After discussing the ideal MOS capacitor we will discuss the accumulation depletion and inversion regimes of operation. Then we will define the threshold voltage and Flat-band voltage and derive formulae for these two. Also, we will see the energy band diagram of the device. Now most of the analysis we do we will be doing without the help of the energy band diagram using the concentrations of electrons and holes within the semiconductor as a function of the voltage or electric field. This is what we had done even in the case of PN junction. And then we will point out exactly at what point we need the energy band diagram.

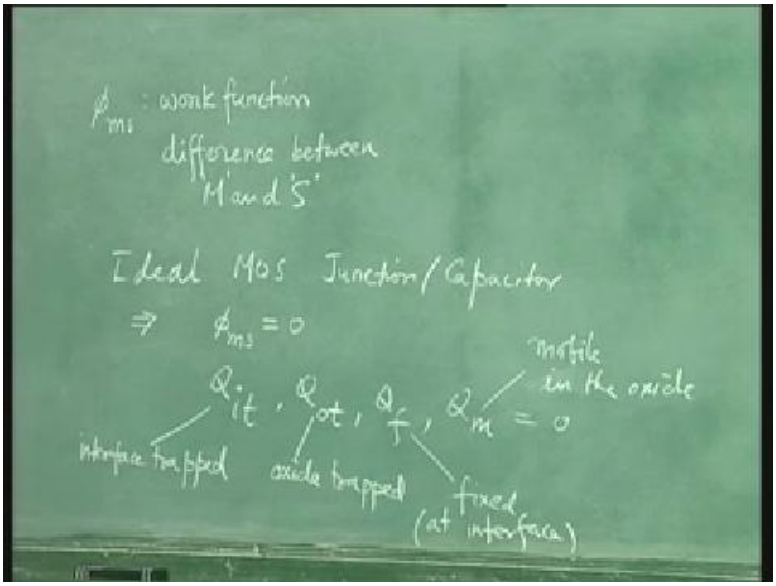
After discussing the Q-V characteristics we will discuss the capacitance voltage characteristics derived from the Q-V or charge voltage characteristics both under low frequency and high frequency condition. And finally we will discuss a three terminal MOS structure and an effect related to this namely the body effect. Now let us start with the device structure and fabrication. Now we said this is the device structure where you have silicon semiconductor and this could be either p-type or n-type. We will consider the p-type semiconductor. On this p-type semiconductor the oxide layer is grown by heating the semiconductor at very high temperatures.

Now generally the oxide layer is grown in the presence of pure or ultra pure oxygen at temperatures as high as 1100 to 1200 degrees C. The time of this oxidation is a couple of hours and it results in oxide thicknesses which range from about 100 Angstroms to 1000 Angstroms. In modern days the technology is advanced so we can have very thin oxide layers

of very good quality. Now it is very important that this oxide layer should be of high quality. That means this dielectric constant [12:36] should be high and it should be very insulating. Now, after the oxide layer is grown on the p-type semiconductor we deposit a metal contact. The metal layer is generally made of either aluminium or in modern days the metal layer is replaced a heavily doped polysilicon layer which also acts like a metal because its conductivity is very high. So, although we use a nomenclature M here, M here stands for either a metal layer in which case it is mostly aluminium or a heavily doped polysilicon layer, metal which is usually aluminium or heavily doped polysilicon.

Now when we say heavily doped polysilicon it obviously means that the doping can be either n-type or pminustype. So you have MOS capacitors with N plus polysilicon or you have MOS capacitors with P plus polysilicon. So, just as substrates can be p or n-type if the gate layer is made of polysilicon it can also be n or p-type. So this metal layer is also called the gate layer and this is the substrate.

(Refer Slide Time: 20:07)



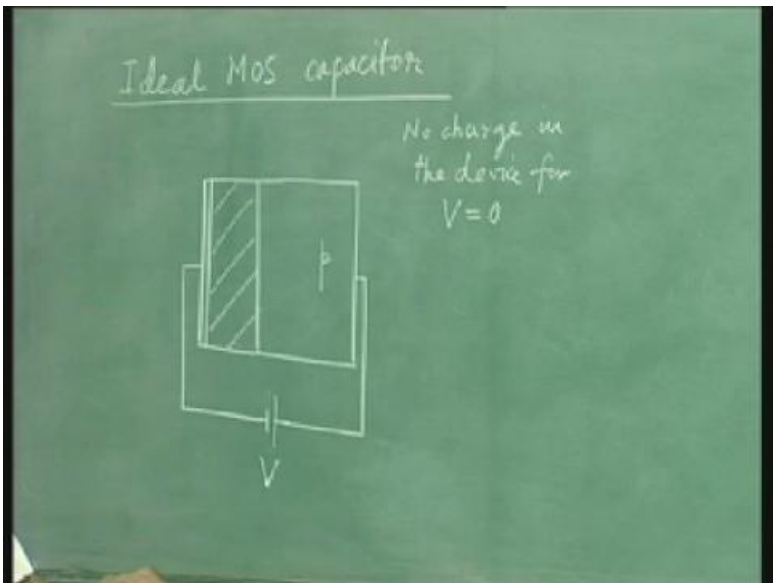
Now some important things we need to understand about this structure is that generally the gate is of a different material than the substrate. Even when the gate is made of polysilicon doping in the gate is different from doping in the substrate. Therefore there is what is called a contact potential between the metal and the substrate. In the case of the gate being a metal the contact potential between the metal and the substrate is also represented using the work function difference and denoted normally as phi_{ms} that is work function difference between metal and semiconductor. So M stands for metal and S stands for semiconductor. As we said earlier that symbol M will be used in general whether it is a metal or whether it is heavily doped polysilicon. So in general there is a contact potential between the gate and the substrate. This is one point.

