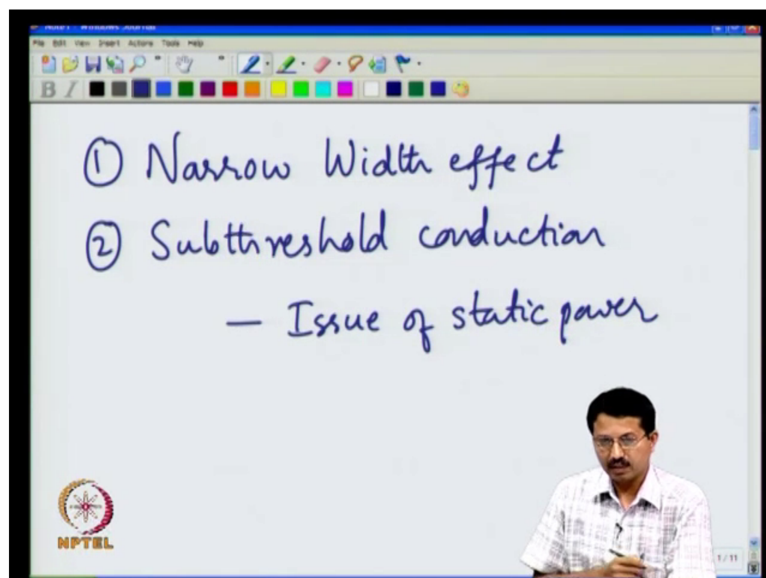
Nanoelectronics: Devices and Materials
Prof. Navakanta Bhat
Centre for Nano Science and Engineering
Indian Institute of Science, Bangalore

Lecture – 04
Subthreshold Conduction

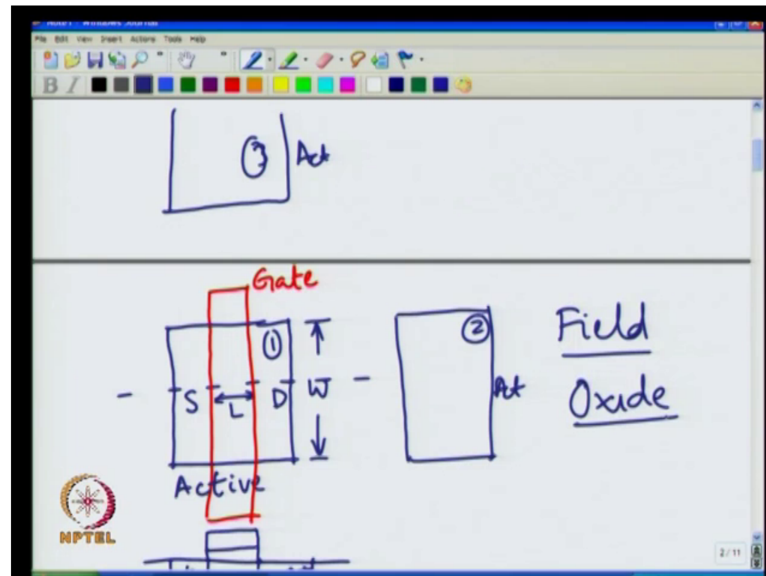
Today we will look at two important characteristics in the context of nano transistor one is what is called narrow width effect, the other important aspect is what we call sub threshold conduction.

(Refer Slide Time: 00:28)



And this in turn will lead us to the discussion of the issue of static power in today's state of what CMOS technology it has really surfaced as one of the very important challenges to deal with right. In the last lecture we have looked at short channel effect that is when you decrease the channel length of the transistor what happens to the threshold voltage of the transistor. So, first let us start with the discussion today that is narrow width effect.

(Refer Slide Time: 01:26)



So, when we are talking of a width of the transistor let me just refresh you with the top view of a transistor as you may recollect we have seen in one of the earlier lecture we call this as an active area and typically whenever you have 2 rectangles one called active and the other one called gate in top view which crisscross each other that indicates a transistor. So, now, in top view this is a transistor right you have a gate here and you know active area and in turn you see this side of the transistor resource this side of the transistor is drain and this distance is what we have been calling as a channel length which is essentially the distance between source and drain junctions along this direction this is our channel width right.

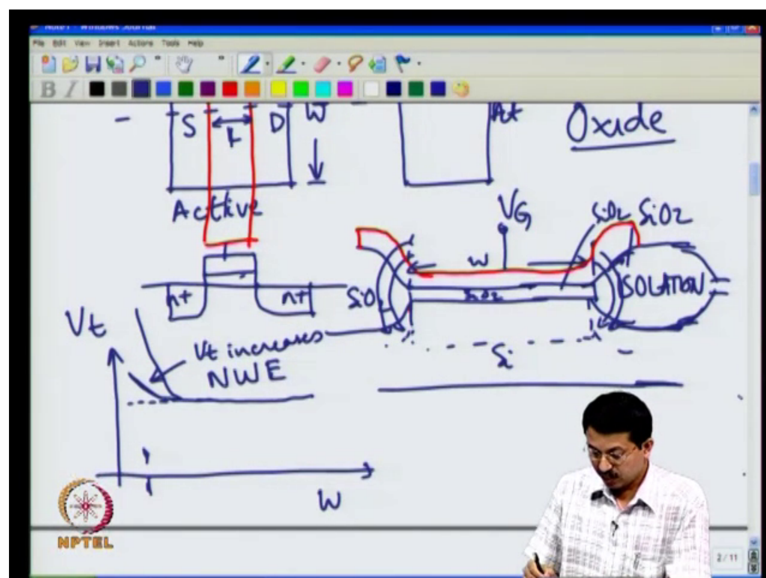
Now, last lecture we saw the transistor cross section by essentially taking cross section in this direction right when I do a cross section here and I take a side view as you know we essentially see something like this right, here I have the source junction and here I have this drain junction right and on top I have gate insulator and this is my gate electrode right. And looking from the top what you see is essentially the gate electrode which is sort of crossing this active area, the active area then includes in this direction the source channel and the drain region all these put together are active area. When we talk of integrated circuits with millions of transistor 2 neighboring transistors are always isolated from one another right. It used to be in olden days junction isolation, but junction isolation is just not feasible and we have been doing oxide isolation for last several generations of the technology.

