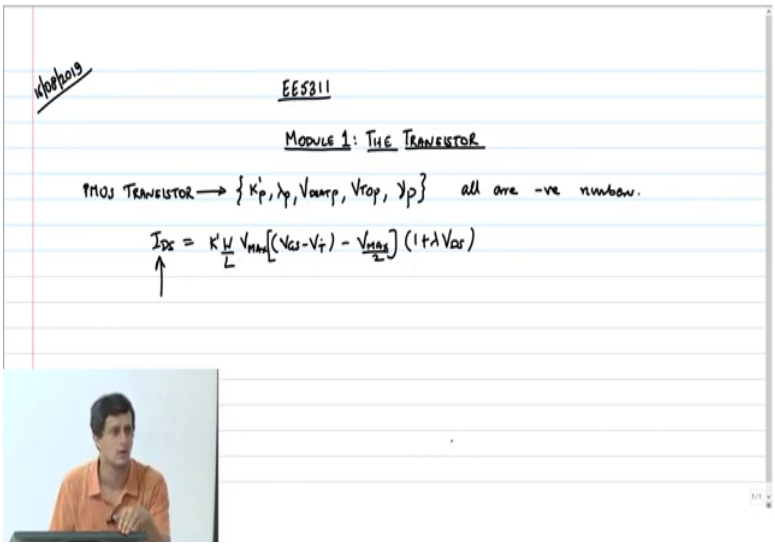
Digital IC Design
Prof. Janakiraman Viraraghavan
Department of Electrical Engineering
Indian Institute of Technology, Madras

Lecture - 12
Introduction

So, good morning everyone. So, in the last class we discussed about the PMOS transistor, we concluded some of the short channel effects first right, substrate leakage and gate leakage and all that, then we discussed about the PMOS transistor.

(Refer Slide Time: 00:37)



And the key conclusion, right for PMOS transistor is K p prime and then what is the other parameter; lambda p prime, V DSAT p prime no prime there, there is no prime here either

right and V T naught p and gamma p. These are the five parameters for the level one spice model, all are negative numbers, ok.

So, we discussed in detail why each of these parameters should be a negative number, some of them are just because of the convention. I am looking at I DS, and remember we said that I DS was K prime W by L into V MAX into V GS minus V T minus V MAX by 2 into 1 plus lambda V DS.

So, we were looking at current flowing from drain to source and because PMOS current flows from source to drain which is in the same direction as that of the charge carriers because it is now because of holes, drift is because of holes and not because of electrons. I DS will be a negative number, it is just a sign convention and therefore, K p prime has to be negative then V D SAT p.

Of course, because now the potential has to be from source to drain higher potential has to be from source to drain by definition and only then current can flow from source to drain because current always has to flow from higher potential to lower potential no change in that. So, therefore, the voltage is become negative because of that right, then the lambda p is again because V DS is the negative number, right for current to flow, lambda also has to be negative only then the magnitude of the current can increase, right.

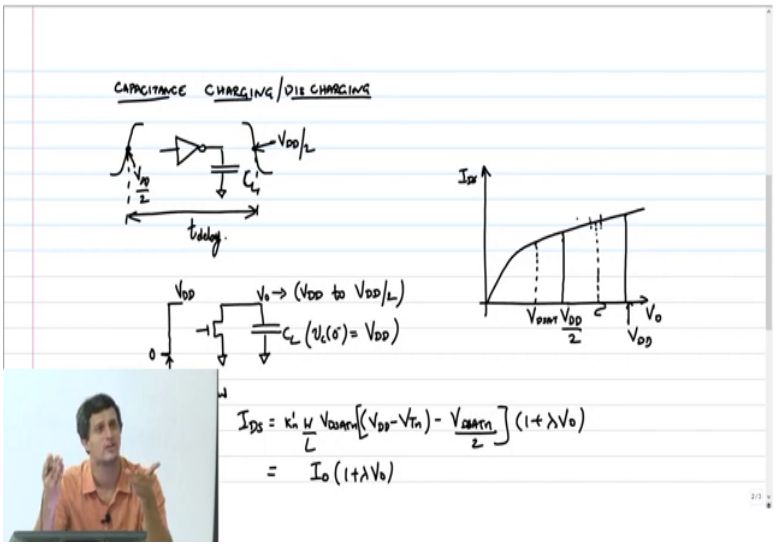
Then similarly the threshold voltage has to be a negative number and even gamma p. For gamma p remember we had to take care of the modular sign; mod of psi S minus psi it will be minus 0.25 plus V SB kind of thing that we got. If you have to take care of that and once you are done with that you will conclude that gamma p also has to be a negative number, ok.