The second point to note is regarding the interface between silicon dioxide and silicon. This interface that is this region is of great complexity. There are different types of charges at this interface these are called interface charges. Now in a simple analysis what we do is we neglect the effects of various charges at the interface. We also neglect the contact potential difference between the gate and the substrate. Now such a structure does not have any contact

potential difference between the gate and the substrate and which does not have any interface charges or even charges in the oxide layer.

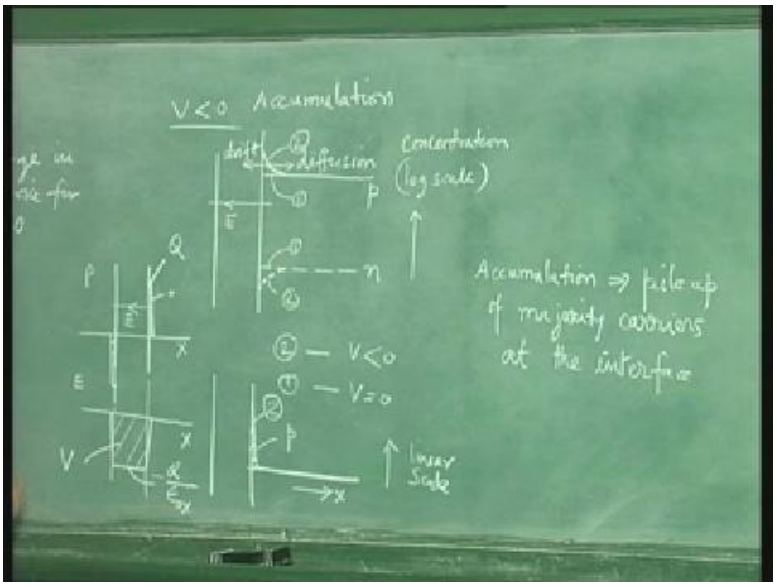
For example, this oxide layer may contain what are called mobile charges. Although we do not want them but during the fabrication some how these charges may get into the oxide. So a device without any interface charges or mobile charges in the oxide and without any contact potential difference between the gate and the substrate is called an ideal capacitor. So ideal MOS junction or capacitor implies phi_{ms} is equal to 0. And the interface oxide charges that is Q_{it} that is interface trap charge, Q_{ot} that is the oxide trap charge, Q_f that is the fixed charge at the interface and Q_m that is the mobile charge in the oxide layer all these are 0. So suffix t stands for trap charge, suffix i stands for interface and o stands for oxide. So Q_f is also an interface charge, suffix f stands for fixed, Q_m is a mobile charge m stands for mobile and this is in the oxide. So let us write: this is the interface trapped charge, this is oxide trapped charge, this is fixed charge at the interface and this is mobile charge in the oxide. So we will be coming to our analysis considering an ideal MOS junction or the capacitor.

(Refer Slide Time: 22:35)



In terms of the terminal characteristics or the link between the charged conditions in the device and the external applied voltage an ideal MOS capacitor means that when you apply 0 voltages between the metal and the semiconductor there is no charge in the semiconductor. So let us consider the charged conditions in the device. This is the oxide and this is the p-type semiconductor. When V is equal to 0 there is no charge in the device either in the semiconductor or in the metal. So ideal MOS capacitor means no charge in the device for V is equal to 0. Now we will start with the condition as to when the voltage applied is negative. What happens when the voltage is negative? We will start with this condition then we will go to the positive voltage. When voltage is negative basically the power supply is like this so it is V less than 0.

(Refer Slide Time: 33:18)



Let us try to sketch the electron and hole conditions in the device. So this is oxide layer and this is the hole concentration because this is p-type semiconductor the hole concentration is more than the electron concentration and this is electron concentration. This is for V is equal to 0. Now as compared to this when you apply V less than 0 what will happen? There is going to be a negative charge on the gate and correspondingly there is going to be a positive charge here. That is, you are going to have an electric field directed like this. Now how do you create a positive charge here? This is created by raising the hole concentration near the interface.

Now when you apply a voltage even then there is no current in the device therefore the semiconductor is always in equilibrium. This is a very important situation associated with MOS capacitor that distinguishes the MOS capacitor from the PN junction. So this distinguishes the MOS junction from PN junction that even under applied DC bias the semiconductor is in equilibrium because there is no current flow. Therefore p into n should be constant and electron concentration will dip down so that this p into this n is equal to n_i square. Now we are plotting the concentrations on a log scale because otherwise we would not be able to show both holes and electrons.

Now this is one and this is two where two corresponds to V less than 0 and one corresponds to V is equal to 0. So these are the conditions for holes and electrons in the device. Let us understand the shape of the hole concentration and electron concentration in a little bit detail. Now what does this kind of shape mean? It means that there is a tendency for the holes to move from left to right because of diffusion current because concentration is more on the left than on the right. Now this tendency for diffusion rightward is balanced by a drift tendency which is present because of the electric field created by the negative applied voltage. So electric field in this direction tries to oppose this motion because of diffusion.