In other words you have this transistor and you have a neighboring transistor let us say about here I am just going to sketch only active area I am not going to sketch the field area and similarly in this direction also you may have another transistor right which will be going like this and this transistor one transistor 1 transistor 2 and transistor 3 have to be isolated if they do not have to talk to each other there has to be perfect electrical isolation.

So, in other words when all these are called active areas and where ever you do not have these active rectangles rest of the area is called field area and in this field area you essentially have grown silicon oxide or deposited silicon oxide to create isolation. In other words if I were you know look in this direction right I have silicon right underneath is active region, but underneath this regions I will initially see an oxide if I see the you know in a depth profiling and then later on I will see silicon coming in I will draw that for you in a minute.

So, in other words if I were to take a cross section along this direction here and look at it from the side because I have taken a cross section and I am looking at it from this side. So, then what you will see is the following right I will keep that out here. So, you have that for your reference.

(Refer Slide Time: 05:04)



So, I am taking cross section and looking at from the side view. So, what you have then is this is your active area that corresponds to your width of the transistor and here

immediately under this silicon oxide. Remember this is silicon oxide this is the same silicon oxide which is this gate oxide immediately under that you have this silicon and this is the edge of your active area this is the edge of the active area which essentially corresponds to this area correct edge of the active rectangle that you have.

So, from here on if you see here what you will essentially see is something like this, this is also silicon oxide this whole thing is silicon oxide and this is also silicon oxide you see, but this is what is called a gate oxide of the transistor and this is a field oxide of the transistor and the field oxide is typically very thick there by if there is a transistor here another transistor this field will extend like this you see this is what we call a field region.

And when there is a next transistor that comes in you again have a thin oxide under next transistor which you know we looked in this direction this is the next transistor. So, between the edge of this active and edge of this active you will have a thick field oxide as is indicated out here and you know this is what is called isolation. Now, you know in CMOS technologies there are 2 different ways of doing isolation you know we will come to that in a minute, but the point I am trying to raise right now is you have this isolation oxide here as well SiO 2 and this is your gate oxide a very thin SiO 2 and remember you know what you see on top here is this poly silicon correct because you see this poly silicon red area rectangle goes beyond this activate edge right this is the active edge it is going to go beyond this activate edge by this amount that is why I have shown this poly silicon going on top of the field oxide as well.

So, this is what you see you right in the in the side view when I take a cross section along the drain along the width access of the transistor. Now, you say there is something interesting that comes out. Typically again when we apply a voltage right we had to create a depletion and then we have to create an inversion right, earlier we essentially said that whenever we apply a voltage you will have a depletion region created in this area and then of course, this region will be inverted and that is when you have a inversion and current will flow when the transistor is on. But you see when I apply a voltage here to this gate V G, this V G will set up certain charge on gate our earlier assumption was that the charge was balanced by an equal and opposite charge which is found only in this area, but in reality you will have what are called fringing fields here

correct at this edge even though that oxide is little bit thicker and hence you will also start creating the depletion region here right.

In other words you see whenever I apply a voltage on the gate the gate not only have to deplete this region, but it will also have to deplete this region before you start talking about inversion. In other words gate has to do extra effort as opposed to what we had in a very simplistic model right which is just this rectangle it is exactly opposite of charge sharing effect that was coming in short channel effect because in short channel effect source and drain charge sharing was helping you to really invert the transistor because part of the depletion width was taken by source and drain region, but here gate will have to take care of additional depletion area right. So, then what would you expect? You would expect that V t should start increasing because of this effect.

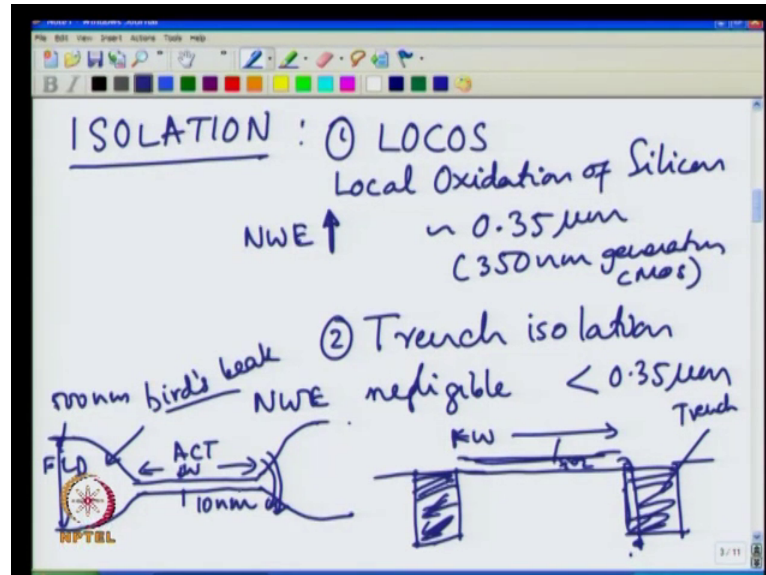
However, again when the width is very large this small area which is essentially due to the fringing field is so small that it is just insignificant, but when width is shrinking when width is becoming comparable to this area which is being depleted because of the gate now gate is doing extra effort that is visible externally and hence you need to apply more voltage to invert the transistor. In other words if you were to now sketch your V t as a function of channel width it should be flat right as per our simple one dimensional derive equation, but when the channel width decreases to the order of 500 nanometer, 100 nanometer then you start seeing an increase in V t. So, V t increases and this is what we call narrow width effect.

This effect shows up only when the width is very narrow very narrow in this region not in this region when the width of transistor is very wide this area is insignificant compared to the actual active area of the transistor. So, this is a very very very important point that you need to also remember when we talk of very small feature sizes of the transistor. Now, narrow width effect is a very strong function of how we make this oxide, how we make this field area you see because this field essentially depends on you know this oxide thickness. Just imagine two cases in one case where this oxide thickness is incrementally smaller than this gate oxide then this fringing field is as strong as this vertical field.

in another case where it is a very thick oxide then that fringing field is just negligible correct and you would imagine; obviously, that the structure of this isolation will have a

very very strong impact on how this narrow width effect will look, whether it looks like this, whether it looks like this or almost flat it does not depend on width at all.

(Refer Slide Time: 11:28)



So, in this context I just want to sort of highlight that when we talk of isolation technology there are two kinds of isolations that we talk about one is what is called abbreviated as locos this stands for local oxidation of silicon. This was a technology isolation technology that was used from you know early days of CMOS you know way back in 70s all the way to I would say mid 90s, till about 0.35 micro meter generation or 350 nanometer generation, CMOS.

