

**VLSI Physical Design**  
**Prof. Indranil Sengupta**  
**Department of Computer Science and Engineering**  
**Indian Institute of Technology, Kharagpur**

**Lecture – 58**  
**Low Power VLSI Design**

So, in this week we shall be starting with some discussions on low power design. As you know designing a circuits for low power has become so much very important in the present day scenario, and it also affects the way VLSI chips are designed and also in particular the physical design process the so called back end design of the VLSI chips. The way they are done today some techniques and some rules you can say design rules are incorporated at various stages of the design. So, that out final product the chip consumes less power.

(Refer Slide Time: 01:10)

**Background**

- Power consumption is a very big challenge in modern-day VLSI design.
  - Almost all portable devices run on battery power.
- Various strategies have been proposed to control the power dissipation.
  - Adds a fourth dimension to design.

Area  
↑  
Delay ↔ Power  
↓  
Testability

IT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, that is the topic of our discussion today, low power VLSI design. So, as I said, power consumption is very big challenge perhaps the most important challenge in modern day VLSI design, which is pushing performance to a secondary level. The primary reason is that almost all the devices we use today they are portable in nature other then of course, the desktops which sit on our tables. This portable device all run on battery power. You think of our mobile phones, our laptops, our notebooks our tabs everything, these are the devices which we use they all get the energy from the battery that is built inside it.

So, if you can design the circuits that get embedded into these devices, in a way that they consume less power or energy whatever you call. So, the battery life of the devices tends to increase which is extremely desirable from the users prospective. So, there have been various works and strategies that have been proposed over the years, which try to reduce the power dissipation. So, we sometime say that we have added a fourth dimension to the design process. What are the other 3 dimensions let us see? You see this is a typical picture that will find in several places in some textbooks also. Traditionally in this 70s and 80s area and delay were the most important parameters.

So, when someone tried to design a circuit, a chip the primary emphasis was how much less area is occupied by my circuits, on the chip, on the silicon floor, and what is the maximum speed. So, we try to reduce the delay now often, this area and delay are mutually conflicting. So, some typical figure of merits was proposed and used. One of them was the product of area and delay square, a  $d^2$ . So, there are number of such have a matrix which you are used and proposed. So, you can have a figure of merit.

But as the circuits started to become more and more complex, so what happened was that, you see area and delay were not the only parameters. Now after a chip is manufactured mean one very important task of the designers or you can say the text engineers was, to ensure that the chip does not contain any apparent fabrication defects. So, the chip needs to go through a process of testing, more the complexity increases in a chip the task of the testing becomes more and more difficult.

So, there has to be some design rules like the design for testability techniques that where discussed earlier. Those have to be incorporated or embedded in the design process itself. So, that whatever circuits we design they are easily testable. So, that is why we say that in the 80s and 90s, this third axis of this in this diagram testability came into the picture, but now 2000 onwards with the advent of battery operated mobile devices portable devices, power as become a very important fourth parameter in this design space.

So, not only we have to address area delay and testability, also you have to address this power consumption issue.

(Refer Slide Time: 05:16)

**Power and Energy**

- Power is drawn from a voltage source attached to the  $V_{DD}$  pin(s) of a chip.
- Terminologies:
  - Instantaneous Power  $\rightarrow P(t) = i_{DD}(t)V_{DD}$
  - Energy  $\rightarrow E = \int_0^T P(t)dt = \int_0^T i_{DD}(t)V_{DD}dt$
  - Average Power  $\rightarrow P_{avg} = \frac{E}{T} = \frac{1}{T} \int_0^T i_{DD}(t)V_{DD}dt$

IT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, let us look at some of the terminologies. So, in a typical chip which in today's technology is normally built using CMOS technology. So, we say that power is dissipated whenever there is a current flowing from the power supply which you called the  $V_{DD}$  to the ground. So, power is drawn from the voltage source, and this source I mean sources of these currents can be various. These we shall be seeing subsequently.