So, this is drift and this is diffusion. They are exactly in balance so that under steady state conditions you have a shape like this. At each point the drift and diffusion would be in balance. And as you move away from the interface the effect of the field will not be felt and therefore you have a flat concentration. Here p into n should be equal to n_i square because

this semiconductor is under equilibrium. Another explanation can also be given in terms of the balance between drift and diffusion.

Now what we need to do is we need to see how these variations in hole and electron concentrations translate to a charged distribution. So we will sketch the charged distribution below that is the space charge. Now to get this space charge you recognize the fact that what is shown on the log scale if you draw on the linear scale as done in the case of PN junction you will have a situation like this. So this is p, this is linear scale. So it is the hole concentration for voltage less than 0 for condition two.

Now we cannot show the electron concentration here really on a linear scale so this shaded area close to the interface indicates a pile up of positive charge. Now this charge is almost like a sheet charge. A sheet charge is a charge that is very close to the interface and it is almost like a sheet and thickness of this charge is very small. Generally such sheet charges are because of mobile carriers because mobile carriers can all remain attracted to the surface. They can move to the interface and then they can be held there by the electric field. Now, notice also the fact that although on a log scale the variation does not appear to be that much but on a linear scale the variation is rather rapid and restricted very close to the interface. It is understandable because on a log scale this shows several orders of variation.

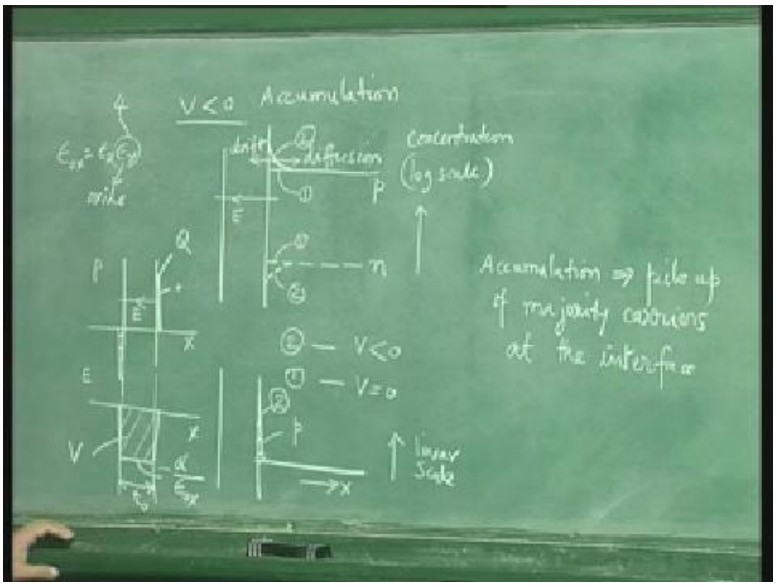
On the linear scale a variation by a factor of 10 appears here and the variation that occurs thereafter really cannot be shown. This point was discussed even in the case of PN junction as the difference between the appearance of a logarithmic plot or a plot on a log scale and same function plotted on a linear scale. Now when you translate this information to a space charge picture it would look like this. So you have a positive charge and an equal negative charge on the gate.

On the gate it is again a sheet charge and occurs because of accumulation of electrons here and this is the direction of the electric field. So if you sketch the electric field that would like something like this. So this is space charge versus distance, this is the X direction. If you sketch the electric field the electric field will be shown on the negative side because it is directed from right to left. And it would be constant in the oxide layer because there is no space charge and in the semiconductor it will fall rapidly so it restricted almost to the interface and similarly in the metal also it falls rapidly. So this is the electric field in the semiconductor.

Now this condition of piling up of holes in the p-type semiconductor at the interface when you apply negative voltage to the gate is termed as accumulation. So the capacitor is said to be accumulated. That is the pile up of the majority carriers at the interface is what is meant by accumulation. So this condition is referred to as accumulation condition. Now accumulation means pile up of majority of carriers at the interface. Now you can easily relate the applied voltage to the amount of charge here because our interest is in relating the charge in the capacitor to the DC voltage that is applied. This will be the first step in our analysis.

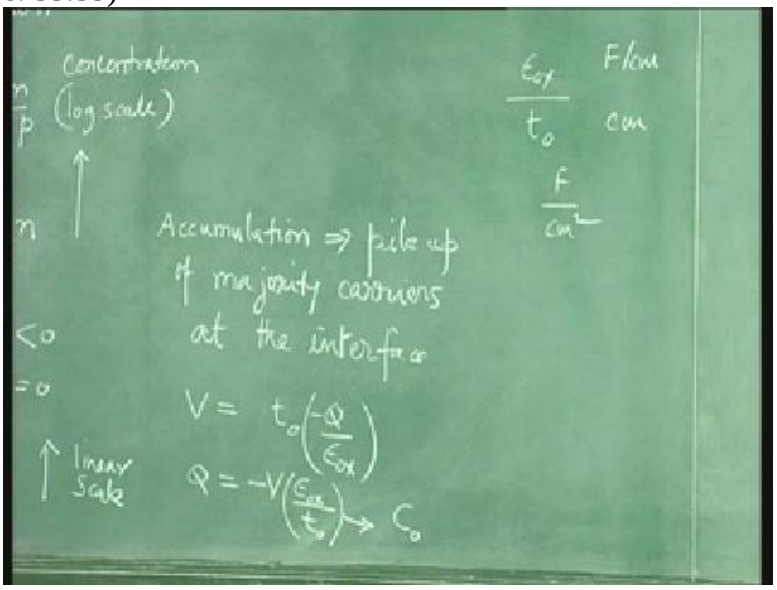
So, if this area under the curve is the charge so this charge is Q and the voltage is given by the area under the electric field distribution so this is V. We can therefore relate the charge to V because by Gauss's law the electric field or this value can be related to the charge as minus Q by E_{ox} that is the permittivity of the oxide. This permittivity E_{ox} is equal to E_0 into E_r of the oxide, the silicon dioxide.