In all these technologies we used to have locos isolation, but all modern technology is used what is called trench isolation and it turns out narrow width effect is very savior very bad in locos isolation and narrow width effect is negligible in these technologies technologies where we are talking of less than 0.35 micron. Now you can appreciate why we moved from locos to trench isolation right if narrow width effect starts becoming very very strong it is very difficult to control the threshold voltage of the transistor right.

So, we want to minimize the narrow width effect and that is why we went to trench isolation technology. So, what is different between these technologies is the following? In locos isolation that is number one which I showed earlier your isolation area looks like this, this is what we call active area which is where your transistor width is defined lithographically the by printing and this is your field area typically you know this oxide

thickness is of the order of let us say 10 nanometer which is about 100 hamstrung. This could be as large as 500 nanometer which is 5000 hamstrung you know almost 50 times more right and that is what is going to give you the isolation from one transistor to the neighboring transistor.

But very interestingly when we try to actually define this field region by this process we always end up with this kind of a transition region in fact, sometimes in locos technology you know nomenclature this is actually called birds beak because when you look at it looks like a birds beak coming in here. This birds beak is a major problem in locos technology right you know no matter what you will have this effect whereas, if you have a trench isolation technology, trench isolation would look like this if this is your silicon and if this is your active area this is your silicon oxide you actually have what is called a trench dug in silicon which is completely filled with oxide this whole thing is filled with oxide and this is why it is called trench isolation. This is your active area which is your width and outside the width is this region here.

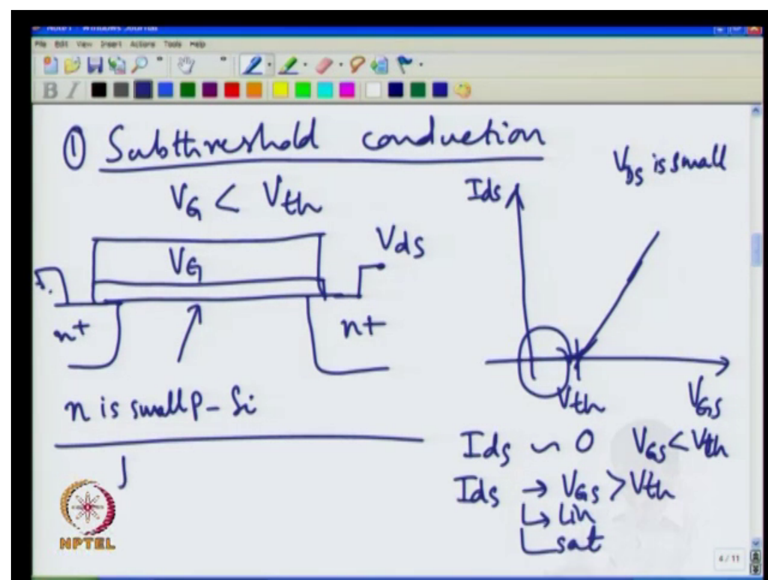
If you have very thick oxide right at the edge of this active your fringing field here will be very very small right, as a result of that you will not see much narrow width effect whereas, although you have thicker oxide here right where it matters you know as far as narrow width effect is concerned it is really thin, the oxide is very thin and hence your fringing fields can be very strong here whereas, here your fringing is not very strong it can be neglected for all practical purposes. And that is why your trench isolation technology is what is used for all technology is less than 0.35 micron it minimizes the narrow width effect it minimizes the birds beak problem as well.

We will not really go into the details of how one would actually get this structure because that is really the details of the semiconductor processing right we will not really talk. So, much about this in this part of the course we will only focus mostly on device physics part, but there are ways to do that. For example, as the name suggests here local oxidation of silicon meaning oxidation can be done locally in silicon wafer by I am asking for certain regions of silicon wafer right. So, you mass that active region of the silicon wafer by a dens material such a silicon nitride. So, oxygen does not diffuse through silicon nitride and the rest of the area is exposed and oxidation takes place in those area that is conversion of the silicon into silicon oxide where is in the active region since they were protected by silicon nitride they do not get converted into silicon oxide

and that is how selectively I can convert part of the region into a thick field oxide region rest of the region is what I used to build a transistor.

Similarly, there are ways to build a trench isolation technology we will not really going to the details. So, that completes the first target that we had for today's lecture what is meant by narrow width effect. The fact that as I start miniaturizing the width of the transistor I need to start considering the fringing fields which would be there typically you know at the edge of the gate along the width direction specifically where gate electrode is going on top right beyond the active area and because of that your threshold voltage will have a increasing trend then that is narrow width effect by going to trench isolation technology as opposed to locos isolation technology we have been able to minimize the narrow width effect .

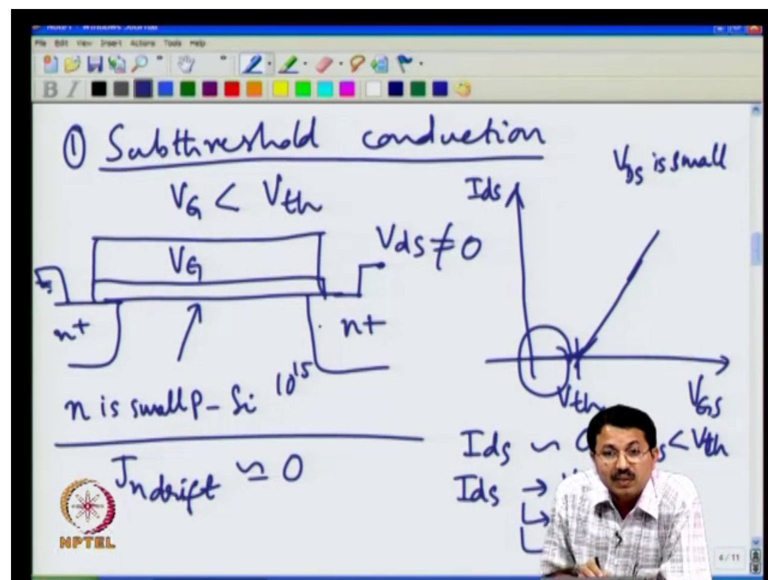
(Refer Slide Time: 17:58)



Now, let us go to another very important aspect that we wanted to discuss for today's lecture and that is sub threshold conduction. This essentially means this is a condition when my gate voltage is less than threshold voltage right, that is you have a transistor and this is your V G gate electrode and these are your source and drain junctions I apply a drain voltage Vds and we say that typically if we look at Ids drain to source current as a function of V GS this being an n channel transistor we say it would look something like this especially for when Vds is small that is when the transistor is in linear region, otherwise transistor regions in saturation regions.

