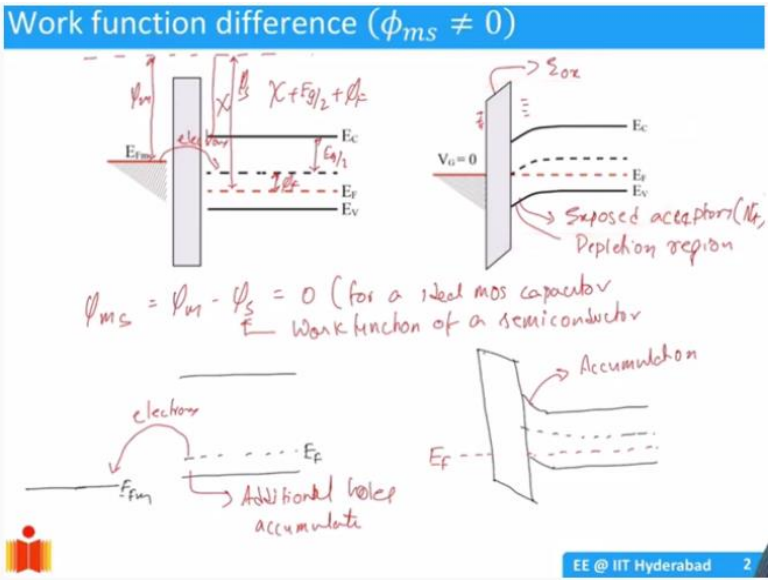
Introduction to Semiconductor Devices
Dr Naresh Kumar Emani
Department of Electrical Engineering
Indian Institute of Technology – Hyderabad

Lecture - 8.1
Non-ideal MOS Capacitor

Hello, everyone. Welcome back to Introduction to Semiconductor Devices. So, in the last week, we were studying how the MOS capacitor works. We studied the electrostatics of a MOS capacitor in quite a bit of detail. And we have also explored you know what happens you know when the depletion charge changes and you know when you change the gate voltage from accumulation to depletion to inversion and so on.

While doing that we have made a series of assumptions right at the beginning. We refer to that as you know the assumptions for an ideal MOS capacitor. It turns out that those assumptions are not strictly valid. They are broken many a times. So, today we will look at 2 of the assumptions that are broken. And they have significant impact on the threshold voltage of MOSFET. So, we will study that today.

(Refer Slide Time: 01:00)



So, to start with, if you remember we have assumed that the metal and semiconductor Fermi levels are at the same point. That means the work functions are same. In other words, now, we define Φ_{ms} as a quantity.. So, Φ_{ms} is a quantity which is the difference of work function in the metal and the semiconductor.

$$\Phi_{ms} = \Phi_m - \Phi_s = 0 \ (for\ a\ ideal\ MOS\ capacitor)$$

But you know in real life, we use metals. Like for example, let us say, we use aluminum. It is got a fixed work function. And what happens to the work function of the semiconductor? Well, it can change. It can depend on the doping density. We know that because the definition of work function is the amount of energy required to take an electron from the E_F to the vacuum level free.

To make it completely free of the semiconductor. So, of course, that depends on the doping density of the semiconductor. For example, if I define this let us say there is a vacuum level.

$$\Phi_s = \chi + {}^{E_g}/_{2} + \Phi_F$$

So, we can calculate. You know the semiconductor's work function would be dependent on doping. And we can actually compute the Φ_{ms}. And it turns out that many a times or most of the cases it is not equal to 0. So, the assumption that the work function difference is the same is strictly not true.

So, what happens if you have a work function difference? So, it turns out that let us say if you have a situation like this where your phi m the work function in the metal is smaller than the work function of the semiconductor. Electrons will be transferred over this barrier into the semiconductor. There is always finite probability of electrons getting transferred over this. At equilibrium, what happens is there is a probability this way.

And then there is a probability this way. They will balance out. So, there is no net transfer. But right now, in the beginning, you know when you have E_{Fm} of the metal substantially higher than semiconductor. There is a positive. There is a nonzero probability of electron transfer this way. So, when electron is transferred into the p-type semiconductor, it recombines with a hole. And that results in reduction in the number of holes.