(Refer Slide Time: 34:16)



Generally the value of this is about 4 if the oxide quality is good. We can now relate the applied voltage to the charge in the semiconductor using this diagram. Since the area is V, clearly V is equal to minus Q by E_{ox} into this distance that is the oxide thickness.

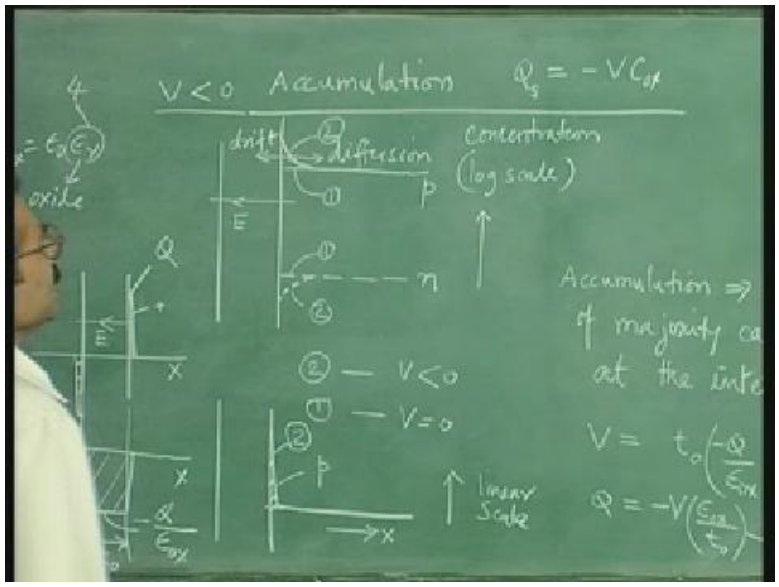
(Refer Slide Time: 35:35)



So this is nothing but t_0. So we write V is equal to t_0 (minus Q by E_{ox}) with a negative sign which you can translate to Q is equal to V into E_{ox} by t_0. Now this quantity E_{ox} by t_0 has dimensions of capacitance. In fact, strictly speaking, it is dimensions of capacitance per unit area. You can see this as follows: The unit for E_{ox} is F by cm and the unit of t_0 is cm so this is F by cm square.

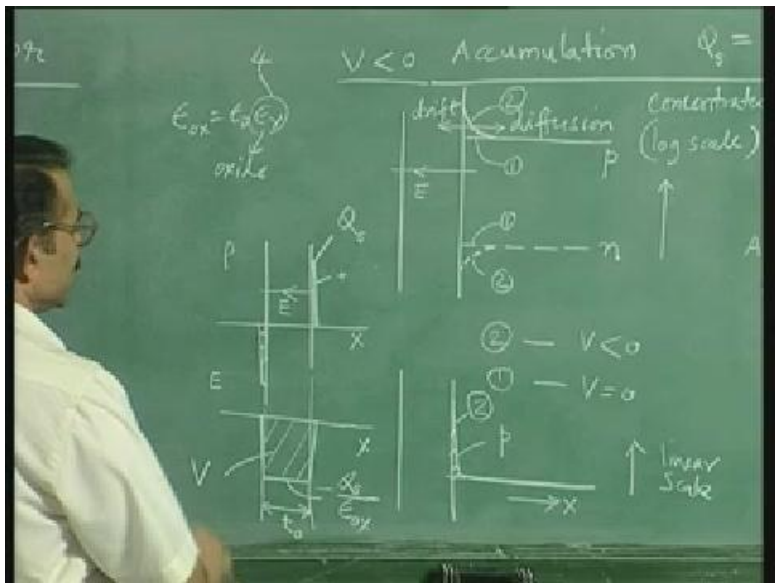
In other words, it is this capacitance per unit area. So normally this term is given the symbol C_0 because this represents the capacitance of a parallel plate capacitor in which there are two metal plates separated by the oxide layer of thickness t_0.

(Refer Slide Time: 36:24)