Now, this point here is what we call threshold voltage we say that Ids is equal to 0 for V GS less than Vth correct then Ids will be nonzero right for V GS greater than Vth and then it could either be in linear region or in saturation region depending on what is your drain value voltage value that you have. But what we are really now interested is this region what happens when V G is less than Vth is current indeed 0 or there is a nonzero current if there is a nonzero current what is the basis of that nonzero current remember this is a p type substrate. When I have reach V G is equal to Vth that is when I have created a channel here correct and when there is a channel you have current which is a drift current correct the current flows because there are lots of carriers and you have applied drain voltage that sucks up electric field and hence you have a current flowing through the transistor, but when I have V GS less than Vth I do not have a channel really.

(Refer Slide Time: 20:32)

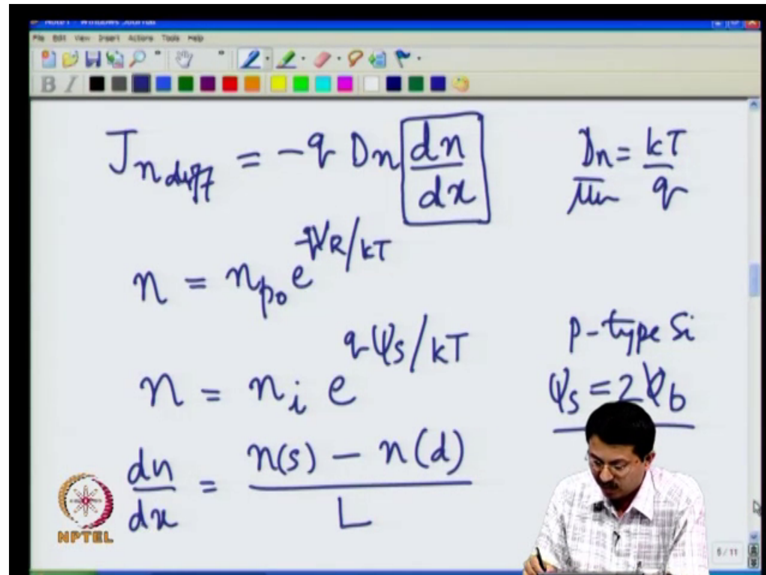


In other words in this region n is small and hence you were drift current also fields are small your drift current can be ignored really. However, n is small, but nonzero you see n cannot be 0 even when it is a p type silicon which is doped between 10 to the 15 acceptor impurities my n here is still 10 to the 5 right which is minority carrier concentration.

Now, we are talking of a situation where I have applied voltage here nonzero voltage and I applied ground potential here now I am asking the question there is no drift current, but could there be a diffusion current if there is a concentration gradient of electrons from

source to the drain we could certainly except diffusion current to flow through right because.

(Refer Slide Time: 21:27).



if you recall the diffusion current density is given by correct where you q is the electron charge d n is the diffusivity of electrons of course, D n as you know is given by D n by mu n is k T over q thus so called Einstein's relation where k T over q is a thermal voltage mu n is the mobility right d and a is essentially diffusivity of the carriers. So, this is diffusion current. The diffusion current exists whenever there is a concentration gradient n can be small, but d n by d x can be there. So, we need to really worry about what is d n by d x.

So, let us really find out what is d n by dx now. If drain voltage where to be 0 source voltage where to be 0 then there would be absolutely no gradient in electron concentration the electron concentration gradient will come in only when I create an a symmetry by applying a unequal voltage on the source and drain terminal as is the case here is 0 and Vds which is not equal to 0. In fact, we will see little later that the electron concentration here will be more compared to the electron concentration here. In fact, we will derive the expression for the electron concentration but qualitatively you can think about this in the following manner this is a reverse bias n plus p junction and this is a reverse bias n plus p junction right. But here I have a nonzero reverse voltage right if you recall your p n junction theory that is something called a law of a junction at the edge of

the depletion region your carrier concentration is really governed by what is the reverse bias voltage that you are applying. In other words if you recall your n in general is given by n p naught times e to the whatever reverse voltage that q V R by k T in a diode in this is in a simple diode right. The minority carrier concentration gets modulated based on what is the applied reverse voltage.

Now let us actually come from the other perspective that we have been talking about right from the MOS theory right if you recall we can write electron concentration in general as e to the q psi S by k T where psi S is what we call surface potential and in fact, we have seen that when I reach in version psi S is equal to 2 phi b. In a p type silicon the you know surface potential to begin with psi S when you do not have any applied voltage its flat band condition you are electron concentration will be much lower than the hole concentration.

But as you start applying the forward bias voltage you know your psi S starts increasing right and eventually it becomes equal to 2 phi b and that is when we have a inversion condition right you remember that right. And essentially what you know we are talking about now when we are talking about this diffusion current d n by d x is to really find out n at source reason I call it n at S minus n at drain divided by the distance between source and drain which is the channel length. So, if you can find out you know what is the electron concentration at the source and electron concentration at the drain that in turn will you know give me what is you know your diffusion current because from there you can get the carrier gradient right.

(Refer Slide Time: 26:16)

$$n(0) = n_{p0} e^{q\psi_s/kT}$$

Flatband $\psi_s = 0$

$$n(L) = n_{p0} e^{q\psi_s/kT} e^{-qV_{ds}/kT}$$

So, here we can essentially write the electron concentration at the source side as you know something like this right n p 0 e to the q psi S by k T right with an assumption here that under flat band condition psi S is equal to 0, that is how we set up the convention for measuring the surface potential right. You know flat band condition for that is psi S is equal to 0. In fact, when psi S is equal to 0 remember the bands are flat in silicon and as a result of that your whole concentration is really n p naught, n p naught stands for I mean electron concentration electron concentration in p side under thermal equilibrium condition right that is not at all altered.