So, that is why even if you do not apply any voltage, you will end up getting a depletion. So, here basically there are exposed acceptors. So, the charge is going to be you know Na minus the density the negative charge associated with an acceptor which is ionized. So, you have less of holes. And that is why you have this you know depletion region. You also call it the depletion region.

The holes are reduced in number. So, it is depleted. So, without any external voltage, we are having the semiconductor to be depleted. That is one of the consequences of having a non-zero work function difference. So, because and also you see that you know there is a electric field in the oxide which is constant. Why does that happen?

Because well, you have negative charge here and there will be a positive charge on this side automatically because electron is removed. So, on overall, there is going to be electric field across oxide as well across the semiconductor. It is like a depletion. So, this is what happens when there is a work function difference. You can have another scenario. For example, you can have a scenario where you have a, this semiconductor E_F is somewhere here.

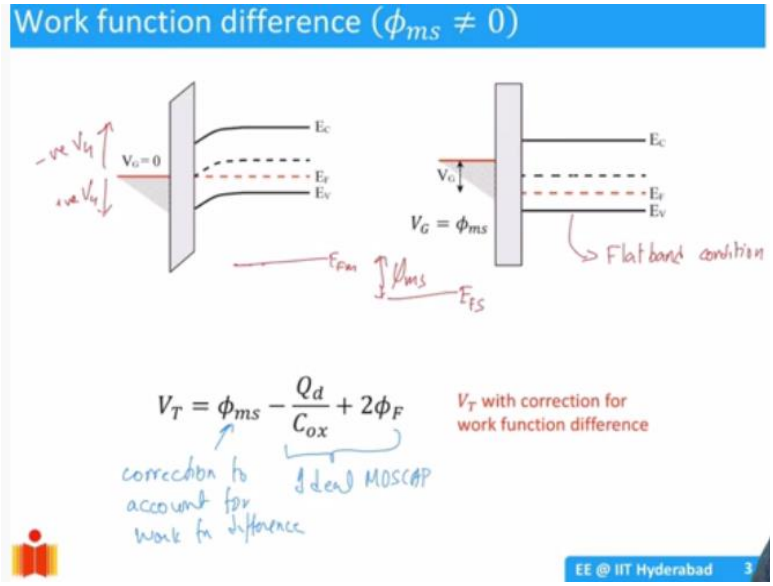
And then you have a metal which is having E_F here. This is also possible because we saw that you know for if you take metals like platinum they are going to have reasonably high work function. Aluminum is only 4.35 or so, but platinum I think is much higher than that. So, in that scenario, you will have electron transfer in this fashion. In the earlier case, electron transfer was this way.

Electrons always move from regions where the Fermi energy is higher to the regions where Fermi energy is lower. When you have electrons moving this way, you will see that holes are getting created here. You know additional holes accumulate here because this is only at this surface this effect is valid. So, because of that, the band diagram you could draw. And it might look something like this.

So, E_F is still uniform because there is no applied gate voltage. But the semiconductors already accumulated. So, any, I know, either of this scenarios can happen or it might happen that we have metal work metal and semiconductor work function difference is very large. Then it really go to even inversion sometimes. So, this is an important aspect that we need to study. So, how do we account for this?

So, let us say you have this sort of a depleted semiconductor. What can you do to make it go back to the original state? So, to do that, we can essentially apply.

(Refer Slide Time: 08:02)



Like for example, I just we have read on here this is the depleted state. And we have said that whenever you apply positive gate voltage to the metal, it will move down move down positive V_G. If I apply negative V_G, it moves up, the Fermi level of the metal moves up. So, now, what we can do is apply a negative voltage such that you pull the metal Fermi level up. So, that will force the bands to become again it come into the flat band condition.