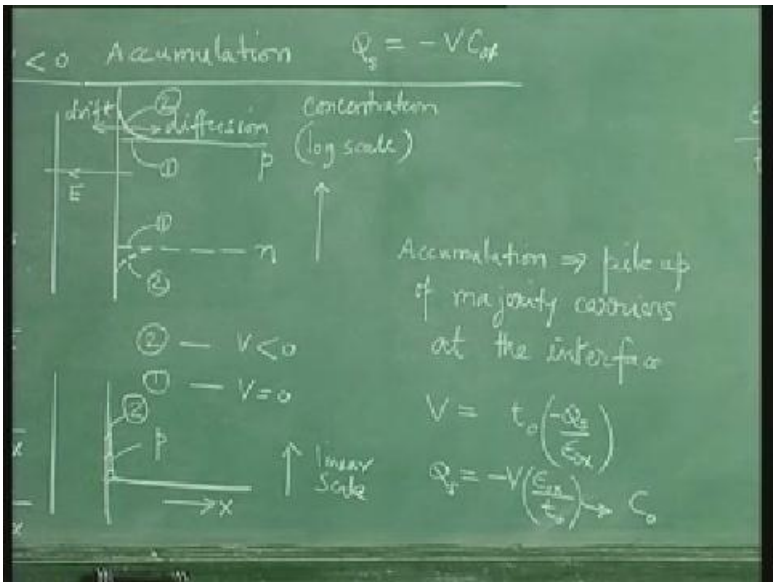
So we can therefore write the equation for accumulation as: The charge per unit area in the capacitor is Q is equal to minus V into C_{ox}. Please note that the voltage itself is negative for accumulation so the charge in the semiconductor is positive. Always we are writing an equation for the charge in the semiconductor. What we can do is we can put a suffix s to indicate that we are referring the charge in the semiconductor.

(Refer Slide Time: 36:31)



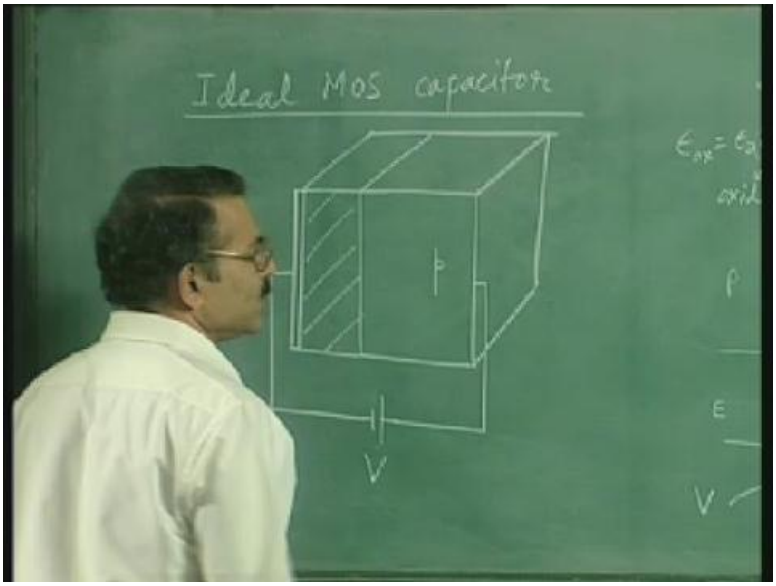
So accordingly we modify our diagram here and we will put a_s here and a_s here. Similarly we put suffix as s in this equation.

(Refer Slide Time: 36:38)



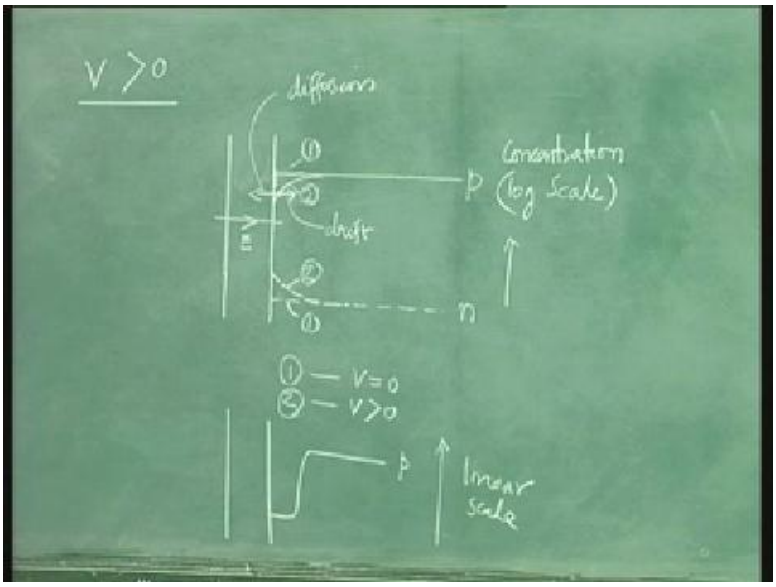
Now we must clarify what is meant by the unit area. When you look at this capacitor what is the unit area that we are talking about. Please note that the area we are talking about is, if you take the 3D picture this is how it will look.

(Refer Slide Time: 37:40)



Now this is the area we are talking about. So, per unit area means this particular area which is perpendicular to the board is of unit dimension so may be 1 cm square or maybe some other unit but that is the area. Capacitance is charged by unit area, if you take a unit area of this kind. Therefore this completes our discussion on the accumulation region of how much is the charge. MOS junction or capacitance which occurs for gate voltage is less than 0. Now let us move to gate voltage greater than 0 and see the conditions there.

(Refer Slide Time: 43:37)



Now we are discussing V greater than 0 this means here the power supply has this polarity. Now let us draw the distribution of holes and electrons for this condition as we have done earlier. So this is the oxide layer and now for V is equal to 0 we know this is our condition, holes and these are the electrons on a log scale. So always condition number 1 is V is equal to 0. Now starting with this when you apply voltage to the gate more than 0 you are going to create an electric field in this direction. Now this electric field obviously needs a negative charge here and a positive charge here. Positive charge in a metal is created by removing of electrons.