This is what we discussed with respect to PMOS devices and of course, the all the plots will now sit in the third quadrant because everything is going to be a negative number voltages are negative currents are also negative. So, all the plots will be in the third quadrant, right. And this is a topic I started last time and there were a lot of questions. So, let me go through this

resistance thing once more, ok. So, that there are no doubts after the class I had a lot of questions, ok.

So, the idea is we have derived equations for the currents of the transistor, we know how the transistor behaves with various voltages and stuff like that that is fine, but how are we going to use this transistor. Invariably, we are going to only use this transistor either to charge a capacitor or discharge the capacitor. In this entire course is all about that how fast you can charge the capacitor, how fast you can discharge the capacitor, right. This capacitance can be because of it is own parasitic capacitance like I will discuss today or it can be because of a load capacitance, right.

(Refer Slide Time: 04:29)



So, I have a capacitance discharging. So, if I have an a logic inverter driving a load capacitance, right. This inverter is going to either charge this capacitor or discharge this capacitor, and the delay of the gate is determined by how fast this action can happen, ok.

So, I also told you that delay is defined as the time it takes for the input to go to 50 percent, right from there till the time the output reaches 50 percent, ok. So, if this is my V DD by 2 and this also is V DD by 2 when the delay of the gate is defined exactly as this delay ok, this is the delay and we want to estimate this, ok. So, the simplified picture is we have a capacitor C L and an NMOS is going to discharge it. We will see why and NMOS cannot charge the capacitor or a PMOS cannot discharge the capacitor, we look at it later ok, but for now let us just look at this picture the capacitor is charged V C of 0 minus equals V DD.

So, this inverter is going to now discharge the capacitor through an NMOS transistor, ok. So, this voltage V naught is going to fall from V DD to V DD by 2 because of the delay definition. That is why, V DD by 2 is a very sacrosanct number, I am not worried about taking it all the way to 0, ok. Now, we have to make some assumptions to simplify our analysis otherwise, it becomes very difficult. So, if the for example, you the input is going to rise slowly like this then that delay becomes a little hard to handle right, if it becomes a ramp function then I have to see when the delay when the device turns on when it goes into linear, saturation all that it is its very painful to do that.

So, we have to make some approximations to get a first order number and then we will incorporate this input slew also later on. So, the assumption we are going to make is the input rises instantaneously to V DD ok, 0 to V DD instantaneous, ok. So, if at 0 plus the gate is at V DD my V out is also at V DD right, and 0 plus just after I turn it on in what region of operation will my transistor be in saturation region or velocity saturation, right because V G S is equal to V DD. Obviously, V DS has to be greater than V GS minus V T you are removing some quantity from there; therefore, the transistor is in saturation region ok. So, the currents look like this I D and this is V naught.

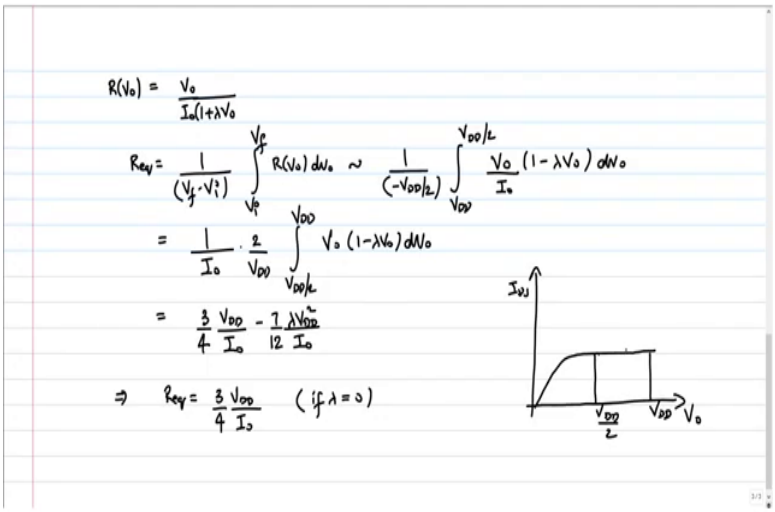
So, right, this is my point where it is V DD, this is V DD and somewhere here is V DD by 2. I am assuming were it is large enough such that it is to the right V DD by 2 is to the right of V DD V D SAT, ok. So, that I do not have to deal with the change in region of operation, ok. So, this is an assumption.