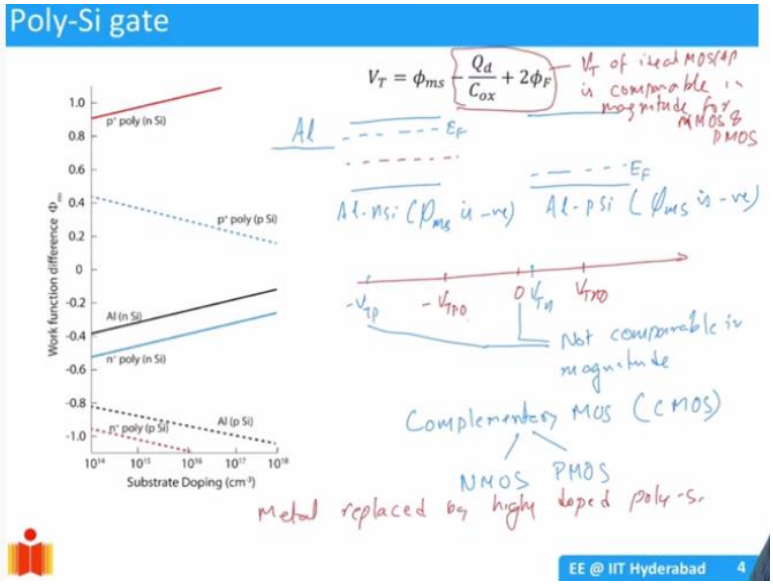
So, this is basically it is called as flat band condition. I mean it is not very complicated. So, bands are flat. That is it, flat band. So, essentially, this was what we wanted initially because this is what we studied. We, in a ideal semiconductor, this was happening at no applied gate voltage. But if you have a work function difference then by default I mean there has to be certain applied voltage to go back to the flat band condition.

So, how much is this voltage? Well, the amount of voltage is going to be you know you have this E_{Fm}. And then this is going to be E_F of semiconductor. So, the difference between them is going to be Φ_{ms}. So, if you apply Φ_{ms} of voltage, then you can actually take the MOS capacitor back into flat band condition. So, that is why we have considered this in the last week where we showed that this is a threshold voltage for an ideal MOS capacitor.

Why is it this threshold voltage? Well, I mean we have derived a relation between V_G and phi s. And then we said that this is when the surface potential is equal to 2 times phi F. Then it is going to be threshold voltage. So, that is why this is the ideal MOS capacitance threshold voltage. But because of this difference, Φ_{ms}, there is a correction to the threshold voltage correction to account for work function difference. This is going to happen.

So, we have the new V_T. So, it does not look so bad. So, all we have to do is compute, what is work function difference metal semiconductor work function difference? There is another problem here.

(Refer Slide Time: 10:42)



That is you know this is essentially you know what was done in the beginning. You know when the technology was you know in the 60s and 70s, people used metal gates. So, one of the consequences of that was let us say you have an n-type semiconductor and you have a p-type semiconductor. The metal work function is going to be let us assume somewhere here. Aluminum is about 4.35 or so. So, work metal work function is here.

So, for most of the semiconductor dopings and for all the semiconductor dopings, if you look at metal let us say aluminum. So, if you look at aluminum and p-type silicon, the work function is always Φ_{ms} is negative. And in this case of aluminum and n-type semiconductor because you know Ei is going to be somewhere here. Only for large dopings, the E_F may actually you know go towards E_C. And then it the Φ_{ms} might be positive.

But in most of the cases, Φ_{ms} is negative again. So, this is what you can verify when you can try to put in the exact work function of the aluminum. And then compute and see what happens for different doping densities. You will see that it will be like this. So, this is also another you know way of showing that. So, here is a substrate doping on the x axis and work function difference on the y axis.

The various things will come to that. The first thing we want to notice is when you have an aluminum and p-type junction or you know sorry aluminum and aluminum gate and then p-type substrate. Φ_{ms} is turning out to be in the range of minus 0.8 to minus 1 or so. So, it is quite large. And for n-type case, it is negative for most of it. Only when you go to 10^{19} or so, then this become may be positive, a small positive.

So, what does, why is this important to us? Well, one of the important, one of the useful implications of an ideal MOS capacitor is that when you have this quantity here. This is essentially V_T of a ideal MOSCAP. So, V_T of ideal MOSCAP is comparable in magnitude for nMOS and pMOS. Why? I mean take a substrate of similar doping 10^{15}, p type 10^{15}, n type. When you compute the V_T by this it will turn out to be same.