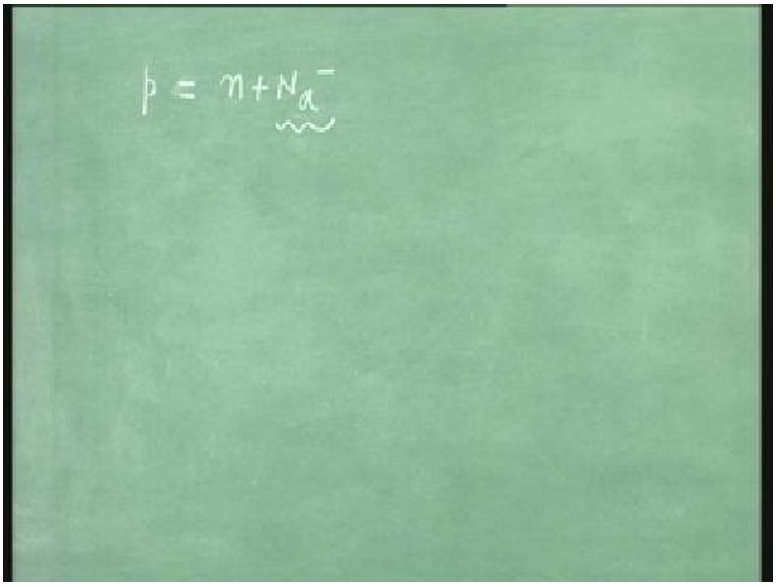
Now how do you create a negative charge in the semiconductor?
Clearly negative charge in the semiconductor will be created by attracting electrons or repelling the holes. This electric field will tend to repel the holes and attract the electrons. Therefore what happens is the hole distribution dips down while electron distribution goes up when you apply voltage greater than 0. As we have discussed earlier even under applied bias the semiconductor is under equilibrium therefore the p into n is always equal to n_i square so this is the concentration on a log scale. Now, again arguing, as we have done for accumulation case this electric field drives the holes away so this is the drift current in this direction. And this drift current is balanced by a diffusion tendency in this direction because holes or more on this side than on the left so they tend to diffuse so there is a diffusion tendency here.

Now we can translate this information onto a linear scale. First let us write, 2 means V greater than 0. When you translate this information to a linear scale it will look like this. The hole concentration falls rapidly so by the time you reach this point it has already fallen by a factor of 10, this is a log scale, just like what we did in a PN junction. Here you have a depletion of holes so in the depletion region if you recall the majority carrier concentration falls like this. This is a fall by a factor of 10 over a small region of width as shown here. And then beyond that the fraction of this concentration is very small.

Again we are not able to show the electron concentration on the linear scale because even though the electron concentration is raised still this maximum electron concentration is orders

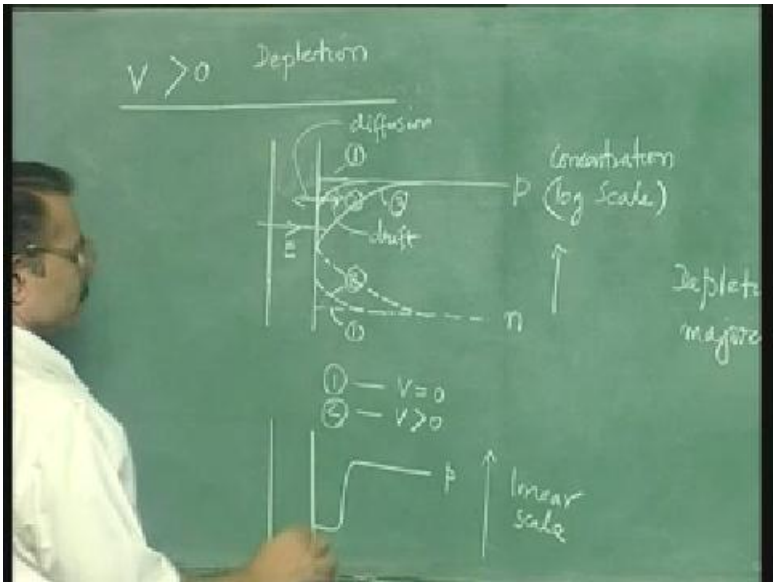
of magnitude lower than this hole concentration. Therefore as a result what you find is the negative charge that is required is created in the semiconductor by a depletion layer. So how does depletion of holes create negative charge?

(Refer Slide Time: 44:22)



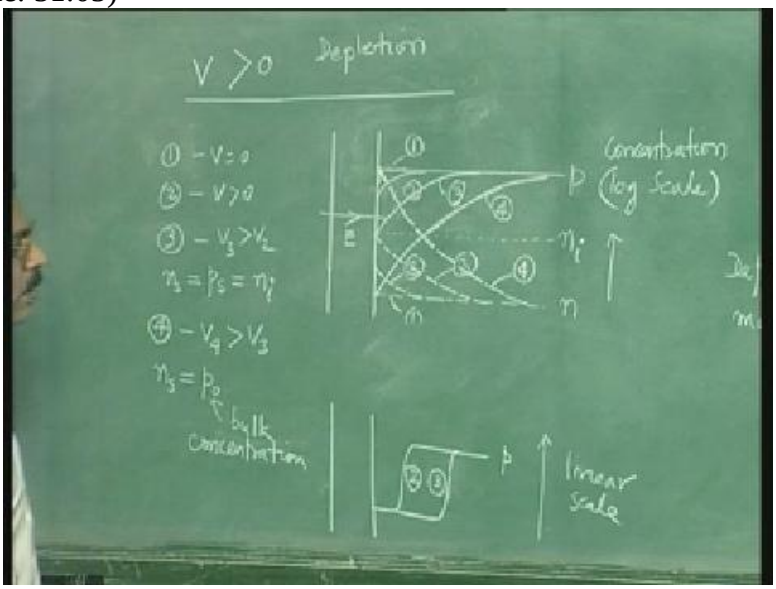
In a p-type semiconductor in which there is no charge the charge balance equation is given by p plus hole concentration that is positive charge is equal to the negative charge that is n plus N_a to the power minus where N_a to the power minus represents ionized impurity concentration. Now this ionized impurity concentration is responsible for negative charges in an acceptor type semiconductor or acceptor type doping, so that is what is happening here. When the holes are removed the ionized acceptors are exposed and they contribute to the negative charge. That is why when voltage is greater than 0 and it is not very large the condition in the capacitor is referred to as depletion.