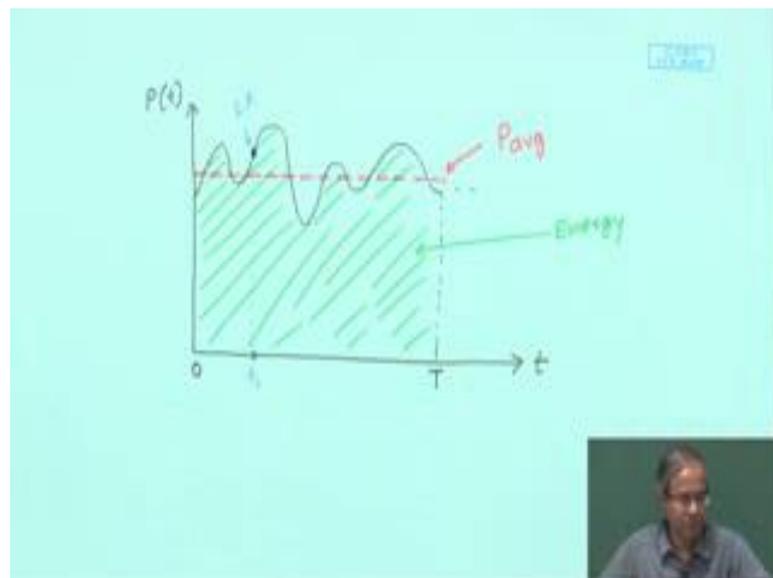
But let us look at how we typically measure the power. So, I shall be explaining this very clearly. So, we distinguish between instantaneous power energy and average power. So, how are they defined? See instantaneous power is defined simple by the product of the current and the voltage. The current flowing from  $V_{DD}$  to ground we call it as  $I_{DD}$ . This is a function of time  $t$  multiplied by the voltage  $V_{DD}$ . Now instantaneous power does not give us a very clear picture of means how much our battery is getting drain because it may so happen that sometimes the instantaneous power is very high, but over long period of time we are consuming very low power, means our currents is very less.

So, some sort of average is more meaningful. Instead of the power drawn at a particular point in time, so we talk about energy, so how is the energy defined? So, I shall be explaining these terminologies graphically very shortly. This energy is defined as the integral of the instantaneous power over a period of time let say capital  $T$ , 0 to  $T$ . What is the sum of the instantaneous power drawn that aggregation is defined as the energy? This

is how you define. And of course, you are computing this energy over a time period  $T$ . So, if you calculate the average simply divide this  $E$  by  $T$  you get the average power.

So, normally when we evaluate the power consumption of a circuit or a device or a chip, we talk about the energy or the average power. Now let us see with the help of a diagram how these 3 things are related.

(Refer Slide Time: 07:57)



So, let us show a typical diagram. This is the access of time and here we show the instantaneous power consumption, which means the product of the current and the voltage. Let us say we the value of  $P$   $t$  varies with time like this, let us say. So, we are interested to measure the power between time 0 and capital  $T$ .

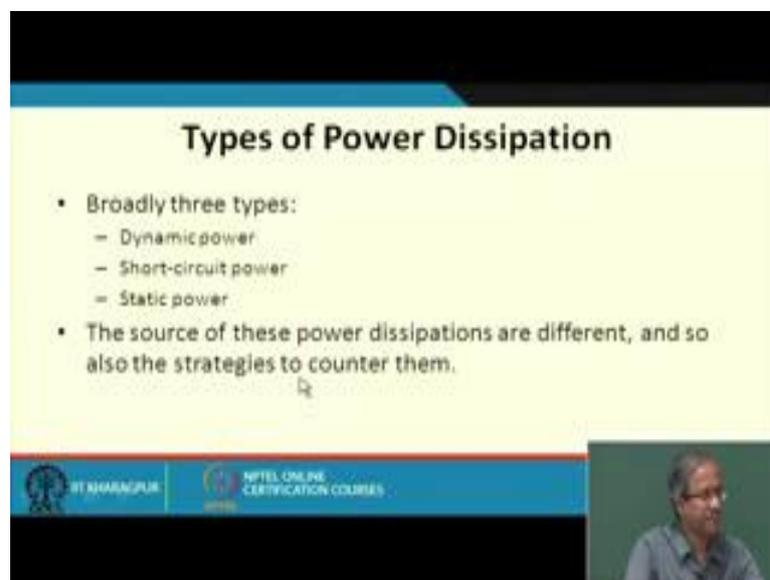
Now, just recalling the definitions, for a particular value of  $t$  let us say here. So, if we try to see what is the corresponding value of  $P$   $t$  it is here, this is the value of the instantaneous power  $I$   $P$ , instantaneous power at this particular time let us say  $t$   $i$ . Right now if I want to calculate the energy. So, energy as I said is the integral of  $P$   $t$  from 0 to capital  $T$ . So, so what is the physical interpretation of the integral? Integral means the area under this curve, this total area under the curve. This is defined as the energy, that is  $\int P$   $t$   $dt$  between 0 and  $T$ .