So, there is a certain amount of symmetry in the type of transistors you can make. Whereas, if the moment you put a Φ_{ms}, Φ_{ms} is a large negative voltage, then the V_T will shift. For example, I could draw this axis voltage axis. There is a 0. So, I will have let us say V_{Tn0} and minus V_{Tp0}. So, basically what I am trying to say is this difference of V_T from 0 is going to be symmetric.

Basically, this is going to be similar you know comparable for ideal MOS capacitor because Qd is going to be comparable and phi s is going to be, the surface potential is going to be comparable. So, you are going to have a symmetric V_T. But the moment you add a Φ_{ms} because of let us say metal, the whole V_T shifts down. If you are adding an aluminum gate, we are seeing that for most of the voltages it is basically negative.

For n type, it is going like this. For p type, it is going like this. So, the V_T is going to now become something different. So, V_T might be somewhere here, V_{Tn}. And this might be some minus V_{Tp}. So, n MOS has a very small V_T maybe. And p MOS might have a large negative V_T. So, it is they are not comparable in magnitude. And this has actually significant implications.

We like the V_T to be comparable because you know in our technology today we call it actually you might have heard about it. We call it complementary MOS technology. CMOS, we call it. CMOS technology which essentially has n MOS transistors and p MOS transistors fabricated on a single chip. And we like both of them to be sort of complementary devices. That means the drive current the threshold voltage everything is symmetric.

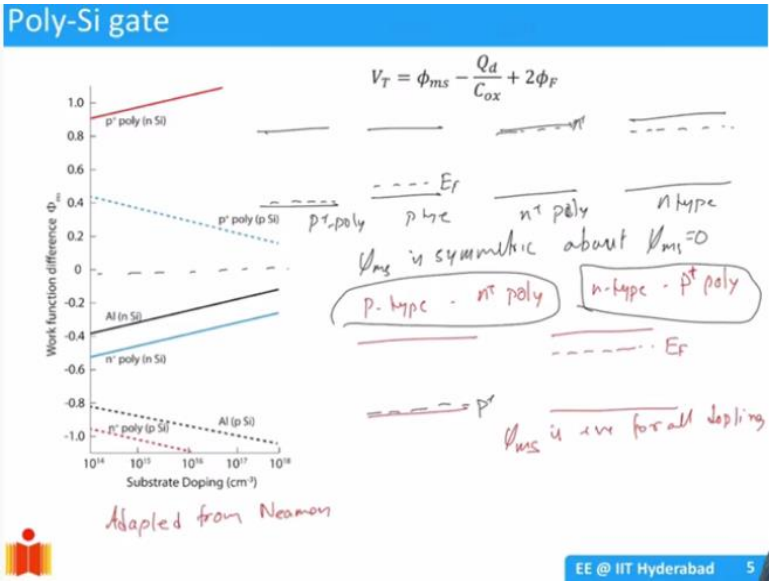
It is easier for us to design circuits with when they are complementary. And there are also some very nice interesting circuit implications for this. So, we like this to be complementary. But if you just have a pure metal. It is not so. So, you might ask you know, why was this done in the beginning? Well, I mean that actually to be honest there is only one type of technology which is nMOS which was available in the beginning.

And the threshold voltages were quite large. So, it did not really matter how much this small difference in the work function was. But over time, the supply voltages you know we use what is called as scaling of voltages. So, in the 1960s and 70s and maybe even 80s, the transistor the chips used to work at 3.3 volt technology and even 5 volts in the beginning and then 3.3 volts.

But with time, we started reducing the supply voltages. So, then the electronic chip started working at 1.8, so, 2.5, 1.8, 1.1 and so on. So, the supply voltages were gradually reduced from the 70s and 80s to present day. The reason this happened was it has some benefits. We will talk about it when we discuss the scaling of MOSFETs. But as you go to lower supply voltages, it is, it became important to realize symmetric sort of a work function.

And the way this was achieved is by replacing metal with what is known as polysilicon. So, metal replaced by highly doped poly highly doped polysilicon. So, this has some benefits. You know we will discuss that in a moment. So, because it is highly doped it like behaves like a pseudometer and it can be used.

