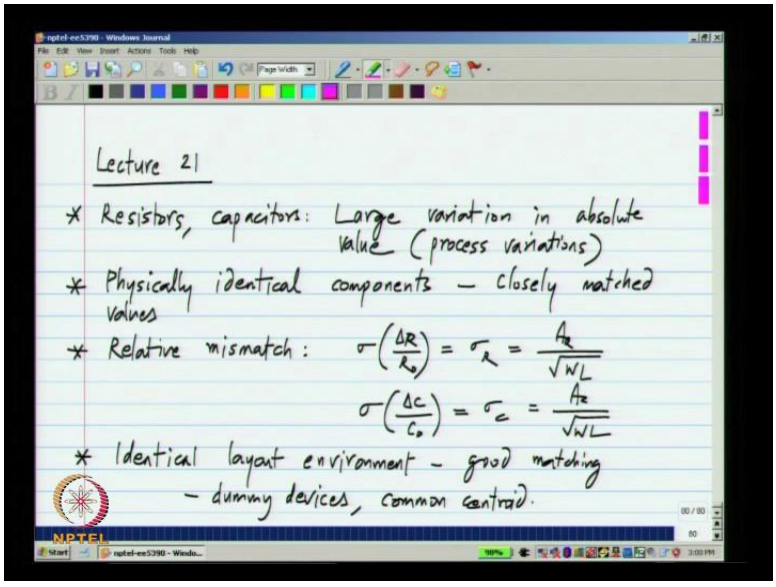
Analog Integrated Circuit Design
Prof. Nagendra Krishnapura
Department of Electrical Engineering
Indian Institute of Technology, Madras

Lecture - 21
MOS transistors

Hello and welcome to lecture 21 analog integrated circuit design. In the previous class we have looked at components available on a CMOS process; that is the commonly used components like the resistor and the capacitor. We will be studying the transistor in this class; what we learnt was that the resistor and capacitor parameters can vary significantly over process; these are known as process corners. But if you make 2 identical resistors; they will have identical value and if you make two physical identical capacitors; they will have identical value and so on.

Now, we also saw that despite this being nominally identical there will be some small mismatch. And, the amount of mismatch is related to the absolute physical size of the device. If you make a pair of physically large resistor then they will match very well to each other; if you make them very small then the matching will be rather poor. Now, we saw quantitatively how to relate the standard deviation of mismatch to the physical size of the resistors and capacitors?

(Refer Slide Time: 01:33)

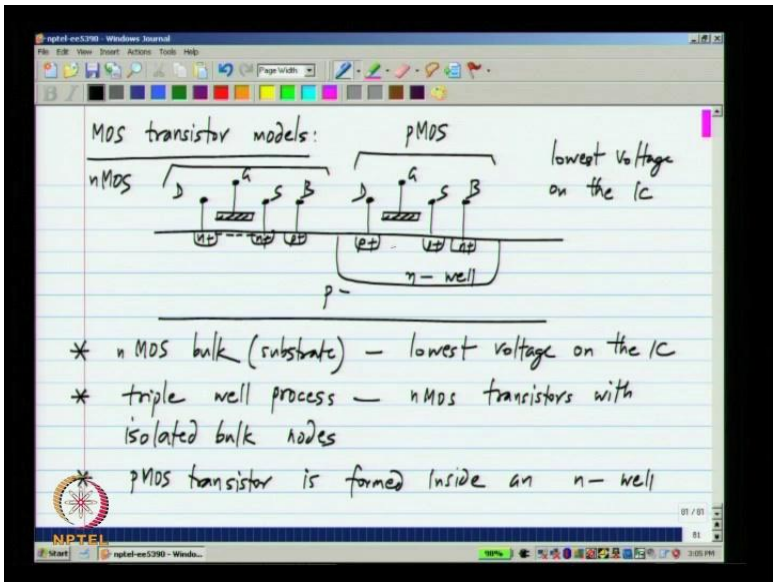


These large variation in absolute value is known as process variations. And, physically identical components will have closely matched values. And, the relative mismatch is related to the physical size; the standard deviation of delta R by R that is the difference in the value of between a pair of resistors to the nominal value of the resistor R naught; which is usually called sigma R which is given by A R by square root of W L; where W and L are dimension of the resistor. Similarly, delta C the difference in values between a pair of nominally identical capacitors to the main value of the component is sigma C and that is given by A C by square root of W L.

And, A C and A R will be given to you by the process boundary; based on this you can figure out exactly how big to make the resistor or the capacitor; will see later that this is to not only of resistors and capacitors. But also of all other components; they all have a similar behavior with mismatch. If you want closely match components you need to have physically large components.

Finally, we need to have identical layout environment for good matching; this means that you may have to use dummy devices; orient all the devices in the same way and perhaps use even common centroid layout and so on. So, to get the very best matching you need to do all of these things; what will do in this class is to look at the most sophisticated component which is available in CMOS process that is the MOS transistor itself. We will discussed the model of the MOS transistor both large and small signal models look at the mismatch; and see how to use it in the circuit that we are going to design.

(Refer Slide Time: 04:41)



Now, in the CMOS process will have 2 types of MOS transistors; first of all most CMOS processes are built starting from a P minus substrate. And, on that you built an n MOS transistor like this; this is the gate, drain, source and the bulk itself for the body will be like that. Now, if you have 2 transistors; 2 n MOS transistors you can see that the bulk is common to both of them. And, we need to keep all of the bulk source and bulk drain junctions reversed biased.