Now, if you just take the average energy over this time,  $P$   $t$  it is energy is varying. So, possible the average will be somewhere here. You show it as a straight line. So, this will

give you the average power  $P_{avg}$ . Now as I said  $P_{average}$  can be computed by dividing the energy divide with this time  $T$ . So, we understand this  $P_{average}$  is this level; that means the height of this rectangle and energy is the total area height multiplied by the width and width is capital  $T$ .

So, if we divide this area by capital  $T$  you will be getting this height; that means  $p_{average}$ . So, diagrammatically I am showing you the difference between instantaneous power the energy and the average power. So, let us come back.

(Refer Slide Time: 10:43)



The slide is titled "Types of Power Dissipation" and lists three types: Dynamic power, Short-circuit power, and Static power. It also states that the sources and counter-strategies for these power dissipations are different. The slide is part of an NPTEL Online Certification Course, as indicated by the logo at the bottom.

- Broadly three types:
  - Dynamic power
  - Short-circuit power
  - Static power
- The source of these power dissipations are different, and so also the strategies to counter them.

So, talking about power dissipation in a typical CMOS circuit, we can classify the power dissipation into 3 types very broadly. Dynamic short circuit and static, we shall see what are the sources of these different kinds of powers and what are the typical methods that are adopted to reduce the effects of these.

(Refer Slide Time: 11:17)

**Dynamic Power**

- Quite predominant in CMOS design style.
  - Dynamic power is required to charge and discharge load capacitances when transistors switch.
  - One cycle involves a rising and falling output.
  - On rising output, charge  $Q = CV_{DD}$  is required to charge the output node to  $V_{DD}$ .
  - On falling output, the load capacitor discharges to  $GND$ .
  - This repeats  $T \times f_{sw}$  times over an interval of  $T$ .

ST KVARAGIRI | NPTEL ONLINE CERTIFICATION COURSES

So, we shall see this subsequently. Let us start with dynamic power. Well as the name implies, dynamic power is something which is depending on the dynamics of the circuits. What is the definition of dynamics? Something which changes with time, dynamics power by its definition means, whenever some signals in the circuits are changing those changes cause some dissipation of power that is basically what is meant by dynamics power dissipation. So, let us try to understand this. So, basically dynamics power is required for what purpose for charging and discharging load capacitances with transistor switch.

(Refer Slide Time: 12:16)

$i = C \frac{dV}{dt}$

$T \times f_{sw} \Rightarrow \text{no. of switching}$

$f_{sw} = \text{frequency of switching}$

$C \downarrow$  discharged,  $C \uparrow$  charged

So, let us look at very simple example of a CMOS inverter. So, here I have a p mos transistor and n mos transistor. This is my input let say A and this is my output let us say f. So, f is given by the complement of A and this is  $V_{DD}$ . Now what I am saying is that this output as I said earlier may be driving the inputs of other gates. So, we can equivalently assume that it is driving a load capacitance. Let us say the load capacitance is C. So, let us say when the value of a switches from 0 to 1, let us say the value of f will be switching from 1 to 0.

Similarly, when a switches back to 0, the value of f will again switch to 1. Now whenever this output node f changes from high to low and low to high what does this mean whenever it is changing from high to low this means, C is getting discharged and whenever it is rising from low to high, we say C is getting charged. Now in this circuits the discharging path of the capacitor is this, through the pull down transistor. And the charging path of the capacitor is this through the pull up transistor.