(Refer Slide Time: 17:46)



So, how does polysilicon work? So, what we will do in this case is let us say you have the same you know p-type substrate. There is an E_F. What if I use a p plus doping for my gate p plus poly in p type? Or, I could use n plus poly I mean n plus means E_F is close to E_C. p plus means E_F is close to E_V. There is a sorry I will take an n type here. n type and then this is n plus poly.

Since it is highly doped , for practical purposes is that E_C or E_V. So, now you will see that the work function difference is quite small, correct. So, because of that if you see here the other curves if you have an n plus poly on n type silicon, it is increasing like this. And if you have p plus poly it is decreasing like this. So, whatever you do if you have a certain amount of symmetry here this if you take as a 0 Φ_{ms} is 0, there is a certain amount of symmetry with the way the Φ_{ms} is changing.

And that has very good I mean properties because you know if you have comparable dopings then your V_T s will be comparable. This was not the case with aluminum you remember. So, you see here if I take my aluminum at this doping it when you increase it, it is you know Φ_{ms} was increasing here. But for the p type case, substrate it was decreasing.

So, it is actually having opposite trend whereas if you use polysilicon gates, there sort of you know have symmetric trains. So, I will say that you know Φ_{ms} is symmetric about Φ_{ms} equal to 0. So, this was useful. And that is why you know slowly by 90s late 80s and early 90s, we started using polysilicon gates instead of metal gates. There were also many other technological issues that we will not get into today.

But it was useful to actually get use this. You can also use the opposite condition. It has some benefits. For example, instead of using if you have a p-type substrate, I will use an n type n plus poly. If you have a n-type substrate, I will use a p plus poly. I can use that also. That combination also I can use. So, if you use, what happens? If I use p-type substrate and n plus poly or n-type substrate and p plus poly, use this combination.

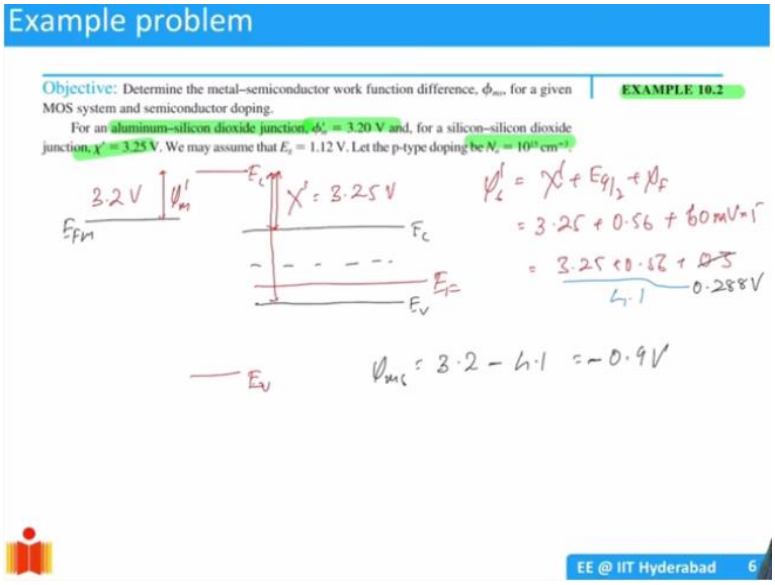
If you do, what happens is you will see that again the difference is almost like close to band gap. That is why you see here the Φ_{ms} if you have a p plus poly that means you know you are using poly to be p plus. And you are having an n-type semiconductor E_F. So, Φ_{ms} is basically positive for all doping actually. That is why you see that it is sort of going in this way. As you increase doping, Φ_{ms} is increasing.

Eventually, it will come to 1.1 eV. That is where the band gap is. Similarly, if you use the opposite case n plus and p-type silicon you will see that this is increasing. So, these numbers are strictly you know this graph is taken from adapted from one of the references. I should have mentioned it. The textbook has one graph adapted from I do not remember the chapter. You can look up. I adapted from Neamen.

The only issue is that here these are also including additional correction factors. Anyway, for now, this is a good enough graph to explain the behavior of how the polysilicon works. So, once we do this sort of a thing, we have a large Φ_{ms}. But it is useful in certain situations. You know I will let you analyze this by. You know you can actually in a nanoHUB, we will show you that you can use aluminum gate or n plus gate or p plus gate.

