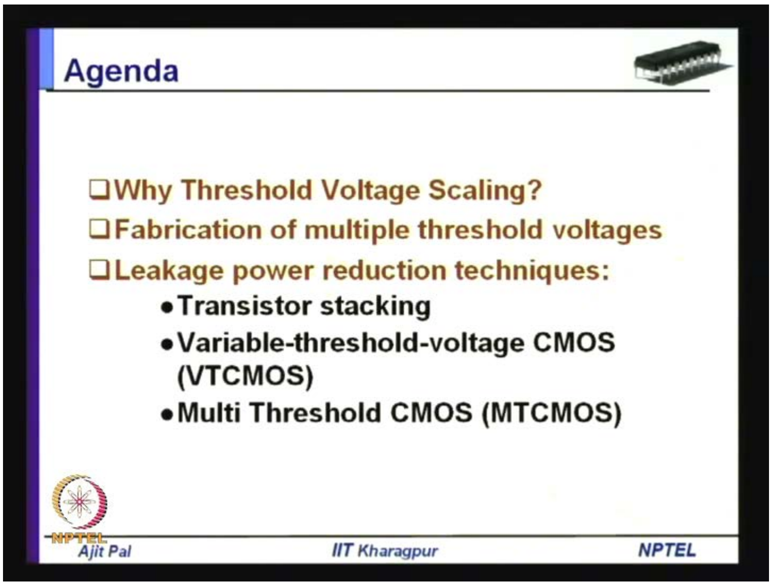
Low Power VLSI Circuits and Systems

Prof. Ajit Pal
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Module No. #01
Lecture No. #32
Minimizing Leakage Power – I

Hello and welcome, to today's lecture on minimizing leakage power. We have already discussed various techniques about minimizing dynamic power, as you know the leakage power is very important, and it is increasing as you are going from one technology generation to next technology generation. So, from today onwards, in few lectures I shall discuss about how leakage power can be minimized.

(Refer Slide Time: 00:48)



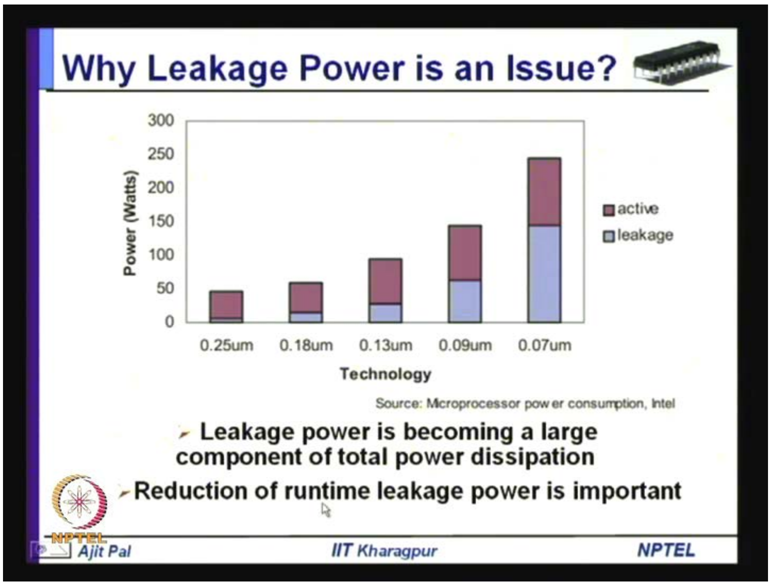
Agenda

❑ Why Threshold Voltage Scaling?
❑ Fabrication of multiple threshold voltages
❑ Leakage power reduction techniques:
- Transistor stacking
- Variable-threshold-voltage CMOS (VTCMOS)
- Multi Threshold CMOS (MTCMOS)

NPTEL
Ajit Pal
IIT Kharagpur
NPTEL

So, here is the agenda of today's lecture; first I shall discuss about why threshold voltage scaling, as we shall see threshold voltage plays a very important role in minimizing leakage power. So, we shall discuss about the role of threshold voltage, and we shall see how threshold voltage scaling can be done to reduce leakage power. Then, we shall discuss about various techniques for fabrication of multiple threshold voltages on a single chip, and that is necessary for realizing the low power circuits, particularly for minimizing leakage power. Then, we shall discuss about leakage power reduction techniques; like transistor stacking, variable-threshold-voltage CMOS, VTCMOS and multi threshold voltage CMOS, MTCMOS. These techniques we shall discuss in today's lecture.

(Refer Slide Time: 01:43)



Why Leakage Power is an Issue?

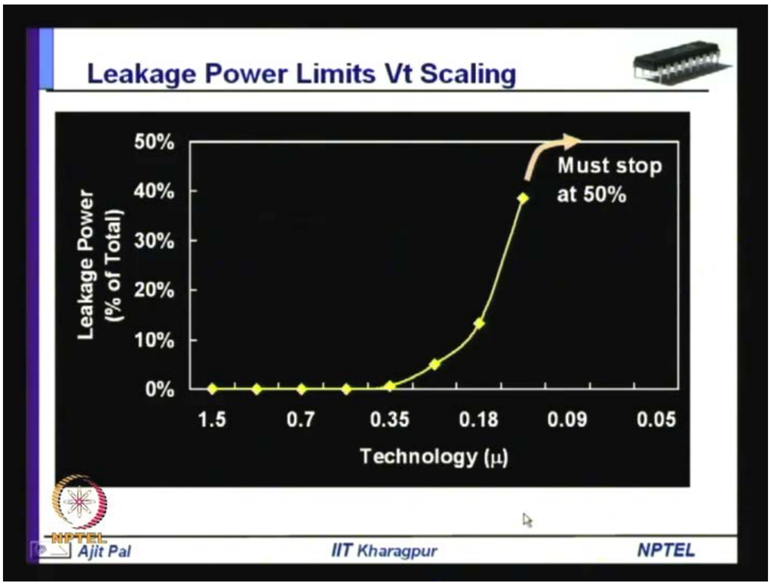
Source: Microprocessor power consumption, Intel

- Leakage power is becoming a large component of total power dissipation
- Reduction of runtime leakage power is important

Ajit Pal IIT Kharagpur NPTEL

Let us start with why leakage power is an issue, I have already discussed about it, as you are going from one technology generation to the next technology generation, the leakage power component is increasing and as you can see in 70 nanometer technology, the leakage power dissipation has overtaken the dynamic power. So, not only that leakage power is becoming a large component of total power dissipation, as it is clear from this, so more than 50 percent. And also another important issue is runtime leakage power, earlier the leakage power was minimized only when the circuit was in the standby mode, because in the active condition, the percentage of leakage power was insignificant compared to dynamic power. So, at runtime people were primarily considering the reduction of dynamic power, but now that is not so. So, even during runtime when the circuit is in operation, it is essential to reduce the leakage power; that is why the reduction of runtime leakage power has become important.

(Refer Slide Time: 02:57)

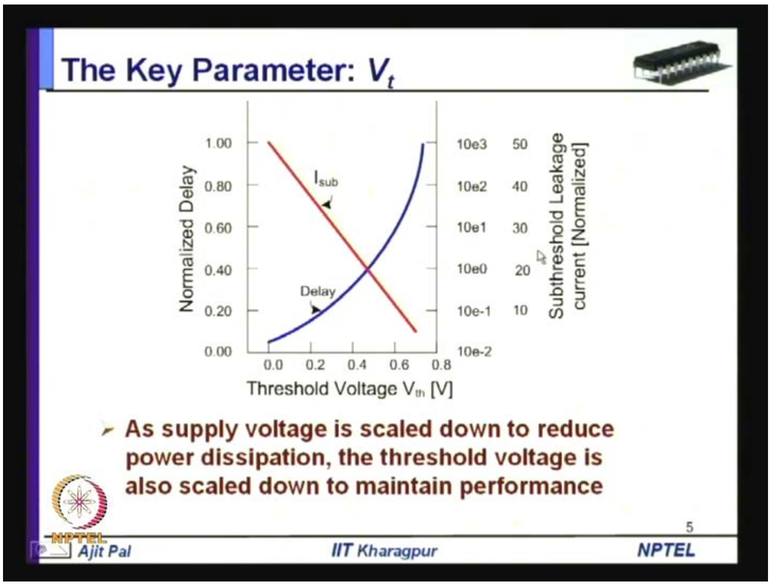


Leakage Power Limits Vt Scaling

Ajit Pal IIT Kharagpur NPTEL

And I have already shown you this particular diagram; you can see the leakage power is increasing as you are going from one technology generation to the next technology generation. And requirement is that as it is crossing 50 percent of the total power, you must stop, somehow the increase in leakage power has to be stopped somehow, and what is that somehow, we shall discuss starting from today’s lecture.