So, as you know in a CMOS circuit, depending on the value of A so either the pull up or the pull down one of the networks is conducting. So, depending on that this C will be either discharging or it will be charging. Now every time there is a charging discharging of a capacitor, it will consume power because you recall the basic definition from circuit theory, current is defined as  $C \frac{dV}{dt}$ . So, you can define the current flowing as the product of the capacitor and the rate of change of voltage.

So, whenever there is this kind of charging discharging going on there will be a current flowing and the product of currents and voltage will be the power consumption. So, let us come back to this slide. So, I have just shown in one cycle we can have a rising output and also falling output right. So, whenever there is a rising output; that means, the output node is getting charged charge means what the capacitor is getting charge to  $V_{DD}$ . So, what is the total charge? Charge is the product of capacitance and voltage. So, C multiplied by  $V_{DD}$ . This much charge is required to charge the output node to  $V_{DD}$ .

Similarly, when the output goes to 0, the load capacitor needs to discharge to ground, right now it can. So, happen that let us say we are considering a certain time period T right. Now within this time period T the output f may be going up and going down several times. This may so happen be depending on the behavior of the circuit. Now if this kind of thing happens which means, in the same interval of time T the capacitor is

getting charged and discharged multiple number of times. See charging discharging, charging discharging, charging discharging.

So, if I define a parameter  $f_{SW}$ , this is the frequency of switching then within a time period  $T$ . So, if you multiple  $T$  with  $f_{SW}$ , this will be number of switching that is taking place within this time period  $T$ . The product of the time and the frequency of switching this you can call as the frequency of switching  $SW$  right.

(Refer Slide Time: 17:10)

• If the frequency of output switching is  $f_{sw}$ , the charging and discharging cycle will repeat  $T \times f_{sw}$  time over a time interval  $T$ .

VDD  
 $I_{DQ}(1)$   
C  
 $f_{sw}$

IIT KHARAGPUR  
NPTEL ONLINE  
CERTIFICATION COURSES

So, this already I have explained, that if the frequency of output switching is  $f_{SW}$ . That charging and discharging cycle will be repeating so many times over a time interval  $T$ .

(Refer Slide Time: 17:31).

• Calculation of average dynamic power:

$$P_{dynamic} = \frac{1}{T} \int_0^T i_{DD}(t) V_{DD} dt$$

$$= \frac{V_{DD}}{T} \int_0^T i_{DD}(t) dt$$

$$= \frac{V_{DD}}{T} [T f_{sw} C V_{DD}]$$

$$= C V_{DD}^2 f_{sw}$$

Proportional to load capacitance.  
Proportional to square of  $V_{DD}$ .  
Proportional to frequency.

IT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

Now, let us see that how we can calculate the average dynamic power in a circuit. Now our assumption is that we have a circuit like this, where there is a period time T and the inputs are changing at some rate.

(Refer Slide Time: 17:54)

Charging:  $0 \rightarrow V_{DD}$   
Discharging:  $V_{DD} \rightarrow 0$

$$\frac{V_{DD}}{T} \int_0^T i_{DD}(t) dt$$

$$= \frac{V_{DD}}{T} \int_0^T C dV$$

$$= \frac{V_{DD}}{T} (C V_{DD}) (T f_{sw})$$

$$= \underline{\underline{C V_{DD}^2 f_{sw}}}$$

$f_{sw} \times T$

So, I am again showing that picture that I am concentrating my calculation within a time period of T and within the time period T the input or the output whatever we call in inverter both will be the same will be changing certain number of times. So, I am calling it as the frequency of switch.

Now, let us look at this calculation. So, how do we calculate the dynamic power? This is actually the average power consumption. So, whenever the output is switching. So, between time 0 and T, I simply multiply the current with the voltages, this is the definition of the instantaneous power, you integrate it over a time T you get the energy take the average you get the average power. Since  $V_{DD}$  is a constant you take it out. So, you have an expression like this. So, let us see how from here I get this. So, we get an expression like this I am just working this out. So, you get  $V_{DD}$  divided by T integral 0 to T current dt.