I will let you analyze the usefulness of this situation. Wherein you use a opposite type of poly you know. Like, for example, p type and n plus poly, what is the advantage of this? You know it looks like Φ_{ms} is either large negative or large positive. So, why is it useful? I will let you analyze this. Please run it. And we will discuss at some point down the line. So, this is the theory part of it. Let us try to solve one problem.

(Refer Slide Time: 22:59)



So, the problem given to you is that determine the metal semiconductor work function difference Φ_{ms} for a given MOS system and semiconductor doping. So, the aluminum silicon dioxide junction is given to you with a phi m. This is s and asked to calculate the Φ_{ms}. Fine. So, the textbook uses a slightly different notation that you know if you have a, this is what type of gate aluminum. So, aluminum let us say this is here E_F. And then you have let us say E_C, E_V. So, when we say that it is phi m, so, basically there is an oxide in between. I am so there is an oxide here right in between. So, let us assume this is your oxide E_C and this is your oxide E_V.

The way textbook defines phi m prime. Phi m prime is this quantity. The difference of metal work function with the E_C not the vacuum level, but E_C of the oxide. Similarly, chi prime it is defining as this. Remember, original chi was the distance of E_C from vacuum level. But now we are defining chi prime as something from ECH of oxide. So, this is given to us 3.25 volts. And this is given to us 3.2 volts.

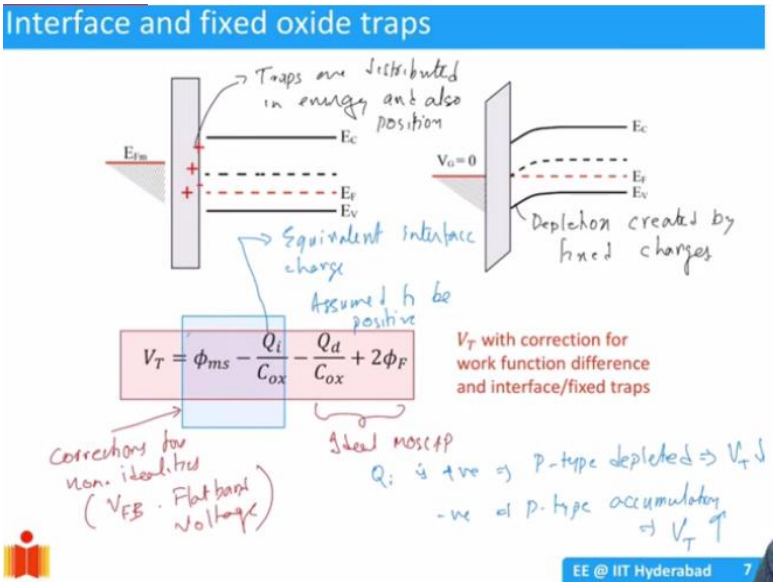
So, that is why I have drawn E_F to E_C to be slightly lower than E_{Fm} here. So, now, what is the work function difference, Φ_{ms}? Well, for that, you need to find, what is this work function of the semiconductor? Work function of semiconductor will simply be we have the Ei. And it is given to be a p-type semiconductor Na doping. This is Na of 10^{15}. So, there is going to be some I should do with maybe red. This is my E_F.

$$\Phi'_s = \chi' + {}^{E_g}/_2 + \Phi_F = 3.25 + 0.56 + 0.288 = 4.1V$$

$$\Phi_{ms} = 3.2 - 4.1 = -0.9$$

So, this is how you can calculate the work Φ_{ms} work function difference. And that will modify your V_T.

(Refer Slide Time: 27:53)



So, this is one form of non-ideality. There is another form of non-ideality that we would like to study which is basically the formation of traps in the oxide. So, in the assumptions for an ideal MOS capacitor we said that there are no traps in the oxide and the interface between silicon and silicon dioxide is perfect. You know there are no traps there as well. But again this is not a very accurate assumption.

There are certain limits to it because always when you form an oxide in any process there is going to be some traps. And even in interface, there are going to be some dangling bonds that we discussed sometime back. So, these traps basically I can represent them as you know some positive or negative charges. So, basically traps are distributed in energy and also position, energy and position.