Because this drain and source junctions are there only to make connection to the channel that will be formed between them; these p n junctions should not be forward biased. So, this means that typically you are constrain to connect the bulk of the n MOS transistors to the lowest voltage on the integrated circuit; n MOS bulk which is basically the substrate, common substrate connected to the lowest voltage on the I C. Now, because of these constraint we also have to look at what happens to the transistor when the bulk voltage is different from the source voltage ok?

Now, there is a variant of the process which is known as a triple well process; in this case you can have n MOS transistors with isolated bulk nodes. So, if you do have a triple well process what happens is you will have some situation like this; you will have an n well and that you will have a p well. Now, because this p well inside has no connection to this one; the bulk of the inner transistor can be connected independently of what happens to the substrate of the outer transistor? So, in this case this bulk here need not be connected

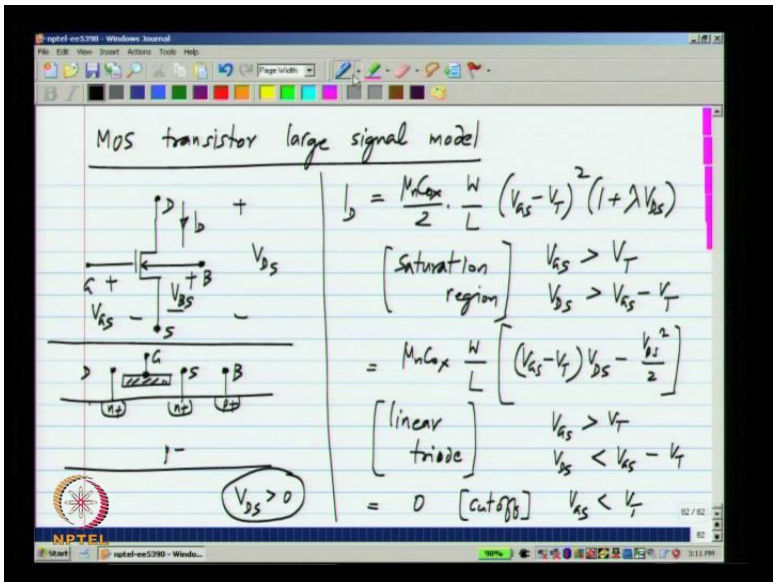
to the lowest potential; you can connected to whatever potential is required for the proper operation of that transistor.

But most of the processes that are commonly used are not triple well transistors. And, you will be forced to connect the n MOS transistors bulk to the lowest voltage on the chip; that something in keep in mind while designing the integrated circuit. Now, we cannot design with only n MOS or we could but design would be rather inefficient it is better to have both n MOS and p MOS transistors; to have p MOS transistors we have an n well. And, inside that you make the p plus drain and source connections; and you will have a p MOS bulk connection to the n well. So, this is a p MOS transistor; that is an n MOS transistor. So, the p MOS transistor is formed inside an n well.

Now, because you make all the components on a common substrate there is something common electrically to all of them; those are the things we need to understand. In discrete component circuits every component is separate; if you have 2 transistors there will be in separate packages. And, the really is no communication between them so to speak where as on an integrated circuit if you have 100 transistors they all share a common substrate.

So, they will in some way talk to each other; there is something common to them. Now, in order to isolate them we need to take care of something; one of the things that we do is to connect the substrate to the lowest potential on the chip ok. Now, this also means that the model that you use for the MOS transistor; should model effect of this substrate voltage or the bulk voltage on the transistor.

(Refer Slide Time: 09:39)



We will first review the MOS transistors large signal model; the model is similar for n MOS and p MOS transistor and then will go on to the small signal models. So, if I consider an n MOS transistor there are 4 terminals really and physically they correspond to these terminals. And, this is the bulk; it is not our goal we have to discussed in detail the physics of the MOS transistor; we will only study the models to the sufficient extent that we can use them for circuit design.

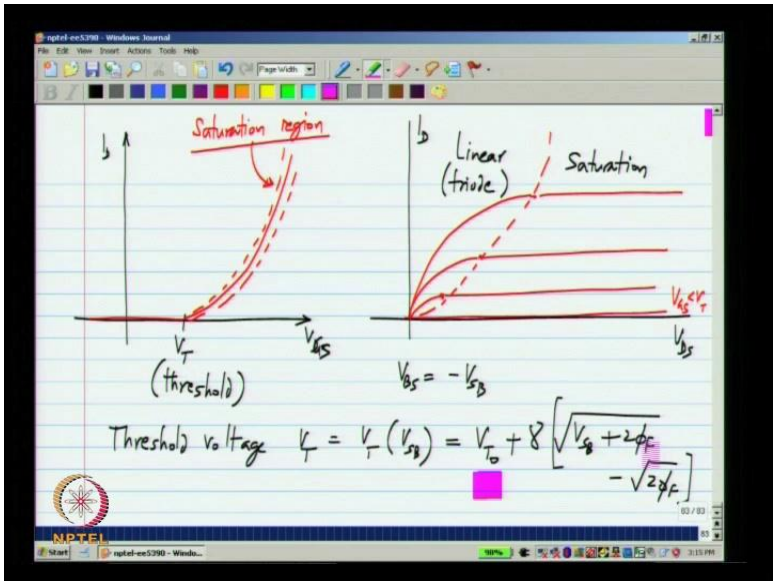
Now, one thing you all observe is that from the layout as well as from the symbol; the drain and source are interchangeable the device is cemetrical with respect to drain and source. And, that is true in reality; we do label one of them to drain and other one the source. And, what makes the distinction is which voltage is higher; it is not related to how the device is fabricated.