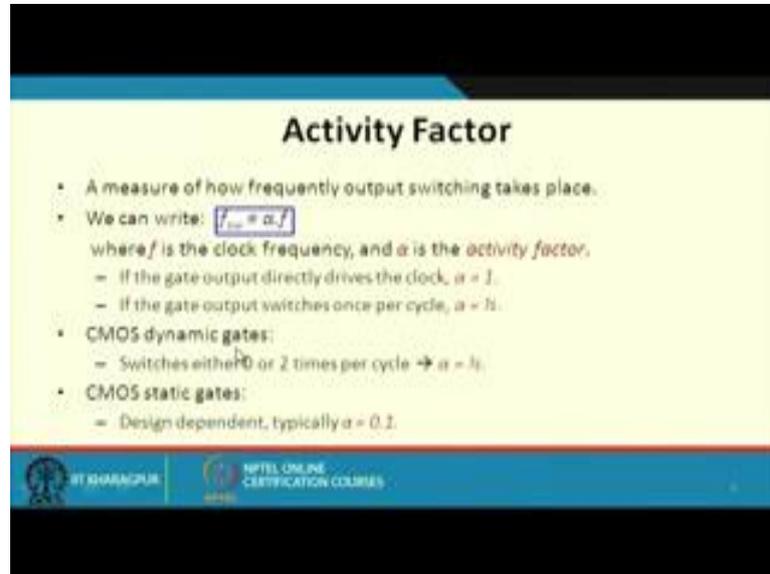
Now, let us make an observation. So, our observation is the value of current is defined as the product of the capacitance and the rate of change of voltage. So, if you take this into account. So,  $I_{DD} dt$  becomes  $c dV$ . So, I can write it as  $V_{DD}$  by T integral  $c dV$ , now the independent variable is V. So, what will be the range of v see I am working from 0 to T? So, this will be from 0 to what something I am writing a question mark here. Because you see a within this time T as I said charging and discharging will take place multiple number of times, right multiple number of so, for each charging, I can say each charging the voltages is going from 0 to approximately  $V_{DD}$ , in the load capacitor and for discharging it is the reverse  $V_{DD}$  back to 0.

Now, in each such cycle I can write it like this.  $V_{DD}$  by T see  $C dV$  is nothing but C into V now since this charging up to  $V_{DD}$ . So, for every charging it will be c into  $V_{DD}$ . Not only this how many times it is charging within this period T it is charging if SW multiplied by T times as I said. So,  $f_{SW}$  multiplied by T, this will give you at how many times this charging and discharging is taking place within this time period T. So, you will have to multiply this by this T into  $f_{SW}$ , this will give you the total value of the integrate. See here basically this integral you are separating out between these each of these segments 1 2 3. So, that is why if you just add them out there are. So, many such segments, you multiply this by this right.

So, how much this comes to if you make a simplification.  $C V_{DD}^2$  square capital T cancels out and  $f_{SW}$ . This is the final expression for dynamics power right. So, in this expression also this same thing is shown finally, the value of the dynamics power is this and the point to observe here is that, the dynamics power is proportional to load capacitance. So, to reduce it you can try and reduce the value of C proportional to square

of  $V_{DD}$ . So, if it is possible to reduce the supply voltage that will also be good and of course, proportional to the signal switching frequency.

(Refer Slide Time: 22:39)



**Activity Factor**

- A measure of how frequently output switching takes place.
- We can write:  $f_{sw} = \alpha \cdot f$   
where  $f$  is the clock frequency, and  $\alpha$  is the activity factor.
  - If the gate output directly drives the clock,  $\alpha = 1$ .
  - If the gate output switches once per cycle,  $\alpha = 0.5$ .
- CMOS dynamic gates:
  - Switches either 0 or 2 times per cycle  $\rightarrow \alpha = 0.5$ .
- CMOS static gates:
  - Design dependent, typically  $\alpha = 0.1$ .