Now as you start applying more and more gate voltage you know this psi S starts increasing eventually psi S becomes 2 phi b when phi S becomes phi b n p n 0 is equal to n i correct because that is when your intrinsic level will bend and come to the Fermi level location because that is a band bending of phi b. And when psi S is equal to 2 phi b intrinsic level will fall below the Fermi level and that is when it has become n type. So, that is when you go from n p naught all the way to the carrier concentration which is equal to hole density in a p type dope semi conductor right. So, that is how we have this carrier concentration.

On the other hand at the drain junction because my drain voltage is nonzero your electron concentration would actually go down below the source electron concentration by this modulation factor which is q Vds by k T. When Vds is equal to 0 as you can see

here your drain concentration is equal to source concentration that is what you expect in a symmetric device, but here you have created a symmetry applying voltage and hence you know you have a carrier gradient and that will set up a diffusion current. So, now, given this you know we can write the expression for d n by d x which is simply n p naught e to the q psi S by k T into 1 minus e to the minus q Vds by k T divided by length.

(Refer Slide Time: 28:58)

Now, I can use this in my drain current equation that I had here replace d n by dx with that expression and multiply that with q d n then I get a current density multiply that with area you get current eventually right. So, then let us find out what happens to your current right, so your current then J n diffusion is minus q d n you know we could replace the d n as mu n times k T over q because d n by mu n is k T over q remember that so I have just made that substitution there and then n p naught e to the q psi S by k T by L into one minus e to the minus q Vds by k T. A very important observation that you can make right away in terms of the sub threshold current and it depends on drain voltage is that when Vds is a few times thermal voltage you see q Vds by k T. So, q bring it to the denominator its k T over q in the denominator at room temperature it is about 25 millivolts and if your Vds is say 75 100 millivolt then e to the minus 4 is negligible compared to 1.

In other words when Vds is little more than thermal voltage your sub threshold current is more or less independent of drain voltage in other words it just sort of flat comes out it

only is dictated by what is the psi S value and psi S value we will see in a minute is governed by your gate voltage.

So, you know for we will just look at this condition for Vds greater than a few k T over q you can approximate this J n diffusion as this right this is what we have. Now what I am interested really is to replace this psi S in terms of gate voltage because I my interest eventually is to find out when V G is less than V t for any given value of V G what is the corresponding drain current.

(Refer Slide Time: 31:49)

$$V_G = V_{FB} + \psi_s + \frac{\sqrt{2\epsilon_s q N_a \psi_s}}{C_{ox}}$$

$$\frac{dV_G}{d\psi_s} = 1 + \frac{C_d}{C_{ox}} = \frac{C_{ox}+C_d}{C_{ox}} = m$$

$$d\psi_s = \frac{C_{ox}}{C_{ox}+C_d} \cdot dV_g$$

Now, I already see that it is nonzero current. So, now, let us find out what it is then, psi S you see we can write psi S in general as V FB I am sorry not psi S we will just re write this. So, what I mean here is V G that is applied gate voltage is really V FB plus drop across silicon and that is what we are calling psi s. In fact, when I reach inversion that is when I replace left hand side V G as Vth and write 2 phi b for psi S correct and what is this here its 2 epsilon q times n a psi S again when I reach inversion I replace psi S by 2 phi b and this becomes 4 epsilon q Na phi b the classical equation that we had earlier right divided by C ox. Let us see the dependency here let us let us look at this d V G by d psi S V FB is a constant flat band voltage you know that depends only on the functions difference that you have so that is 0 and here d we are differentiating left hand side and right hand side with respect to psi s. So, here what you essentially get is this is 1 plus differentiation of this.

And you know if you do that and if you simplify you know I will just skip a few steps just for you know gravity here, eventually you can show that this is given by this expression you differentiate this psi S is psi S to the power half then half you know psi S will come to the denominator as root psi S when you differentiate psi S to the power half with respect to psi S right. You can do all that algebraic simplification and eventually you write Cd depletion capacitance as epsilon S divided by w d and wd in turn depends on psi S because psi S is the voltage across the depletion width.

If you do all that you will come up with this very interesting expression and you can even sort of rewrite it if you want as you know C ox plus Cd divided by C ox both are one and the same. In fact, sometimes we call this factor here as m where m is greater than one right which is very obvious because its C ox plus Cd divided by C ox, now you can re write this to get d psi S as C ox divided by C ox plus Cd times d V G yes. So, what is this telling you, you know this is telling you something very interesting and that is it appears now as a series capacitance circuit that we have. What is there in the series capacitance circuit? I have V G here, I have C ox here and I have C depletion here which is silicon capacitance and this is my psi S.

And when there is a voltage change on the gate correspondingly there is a voltage change at psi S meaning that is a voltage drop across this Cd as you know in a capacitive divider the voltage across this capacitor is the total voltage that is being applied or the change in voltage divided by other capacitor divided by the total capacitor correct and that is exactly what we have here. In other words what this is telling you is that the only fraction of the change in gate voltage appears as a change in surface potential and what is that fraction is determined by this ratio.

As we will see this ratio becomes extremely important and that in turn determines what is your sub threshold volt.

(Refer Slide Time: 36:11)

And you know given this as I said already you know you have this these psi S is you know a d V G by m correct and because you know C ox play Cd divided by C ox is m and I do know that when psi S is equal to 2 phi b V G is equal to V t. And now we are asking the question when psi S is not is equal 2 phi b accordingly V G is not is equal to Vth what happens right. So, based on that I can sort of rewrite this equation psi minus 2 phi b is equal to V G minus Vth divided by m right.

The particular case when you know as I just we are now talking of psi being less than 2 phi b that is V G being less than Vth right and that is when we are trying to ask the question what is the current right. In other words your psi again here is 2 phi b plus V G minus V t h divided by m.

Now let us go back to the situation that we had your drain current J n diffusion is all this factor times e to the q psi S by k T divided by L. Now I can replace this psi S because I have an expression for psi S below V t psi S as a function of V G is now derived right let us make use of that equation and also let us you know instead of current density we will write the current I n or Ids simply as J n diffusion times area.

(Refer Slide Time: 37:55)

We will not worry about what that area is you know area is the width of the transistor times the thickness or which the current is flowing because that is a cross section for you available for the current to flow. But let us just keep it as a and also let us club all this pre factor that we have as some constant some constant k because we are only interested in this exponential factor right that that is our region of interest right now. So, in other words you know your current is some constant k which will absorb this A area and all other parameters such as mobility minority carrier concentration temperature and all that right and this exponent e q psi S by k T that I am going to re write now as e to the you know q 2 phi b by k T.