So, what is the equation of the current I DS, let us say it is velocity saturated it is K m prime W by L into V D SAT n into V; what is V GS? V DD right, VDD minus V Tn minus V D SAT n by 2 into 1 plus lambda V naught. Suppose, this is a long channel device then this would be in saturation where this would just be K n prime by 2 W L V GS minus V T VDD minus V T whole square, ok.

So, let me not worry about all that I will just write this as I naught into 1 plus lambda V naught ok, whether it is velocity saturated or not or it is in saturation I naught is independent of V naught, that is the only thing I am worried about, ok. So, now, just look at the picture the V naught is coming down it is being the capacitor is discharging right, V naught is just the voltage across the capacitor and the voltage is coming down. So, if you look at the resistance it now becomes a function of V naught, ok.

So, last class somebody asked me why do not you do a derivative out here. See in analog it is to do a derivative because you fix your bias somewhere here and then you will do some sort of a small signal analysis over that bias point. So, in some sense we have linearized this whole thing here. So, when you do a Taylor series there you want to take a derivative and then you expand it in terms of the input variables. So, therefore, you take derivatives there, but here I am actually discharging it from VDD to VDD by 2 it is such a large range, right. So, I cannot do that small signals analysis of taking a derivative here, I have to find some other way finding the resistance, average resistance now which is a function of V naught, right.

(Refer Slide Time: 11:22)



So, R of V naught is simply what, V naught by I naught into 1 plus lambda V naught. Now, I have a function f of x that is going from a to b I want to find what is the average value, what do you do you just integrate and divide by the x x 2 minus x 1, right that is all I am doing here. So, I am saying that r equivalent is defined as V final minus V initial into V initial to V final of R of V naught D V naught, that is why that is what I am defining as the average resistance, ok.

So, this I will now write as minus VDD by 2, because V f is VDD by 2 V I is VDD and so, I am going to write this as V DD to VDD by 2 into V naught by I naught, right that what 1 by 1 plus lambda V naught I am going to write as a Taylor series because lambda is very small, ok. I can expand this approximately as 1 minus lambda V naught ok, maybe I should put an approximate here d V naught.

Of course now I naught is a constant; so, I will just write into 2 by V DD into V naught into 1 minus lambda V naught d V naught, right. So, what did we get the answer here you solve the integration and you just get what 3 by 4 V DD by I naught minus 7 by 12 lambda into V DD squared by I naught right, this is what we got, fine.

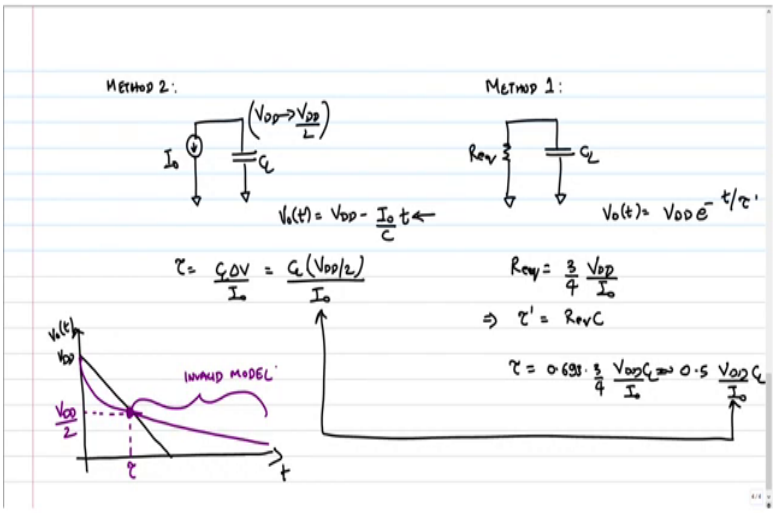
So, if it is going to have channel length modulation then this is the only way you can do it, because there is a variation of resistance all through because of that 1 plus lambda V naught, right. So, now, the interesting thing is if it is a long channel device then lambda is equal to 0, ok. So, implies R equivalent is 3 by 4 V DD by I naught if lambda equal to 0, ok.