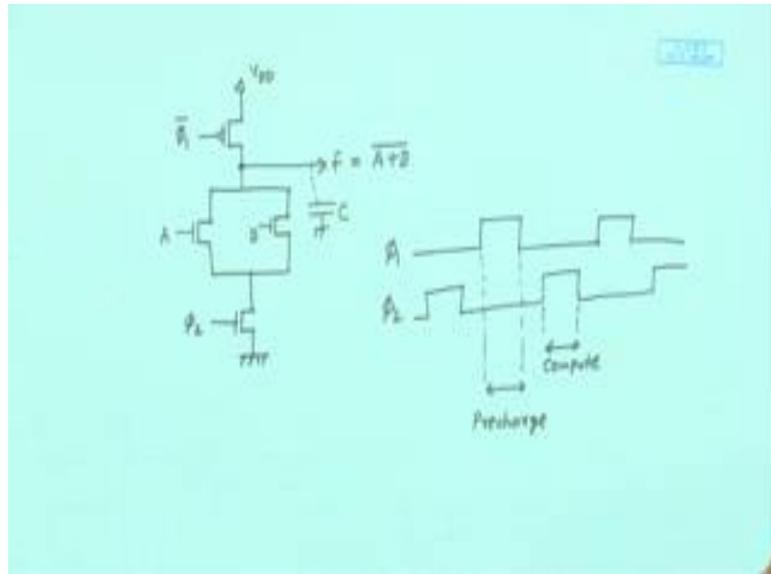
So, if you can reduce the frequency that will also result in less dynamics power. Now taking the concept of dynamic power one step forward, let us define something called an activity factor. Now see in the previous expression we talked about this  $f_{SW}$ , which was the frequency of switching of the input or the output signals right. Now here we are talking about the clock frequency  $f$  is our clock frequency, now we are trying to relate this  $f_{SW}$  with  $f$ , by multiplying it with the parameter alpha, which is defined as the activity factor.

You see some signal can change state at a rate which is at most equal to the clock frequency. Because it is the clock the edges of the clock that defines the various events in circuits, but all edges of the clock need not result in a change in the signal state. So, only a fraction of the clock edges will actually result in an activity that will result in consumption of dynamics power. So, we define something called an activity factor, which if we multiplied by the frequency we get  $f_{SW}$ . This is how we just estimate  $f_{SW}$  now. If you are directly connecting the clock with the gate input, so the input as well as the output will be changing at every edge of the clock.

So, there we say alpha equal to 1, but if the gate output switches once per clock cycles; see once per clock cycle means or in every clock cycle the clock is transiting twice low

to high and high to low. I am saying that the output is changing once for every 2 transitions of the clock.

(Refer Slide Time: 24:51)



So, here the activity factor is defined as half. And we have something was CMOS dynamics gates I have not talked about it. CMOS dynamics gates are something like this. Say you can implement a out nor. This is a p type transistor.

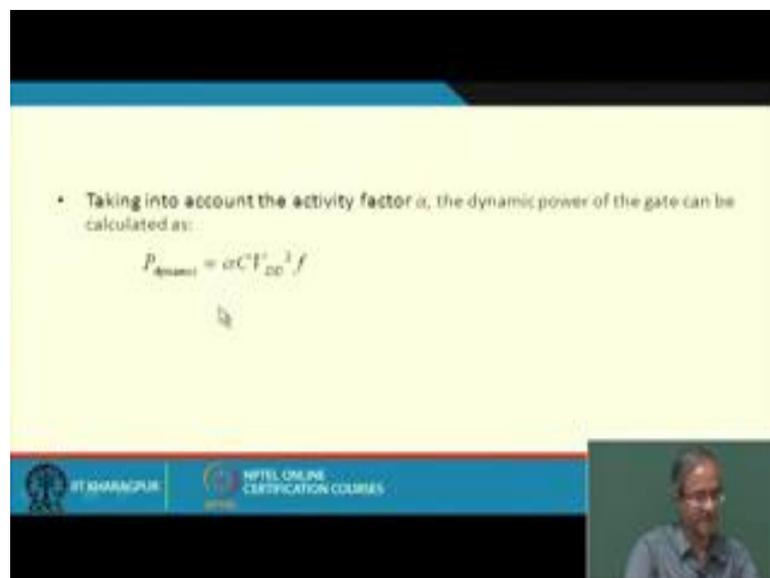
So, you see in a dynamics CMOS, dynamics logic, we are not using many transistors in the pull up, this is the output. This is a NOR gate. So, here in the pull up network we are using very few number of time here only one. And you see you are using a clock phi 1 and phi 2. It is like a 2 face clock. So, how it works, let me just briefly tell you, let us say phi 1 is running like this, and phi 2 is running like this. They are non-overlapping right. So, when phi 1 equal to 1, phi 1 bar will be 0 and this will be conducting. So, it is here. So, the output load capacitance will be here this is your C. So, it is during this period when phi 1 is high, you call this stage as the precharge stage, during this period phi 2 is 0. So, this transistor is off.