(Refer Slide Time: 46:16)



Depletion means removal of majority carriers. So, in accumulation it is the pile up of majority carriers but here it is the removal of majority carriers. Now, as you go on increasing the voltage further what is going to happen? If you increase the voltage further obviously more of holes are going to be depleted and electrons are going to pile up. Now the voltage may be sufficiently large so that the hole and electron concentrations at the interface become equal. So it can be something like this. This is condition 3.

(Refer Slide Time: 51:05)



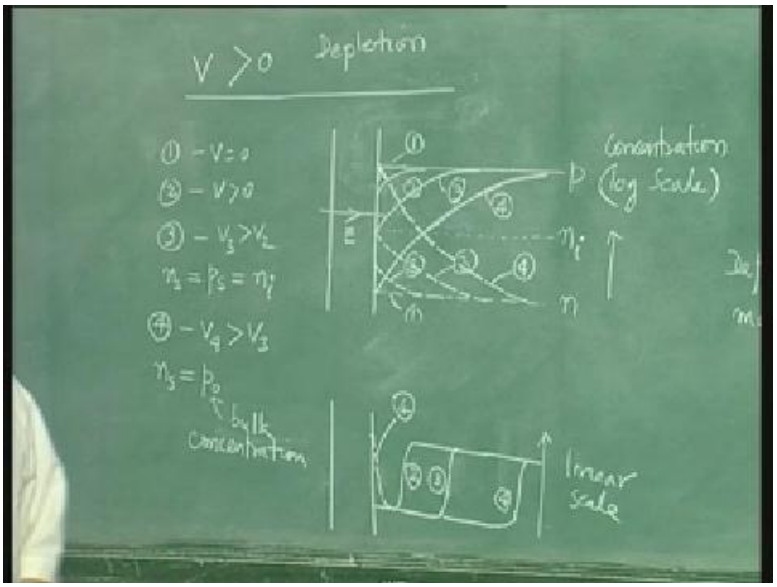
Let us remove these arrows so that it looks simple. Now condition 3 corresponds to voltage when the surface concentration of holes and electrons are the same. So let us write this here, we will list all the conditions 1, 2 and 3 on the left hand side. So 1 is voltage equal to 0, 2 is voltage greater than 0 and 3 is a higher voltage so V_3 is greater than V_2 and voltage is such that n_s is equal to p_s that is the surface concentration of electrons is equal to the surface concentration of holes.

Surface concentration is the concentration at the interface. The terms interface and surface are used interchangeably in this semiconductor literature. Obviously when n_s is equal to p_s since there is equilibrium condition in the semiconductor both will be equal to n_i so this is the condition. Now for this condition if you want to draw the concentrations on the linear scale how will it look?
You will have the hole concentration coming down much earlier so this is 2 and this is 3. So hole concentration starts here and then it falls on the linear scale. Again for condition 3 we are not able to show the electron connection because you know n_i in silicon is of the order of 10 to the power 10 whereas this doping would be more than 10 to the power 14. So there are four orders of magnitude difference from here to here. So we really cannot show the electron concentration scale though it is piling up. Now you can go to even higher voltage and at some voltage what will happen is this electron concentration will become equal to the hole concentration in the bulk. So let us draw that condition.

This is 4 and corresponding to this 4 electron concentration the hole concentration would be something like this. This is 4 and now this means that n_s is equal to p in the bulk. Condition 4 is, the voltage for this condition is more than the voltage for condition 3 and the condition on concentrations inside is that n_s is equal to p_0 where p_0 refers to the hole concentration in the

bulk. Bulk means this region inside that is here away from the interface where the concentration is decided solely by doping and there is no electric field. So this p_0 is the bulk concentration. Now how do we draw the hole and electron concentrations on a linear scale for this case?