So, the point I was trying to make last time is we found an equivalent R right; why are we finding equivalent R because I can now treat this as an R C discharge and I can evaluate delays in terms of the time constant of that R C network, ok. But, if lambda equal to 0 then this is just like a constant current source discharging my capacitor. Why is it a constant current source, because there is no dependence on V naught.

Now, if lambda equal to 0 the current basically just becomes I DS versus V naught. This simply becomes the flat line and this is my region of operation from VDD to VDD by 2, ok. And of course, now I do not have to worry about V D SAT because this will always be in this thing no you have to still worry about it you have to make sure the V T is above some number below some number for it to be in saturation region all through, ok.

So, assume V D is high enough there is no problem. So, this is now like a constant current source discharging the capacitor, ok.

(Refer Slide Time: 15:37)



So, we did the same calculation in method 2. So, it is a constant current source discharging my capacitor, method 1 is you had a thing like this R equivalent into and C L. Now what is the time it is going to take to and what is this this current is I naught, by the way if lambda equal to 0 this current is just I naught. What is the time it is going to take to discharge from VDD to VDD by 2? If it is a constant current source discharge C delta V by I, right.

So, the time is C delta V by I naught C L right, C L into VDD by 2 divided by I naught. Here, the R equivalent was 3 by 4 V DD by I naught which implies that the time constant of this RC network right, let me call it tau prime is R equivalent into C. Now, the delay of this thing to come from VDD to VDD by 2 is 0.693 times the time constant. So, this is 0.693 into 3 by 4 V DD by I naught, right. So, this is approximately again we will get 0.51 or 52 or something like that I naught into C L.

Now, these two answers indeed do matc right. Now; obviously, if lambda is 0 then the constant current source model is the right model because the transistor is a current source, ok. So, if you look at the model that we are using for delay ok, this guy is if you look at this V naught of t it will start from VDD and then go linearly right. What is this C D V by d q equal to I. So, I naught by C into t whereas, this guy if you see V naught of t this V DD e power minus t by tau prime.

So, the two models actually follow different equations with respect to time, but this one is more accurate, ok. So, if you look at V naught of t versus t, you have a straight line like this for method 2. The method 1 is going to look like this asymptotically if you go down to this thing.

So, this is my delay tau and this is VDD by 2. So, the two models give you the same answer at that particular instant of time, any other instant the two models are different, right. The delays are the same question why do you why do you even want to do it to the resistance method 1 is of course, if there is channel length modulation then this current source thing fails.

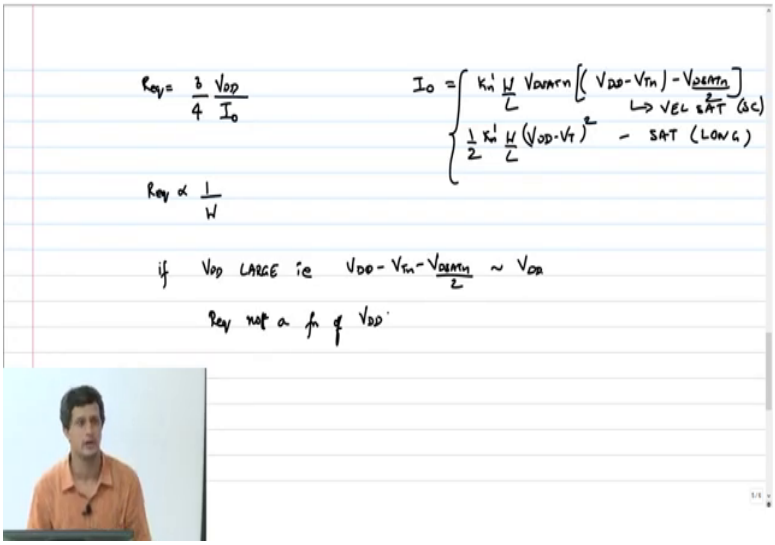
I have to now you can of course, solve this from first principles by saying that I e equal to minus C dv by dt and use the equation of the you know make I dependent on V and solve the differential equation then it is fine; otherwise, I need to do it through this resistance method which is a reasonable approximation because all I am worried about is that delay at when it comes to VDD, this is the only point of interest in this model for me.