Now, it is quite common to model current that goes from the drain to the source; as a function of the gate source voltage, the drain source voltage and the bulk source voltage or sometimes the opposite of it the source bulk voltage; we will see what the model is. Now, the cemetrical structure of the mosfet means that perhaps it would have been more sensible to model the device while referring all the voltages to the bulk. And, there are such models which are cemetrical and nice; but it is been long hell convention that things are usually model by a referring to the source in unlock circuit design. So, will continue to use that.

So, the current I D which goes in to the drain and comes out as a source; it turns out that the gate current and bulk current are negligible in most cases is given by mu n C o x by 2 W by L V G S minus V T square 1plus lambda V D S; this is the model V G S happens to be more than some parameter V T the threshold voltage. And, the drain source voltage is greater than V G S minus V T; this is known as the saturation region the current equals mu n C o x W by L V G S minus V T times V D S minus V D S square by 2; this is the case when V G S is more than V T. And, V D S is less than V G S minus V T this is known as the linear or the triode region and also the current will be 0; when V G S is less than V T ok.

And, implicit in all of these things is that V D S is greater than 0; that is why we identify one term as the drain the other one as the source. If this happened to be more than that then we would call this the drain that the source. So, V D S greater than 0 is common to all of these things. And, the current is 0 this is called cut off; when V G S is less than V T; you are probably already familiar with this model from the basic electronics course this will be a quick review of the mosfet model.

(Refer Slide Time: 14:13)



Now, graphically what does this look like? We can plot I D verses V G S for a constant V D S. So, this model says that the current is 0 up to a voltage called a threshold voltage; and after that it increases as a parabola. And, if you plot I D verses V D S what happens is when V G S is less than V T; the current is almost 0. And, when V G S is greater than

V T in the triode region the current increases and in the saturation region it increases very slowly something like that. And, this boundary between the triode and saturation region follow the parabola. In fact, it is the same curve as we shifted to the origin. by the way this is the current in saturation region.

Now, because it is in saturation region the current does not vary too much with V D S; it does vary but not greatly because the sloop here sloop in this region is very small. So, on this curve this is the saturation region and this is the linear or triode region; I am assuming you are familiar with these MOS characteristic; I will only discuss things that are slightly unfamiliar from basic courses.

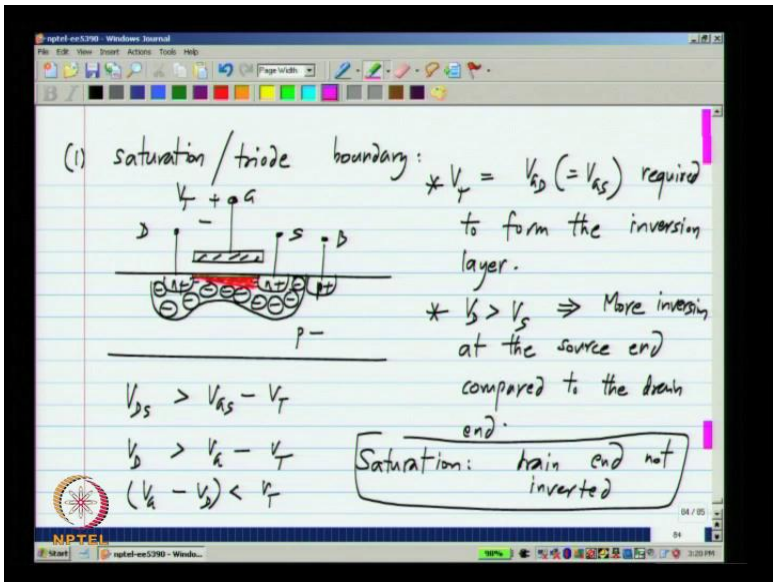
Now, what it says is that in the saturation region the current is almost constant it is independent of V D S, it is dependent on V D S through this parameter lambda but this lambda is a small number. So, the current can be thought of its almost independent of V D S. And, that is why we use it as a current source later we see that for most amplifier; we would like a transistors in the saturation region.

The linear or triode region is also useful; and for a certain portion of the region when V D S square is very small compare to the other term; we will have linear dependence between I D and V D S. So, it can be used as a resistor. And, also when you use the MOS transistors switch it will operate between deep triode region where V D S is almost 0; and cutoff when there is no current at all.

Now, one of the things we should keep in mind is that the threshold voltage V T itself is dependent on V S B or V BS; V BS is nothing but minus V S B you see that so far in our expressions the bulk source voltage V S B has not appeared in the picture. But it will indirectly through the threshold voltage V T; and V T is given by some V T 0 which is a threshold voltage when V S B equal 0 plus some parameter gamma times square root of V S B plus 2 phi F minus square root of 2 phi F.

We will not discuss these equation in detail V T 0 is nothing but the threshold voltage with zero V S B. And, you can see that the threshold voltage increases with V S B; you can referred to some device text book for the meaning of gamma and 2 phi f and so on. But for us right now will only focus on the fact that the threshold voltage increases with V S B. Now, what does this mean its try and make introduced sense of these model equations that we have written so far.