(Refer Slide Time: 03:31)



The Key Parameter: V_t

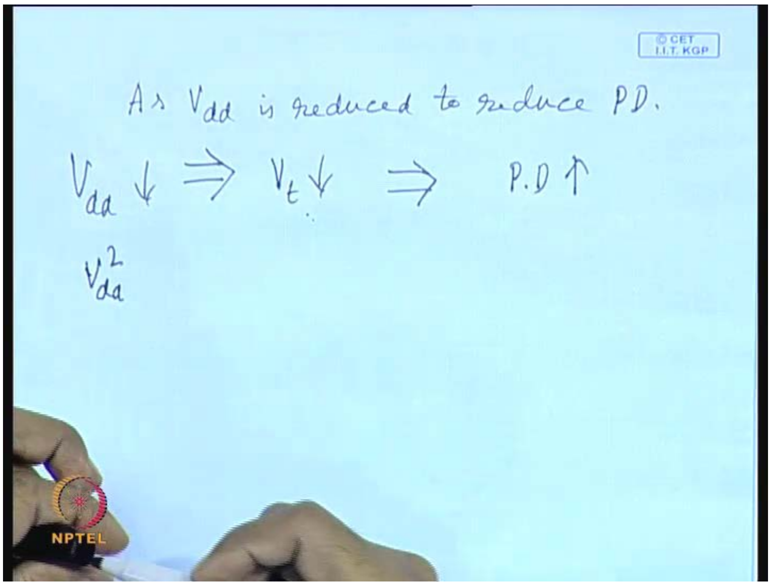
- As supply voltage is scaled down to reduce power dissipation, the threshold voltage is also scaled down to maintain performance

5

Ajit Pal IIT Kharagpur NPTEL

As I mentioned, in minimizing leakage power, the key parameter is a threshold voltage V t, and the threshold voltage is actually influences two important parameters; one is your non-delay, here the normalized delay has been plot, as you can see normalized delay reduces more or less quadratically, as you reduce the threshold voltage; that means, smaller the threshold voltage, smaller is the delay. So, from the view point of performance, it is essential to have smaller threshold voltage. So, smaller threshold voltage in a circuit leads to faster operation and higher performance. On the other hand, if you look at this particular diagram, this curve the red line, that is your sub threshold leakage current, which is one of the leakage current components, and you notice that, you must notice that on this side, that sub threshold leakage current has been plotted in the logarithmic scale. Even by plotting in the scale, you can see it is increasing linearly, as you are reducing the threshold voltage; that means, as the threshold voltage is reduced, the sub threshold leakage current increases expansively.

(Refer Slide Time: 05:00)



As V_{dd} is reduced to reduce PD.

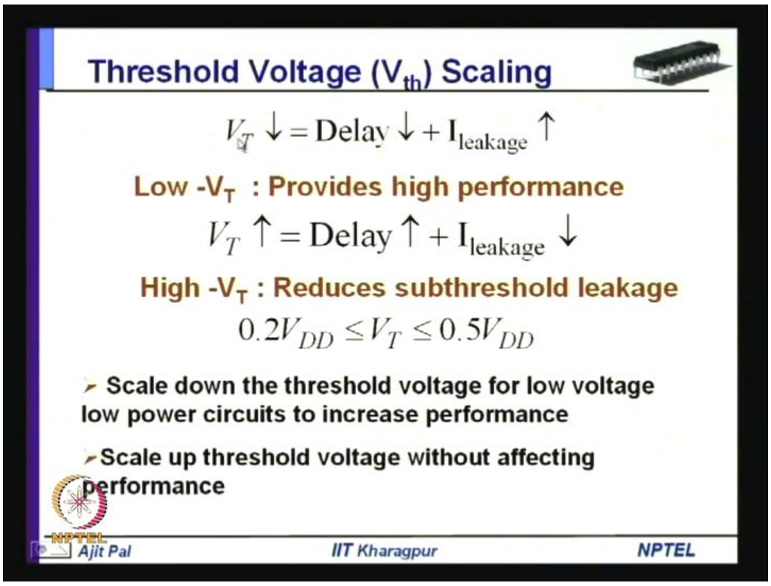
$V_{dd} \downarrow \Rightarrow V_t \downarrow \Rightarrow P.D \uparrow$

V_{dd}^2

© CET
I.I.T. KGP
NPTEL

And as you know, as you are reducing the supply voltage V d d, as V d d is reduced; reduced to reduce power dissipation, what you have to do. You have to, this implies that you have to reduce V t V t has to be lowered; that means, as V d d is lowered, you have to reduce threshold voltage, you have seen that in constant field scaling, but as you do so, it increases the power dissipation. So, it has become a kind of vicious circle, you know threshold supply voltage is reduced to reduce the power dissipation, particularly the dynamic power dissipation as you know the dynamic power dissipation is proportionally to V d d square, leakage power is also reduced. So, supply voltage reduction, supply voltage scaling is a very important technique in reducing power dissipation, as you do so, you are reducing the threshold voltage, and as you reduce the threshold voltage power dissipation is increasing. So, it has turned out to be a vicious circle, what is the way out from this vicious circle.

(Refer Slide Time: 06:14)



Threshold Voltage (V_{th}) Scaling

$$V_T \downarrow = \text{Delay} \downarrow + I_{leakage} \uparrow$$

Low -V_T : Provides high performance

$$V_T \uparrow = \text{Delay} \uparrow + I_{leakage} \downarrow$$

High -V_T : Reduces subthreshold leakage

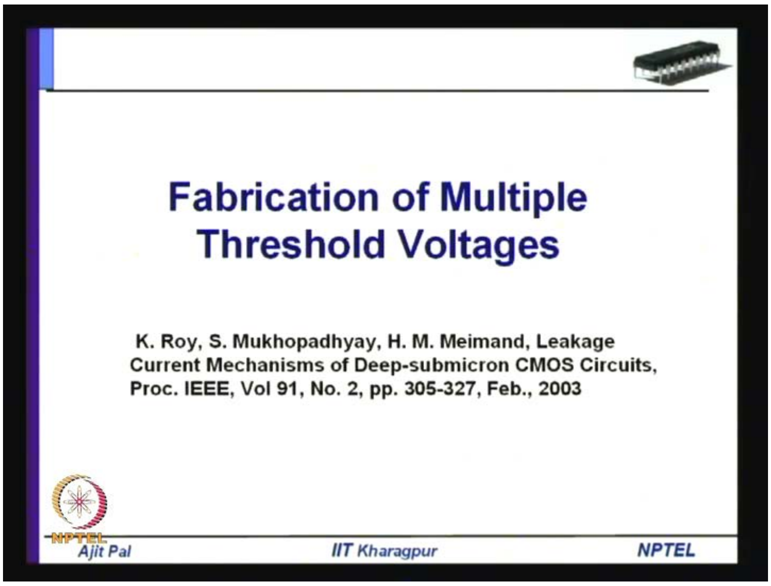
$$0.2V_{DD} \le V_T \le 0.5V_{DD}$$

- Scale down the threshold voltage for low voltage low power circuits to increase performance
- Scale up threshold voltage without affecting performance

Ajit Pal IIT Kharagpur NPTEL

You can see smaller the threshold voltage, smaller the delay, larger is a leakage current. On the other hand, I mean you can say that low V t provides high performance, on the other hand as you increase the threshold voltage; larger is the delay lower is the leakage current. So, high V t reduces sub threshold leakage current, and as you know in a circuit the threshold voltage has to be kept in the range of point 2 V d d to point 5 V d d, to get you know reasonable performance at the time reasonable leakage power dissipation. So, from this, our observation is we can scale down the threshold voltage for low voltage low power circuits to increase performance. On the other hand, we can scale up threshold voltage without affecting performance. So, our objective will be to judiciously use, both high threshold voltage and low threshold voltage. Low V d d and high low threshold voltage and high threshold voltage in such a way that we shall not compromise on performance, at the same time we shall reduce the leakage power dissipation, that is the main challenge of that leakage power minimization or we can say threshold voltage scaling.