So, capacitance if getting charged through V DD, this is the precharge state, but when phi is 2 this part then phi 1 is back to 0. So, this transistor is off the capacitor was already charged, the phi 2 is high means this is conducting, now depending on A and B this capacitor can discharge or may not discharge. So, we call this as the compute phase. So, depending on the phases of phi 1 and phi 2 you can alternate. So, depending on the value

of A and B the output value may change it may not change. While in every cycle worst case it can change twice, during precharge state it can go from 0 to high, again during compute state it can go back to from high to low right.

So, this is actually mentioned here, that for a CMOS dynamics gate switching takes place either 0 to 2 times for an average of alpha equal to half. Because 2 times per cycle means alpha equal to 1, 0 means 0. So, average is half, but for CMOS static gates conventional CMOS gates. The activity factor is very much dependent on the design; the typical value is around 0.1.

(Refer Slide Time: 27:50)



• Taking into account the activity factor  $\alpha$ , the dynamic power of the gate can be calculated as:

$$P_{\text{dynamic}} = \alpha C V_{DD}^2 f$$

So, if you take the activity factor into account, our dynamics power equation changes to this square. You replace  $f$  SW by alpha into  $f$  right. So, alpha is also a factor here.

(Refer Slide Time: 28:07)

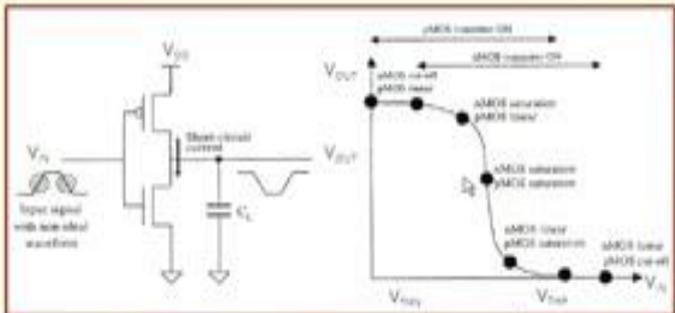
### Short Circuit Power

- Short circuit current occurs in a CMOS gate during signal transitions when both the nMOS and pMOS networks are ON and there is a direct path between  $V_{DD}$  and GND.
  - Also called *crowbar current*.
  - May account for more than 20% of the total power dissipation.
- Directly related to frequency of switching.
  - As clock frequency increases, the frequency of transitions increases.
  - Thus short circuit power dissipation also increases.



Now, we look at short circuit power. Short circuit power this is some kind of power dissipation in a CMOS gate, when the signals transitions are taking place. Because they are periodically for very small amount of time, both n mos and p mos networks can be conducting. This is also called as the crowbar current. And in modern day designs this short circuit power can consume so means of the tune of 20 percent of the total power dissipation right. And faster is the frequency of switching more will be the short circuits power because as you increase the clock frequency the transitions frequency will also increase.