(Refer Slide Time: 18:34)



First of all the saturation triode boundary; what does it mean? I think again you will be familiar with the operation of the MOS transistor from basic devices courses; this is the n MOS transistor you know that if you tied drain and source together. So, that these two are at the same potential. And, then you go on increasing the potential of the gate; what happens is below the gate you will form an inversion layer. And, around here you will have a depression layer with fixed charges.

And, in this you will have the inversion layer; and the inversion layer is what is responsible for conduction of current in the MOS transistors. Now, the amount of voltage you need let us say we have the gate and source tied together; the amount of voltage you need between the gate and drain or source to start the process of inversion is the threshold voltage V T is nothing but V GD which is the same as V G S with this connection required to form the inversion layer. Now, what is happening when you have V D naught equal to V S; let us say we do not have this connection; this is the normal case in a circuit the drain voltage can be quite different from the source voltage.

Now, let us imagine a case where drain voltage is larger than the source voltage; what happens? The inversion layer will be thicker at the source side and thinner at the drain side; it will look something like that. Now, V D greater than V S means that there is less voltage across the oxide at the drain side than at the source side; it also means more inversion at the source end compared to the drain end. Now, what was the condition for

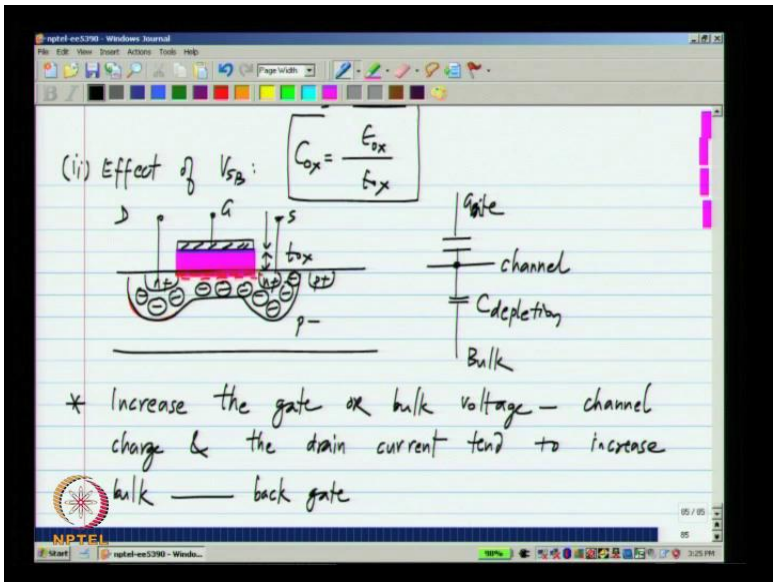
the saturation we said that V D S has to be greater than V G S minus V T; in other words V D is greater than V G minus V T. And, I will remove the common source reference from V D S and V G S both or V G minus V D is less than V T; basically the voltage across the gate and the drain which is basically the drain end of the channel is less than V T.

So, what it means is that in saturation region the drain end of the channel is not inverted whereas, the source end is inverted; that is the channel is pinched of at the drain. In fact, if you recall from the device physics courses that is how we get the square law model; you assume that there is a pinch of. And, then from there you can get the square law dependence of the current on V G S.

Now, what happens in triode region? In triode region both the drain and the source end are inverted and the channel is continues all the way from source to drain. And, when V D S equal 0 which is the lowest V D S you can have or you can say that this is the deepest you can going to the in the triode region. Then, we will have a uniform channel from the drain to the source; in saturation the drain end is not inverted and in triode region both the ends are inverted.

So, that is one thing to keep in mind. Now, this also tells you once the drain end is pinched of the drain will not have a significant effect on the amount of charge in the channel. And, that is why we get this constraint current in the channel when the trances region saturation region. So, you start by inverting both ends and then you raise the drain potential to a point; where the drain end is no longer inverted. Then, the drain voltage does not have significant effect on the channel charge because it is not even connected to the channel. So, now you will have charge and consequently a current; that is independent of the drain voltage. Now, it does turn out that it has a minor influence and that is we have small slope lambda.

(Refer Slide Time: 23:57)



Effect of the source bulk voltage; so first of all again if you recall the basic operation of the MOS transistor; what happens is that you start with the piece of semi conductor which means that it does not really normally conduct; there is no conduction layer between the drain and the source sides. Let me call this the drain and this the source, that is the gate. Now, very basic principles can be thought of us you have a capacitor between gate and this point; gate and this region of the MOS transistors. If you raise the voltage across the capacitor sufficiently you will form charges in the semi conductor and that can be used for conduction.