So, if you ask me which is better it does really does not matter and I also pointed out that beyond this both the models are actually not valid because when it drops below VDD by 2 and eventually goes below V D SAT. The transistors goes into a linear region of operation right and there the equation itself changes. So, when I do that resistance calculation it will go into a different region of operation all together. So, that is not even worth sort of going through all the way down to 0 with that equation because you are not of interest to us, ok.

So, beyond this invalid model, the reason I am pointing this out is you will see that we use certain models in this course which are valid only in certain regions of operation outside that region of operation. If you ask me what the value of the model is I would say I do not care because the model is not even meant to handle that region of operation that is the point I am going to drive home here ok, any questions.

So, the other thing is of course, later when I start putting transistors in series C Vs or in parallel, then I it is very easy for me to reduce that network using series parallel resistance combinations if I know R equivalent current source model I cannot do that any questions here, ok.

(Refer Slide Time: 21:39)



So, let us look at this equation a little more closely R equivalent is 3 by 4 I leave out this lambda parameter I naught, if it is velocity saturated then I naught is equal to K n prime W by

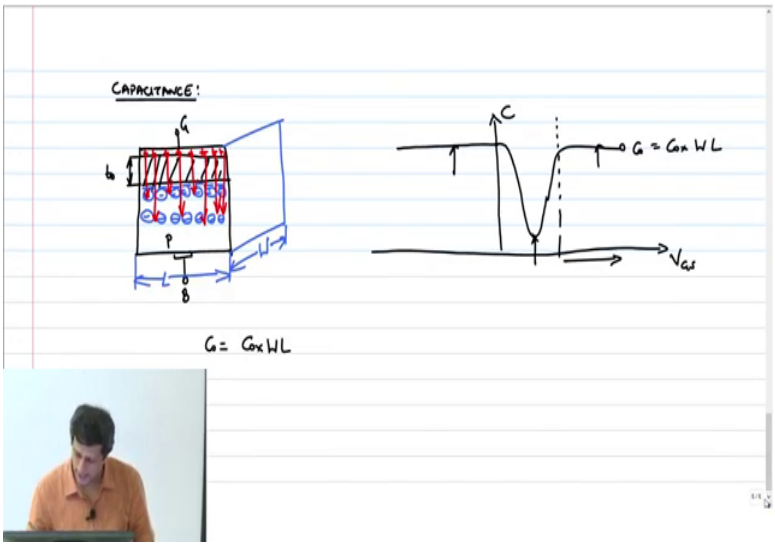
L V D SAT n into V DD minus V Tn minus V D SAT n by 2 ok. If it is a long channel device and it enters saturation region then it is just K n prime W by L into vg VDD minus V T whole squared and half right, this is saturation velocity saturation it does not matter. So, this is a long device, this is a short channel device, ok.

So, look at this equation of R equivalent now right clearly it is going to be inversely proportional to W, 1 over W because I naught is sitting below, right. So, therefore, if I double the width of a transistor then the resistance of the transistor drops by a factor of 2 right, all the resistance becomes 50 50 percent of what it was earlier.

So, this is 1 key technique that we will use if you want to reduce your resistance then you upsize the transistor and that is what this equation is also telling us, right. Now, what happens if VDD is very large; suppose, V DD is very large such that VDD minus V Tn minus V D SAT n by 2, right or VDD is very large compared to V T. Then what will happen, yeah. If it is velocity saturated and VDD is very large, what happens to the R equivalent as a function of VDD yeah, but there is also an inverse dependence on 1 by VDD minus V Tn minus V D SAT and. So, is we are VDD is large right what does large mean that is VDD minus V Tn minus V D SAT n by 2 is nearly VDD right, this is just a first order thing to show you that are equivalent then is not a function of VDD.

So, if I have a 180 nanometer device and I operate it from 1 you know 1.8 volt down to 1.5 volt then the R equivalent will not change much 3 ok. On the other hand if your VDD is very small then of course, these things cannot be neglected and that has a more complicated relationship with really VDD. In this course we are going to deal only with higher V DDs, we are not going to bring down the very close to suppress hold or even low voltage operations, ok.