(Refer Slide Time: 07:41)



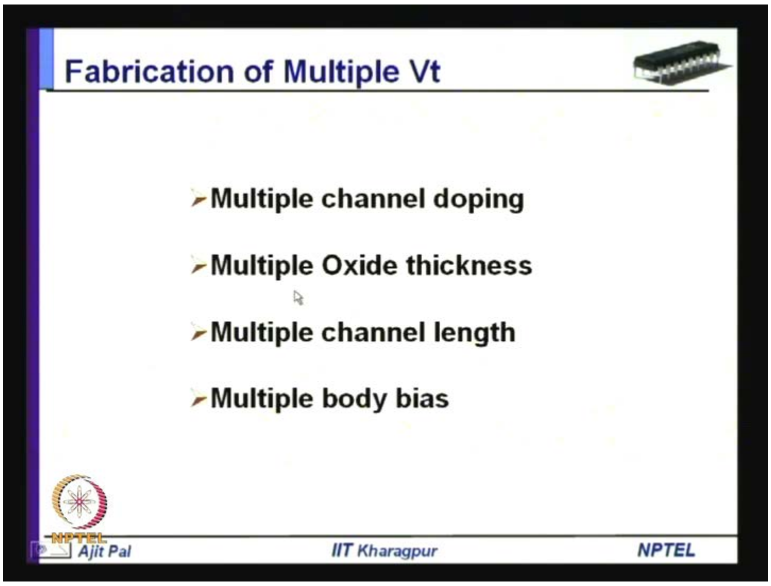
Fabrication of Multiple Threshold Voltages

K. Roy, S. Mukhopadhyay, H. M. Meimand, Leakage Current Mechanisms of Deep-submicron CMOS Circuits, Proc. IEEE, Vol 91, No. 2, pp. 305-327, Feb., 2003

Ajit Pal IIT Kharagpur NPTEL

So, before we discuss various techniques, we shall discuss about fabrication of multiple threshold voltages, because as you have seen in a single circuit, you will be fabricating multi transistors of multiple threshold voltages, how can it be done. So, there are various approaches which I shall discuss, and this has been taken from a paper by Kaushik Roy and his group, published in proceedings of the I triple E in the year 2003. The title of the paper is leakage current mechanisms of deep-submicron CMOS circuits.

(Refer Slide Time: 08:22)



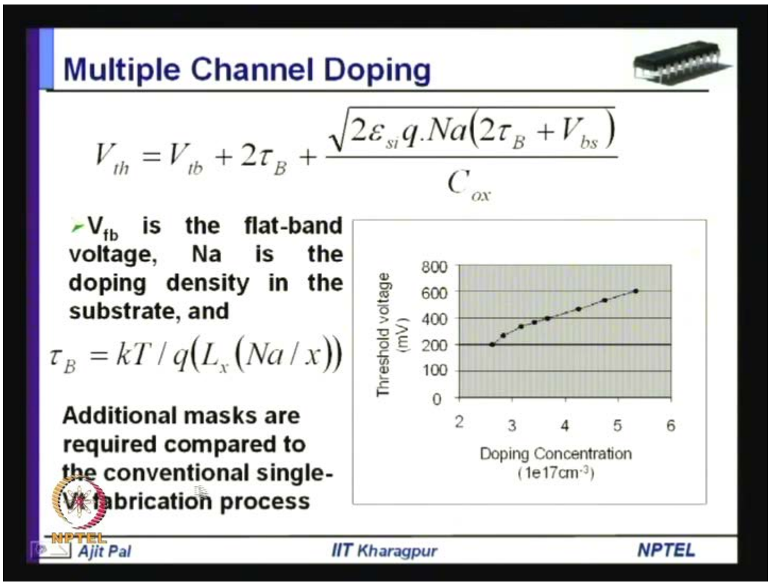
Fabrication of Multiple Vt

- Multiple channel doping
- Multiple Oxide thickness
- Multiple channel length
- Multiple body bias

Ajit Pal IIT Kharagpur NPTEL

So, in this paper we will find various techniques of realizing multiple threshold voltages, and the techniques are; number one is multiple channel doping. Second is multiple oxide thickness. Third is multiple channel length and multiple body bias. These are the four basic approaches used for fabricating multiple threshold voltages. Let us see how can it be done.

(Refer Slide Time: 08:41)



Multiple Channel Doping

$$V_{th} = V_{tb} + 2\tau_B + \frac{\sqrt{2\varepsilon_{si} q.Na(2\tau_B + V_{bs})}}{C_{ox}}$$

- V_{fb} is the flat-band voltage, Na is the doping density in the substrate, and

$$\tau_B = kT/q(L_x(Na/x))$$

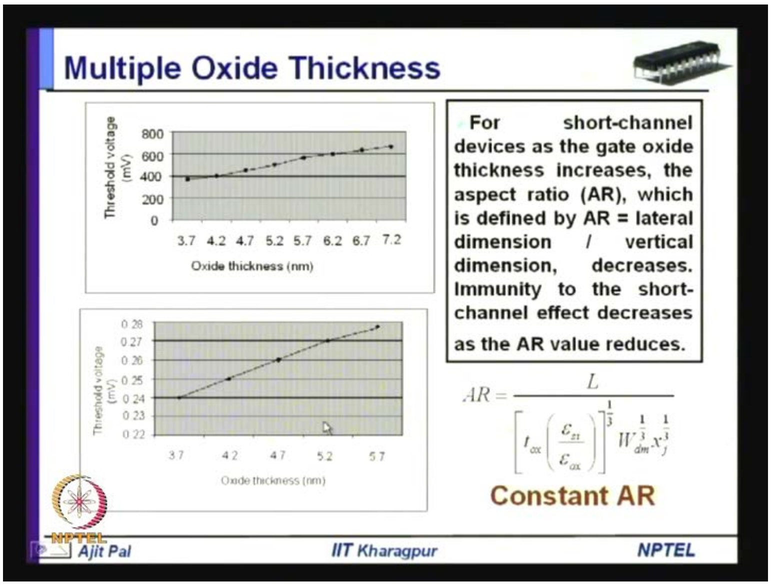
Additional masks are required compared to the conventional single-Vt fabrication process

Ajit Pal IIT Kharagpur NPTEL

First one is multiple channel doping, so multiple channel doping is a technique, where you know within the channel you have to perform a kind of ion implantation, to achieve some kind of some level of doping, and you can see how the threshold voltage is related to the doping concentration N a. So, N a factor is here and N a factor is also in this factor tau b. So, V t b here is flat-band voltage, N a is the doping density in the substrate, and you can see tau b is equal to k T by q, into L x into N a by x, where x is the kind of channel presentation. So, the threshold voltage, how the threshold voltage varies with doping concentration is shown in this particular diagram. Here as you can see as the doping concentration is increased from one to the power ten to the power seven per cubic centimeter, this is the scale in the x axis.

So, that means, here it is two point five into ten to the power seven per cubic centimeter, two if you make it more than five into ten to the power seven per cubic centimeter, it will increase the doping concentration in this way. Then you find that the threshold voltage varies from 200 mille volt to 600 mille volt. So, the threshold voltage can be varied by changing the doping concentration in the channel region. However, if you want to realize two transistors having two different threshold voltages, you have to perform ion implantation twice, using two separate masks. So, that will indeed increase the cost of fabrication. So, additional masks are required compared to conventional single V t fabrication process. So, you can achieve multiple threshold voltage, realized multiple threshold voltage transistors in a single circuit. However, you have to incur additional cost, in terms of additional processing steps, fabrication steps and additional masks.

(Refer Slide Time: 11:18)



Multiple Oxide Thickness

For short-channel devices as the gate oxide thickness increases, the aspect ratio (AR), which is defined by AR = lateral dimension / vertical dimension, decreases. Immunity to the short-channel effect decreases as the AR value reduces.

$$AR = \frac{L}{\left[t_{ox}\left(\frac{\varepsilon_{si}}{\varepsilon_{ox}}\right)\right]^{\frac{1}{3}} W_{dm}^{\frac{1}{3}} x_j^{\frac{1}{3}}}$$

Constant AR

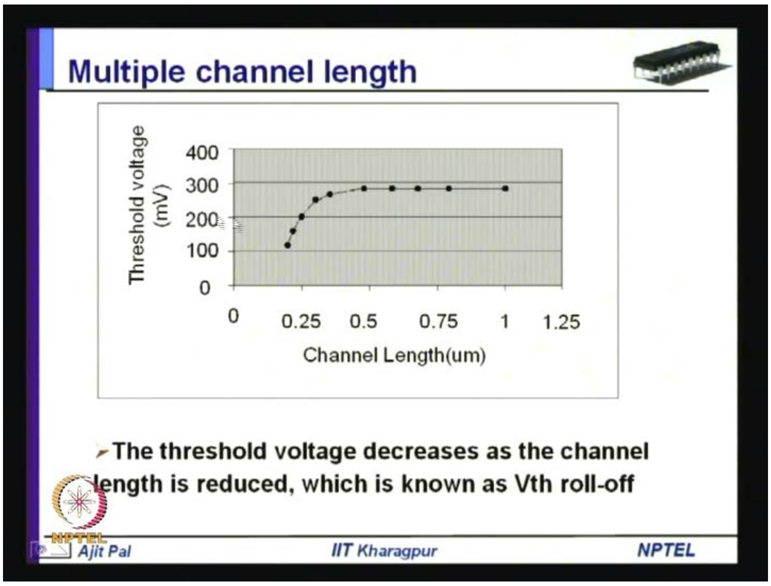
Ajit Pal IIT Kharagpur NPTEL

So, this is one technique that can be used for realizing transistors with multiple threshold voltages. Second technique is multiple oxide thickness, as you know you have got a silicon dioxide before on top of the channel. So, here is your source and drain region, this is source and drain, and here you put a silicon dioxide layer, and on top of that polysilicon is deposited and this forms the gate. So, this thickness of the silicon dioxide plays a very important role in deciding the threshold voltage. So, it is a common sense that smaller is the thickness, you will require lower voltage to create the channel region, and larger the thickness you will require larger voltage, gate voltage to create the inversion region in the channel. So, I mean oxide thickness that you had used the oxide thickness; you can see you require lesser threshold voltage. So, there are two curves are shown in this particular diagram. Here one is with constant aspect ratio, so for top one is without maintaining constant aspect ratio.