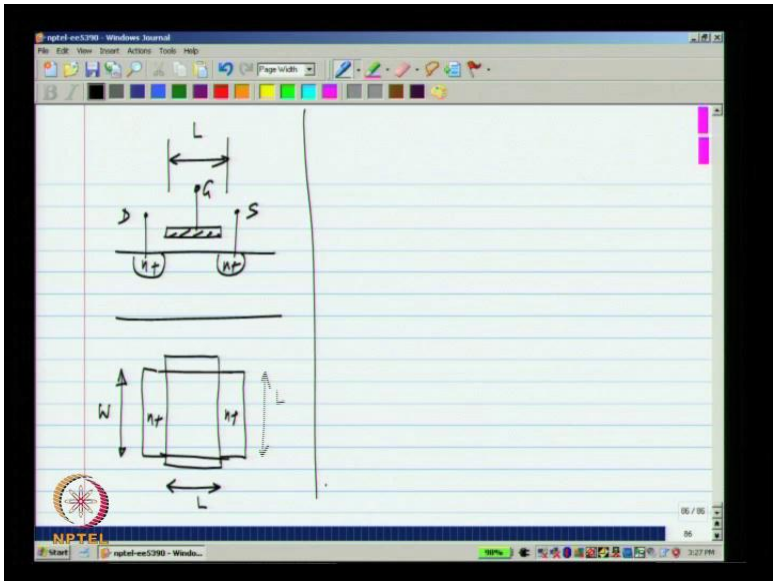
Now, actually if you observe this there is not 1 capacitor but 2; first of all there is the capacitance between the gate and the channel; and also there is a capacitance between the bulk and channel across the depletion layer. So, you can think of it as having the capacitors from gate to the channel. Now, the thickness of this oxide is t ox and the capacitance of this per unit area epsilon ox by t ox is what is known as C ox; this appears in the expression for the current as you well know; also on other side you will have the capacitance of the depletion region basically the capacitance of this region right.

Now, clearly if you raise either the bulk voltage or the gate voltage the channel charge will be influence in the same way. Now, imagine that the channel is fixed to some potential; you increase the bulk voltage that there are negative charges on this plate of the capacitor which is basically increase in the channel charge. If we increase in the gate

voltage again there are negative charges on this plate of the capacitor; which correspond to an increase in the channel charge.

So, whether you increase the gate voltage or increase the bulk voltage qualitatively the same thing is happening in the mosfet; there are more charges in the channel. And, once you have more charges in the channel you will have more charges available for conduction and the current will tend to increase. So, that is why this bulk will also many times called the back gate. And, there are even circuits which use the back gate like the normal gate that we use.

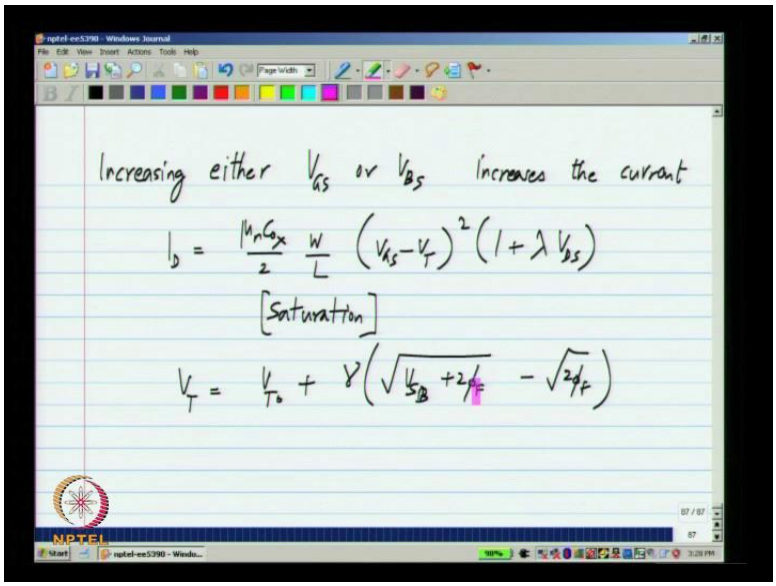
(Refer Slide Time: 27:35)



Now, earlier I did not mention what the W and L in the mosfet where as usual this is the gate source and the drain. And, the distance between the drain and source is the length of the transistor. And, if you look at the top view of this device; there is certain width for the source and drain regions. So, the current flows across this width and that is the width of the MOS transistors.

Now, it turns out that if you double the width it is like connecting 2 MOS fed parallel. Let us imagine a case where you have 2 MOS fed with the same V G S, V D S same V B S the voltages everywhere. And, then you put 2 MOS fed next to each other the current in them just add up; that is why we get the proportionally of the current to the width of the transistor. If you increase the width the current will increase proportionally.

(Refer Slide Time: 29:20)



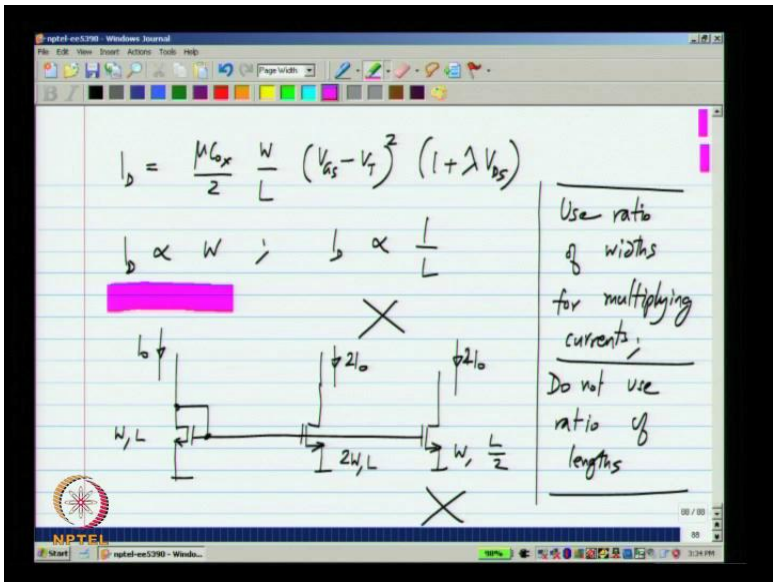
So, that is the basic large signal model of the MOS transistor. Now, I said that increasing either V G S or V B S increases the current and the equations that we had were mu n C o x by 2 W by L V G S minus V T square times 1 plus lambda V D S; this is the current in the saturation region the principle is the same in the triode region as well. And, the threshold voltage itself is given by V T naught plus gamma square root of V S B plus 2 phi F minus square root of 2 phi F. Now, this equation is written in terms of V S B but you see very easily that; if V S B increases the threshold voltage increases. And, if the threshold voltage increases the current reduces.