They have some distribution. They can be in different locations. For example, there could be some distribution of trap plus minus. They could to be on a interface at whatever energies. So, this is going to be there. So, if you have some traps, they are going to hold some charges. You know if they are fixed traps, they are going to hold the charge permanently. So, how will that impact the V_T?

Suppose, if you place a positive charge let us say somewhere in the semiconductor here. That positive charge will have a field which will drive the holes away. If you are having a p-type

semiconductor that they it will drive the holes away. And without any applied gate voltage, you might have a scenario where there is depletion. This depletion created due to, created by fixed charges or interface traps.

You know fixed any you know it can be fixed trap oxide traps or you know oxide means in the bulk of the oxide, interface will, right at the interface between silicon and silicon dioxide. So, you can have either variety of these things. So, when you have such variety of charges then you will have depletion or sometimes if you have a negative charge you can have accumulation. That is also possible.

So, it becomes a very tricky thing to analyze it you know in detail what exactly is happening. So, one of the, we can make an assumption to account for this. So, we will define what is known as an equivalent interface charge. It is you know it is assumed to be at the interface between silicon and silicon dioxide. Let whatever be the distribution of the charges in space, we will assume that all of it is at the junction. And there is an equivalent charge.

We can try to compute. So, that it has an same effect on V_T. So, this equivalent charge I mean there are experimental techniques by which you can determine how much is this. But we do not need to know that. For us, for our practical purposes, we will say that there is this interface charge. And it is assumed to be positive right now. If your equivalent charge is given as Qi you know that is the interface charge.

If it is given as minus let us say some charge coulombs then we have to put that minus. So, right now, in this expression, this is positive here. So, once you have it, your V_T will be modified. How does it get modified? Well, the expression is this. We have the ideal MOS capacitor. And then there is an additional correction because of this. So, this set here is basically corrections for the non-idealities.

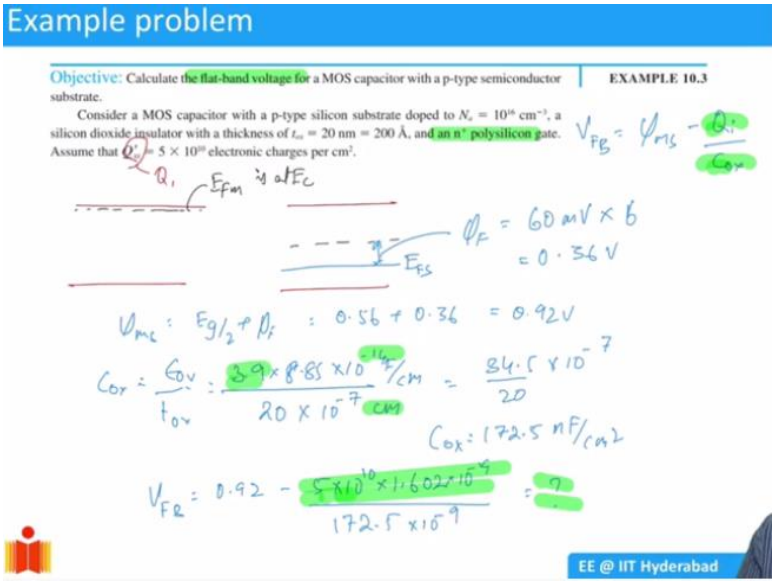
So, this is also called as sometimes work V_{FB} flat band voltage. So, flat band voltage, if you apply this to the gate then your bands will be flat. And the rest of it is ideal MOSCAP. So, now, let us just analyze quickly. Let us say if your inter Qi equivalent interface charge is positive, if your Qi is positive what should happen implies p-type semiconductor is depleted. If you have a p-type semiconductor it will drive the holes away p type depleted.

If your p type is already depleted, it is easier to create inversions. So, V_T is lowered because of this. Implies V_T reduces. That is what is happening because Qi is positive. So, overall V_T will reduce. Formula is telling you and physically also once you deplete it. Otherwise, the gate had to deplete the semiconductor and then produce the inversion. But now, already some part of the, you know job is done by the fixed charge. So, already it is been depleted.