So, for short-channel devices as the gate oxide thickness increases, the aspect ratio which is defined as lateral dimension by vertical dimension decreases. So, this is a problem particularly in short-channel devices. So, if you do not bother about the variation of this aspect ratio, lateral dimension by vertical dimension, then you can get a variation of threshold voltage starting from, you can say a little less than 400 mille volt to little more than 7600 mille volt, as you varied the oxide thickness from 3.7 nanometer to 7.2 nanometer. So, by varying the oxide thickness in this range, you can achieve variation in the threshold voltage. So, what does it really mean, here also you know whenever you have to fabricate two different oxide thicknesses, again it will require additional masks in the oxidation process, as you know that silicon dioxide is deposited by a process known as dry oxidation. So, whenever you do dry oxidation, then that oxide thickness can be controlled by controlling the time of oxidation, duration of oxidation.

So, that means, you have to fabricate two different transistors by depositing different amounts of silicon dioxide, and by using the dry oxidation process. So, again this will require additional step, and the lower curve where the aspect ratio is maintained constant, where aspect ratio is l by this the t o x into epsilon s I by epsilon o x to the power 1 by 3 into w 1 by 3, that is a width and x into 1 by 3 j. So, if you maintain this aspect ratio constant, then of course your variation is less as we can see, as you vary the oxide thickness from three point seven to five point seven. The threshold voltage varies only from point two four to point two eight, you can say close to two point two eight. So, there is small variation in threshold voltage, whenever you are maintaining the aspect ratio constant. However, if you do not maintain the aspect ratio constant then you can have larger variation in threshold voltage, by having different oxide thicknesses. So, this is the second technique.

(Refer Slide Time: 15:05)

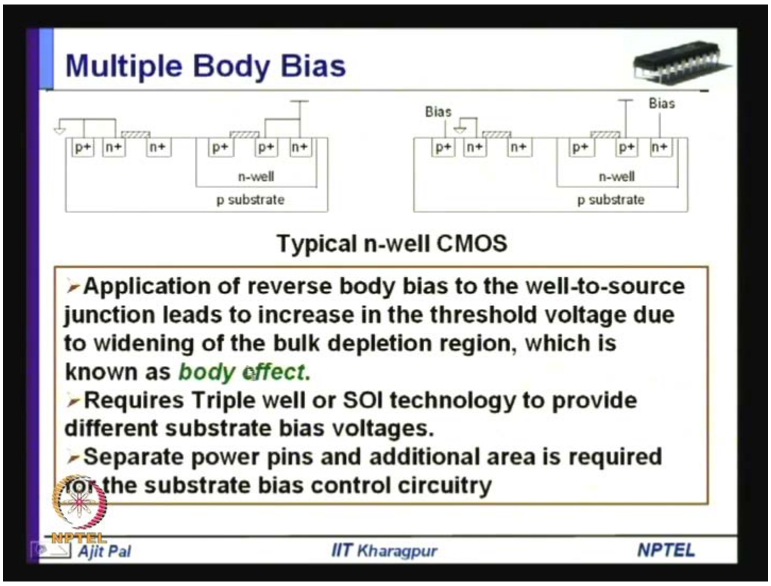


Multiple channel length

- The threshold voltage decreases as the channel length is reduced, which is known as Vth roll-off

Coming to the third technique, where we use multiple channel length, you may recall that I discussed about a phenomenon known as channel length modulation, no sorry channel length roll-off, where what happens you know, as you reduce the channel length and particularly whenever it is less than you can say point two five, a little more than point two five, say point three or so. You can see normally the threshold voltage is independent of the channel length, threshold voltage does not depend on the channel length. However, you can see around this, there is a kind of roll-off, V t h roll-off, this is known as V t h roll-off, the threshold voltage rolls off as you reduce the channel length. So, for small variation of channel length there is a large variation of threshold voltage. So, this phenomenon can be utilized, and this is known as the threshold voltage decreases as the channel length is reduced, which is known as V t h roll-off. So, this happens only in small dimension transistors and particularly represents a MOS transistor. So, this can be utilized to have multiple threshold voltages, by varying the channel length, by small amount and you can see there is a significant change in the threshold voltage. So, say little more than 100 mille volt to about a little less than 300 mille volt, there is a variation in threshold voltage, as the channel length varies from may be point two micron to say point five micron. So, this is a very interesting technique for changing the threshold voltage by varying the channel length. This is the third technique.

(Refer Slide Time: 17:05)



Multiple Body Bias

Typical n-well CMOS

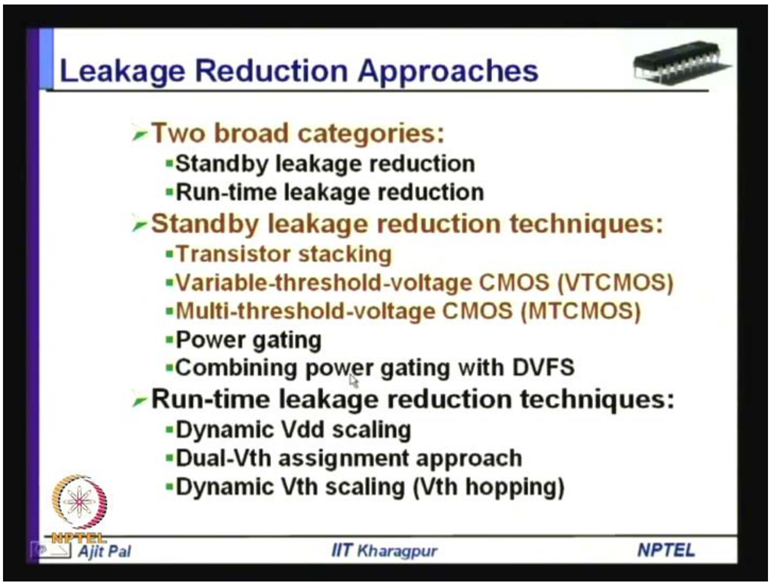
- Application of reverse body bias to the well-to-source junction leads to increase in the threshold voltage due to widening of the bulk depletion region, which is known as *body effect.*
- Requires Triple well or SOI technology to provide different substrate bias voltages.
- Separate power pins and additional area is required for the substrate bias control circuitry

Ajit Pal IIT Kharagpur NPTEL

Fourth technique is multiple body bias, as you know normally the substrate is connected to the source and then it is grounded, as you can see here is a p plus region and that is connected to the source and it is grounded. Similarly, the n-well is connected to V d d by with the help of a n plus region formed in this n-well. So, this is what is done normally, and in such a case we call that there is no bias, or body bias is zero, what can it be done this p plus or the substrate can be separated from this source, as you can see. And, here also this n-well can be separated, and you can take out as a separate terminal. Then you can apply different voltages with respect to source. For example, the substrate can be, you can apply different voltage to this substrate with respect to the source, and it has been found that if you apply a reverse voltage here, you know that increases the depletion region, leading to increase in the threshold voltage. So, application of reverse body bias is to bias to the well-to-source junction, leads to increase in the threshold voltage due to widening of the bulk depletion region and this is known as body effect. I have already discussed about it.

However, whenever you want to apply body bias voltage to the substrate and to the well, you have to use a sophisticated fabrication technique known as triple well technology, or SOI silicon on insulator technology to provide different substrate bias voltages. So, both triple well as well as silicon on insulator technology are costly fabrication technique, and as a consequence you can definitely implement this, but the fabrication cost increases. So, not only that you will require separate power pins for this substrate, as well as for the well and you will require additional area for routing those lines, and again you will require some substrate bias control circuitry to apply different bias voltages depending on your requirement. So, this is how by applying multiple body bias voltages, you can realized transistors of different threshold voltages. Earlier we have already discussed how the threshold voltage varies with the body bias.