So, finally, the conclusion is that if V S B increases the current will reduce which is the same as saying if V B S increases the current will increase; V S B is the variable of choice. Because V S B is always positive in an n MOS transistor. And, this is because we would like to keep that source bulk junction reverse bias; we do not want to turn on those diodes the reason the junction is there to is to make connection to the channel. If we turn on those diodes those current will interfere with the normal operation of the mosfet; they can cause some undesirable effect in the operation of the chip. So, because V S B is always positive the equations are always written in terms of V S B not V B S. But you can easily see that if V B S increases the current will increase.

So, we have to consider the effect of all 4 terminals on the current in the MOS transistor; we already looked at the large signal models. And, we will use this model as a sort of

guideline for a design. Now, it turns out that the actual model of the MOS transistor that reflects the operation of a model MOS transistor its really really complicated; what we will use this very crude model for the level zero model is to gain some incentive to the operation of the circuit.

(Refer Slide Time: 32:05)



At this point again let me rewrite the expression for the drain current. Now, this expression says that I D is directly proportional to W and I D is inversely proportional to L. Now, the current turns out to be directly proportional to W because there is a conduction current flowing across a width of W. Now, if you increase the W the length across which you have a flow of current will increase; that is why if you have W that is doubled this like having 2 transistors in parallel and the current will also double. So, this means that the proportionality to W is more or less exact. Now, if you take a real mosfet and if you double the width of the transistor you will find that the current will approximately double.

Now, the inverse proportionality to length comes under a more respective condition; it comes about because the field in the MOS transistors reduced. But because there are lot of non ideal effects that make the MOS transistor correct; but exactly inversely proportional to L; you cannot rely on this inverse proportionality. So, what I mean is the following right. So, let us say you have a certain MOS transistor; I think you are all familiar with current mirrors. Let me take 2 cases where first of all I have a diode

connected transistor here; we will deal with these circuits in detail later I just want to illustrate a point here. So, let say inject a current I naught. So, let say I want 2 I naught and 2 I naught here ok.

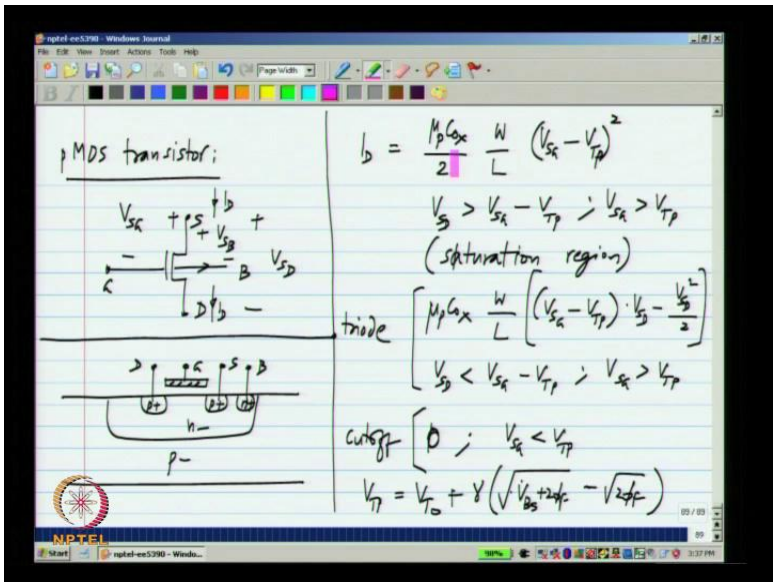
And, let us say the aspect ratio of this is W by L; that is the width of this transistor is W and the length is L. In fact, to make it expressive let me not write W by L; I will write W coma L. Now, let us say I want 2 I naught there are 2 ways of getting it I can make 2 W and L or I can make W and L by 2; assuming that L by 2 is above the minimum possible length of the transistor in the given process.

Now, you will find that this 2 W by L will give you almost 2 I naught whereas W and L by 2 the current in this branch can be very far from 2 I naught. So, you would never ever want to do something like this. If you want let us say current mirror or a number of identical transistor whose currents have to be scaled; when I say identical they are identical in all respects expect for a scaling of current; you should scale only the W and not the length of the transistor.

So, for instance when you make current mirror and you want ratio of let us say 2 to 1 or 4 to 1 or whatever it is between the input current I naught. And, the current that you want in the particular branch; you should scale the width of the transistor accordingly and not the length. In fact, if you want close matching between any pair of transistor; they should have the same width will say how to arrange the transistor layout; so that that happens later.