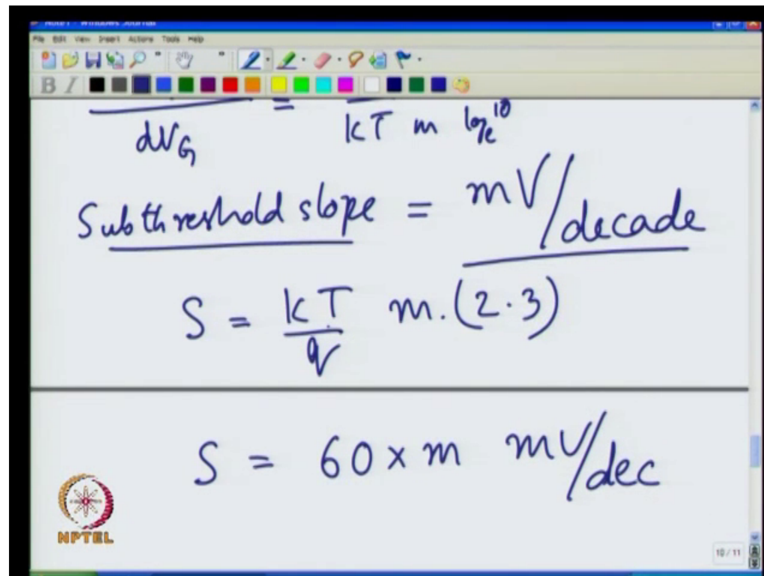
This times e to the q V G minus Vth by m fine I am just looking at that exponential factor that I have here this exponential factor, L is also observed in that constant right do not worry about that and in fact, I make an observation that for a given doping concentration phi b is fixed and this is also constant I will also observed that in a final another constant some k prime. So, if I do that then you know I have an expression here Ids is some k prime e to the q V G minus Vth divided by m very very simple equation now. It says drain current has now an exponential dependence on gate voltage you see when the transistor is on in a MOSFET drain current has linear or quadratic dependence when it is in linear or saturation region, but now the drain current has an exponential relationship.

(Refer Slide Time: 39:42)

And more importantly we are really interested in this quantity actually that is log Ids if you were to take here again some log constant some you know constant will not worry about that right then what you really have is q V G minus V t by m by log 10 to the base e or L n e here I am taking the log to the base 10 here log to the base 10 you see right this is what you will get. The quantity that I am interested eventually is this d log Ids to the base 10 by d V g, this is a quantity that I am interested.

So, if you do this differentiation what you get as you can see here is somewhere I have when I substituted that I think I have missed this k T right in the denominator right. So, here if this is q by k T right this is here, but I forgot to write it here right q by k T. So, you know that k T is out here and accordingly and that is also here And what you essentially have is k T m log 10 ten to the base L n its natural log of 10 L n e.

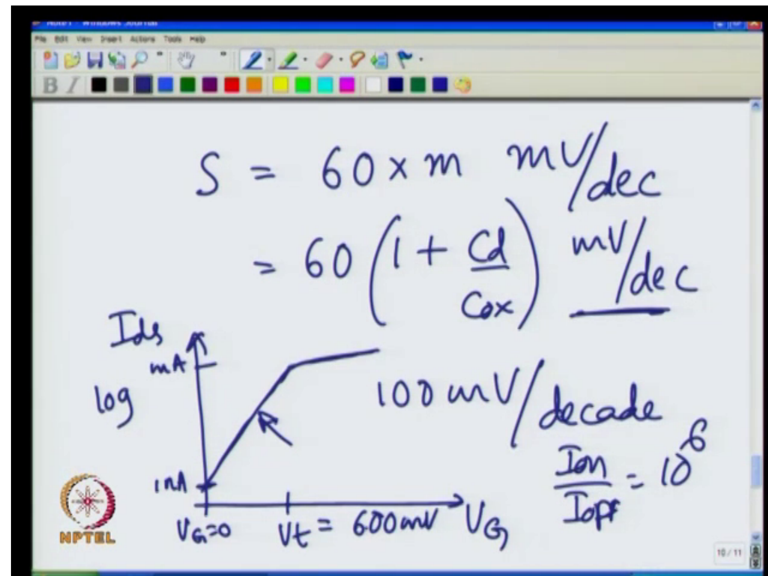
(Refer Slide Time: 41:54)



Now, I define a quantity called sub threshold slope which is inverse of this. In fact, that is why the unit for sub threshold slope is typically millivolt per decade this is you see this is volt and this is log 10 which is decade variation right you know because you have taken log the y axis is you know how many decades you are changing right and that is the unit that you have come up to right. So, now, this essentially one over slope which means you invert this right. So, your sub threshold slope which is extremely important quantity for transistor now we have observed and we have derived this from first principle as m times log 10 to the base e turns out L n natural log of 10 is 2.3 and k T over q at room temperature is 25.8 millivolt approximately. You multiply that with 2.3 you get 60 a good number to remember.

So, in other words your sub threshold slope is 60 times m millivolt per decade because I took thermal voltage in millivolt right, that is why it is coming in millivolt per decade.

(Refer Slide Time: 43:23)



As you know m is really 60 times one plus Cd over C ox millivolt per decade. What is the significance of this sub threshold slope? You see sub threshold slope as you can see depends on a absolute temperature. So, if your has to operate no matter what technology you are using it will always remain at the same value it cannot change whether it is 1 micron technology or 50 nanometer or 40 nanometer k T over q is same, L n 10 2.3 is same, that 60 factor will never change.

What this is telling you is that the very first picture that we saw here below threshold voltage current is small; however, current decreases exponentially it will never go to 0 it will decrease with exponential rate and the rate is fixed and the rate is fixed which essentially says every certain millivolt current decreases by so many decades. In other words if I were to now look at the drain current versus gate voltage with the drain current in log access what you will see is the following this is Ids, now this I am plotting in log scale.