(Refer Slide Time: 19:44)



Leakage Reduction Approaches

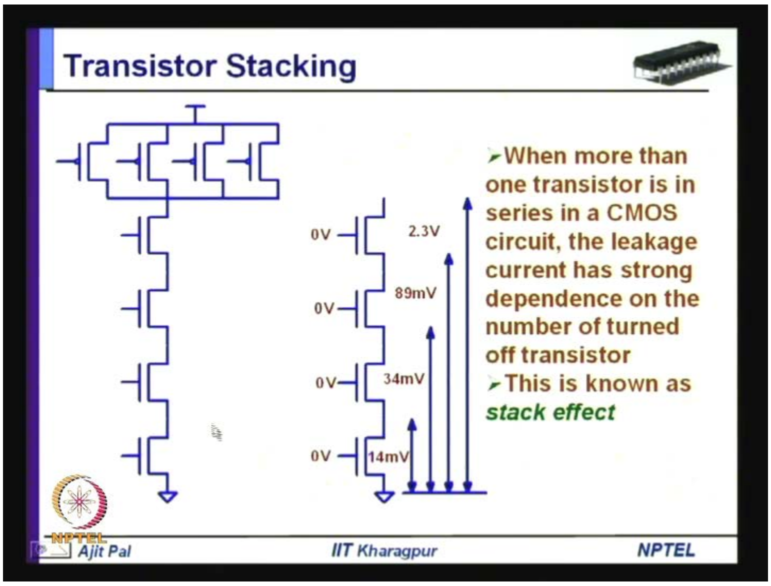
- Two broad categories:
  - Standby leakage reduction
  - Run-time leakage reduction
- Standby leakage reduction techniques:
  - Transistor stacking
  - Variable-threshold-voltage CMOS (VTCMOS)
  - Multi-threshold-voltage CMOS (MTCMOS)
  - Power gating
  - Combining power gating with DVFS
- Run-time leakage reduction techniques:
  - Dynamic Vdd scaling
  - Dual-Vth assignment approach
  - Dynamic Vth scaling (Vth hopping)

Ajit Pal IIT Kharagpur NPTEL

Now, coming to the leakage reduction techniques, we can divide the leakage reduction techniques into two broad categories; number one is known as standby leakage reduction, second one is known as run-time leakage reduction, what do you really mean by them, as I was mentioning earlier when the leakage component was very small, compared to the dynamic power, then leakage power reduction techniques were applied only when the circuit was in standby condition. So, the techniques which are used for reducing the leakage power when the circuit is in standby, that will not in operation, you are not applying clock, you are not applying input. In such a case you apply some technique to reduce the leakage power; those are known as standby leakage power reduction techniques. On the other hand, whenever the circuit is in operation circuit is running, in such a condition you are interested in reducing leakage power along with dynamic power, and those are known as run-time leakage power techniques. So, the standby leakage power reduction techniques.

There are several techniques; number one is known as transistor stacking, second one is variable threshold voltage, CMOS VTCMOS, third one is multiple threshold voltage CMOS, M TCMOS and actually this MTCMOS technique has become more matured, and this is known as power gating or power management. So, in the next lecture I shall discuss in detail about this. Then there are other another technique combining power gating with dynamic voltage and frequency scaling. So, earlier we have discussed dynamic voltage and frequency scaling technique, which can be combined with power gating to reduce standby leakage power reduction which I shall discuss in the next lecture, and then there are run-time leakage reduction techniques and there I shall discuss about three run-time leakage reduction techniques, dynamic V d d scaling, dual V t h assignment approach, and dynamic V t h scaling, which is also known as V t h hopping in a particular situation, when you have got only two threshold voltages. So, we shall discuss these techniques one after the other and today we shall focus on these three approaches; transistor stacking, variable-threshold-voltage CMOS, and multi-threshold voltage CMOS MTCMOS. So, these three techniques we shall discuss today.

(Refer Slide Time: 22:24)



Transistor Stacking

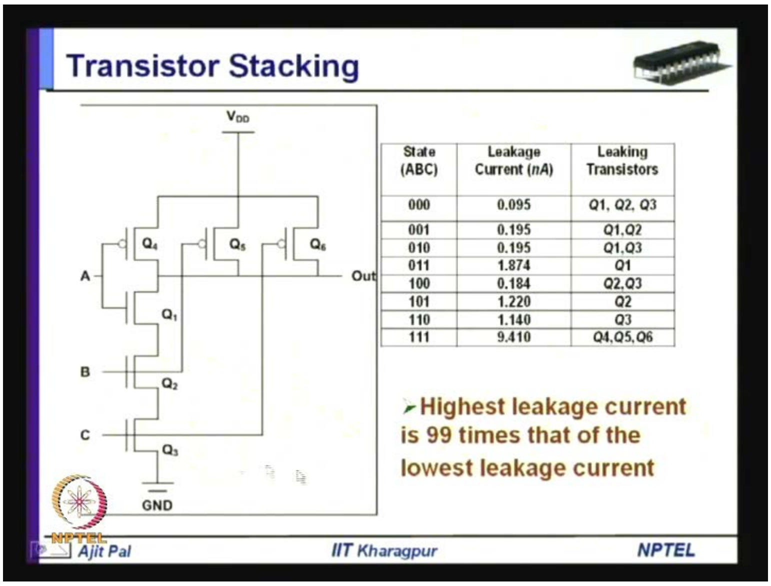
- When more than one transistor is in series in a CMOS circuit, the leakage current has strong dependence on the number of turned off transistor
- This is known as *stack effect*

Ajit Pal IIT Kharagpur NPTEL

Coming to transistor stacking, whenever you are realizing circuit in this manner, say this is a four input NAND gate, so where you have got four n MOS transistors in series, and four p MOS transistors in parallel. Now whenever you do the stacking in this manner, there is a possibility some of the inputs will be 0, and some of them will be one; that means, you have got four inputs. So, there are 16 possible input combinations starting from all zero to all one. So, in when all the inputs are 0 all the n MOS transistors are off, when you are you apply 1 1 1 1 all the n MOS transistors are on. So, in this way you can have sixteen different combinations and those 16 different combinations leads to the different number of transistors off in this cascade. So, let us consider the situation when you have applied 0 0 0 0 to this. So, in such a situation all these n MOS transistors are off. However, all the p MOS transistors will be on. So, when all the p MOS transistors are on and all the n MOS transistors are off, you can see although these transistors are small I mean off, some small amount of leakage current will flow, as you know there are different types of leakage currents like sub threshold leakage current which flows, even when the input voltage is zero or less than zero or less than the threshold voltage. So, this is known as sub threshold leakage current, and because of the flow of sub threshold leakage current there will be small drop across these transistors.

So, the drop will not be much, but there will be a small drop across these transistors. So, you can see 14 mille volt is across the first transistor, which is connected to the ground, then 34 mille volt between the drain of the second transistor to ground, then 89 mille volt from the drain of the third transistor to ground, and two point three volt from the drain of the fourth transistor to ground. So, that means, in such a situation we find this source of this transistor is having a voltage 14 mille volt and this is zero volt, this is having 39 mille volt and this gate is having zero volt; that means, source is having 34 mill volt and gate is zero volt. And as you know for all these transistors, we have assumed that the substrate is connected to ground, because body bias is zero, we have assumed that. So, in such a situation, it has been found that when more than one transistor is in the series in the CMOS circuit, the leakage current has strong dependence on the number of turned off transistors. So, the number of turned off transistors will decide how much leakage current will flow, and this is known as stack effect.

(Refer Slide Time: 25:51)



Transistor Stacking

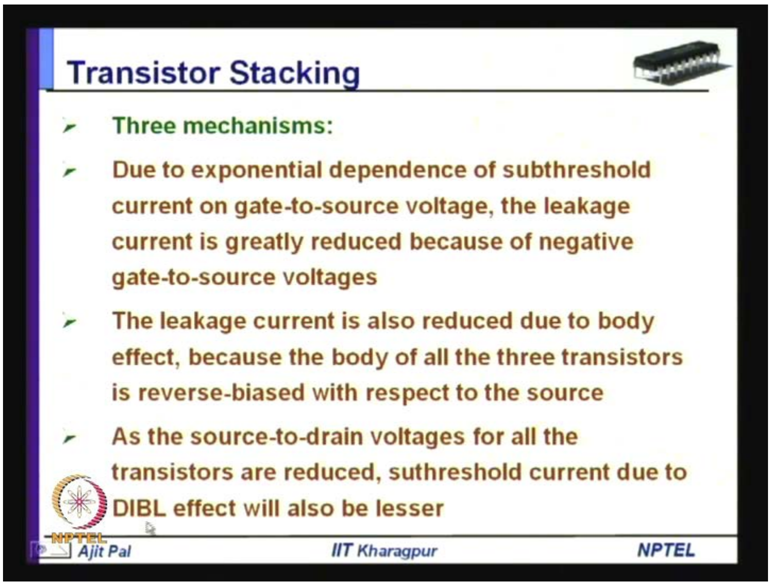
State (ABC)	Leakage Current (*n*A)	Leaking Transistors
000	0.095	Q1, Q2, Q3
001	0.195	Q1,Q2
010	0.195	Q1,Q3
011	1.874	Q1
100	0.184	Q2,Q3
101	1.220	Q2
110	1.140	Q3
111	9.410	Q4,Q5,Q6

- Highest leakage current is 99 times that of the lowest leakage current

Ajit Pal IIT Kharagpur NPTEL

So, whenever you are having different voltages we shall see, what kind of current variation occurs. Say for example, here we have got a three input n and gate and you have got eight possible input combinations, starting with starting with all zero to all one you can see the leakage current is varying from point 059 nanoampere to 9 point 410 nanoampere. So, when all these transistors are on, then you can see the leakage current is maximum and when all the transistors are off leakage current is minimum. So, when you are applying 000, the leakage current is dependent on this Q 1 Q 2 Q4; that means the all the three transistors are in series. So, leakage current is essentially decided by the leakage current these transistors, and when it is 001. 001 means this a is 001; that means, this is off and this is on. In such case you can see leakage current is increasing, but leakage current is much less than whenever it is 111. So, you find that highest leakage current is 99 times that of the lowest leakage current. So, leakage current can be different for different numbers of transistors are off, why so. So, let us try to understand why this occurs.