So, you use ratio of width for multiplying currents you never use ratio of lengths. So, that is a brief over view of the large signal model of the MOS transistor; where we use the n MOS transistors for illustration. But I will very quickly go through the p MOS transistor and show that it is exactly the same way.

(Refer Slide Time: 36:26)



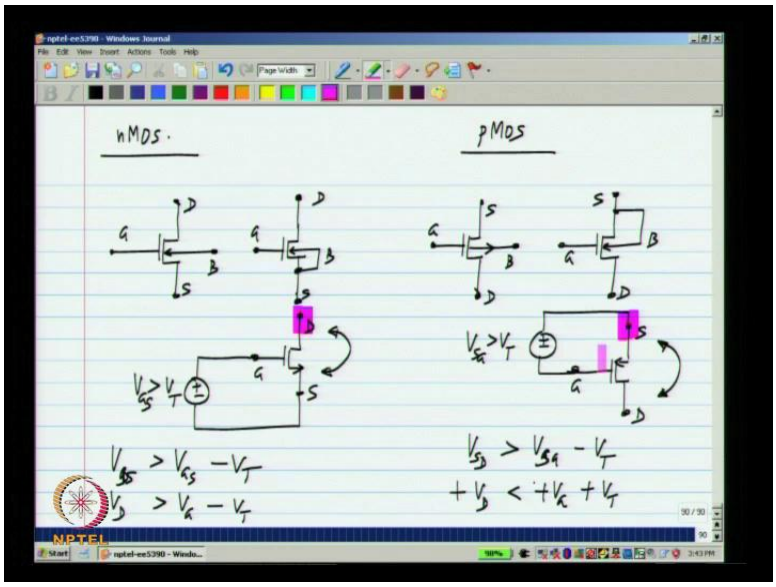
The p MOS transistors symbol is like this when we include the bulk; this is the bulk source, drain and gate usually the source is drawn above the drain. Because the source potential is higher than the drain potential usually in a p MOS transistor. And, in a common integrated circuit you have the p substrate the n well and the p MOS transistor inside and that is the bulk.

Now, the equations are exactly the same; except that we write them in terms of V S G, V S D and V S B; I D which flows from the source to the drain is given by mu p C ox by 2 W by L V S G minus V T P square. And, this is when V S D is greater than V S G minus V T P and V S G itself is greater than V T P; this corresponds to the saturation region in the p MOS transistor. And, in the linear or triode region will have here will have and finally will have zero current; when V S D is less than V T P. So, these correspond to the triode and the cutoff regions respectively right.

Now, the equations are exactly same as that of the n MOS transistor. And, expression for the threshold voltage is also the same; the p MOS threshold voltage will be the threshold voltage with 0 bulk source voltage plus gamma times square root of V B S plus 2 phi F minus square root of 2 phi F. Now, compared to the n MOS transistor there are many ways of writing the p MOS equations in some books and references the threshold voltage is negative.

And, you refer to V D S and V G S which are negative voltages; what we have done is in our notation the threshold voltage is positive. And, we also refer to voltages V S D and V S B which are positive; again for the n MOS transistor the threshold voltage is positive as well. Now, these are known as enhancement transistors; the transistor in which the threshold voltages are positive; and this is the most common type of transistor. So, that is what we will assume we will have.

(Refer Slide Time: 40:24)



Let us little more discussion on the regions of operation on the n MOS and p MOS transistor. Because we have to keep calculating this all the time in circuit design; it is extremely important to become very familiar with this; this is the symbol that is use for the n MOS transistor. And, sometimes the bulk is tie to the source; and in this case many times 3 terminal symbol is use just for simplicity.

And, you can see that if bulk is tied to the source the threshold voltage equal V T 0; the threshold voltage at zero value of V S B. And, we can think of it is a 3 terminal device like this with V T equal to V T 0. Similarly, for the p MOS transistor sometimes the bulk is tie to the source; it is like this. And, in this condition we can represent the p MOS also with a 3 terminal symbol; where this is the source, gate and drain.

Now, what does it mean for the transistor to be in saturation or in triode region; first of all in the normal operation of the transistor I will use the 3 terminal symbol for simplicity here; there will be a V G S which is greater than V T. I do not necessarily mean that

there will be a voltage are connected between gate and source. But whatever is connected will established a V G S that is greater than V T. Now, if this voltage become very small between drain and source; the transistor will go in to triode region; we can imagine that when these 2are squeeze together the current keeps be getting smaller and smaller.

So, as the drain voltage comes closer to the source voltage; the current gets smaller. And, exactly the same as the case in the p MOS transistor as well; we will apply a V S G that is more than the threshold voltage. And, if the drain voltage comes to close to the source voltage if the drain voltage increases and comes to close to the source voltage. Then, it will tend to go in to the triode region.

And, if the drain and source voltages are the same the current will be 0; there will be no voltage across the device; that is somewhat similar to having zero voltage across the resistor which gives you 0 current. The region I am mentioning this is sometimes there is confusion about whether it goes in to triode region if the drain voltage is decreases or increases especially what happens in a p MOS.