(Refer Slide Time: 25:38)



First let me just go back to this picture of this MOS cap that I had ok, MOS capacitance a 2 terminal device that I had metal, oxide and then I apply a V G and I apply a body right, let me call this gate. Now, as I sweep my V GS from a negative value to positive value right, what do you think will happen to this capacitance of this device, ok. I am going to now do a C V plot; C versus V GS. If V GS is negative then what happens by the way let me also put the thing this is p p type substrate, V GS is negative what region of operation is this in accumulation right so; that means, you will have lot of holes sitting here, right.

Now, if the concentration of holes is so high at the surface any negative charge that you put here is effectively going to be terminated right at the surface. The electric field will terminate right at the surface because the concentration of holes at the surface is that much higher. So,

of course, there will be depletion region and all that which is negligible all I am saying is it is negligible.

So, if you look at this capacitance this is going to be just that parallel plate capacitance which is epsilon naught W L by T ox, ok. So, of course, this is my L, W is into the plane right and this is my T ox. So, C naught prime was C ox or let me just call it C naught C ox into W into L, correct. So, this capacitance will just look like this. Now, what do I do I move this into the positive region V GS, V GS is going to be greater than 0 and then which region will it start operating in it depletion mode, right.

So, if you look at this I am now to put positive charges here because V GS is greater than 0, this will start depleting my channel these are immobile carriers again. In the depletion mode; obviously, there is negligible surface charge right. So, the electric field will actually start from the positive charge and terminate deep into the channel also like this, you know it.

So, therefore, if you look at the capacitance in this depletion mode it is going to be a capacitance of the oxide plus not plus in series with a depletion capacitance. Obviously, if you put 2 capacitances in series then it sort of drops it is C 1 C 2 by C 1 plus C 2 and therefore, this capacitance will now drop, ok.

Now, at some point the surface charge density starts increasing again and when you reach inversion mode the surface charge density will be so high right, that I have been able to terminate all the electric field almost at the surface. So, therefore, the capacitance will now simply go up back to this old value that we had, this is just a sweep of capacitance versus V GS.

Student: (Refer Time: 30:15).

Once it goes to the.

Student: (Refer Time: 30:24).

So.

Student: (Refer Time: 30:32).

So, what do you think it will be?

Student: (Refer Time: 30:42).

Ok.

Student: (Refer Time: 30:59).

So, ok; so, maybe I will discuss this offline with you, but typically the picture is little naive here that I have shown you, when you do it things will change in a sort of a gradual manner right. In fact, if you look at this picture capacitance may start dropping like this itself, it is not so abrupt as I have shown here, ok, but there will never be a discontinuity because nothing is abrupt in reality ok.

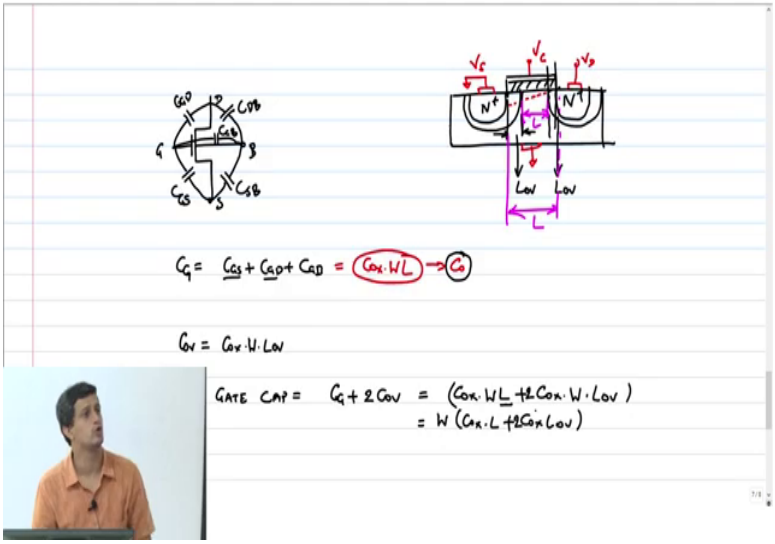
But anyway so, I just want to point out that there is this picture. Now, where are we going to operate this device our transistor in and what capacitance should be worried about should we worry about this value, should we worried about this value or should we worry about this value. So, let us say that the device is in linear region of operation transistor where is it on this picture right side it has already got inverted.

So, really if you look at the region of operation we are worrying only about to the right of this region. Even in linear operation what I do not want you to assume is linear region means it is in depletion mode now, when it enters linear region it is already channel has got inverted that is how the we have exceeded threshold voltage, right.