(Refer Slide Time: 27:32)



Transistor Stacking

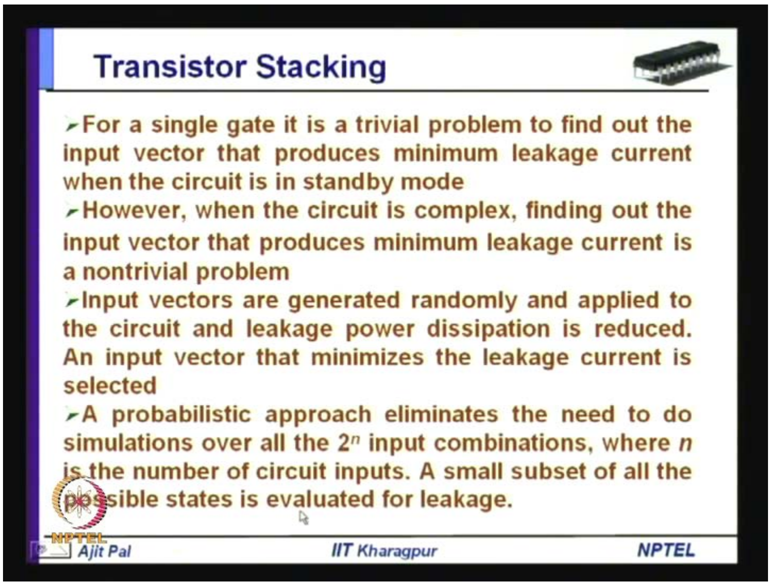
- Three mechanisms:
- Due to exponential dependence of subthreshold current on gate-to-source voltage, the leakage current is greatly reduced because of negative gate-to-source voltages
- The leakage current is also reduced due to body effect, because the body of all the three transistors is reverse-biased with respect to the source
- As the source-to-drain voltages for all the transistors are reduced, suthreshold current due to DIBL effect will also be lesser

Ajit Pal IIT Kharagpur NPTEL

There are three mechanisms involved in it; number one is due to exponential dependence of sub threshold current on gate-to-source voltage, the leakage current is greatly reduced, because of negative gate-to-source voltage, as I was explaining earlier, as you can see here. So, this is the source is having 14 mill volt, and as you know substrate is connected to 0 volt; that means, substrate is having a reverse voltage with respect to the source, because 14 mill volt is positive with respect to zero volt to which the substrate is connected. Similarly, for this transistor the source is thirty is connected to, is now having a voltage of 34 mill volt, but substrate is 0 volt, so again a reverse body bias is applied. So, whenever the source voltage is, I mean substrate is negative with respect to the source or of course, in this case we are considering sub threshold leakage current. So, if we find gate is 0 volt and source is positive. So, source gate is negative with respect to the source in all the three cases. So, this is leading to this sub threshold leakage current reduction.

So, due to exponential dependence of the sub threshold leakage current on gate-to-source voltage, leakage current is greatly reduced, because of the negative gate-to-source source voltage. Second phenomenon is the leakage current is also reduced due to body effect, because the body of all the three transistor is reverse-biased with respect to the source. So, I was explaining, so substrate is connected to zero and you can all the sources are having positive voltage. So, it is it is leading to reverse body bias, this is a second phenomenon and third phenomenon is the source-to-drain voltage for all the transistors are reduced, sub threshold current due to d I b l effect is also lesser. So, we have seen, the voltages across the transistors are very small 34 mill volt as you see, the voltages are 30 14 mill volt, 34 mill volt, 89 mill volt and so on. So, this is leading to a small this will lead to smaller leakage current, because that d I b l effect, you can drain-induced barrier lowering, smaller the drain voltage smaller will be the leakage current due to d I b l effect. So, because of these three phenomenons, there is reduction in the leakage current.

(Refer Slide Time: 30:25)



Transistor Stacking

- For a single gate it is a trivial problem to find out the input vector that produces minimum leakage current when the circuit is in standby mode
- However, when the circuit is complex, finding out the input vector that produces minimum leakage current is a nontrivial problem
- Input vectors are generated randomly and applied to the circuit and leakage power dissipation is reduced. An input vector that minimizes the leakage current is selected
- A probabilistic approach eliminates the need to do simulations over all the 2^n input combinations, where n is the number of circuit inputs. A small subset of all the possible states is evaluated for leakage.

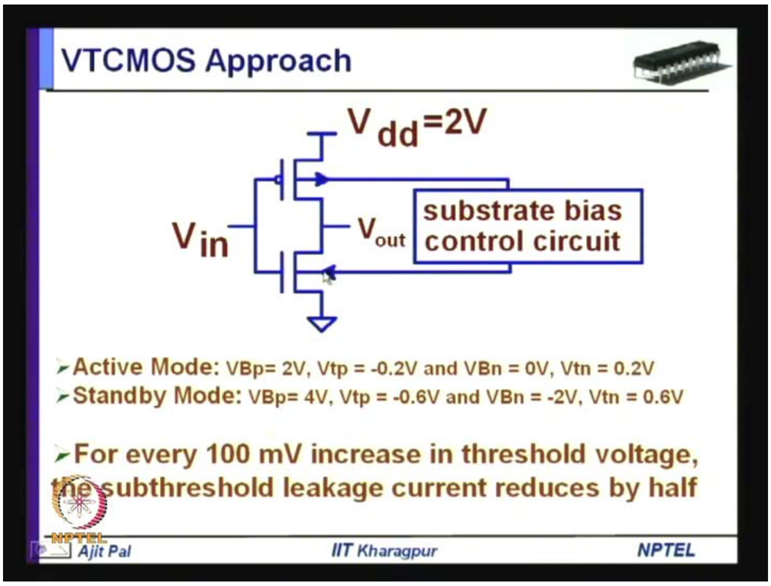
Ajit Pal IIT Kharagpur NPTEL

Now the question arises, how you will apply the input voltage when the circuit is in standby. So, what is the approach now we know, by applying a suitable input combination, you can reduce the leakage current. Obviously, if it is a four input NAND gate, in such a case if you apply 0 0 0 0, that will lead to minimum leakage current; that means, when this circuit is in the standby mode, NAND gate is in the standby mode, you will apply this input combination, SO that the leakage current is minimum. However, whenever you have got a complex circuit, you have got a multilevel implementation of a complex circuit. So, you have got different types of gates and zero to one input may mean one to another input, because of inverters available in within the circuit. So, in such a situation, what input combination will lead to minimum leakage current, how do you find out that. So, for a single gate it is a trivial problem to find out the input vector, that produces minimum leakage current when the current is in standby mode, but when the circuit is complex, finding out the input vector that produces minimum leakage current is a nontrivial problem. In fact, it has been found that it is a n p complete problem, what do you mean by n p complete problem; that means, you have to apply all possible input combinations to find out which one gives you minimum leakage current.

And if the number of inputs is say 50, you have to apply two to the power 50 input combinations. So, that may take very large time to find out which input combination will give you minimum leakage current, so what can be done. Normally, what is done input vectors are generated randomly and applied to the circuit and leakage power dissipation is reduced. So, it finds out the input vector, that minimizes the leakage current, you know that input vector is selected, which minimizes the leakage current. So, what you are doing instead of applying all possible input combinations, you are applying some input combination, a subset of them randomly. And finding out which one gives you minimum leakage current, and that you will use when the circuit is in standby mode. And there is another approach; a probabilistic approach that eliminates the need to do simulations over all two to the power n input combinations, where n is the number of circuit inputs. A small subset of all the possible state is evaluated for leakage. Here this is essentially heuristics based approach.