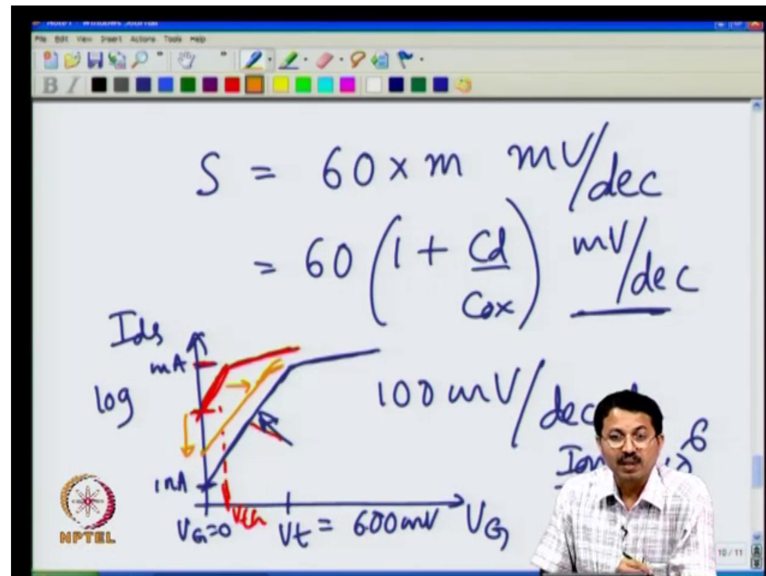
This is your Vth value, very interestingly what you see is this is above Vth, above Vth the variation is only linear or quadratic. So, it looks almost flat, but below Vth current starts dropping very fast, but at V G equal to 0 this is V G which is your off state current is never 0 remember that because this slope is determined by sub threshold slope the typical sub threshold slope is always more than 60 let us say its 100 millivolt per decade. It means that below threshold voltage every 100 millivolt decrease in gate voltage

decreases yours transistor current by 1 order of magnitude in other words if your V t where to be 600 millivolt 0.6 volt which is what you used to have in older generation technology.

This is 100 millivolt per decade and let us say your on current you see on current really does not change once your above [vocalized-noise] t its only a linear dependence right may be here you will have one milliampere and here you may have maybe three four milliampere. Or, let us say this current is in milliampere that is on current of a transistor. If you have a 600 millivolt V t threshold voltage and 100 millivolt per sub threshold slope you can very easily calculate that when I come from threshold voltage 2 0 voltage gate voltage my current should come down by 6 orders of magnitude because every 100 millivolt decrease in gate voltage decreases the current by 1 decade. So, 600 millivolt decrease should decrease the current by 6 decades in other words if you have a milliampere current.

Your off current is nanoampere very good because your I on by I off ratio is now 10 to the 6 on current is milliampere off current is nanoampere that is what we said at least we need to have 10 power 6 on to off current ratio we have a problem today the problem is as per the scaling theory we have been decreasing the dimensions of the transistor. We have decreased the gate voltage when we decrease all the voltages supply voltage today's transistors operated 1 volt. So, we no longer have 600 millivolt threshold voltage today's transistors will have to have threshold voltage of the order of 0.2 volt, 200 millivolt threshold voltage.

(Refer Slide Time: 47:46)

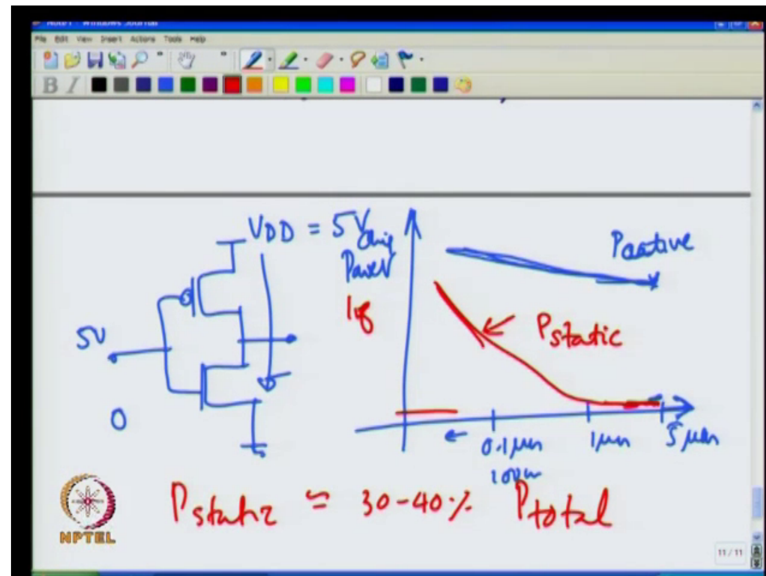


In other words I am actually looking at a highly scaled transistor today which has a threshold voltage of 0.2 volt from 0.2 volt its start dropping, but look it can drop only at this rate. It cannot drop faster than this. Because this is a fundamental limit this is the fundamental limit which is fixed by our derivation that we had k T over q does not change 0.3 does not change and the m no matter what you do is anywhere between 1.1 1.2 you know 1.5 you know in that range.

You may try to do little better and maybe bring it very close to 60, but 60 is ideal you see you can never do it better than 60 right. So, this cannot be improved, so what does it mean? My on to off current ratio is going to degrade phenomenally. So, when transistor is supposed to of be off you will have large current, if fortunately if you are done some good transistor design you have improved this sub threshold slope from let us say 100 millivolt to 70 millivolt volt let us say you know.

Then you know if you have 200 millivolt as your threshold voltage or 210 millivolt just to, so that we get round numbers right. 70 millivolt is a sub threshold slope 70 times 3 is 220 millivolt and hence when I come from 220 millivolt to 0 volt I would decrease the current by 3 orders of magnitude. Milliampere current will only come to micro ampere, my on to off current ratio will suffer when transistor is supposed to be off you will have lot of leakage current right. So, this is a fundamental problem we have today.

(Refer Slide Time: 49:33)



In fact, that is the reason why if you look at that historical trends in CMOS technologies. In fact, we always said that CMOS is so nice compared to bjt because there is no static power dissipation you see this is my CMOS inverter circuit when I apply let us say V DD is older generation technology 5 volt I apply 5 volt here p mos is off, p mos is really off because you know in order to you turn on p mos you have to apply 0 you know when you have 5 its complement right, you really are having very very low current because your threshold voltage is here of the order of 1 volt right.

You have luxury to have several decades of decrease in current and hence this leakage is this current is almost negligible. Similarly when you have a 0 volt which is a static case right when your input logic is 0 n mos is supposed to be off right it is indeed of because n mos will have a positive threshold close to 1 volt 1000 millivolt, even if I have a 100 millivolt per decade which is easily achievable my current will come down by 10 decades 10 orders of magnitude off current is 10 orders of magnitude lower than on current and hence no static power there is no static power we always ignore static power and that is why if you look at the power as a function of C mos technology generation you know let us say starting from 5 micrometer technology going to 1 micrometer and today going to you know 0.1 micrometer which is 100 nanometer and below right. Today we are talking of 60 nanometer 40 nanometer and all that.