Of course, it may not be as strongly inverted as it would be in saturation that is fine, but still the capacitances effectively this capacitance that we are talking about. And of course, as you

just saw in the earlier picture when we discharge the capacitor from VDD to VDD by 2 we are you only in saturation region forget about linear region at all, we are not even entering that region. So, clearly we are only worried about this capacitance here which is equal to C ox into W into L, ok.

(Refer Slide Time: 33:20)



So, with that I can move into the picture of the capacitance for an NMOS transistor or a PMOS transistor actually. It is a four terminal device it turns out there are capacitances all over, from all these terminals this is gate, drain, source, body without loss of generality I am just saying 1 is drain another source, ok. Here I have a capacitance like this C GD, I will tell you the origin of all these capacitance C GS then there is C GB then there is sorry C DB and C SB, ok.

So, now, we first consider the regular device where the manufacturing was perfect in the sense that if I have my by oxide thickness here then wherever this diffusion finishes exactly there the gate will start, ok. This in fact, is manufactured in something known as a self-aligned process, you grow the silicon make everything silicon dioxide then put a mask and then harden 1 region and make the other soft then you etch out the soft regions.

So, because the oxide is the gate oxide is itself acting as it is mask right it is a self-aligned process and typically I will get a nice diffusion region exactly like this, wherever this diffusion finishes exactly there the gate will start, ok. So, therefore, the total gate capacitance C G is basically just C GS plus C GD plus C GB ok, I am writing this has three components because when you now go into saturation region, right.

Suppose this is my source and this is the drain V D, V G and this is V S ok, and body is also grounded when you go into saturation region I told you that the channel will pinch off which means there is more charge or the source side than the drain side. So, there will be as uneven distribution of capacitance actually in this in this kind of region.

So, what I am saying is that total capacitance C G which is a sum of C GS plus C GB plus C GD right is just C ox into W into L some of these, but if you look at the distribution across various regions in cut off. For example, C GB will be hole of C C naught and C GS and C GD will be 0 this is just sort of a distribution model that people have done it is more relevant for analog guys not further it is, ok.

If you look at the saturation region the C GS will be more C GD will be less and the C GB is just made 0. So, all I am saying is this total capacitance which I have called C naught is being distributed across these three components for distal purposes it does not matter because we are looking at the sum of these three components always. For analog as you have to worry about this ac values also and therefore, you look at whether the source has more capacitance or drain has more capacitance and all that, ok.

This is the ideal picture, now what do you do after manufacturing it you form your diffusions implant N plus then you have to anneal this device. When you anneal this device these regions right will expand thermally like this. This is not depletion region, I am talking about the atoms will actually move because I am heating them, I am taking this like 800 degree centigrade. You have to anneal it in order to make all this even out to even out all the manufacturing defects in that process the N plus will now this under the channel as well. So, what will happen is there will be a small overlap capacitance that comes here, this distance is called L overlap, ok.

So, previously if you look at it my channel length was this. Now, after this diffusion because of annealing my channel length has become only this much, ok. So, what is my overlap capacitance, C overlap is it is just like a parallel plate capacitor again, it is this part of the gate how much capacitance is that going to form, the area is L overlap into W, right. So, this overlap will be C ox into W into L overlap, right.

So, the net gate capacitance is C G plus twice C overlap; one for the drain side, one for the source side. So, this I can write as C ox into W L plus C ox into W into L overlap, ok. Again note, what we have done is we have written all of this as some technological constant parameters into the design parameter what is our design parameter W and L for digital design even L is not a parameter right, even that is fixed at the shortest length.

So, this is basically W into C ox into L plus C ox into L overlap. What is the conclusion if you increase the width of the transistor, the gate capacitance goes up ok, any questions here yeah.

Student: (Refer Time: 40:51).

Yeah. So, finally, when I tell you what is L right that is accounting for all this when I say W by L whatever L you substitute there actually accounts for all this ok, they may start for example, at 20 nanometer and eventually it will come down to 40 nanometer or something like that. And in the actual model, if you look at the model file that has been given they do

further calculations to reduce this L; you say L is so much, but internally that L will be reduced and then that L will be plugged into the model for the current equation, ok. Remember, now there is also a depletion region of this nnp junction. So, there will further reduce.

So, all of that will happen ok, yeah.

Student: (Refer Time: 41:46).