So, some heuristics are use to identify a subset, which can be used, which can be applied to the circuit, and from that you know the input combinations which are generated by using that heuristic algorithm, you find out the input combination which gives you minimum leakage power. However, you cannot claim that the minimum possible, only you can say that it may be closer to minimum. So, lot of research papers have been published, for finding out this kind of input combination, and that is known as input assignment, whenever the circuit is in the standby mode. So, this is also known as you know self-biasing. So, you are applying some input combination, such that kind of self-biasing is taking place, which is reducing the sub threshold leakage current, which is reducing the other leakage currents, because of the three different effects that I have discussed. So, this is the technique of transistors taking that can be used to reduce the leakage current, when the circuit is in the standby condition.

(Refer Slide Time: 34:54)



VTCMOS Approach

V_{dd}=2V

V_{in}

V_{out}

substrate bias control circuit

- Active Mode: VBp= 2V, Vtp = -0.2V and VBn = 0V, Vtn = 0.2V
- Standby Mode: VBp= 4V, Vtp = -0.6V and VBn = -2V, Vtn = 0.6V
- For every 100 mV increase in threshold voltage, the subthreshold leakage current reduces by half

Ajit Pal IIT Kharagpur NPTEL

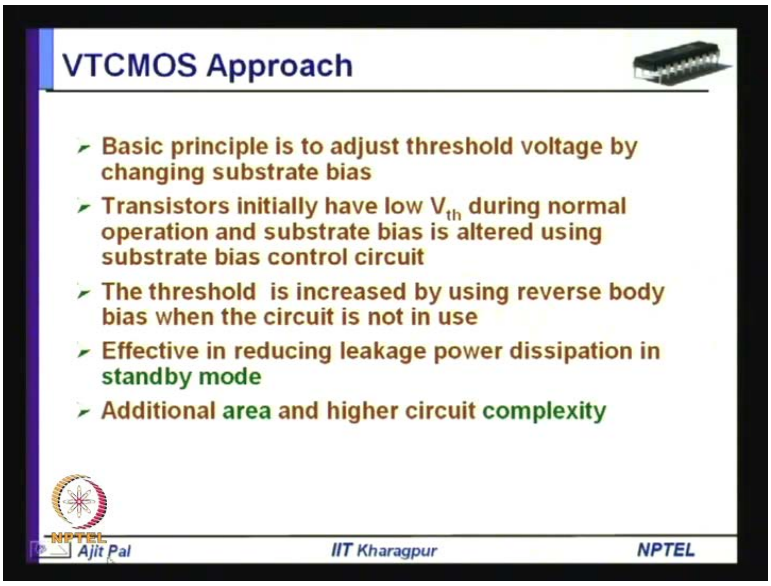
Now, we shall come to the second approach known VTCMOS. So, in VTCMOS, what is done variable threshold voltage CMOS. We have already discussed the technique VTCMOS approach, where you know the, as I told you will be separating out the those pins; that means, the substrate pins; that means, what you will be doing, say you have got a substrate you have fabricated a n MOS transistor. Say let us assume this is p type substrate and this one you will make it p plus to apply body bias, and this is n plus, this is n plus, so this is your source, this is drain and here is the gate. So, even not connecting these two, and similarly you will be having a well, and in this n-well you will be fabricating another transistor, p type transistor, and the well can be connected to, by having n plus region, and this one is your p plus, this is also p plus source and drain, this is your source and drain of the p type transistor.

So, what you will be doing with these connections. You will be taking separately and which are shown here, these two lines. So, this corresponds to the connections taken from the n-well, and this connection is taken from the p type substrate. So, what you have to do, you have to apply substrate bias control signal to these inputs, how can you do that. The circuit can be in two modes; active mode and standby mode. So, in the active mode when the circuit is in operation, what you will do, you will apply a different body bias for example, you will apply V B p; that means, the that p MOS transistor body bias is two volt, and the corresponding threshold voltage is minus two volt. Here it has been assumed the supply voltage is two volt, and when the n MOS transistor, V B n is zero volt; that means, zero body bias. So, that means, this substrate bias is two volt means, it is effectively connected to V d d.

So, effectively 0 body bias, this also 0 body bias. So, whenever both are having 0 body bias, the threshold voltages are minus two volt for the p MOS transistor and 0 volt (()) the 0.2volt (()) the n MOS transistor. So, this is the active mode condition, and obviously, circuit will be fast, because the threshold voltages are small and you will get a good performance. Now what you do when the circuit is in the standby mode, that time you apply V B p is equal to 4 volt. So, you apply 4 volt here, you apply 4 volt here, and here you apply minus 2 volt. So, in such a situation; that means, this voltage is more positive 2 volt positive with respect to this particular voltage, supply voltage. And as a consequence threshold voltage of this transistor is now minus 0.6. So, this is the threshold voltage of this transistor and V p n is minus 2 volt. So, here the threshold voltage of this transistor is 0.6 volt.

So, you find that both the threshold voltages are now increased, and for the p MOS transistors it is increased from minus 2 volt to minus 0.2 volt to minus 0.6 volt, for the n MOS transistor it is increased from 0.2 volt to 0.6 volt. So, whenever the threshold voltages are increased, as you know for every 100 milli volt increase in threshold voltage, the sub threshold leakage current reduces by half. So, here threshold voltage is increased by 0.4 volts, that means 400 milli volts, so for every milli volts it is half, so it is fourth times, half one fourth, one eighth. So, it will become one sixteenth. So, that means leakage current will become one sixteenth whenever it is in the standby mode, compared to the active mode of operation. So, this is known as VTCMOS approach, where you are applying body bias to the circuit, to reduce the leakage current. So, the circuit when it is in a normal mode of operation, you will apply zero body bias, and when the circuit is in the standby condition you will apply reverse body bias. So, that the threshold voltages are higher and leakage current is reduced. So, this is known as the VTCMOS approach.

(Refer Slide Time: 40:02)



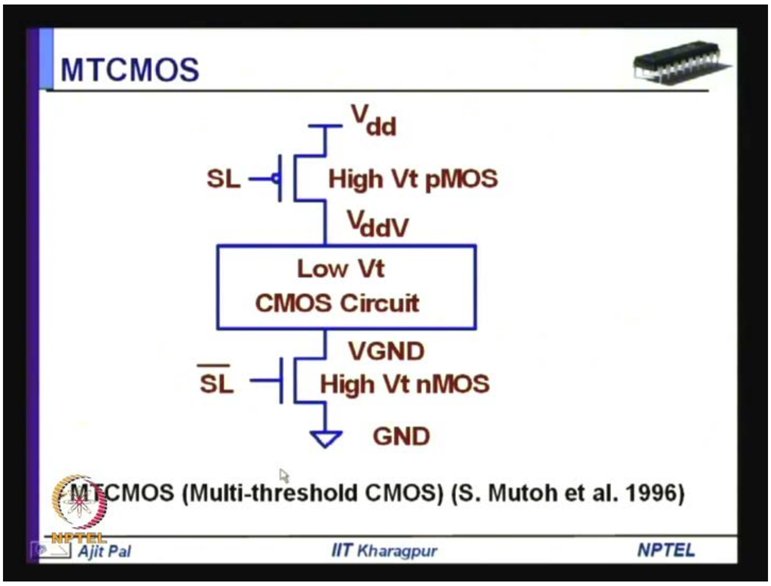
VTCMOS Approach

- Basic principle is to adjust threshold voltage by changing substrate bias
- Transistors initially have low V_{th} during normal operation and substrate bias is altered using substrate bias control circuit
- The threshold is increased by using reverse body bias when the circuit is not in use
- Effective in reducing leakage power dissipation in standby mode
- Additional area and higher circuit complexity

Ajit Pal IIT Kharagpur NPTEL

Coming to the, now I have already explained all these things basic principle is to adjust threshold voltage by changing the substrate bias, transistors initially have low V t during normal operation, and substrate bias is altered by using substrate bias control circuit, and the threshold is increased by using reverse body bias when the circuit is not in use, I have already explained that. So, this is effective in reducing leakage power dissipation in standby mode. However, you know whenever you use this technique, additional area and higher circuit complexity is there. So, it will have some additional area requirement and more complexity, because you have to realize those substrate bias control circuitry. You have to take out the pins to those points, layout cost and other area cost will be more.

(Refer Slide Time: 41:03)



MTCMOS

MTCMOS (Multi-threshold CMOS) (S. Mutoh et al. 1996)

Let us now switch to another technique that is your MTCMOS, this MTCMOS technique, here you have got low VTCMOS circuits present here, and we are adding two additional transistors known as slip transistors. So, one p MOS and one n MOS. So, one p MOS and one n MOS transistors are added in series with these two, and this was published in a paper by sometime in 1996, and they incorporated the term, they introduced the term multi-threshold voltage CMOS; MTCMOS technique. What essentially you are doing. Suppose you have got an inverter, this is an inverter, SO this inverter you can realize by using low V t. So, both the transistors, you can realize by using lower V t transistors, what is being done in addition to this part of the circuit, you will be adding two more transistors in series, this is connected to ground and this is connect to V d d. Now you are applying a signal known as slip control signal, to both these inputs. So, one is complement of the other; that means, when the circuit is in normal mode of operation, this transistor is on, this is transistor is also on.