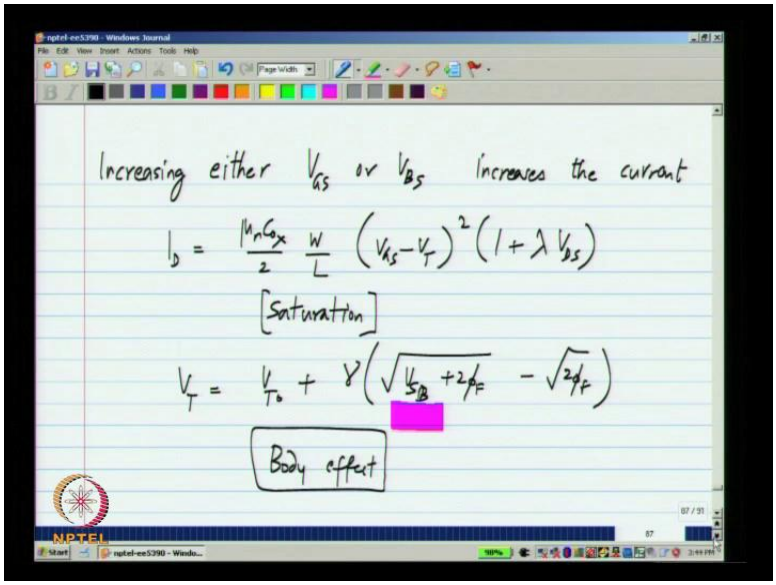
So, if these 2 terminals come close to each other they tend to go towards the triode region whether it is p MOS or n MOS. So, that is one way to think about it; another way is we can write the expression for triode verses saturation region. And, 2 different forms for an n MOS transistor V D S should be greater than V G S minus V T; it can also be written as V D greater than V G minus V T; all I have done is to add the sources voltage V S to the 2 sides. Now, this form a sometimes convenient because we are not calculating V D S and V G S.

And, we are simply comparing the drain voltage compare to the gate voltage. So, what is the same V D has to be greater than V G minus V T. So, this voltage has to be larger than V G minus V T. So, because V T is positive in our transistors; the drain voltage can below the gate voltage but not by more than one threshold voltage that is what it is saying. Similarly, in this case V S D has to be greater than V S G minus V T; which basically says that minus V D greater than minus V G minus V T or V D has to be less than V G plus V T.

So, as the drain voltage goes on increasing it can be above the gate voltage; again V T is positive. But it cannot be above the more than one threshold voltage; that is say also a good way of thinking about triode verses saturation region. So, there should be no

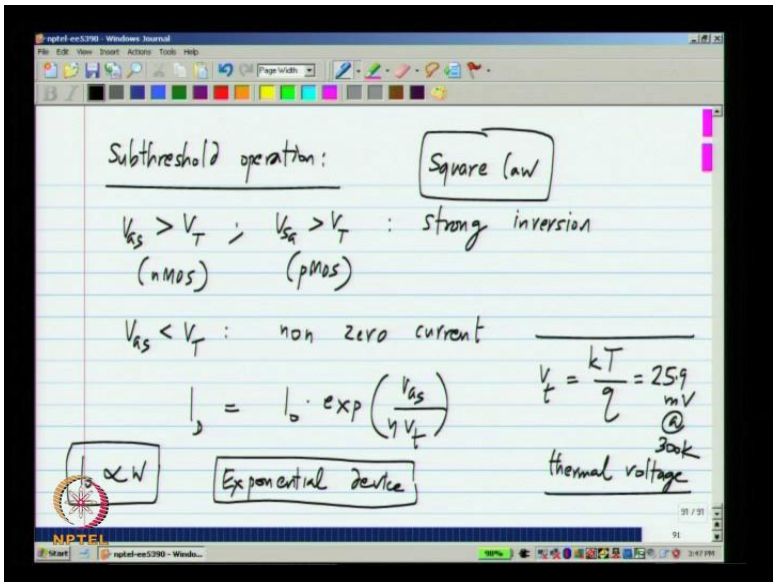
confusion in this regard; because we will soon move to circuit which have many trends of transistors. And, you should be able to quickly tell which transistor will go in to triode region; when? As we will say later after we discuss the small signal parameters; we would like to keep all our transistors in saturation region for proper operation of an amplifiers most of the time.

(Refer Slide Time: 45:56)



This phenomenon of the V B S influence in the current or V S B influence in the current is known as body effect. Now, one last comment on the currents in a MOS transistor. So, far we assume that if V G S is less than V T in an n MOS transistor or V S D is less than V T in a p MOS transistor the current is 0; that is if the gate source voltage is below the threshold voltage; the current will abruptly turn to 0. Now, in reality this is not so.

(Refer Slide Time: 46:41)



The region of V G S being greater than V T in an n MOS or V S G being greater than V T in a p MOS is known as strong inversion. And the square law that we have for the mosfet that we are all familiar with is true; when the mosfet is strongly inverted. But it turns out that when V G S is less than V T in an n MOS transistor the current will not fall to 0; it is still non zero; the current will be small, the current will be smaller in a strong inversion.

But it will not be exactly 0 and it turns out that the current follows an exponential relationship; it is a similar to that of a bipolar transistor; where this V T is K T by q which is 25.9 milli volts at room temperature is the thermal voltage. So, it goes from being a square law device to an exponential device; we will not going to the exact details of what is in I naught. But just like in the strong inversion region I naught happens to be a proportional to the width of the MOS transistor; we will not go into details. But we will just keep in mind that current will be an exponential. And, we will see later that in fact a very useful region of operation for certain applications; will discuss that after we discuss the small signal model of the MOS transistors.

Thank you.