Because, I have 1 overlap on the drain side, one overlap on the source side where, oh sorry this one yeah, this is 2 yeah, yeah.

Student: (Refer Time: 42:09).

Yeah.

Student: (Refer Time: 42:12).

Yeah.

Student: (Refer Time: 42:17).

No, be careful C naught or C ox?

Student: C ox.

C ox is just epsilon by T ox, epsilon naught by T ox.

Student: That (Refer Time: 42:44).

No that is why I am saying that I have called C naught, I am calling that C naught.

Student: (Refer Time: 42:52).

Correct, correct, correct yeah.

Student: (Refer Time: 43:01).

Correct.

Student: (Refer Time: 43:06).

No, no, no, no that is what I am saying, be very careful here this is not because I mean sorry these two are not because of overlap. I am saying this is actually a combination of whether you the transistor is in a saturation mode or linear mode or cutoff mode. If the transistors in cutoff mode then this C GB is C naught, C GS and C GD are 0, if it is in linear mode then each like the charge distribution in the channel is uniform, right.

So, therefore, c gs and C gd are just C naught by 2 each and the third condition when it is in saturation there is an uneven distribution where there is more charge on the source side rather than drain side therefore, that capacitance again changes. So, that is why I said the first picture is even without considering overlap capacitance. I have to write the capacitance of a component of three things right yeah, but for digital circuit purposes just assume that is C ox into W, all sum of all these that is not overlap capacitance, ok, yeah.

Student: (Refer Time: 44:36).

So, it is between gate body, gate drain, gate source. So, the actual picture if you want is if you are doing high frequency analysis then the way the charge changes under the drain and the way the charge changes under the source will be different depending on where you are operating the transistor therefore, the capacitances will change. So, it is not just all 3 of them;

I mean, it is not just C GB. C GB is when you are in cutoff region, right and it is mainly in accumulation mode or something then it is just between the body and this that is all.

Student: Sir.

Yeah.

Student: If the total gain capacitance, the L which we take is the reduced till or (Refer Time: 45:24).

This L here right, this L here.

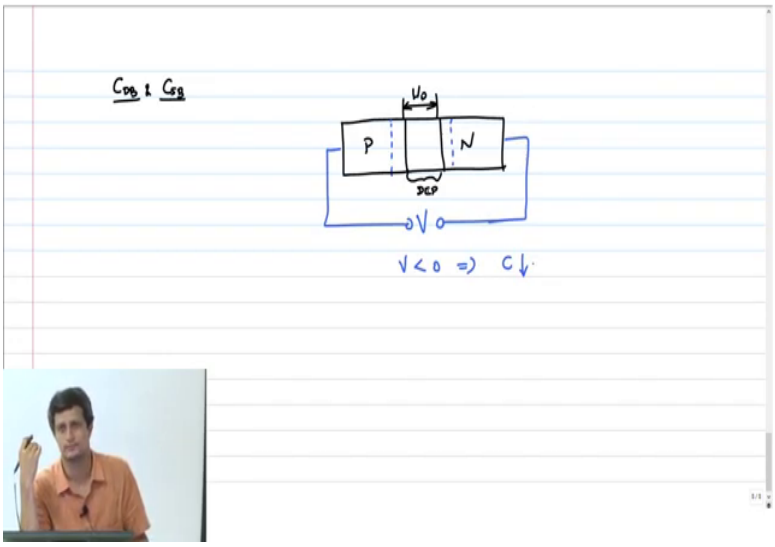
Student: Yes sir.

That is the reduced l.

Student: Reduced.

Yeah that has to be the reduced, ok.

(Refer Slide Time: 45:45)



So, now then other two components which are basically C DB and C Sb the origin of these capacitances are your p-n junction diffusion the p-n junction reverse bias capacitances, ok. So, if I have a p-n junction like this right, this is my P, this is N, I have a depletion region here right, then this there is a total distance of W D, this will now form a capacitance right and give rise to a capacitance in that p-n junction, ok.

If I reverse bias this further what will happen to the depletion region, it will increase right if I increase my reverse bias the depletion region will go up like this. So, what will happen to the capacitance. So, what will happen to the capacitance, it will come down, right. So, the key thing is that it is a function of the applied voltage also ok. So, if this is my applied voltage V then if V sort of is less than 0 it implies capacitance also decreases because depletion region goes further and therefore, it comes down.