Power has active power that is when circuit is switching it is doing some work for you it is a useful power because you are extracting some work from the circuit. Passive power is circuit is not doing anything ideally it should be 0 power, but it could still consume power, but in c mos in older generation technology you know your active power when we look at the active power your active power you know would be somewhere out here let us say and your passive power or a you know off state power or was almost 0 very very small. In fact, you know I can plot this in a log scale right compared to active power this is several orders of magnitude lower that used to be the case.

But today because of the sub threshold leakage of state current is trying to catch up with the on state current you see and what has happened over the technology generation is that with technology scaling power per units transistor decreases, but we also put large number of transistor in a chip and chips are becoming bigger and bigger and hence the total chip power is increasing. Although if you look at one unit gate that is more efficient as we have already seen earlier as per the scaling theory, but now we are talking of total chip power here right this is a chip power which includes active power p active and leakage power leakage power was almost negligible, but today leakage power really looking like this, this is p static. In today's state of the technology especially high performance state of the technology, your static power can be as much as thirty to forty percent of your total power you know among us.

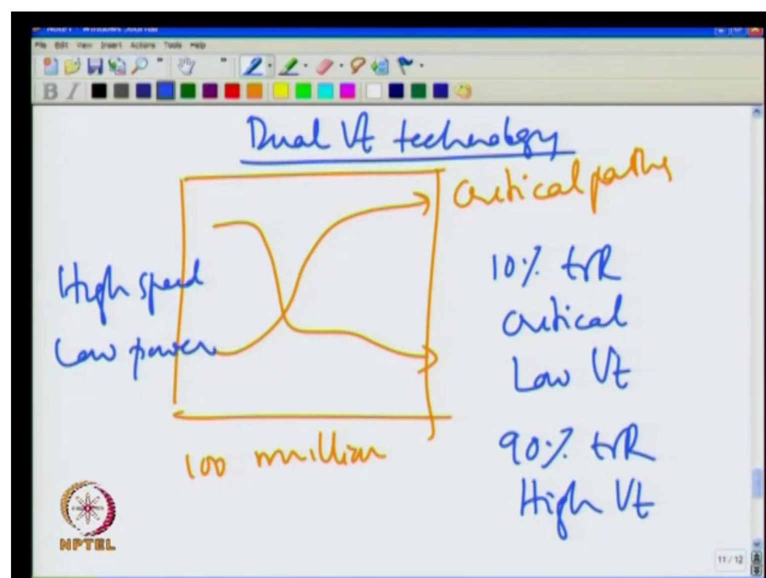
That is of course, you do lot of very interesting things both at the technology level and system architecture level we put the device instrument it should not really do anything your mobile phone screen shuts off right. So, these are some very interesting system architecting that we do, so that you we essentially disconnect the supply voltage because if we supply voltage is there all these transistor will leak phenomenally right, but this is really a very serious problem and especially because today we are talking of mobile devices right and mobile devices are becoming technology driver and we need to really arrest this static power. And this static power increase is simply because of the fact that I have to decrease a threshold voltage you may ask why increase the decrease the threshold voltage you have 1 volts supplied, you still make the threshold voltage point 6 and that is exactly what is done in. So, called low power technologies.

If you really are worried not about performance if you have a microprocessor which goes in your cell phone and a micro processor which goes in a desktop pc both of them

presumably are done on 45 nanometer let us say; however, the transistors that go in desktop pc microprocessor are really called high performance low V t transistors because you do not worry about power there. You get power from your you know wall socket you know its continuously supplies power, but you cannot afford that you know you do not want to recharge your battery of a cell phone every hour right. You are willing to hit get hit on performance you know if the download file download takes maybe 10 minutes instead of 5 minutes that is ok, but you do not want to waste power and hence even though its 45 nanometer technology you may actually say that I do not really need this V t.

I would really increase this V t and hence my on current will accordingly decrease I on current would not be up here on current will be lower, but what you really get benefit is your of current can be decrease. So, you know you do lot of such tricks really I know it also turns out you have today technologies called dual V t technologies you see the idea there is that you know typically you know when you have chip right you know such as the microprocessor chip which is essentially a very high performance chip. It turns out if the speed of the microprocessor chip is really governed by some transistors in the so called critical paths of this circuit.

(Refer Slide Time: 56:04)



In other words you have let me get this correct here, you have this chip this chip may have million transistor few 100 million transistor it turns out at the end of the day when

you actually do lot of circuit design and analysis there may be only specific transistors in. So, called critical paths critical paths from input to the output that matter all others really do not matter it only is important to make these transistors in critical path very fast because these are the bottle necks in your speed and it also turns out typically the transistor number of transistors in critical paths maybe 10 percent of the total number of transistors. So, let us say 10 percent of the transistors of 100 million transistor that you have are in critical paths and these can be done with low V t, whereas 90 percent of the transistors that you have in the rest of the chip can be high V t transistors.

Now, you are trying to get best of both words. You want speed right and you have discovered that speed is only dependent on certain number of transistor in your chip make those transistors very fast. They may also leak more, but that is it is only a part of the total number of transistor right, 90 percent of your transistor a much lower leakage current.

So, there by using the so called dual V t technology in other words you have a n channel transistors you will have 2 flavors of n channel transistor, you have low V t n channel transistors which will be sitting out here very low V t very large current, but very small fraction of the transistor and large number of transistors will be sitting out here which are high V t transistors and as a result of that you get high speed and low power.

By doing this you know some trick intelligent we have doing process technology and transistor design all this is not really there in electric fields scaling theory right I mean it all assumes that all transistors are identical in you know single V t for all transistor, but now we are saying now look this is a problem, but this problem can be solved by having this intelligent solution and hence we adopt this. So, you know that sort of concludes the target that we have today. So, narrow width effect is important that impacts threshold voltage we used trench isolation to overcome that sub threshold conduction is always there right because of the diffusion current your turned off the MOS channel, but there is a diffusion current and you know because of that you have any implication on the of state current of the transistor you have ways to intelligently design technologies to overcome those difficulties as I well. So, let us stop here and we will continue in the next lecture.