So, you can quickly go to inversion. Or, if the Qi is negative, p type will have to have accumulation. So, now, implies this has the overall V_T has to increase because you have to push the accumulation charge away and then deplete it and then invert it. So, the V_T increases. The magnitude of V_T increases. And you can analyze the same thing for the n MOS transistor as well. We can find out what that is.

So, I will leave it to you that leave that as an exercise. So, this is how your traps you know behave. So, let us quickly try to solve a couple of more problems and with that we can you know conclude our discussion on the non-ideality. So far you know we will only discuss these 2. Later on, we will just mention the gate oxide leakage. So, let us take one more problem.

(Refer Slide Time: 33:42)



Wherein we want to find out what is the flat band voltage for a MOS capacitor which is having you know p-type substrate, let us say. So, I will not solve this problem. I will quickly tell you how to do this. So, you are asked to calculate the flat band voltage.

$$V_{FB} = \Phi_{ms} - \frac{Q}{C_{ox}}$$

Well, you will have to live with my you know definitions Qss prime is basically Qi in my definition. So, it is a, you have an n plus polysilicon gate.

And then you have a p-type substrate. So, n plus polysilicon gate. So, I will say that the E_F here is at E_C. E_{Fm} is at E_C because it is highly doped. I will assume that. What about the p type? Well, you have this Ei. And then there is this EFm. So, how much is that? For this particular doping, I need to find out, what is this distance? This is phi F. That is going to be 60 millivolts into 6.

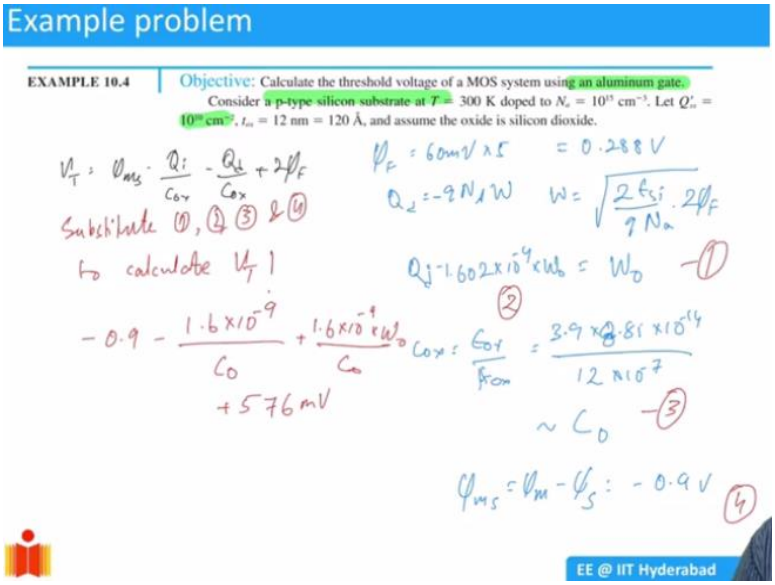
$$\Phi_F = 0.36V$$

$$\Phi_{ms} = {}^{E_g}/_{2} + \Phi_F = 0.56 + 0.36 = 0.92V$$

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{3.9 * 8.85 * 10^{-14}}{20 * 10^{-7}} = 172.5 \frac{nF}{cm^2}$$

$$V_{FB} = 0.92 - \frac{5 * 10^{10} * 1.602 * 10^{-19}}{172.5 * 10^{-9}}$$

(Refer Slide Time: 38:35)



And, just one last example, what is the threshold voltage? Now, with aluminum gate p-type substrate, this all of it is involved here, fine, good.

$$V_T = \Phi_{ms} - \frac{Q_i}{C_{ox}} - \frac{Q_d}{C_{ox}} + 2\Phi_F$$

$$\Phi_F = 0.288V$$

$$Q_d = -qN_AW_0 \ , \qquad W_0 = \sqrt{\frac{2\epsilon_{si}}{qN_a} 2\Phi_F}$$

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$$

$$\Phi_{ms} = \Phi_m - \Phi_s$$

And then discuss various dependencies of the CV. And with that the MOS capacitor discussion will be complete. So, we will do that in the next one, one and a half hour. Then, we will have a brief demo on the nanoHUB. Thank you so much for your attention, good luck.