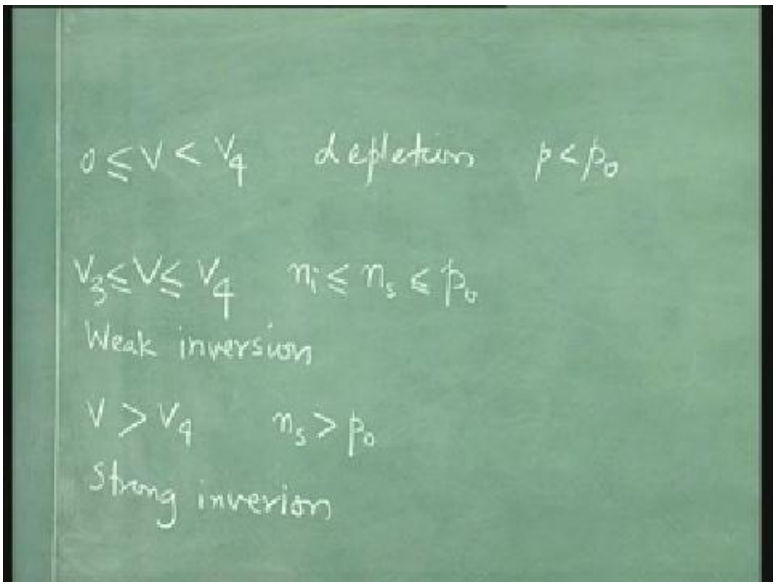
(Refer Slide Time: 52:20)



So moving this point away hole concentration now will fall much earlier than this that is here so this is your hole concentration. But now, since this electron concentration here is equal to the hole concentration inside you can show the electron concentration also. When you plot the electron concentration on a linear scale in a small region it would have fallen by a factor of 10. So here this electron concentration will look like this for condition 4.

The hole concentration is like this showing depletion of holes and there is also an electron pile up here which is now becoming noticeable for condition 4. Therefore beyond this particular voltage when the electron concentration at the surface becomes equal to the hole concentration in the bulk the surface is said to have got inverted and this is called inversion region. That is, in a p-type semiconductor you are able to see at the interface distinctly n-type behaviour because here near the interface the electrons are more than holes. Strictly speaking inversion starts occurring beyond this condition that is V_3 because beyond this condition your surface concentration of electrons will be more than surface concentration of holes. However, since you are not able to see much of an electron charge at the interface normally one talks about inversion occurring only beyond the voltage 4.

(Refer Slide Time: 57:02)

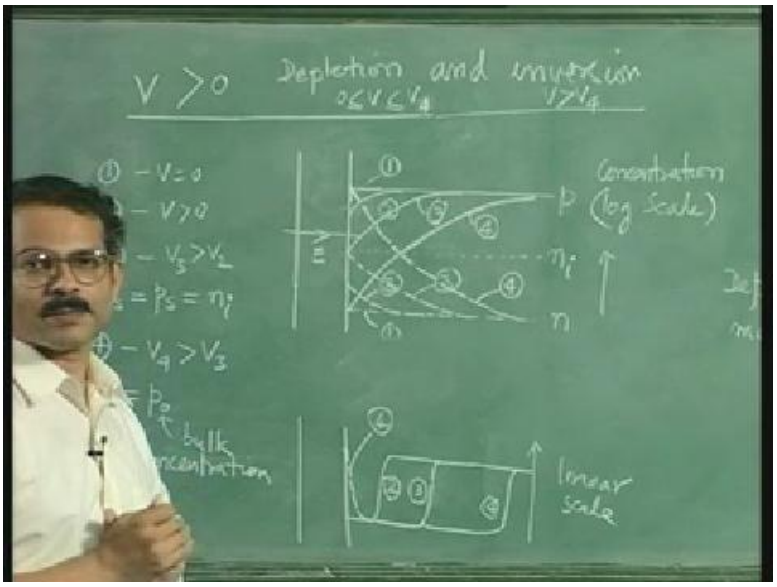


So another nomenclature for the inversion region is that for voltage greater than V_3 but less than V_4 when the electron concentration at the surface is more than n_i or less than p_0 this region is referred to as weak inversion. And the voltage beyond V_4 when n_s exceeds p_0 is referred to as strong inversion. And the condition V less than V_4 you also have depletion so V greater than 0 or less than V_4 you also have depletion of holes. So in this case your p is less than p_0.

Now we need to separate the region V greater than 0 into depletion and inversion. So for small values of positive voltage when there is not much of electron accumulation at the interface it is purely a depletion region. And beyond a voltage V_3 when electron concentration at the surface becomes more than the hole concentration it is inversion but it is weak inversion because on a linear scale it is still not appearing which means the amount of charge because of electrons is still small.

Beyond V_4 that is a condition when the electron concentration starts exceeding the hole concentration in the bulk and then on a linear scale it starts appearing near the interface so a significant amount of electron charge starts appearing and therefore we talk about it as strong inversion. So, in the first course on devices we normally do not treat the weak inversion conditions in detail. Therefore we separate the MOS capacitor regions of operation into accumulation, depletion and inversion. So, this weak inversion region will be assumed to be a part of the depletion region. So you have V greater than 0 depletion and inversion.

(Refer Slide Time: 57:39)



So, the depletion is for 0 lesser than or equal to V and lesser than or equal to V_4 and inversion is for V greater than V_4. Now does this mean that there is no depletion beyond V greater than V_4? No, the depletion region continues to exist but what has happened is the inversion layer has started appearing. Again I want to emphasize why it is called inversion because the electrons are becoming majority carriers near the surface in a p-type semiconductor. So it almost like saying the nature of the p-type semiconductor at the interface has got reversed.

Now, more about this inversion region and the charge conditions in this region we will see in the next class. So, in this class what we have done is, we have considered the MOS capacitor structure and we have set our goal as explanation of the capacitance voltage characteristics of this structure for different frequencies. And then we started with an ideal MOS capacitor for which we have explained the accumulation, depletion and inversion regions of operation when you apply a DC voltage.