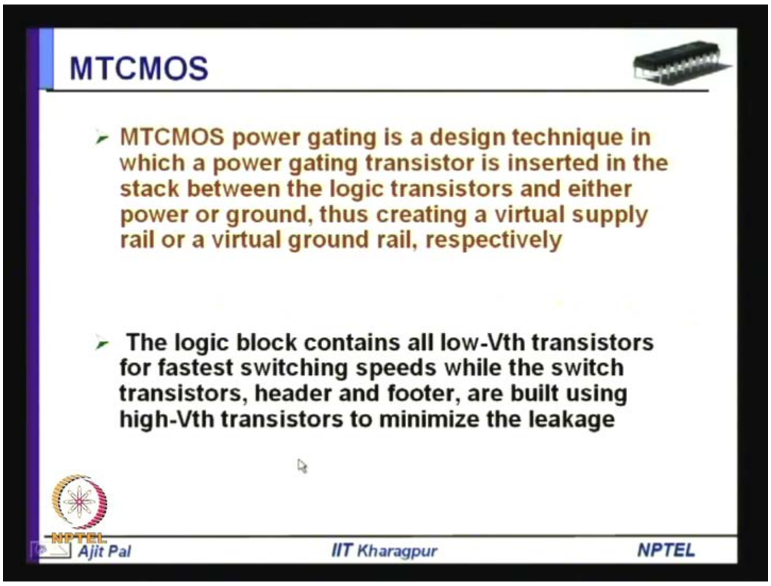
So, in the normal mode you can say these two transistors are on, and these transistors are essentially high V t transistors, this is also high V t transistor, this is also high V t transistor, both are high V t transistors, but these are all on and. So, the current that will be passing through it will be dependent on the current of this particular circuit, because both the transistors are on, and if the input is say 0 volt, then the leakage current will be dependent on the leakage current of this transistor. So, in the normal mode of operation, the current that will be flowing will correspond to the low V t transistors, and these points we will act as kind of virtual V d d. So, this is your V virtual ground, and this point we will act as virtual V d d point.

So, this is virtual V d d point, this is virtual ground and this is in the normal mode. Now in the active mode what you do. You make this transistor off, this transistor also; that means you make this one as 1 and this one as 0. So, when you do that, then this p MOS transistor as n MOS transistor, both are now off. So, when both of them are off, what will happen this point and this point, you know no longer virtual ground and virtual V d d. You may say what current will flow through this. Actually you may say that, now the high V t transistors are series with low V t transistors. So, you may say that, if this is the current of the low V t transistors, and this is the current of the high V t transistors and this is now connected to ground, this is connected to say V d d.

So, current will be dependent on the lower of the two, because they are in series, and obviously when two currents are in series, you know that lower current will flow through the spark, as the high V t transistors are having lower leakage current, smaller current leakage current will flow, when the circuit is in the standby mode; that means, in the standby mode you are making the high V t transistors off, leading to smaller leakage current flowing through this part of the circuit. So, this is the basic idea behind this MTCMOS. So, in MTCMOS circuit, you are having both high V t and low V t transistors, and the design flow is more or less similar, except that you will be adding sleep transistors, which are of high V t in addition to those normal circuit. So, this is the basic idea, and you can say that MTCMOS power gating is the design technique in which a power gating transistor is inserted in the stack between the logic transistors, and either power or ground, thus creating a virtual supply rail or a virtual ground rail.

Here we have seen two transistors, you know this transistor is connected to V d d and this transistor is connected to ground, actually later on we shall discuss particularly in the next class, you will see that anyone will actually serve the purpose. You do not really require two sleep transistors in series, just one sleep transistor may serve the purpose, but when the circuit was purposed by Mutoh et al, they purposed two transistors and whenever implementation is done, sometimes both the transistors are used, but in reality anyone may will serve the purpose; that means leakage current can be reduced by having one transistor, either the p MOS transistor which is known as the header transistor or the n MOS transistor, which is known as the footer transistor. So, either you can have this header, or this footer transistor in series as the sleep transistor, but whatever it may be the case, the leakage current will be reduced, because the high V t transistors in series with the low V t transistors, and you get smaller leakage power in the standby mode.

(Refer Slide Time: 47:26)



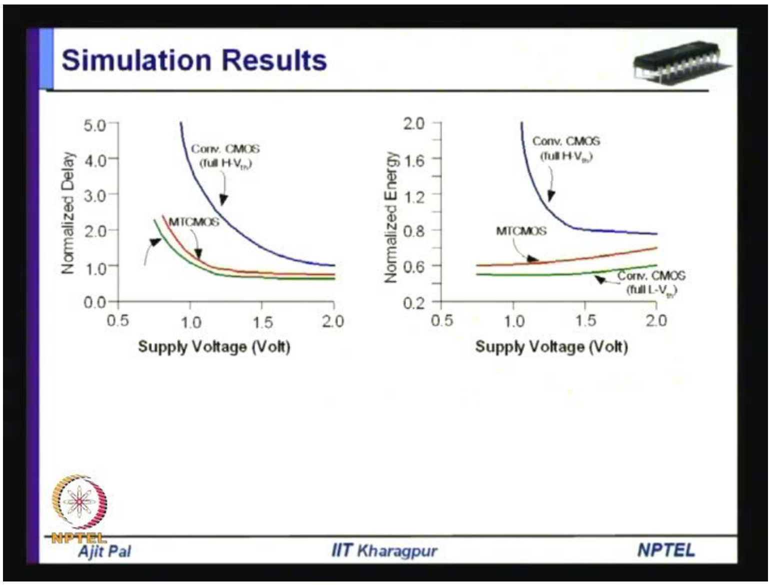
MTCMOS

- MTCMOS power gating is a design technique in which a power gating transistor is inserted in the stack between the logic transistors and either power or ground, thus creating a virtual supply rail or a virtual ground rail, respectively
- The logic block contains all low-Vth transistors for fastest switching speeds while the switch transistors, header and footer, are built using high-Vth transistors to minimize the leakage

Ajit Pal IIT Kharagpur NPTEL

So, the logic block contains all low V t transistors for fastest switching speed, while the switch transistors; that means those sleep transistors header and footer, are built using high V t transistors to minimize leakage. So, basic idea here is you are not really sacrificing the speed of operation, because when the circuit is in normal mode, the sleep transistors are on.

(Refer Slide Time: 47:54)



Simulation Results

Ajit Pal IIT Kharagpur NPTEL

And this was verified by simulation, and the simulation result is shown here. So, here you can see, three lines shown for three different supply voltages. The top one corresponds to conventional CMOS with full high threshold voltage; that means, when all the transistors are of high threshold voltage, may be 0.5 V d d. So, the circuit is realized by using one volt supply voltage. So, if the 0.5 volt is a threshold voltage, this will correspond to the delay, as you know when the circuit is realized by using high V t transistors, the delay will be longer, and as a consequence you got normalized delay higher. On the other hand you can see, when the circuit is realized by using only low V t transistors this is the curve, the green line corresponds to all the circuits realized by using low V t transistors. So, we get much smaller delay. However, whenever you go for MTCMOS, we insert the high V t transistors in series, then it is definitely increasing the delay by small amount, but it is much less correspond to the high V t transistors.

So, we find that, there is some increase in delay, because of the insertion of those high V t sleep transistors, but the increase in delay is not very high, small so, but leakage current reduction will be significant. In other words we can tell that leakage current will correspond to high V t transistors, performance will correspond, if not correspond to very close to low V t transistors. So, you are getting leakage current corresponding to high V t transistors, performance close to the low V t transistors, that is what is being displayed with the help of these lines. Similarly, if you consider the normalized energy, we find that in power dissipation on power delay energy, essentially power delay product that is also very small, whenever it is realized conventional low V t transistors, which is also very close to the conventional low V t transistors. So, in other words you are getting significant reduction in power dissipation, the leakage power dissipation, at the same time without much comprise in performance.

So, this simulation results were published by Mutoh et al as I mentioned in the year 1996. Subsequently this technique has been matured, it has been found that MTCMOS can be easily implemented using existing circuits, and MTCMOS reduces only the standby power, large inserted MOSFETS will increase area and delay as we have seen. And extra V t memory circuit is needed to maintain the data in the standby mode. These aspects, the use of this MTCMOS technique for, this is known as power gating or power measurement technique, this are led to a matured technique known as MTCMOS based power gating, where there will be many aspects you have to consider, particularly you will see how the powering up is done power down takes place, how to insert the flip transistors. There are many aspects which I shall discuss in detail in my next lecture. So, with this we have come to the end of today’s lecture. Thank you.