(Refer Slide Time: 28:59)

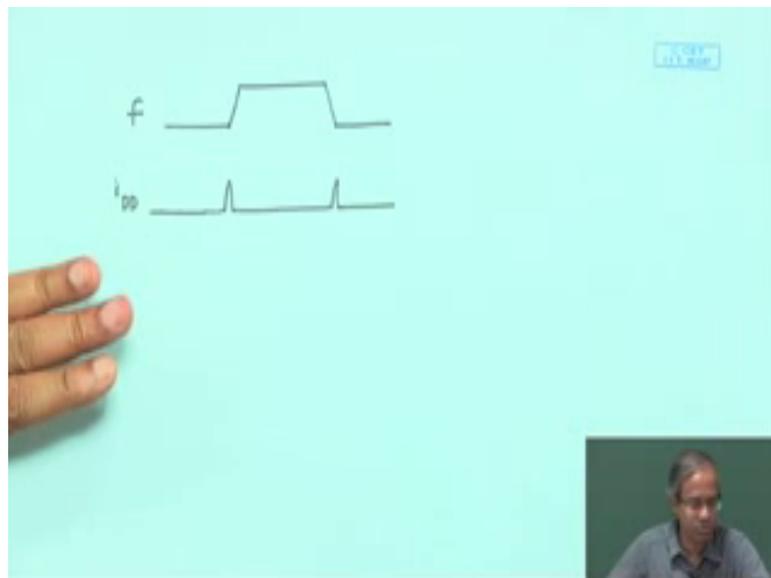


The diagram illustrates the short-circuit current in a CMOS gate. On the left, a CMOS gate is shown with an input signal  $V_{in}$  and an output node connected to a load capacitor  $C_L$ . The output voltage  $V_{out}$  is shown as a square wave. A period where both the nMOS and pMOS transistors are on is labeled "Short-circuit current". On the right, a graph plots  $V_{out}$  against time. The graph shows the output voltage transitioning from  $V_{DD}$  to  $V_{SS}$ . The period where both the nMOS and pMOS transistors are on is labeled "Short-circuit current". The graph also shows the output voltage reaching  $V_{DD}$  and  $V_{SS}$  levels, with the nMOS and pMOS transistors being on during the transition.



So, let us see how it happens. We take simple CMOS inverter, where you say that the input is going from 0 to high and high and low. You see the input 0 to 1 and 1 to 0 cannot be instantaneous; they will be a rise time and a fall time. And this is a typical input output characteristic curve of the inverter. There is a region where the p mos transistor is either fully on or partially on, and there is a region where the n mos transistor is either fully on or partially on. There is an intermediate region where the transitions are very fast a short range of voltage during which both the n mos and p mos transistors may be conducting.

(Refer Slide Time: 29:51)



So, what might happen is that, suppose the output of the circuit. So, it is changing from 0 to 1, it is changing from 1 to 0. So, if I also plot the value of currents drawn from  $V_{DD}$ , normally in a CMOS currents drawn in very low, but during transitions will find that they will be a small spike here because momentarily, both the pull up and pull downs are conducting. And higher the value of the frequency more will be the spikes coming in right. This is the source of the so called short circuits current. And lastly we talk about static power. Static power is something which is dissipated even when no signal transitions are occurring.

(Refer Slide Time: 30:26)

**Static Power**

- Static power is consumed even when the chip is quiescent.
  - Leakage effects draw power from nominally OFF devices.

The diagram shows a cross-section of a transistor with labels: Gate leakage (pointing to the gate oxide), Drain junction leakage (pointing to the drain-substrate junction), and Sub-threshold current (pointing to the channel region). The output voltage is labeled  $V_{out}$ .

IT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

These occur mainly due to the leakage effects. So, even if the transistors are off, non-conducting due to some see the junctions can be modulus some reverse bias diodes which are not conducting. There can be drain junction leakage sub threshold current many such sources are there, gate leakage because of this kind of effects, a small amount of current will continuously be flowing. So, this leakage power will always be there. This cannot be avoided, but of course, you can reduce it by using some techniques.

(Refer Slide Time: 31:14)

- As technology scales, static power dissipation becomes more and more important in terms of leakage power.
- Various causes of leakage in fabricated devices:
  - Sub-threshold leakage.
  - Reverse biased p-n junction.
  - Channel punch through.
  - Oxide leakage.
  - Hot carrier tunneling effect.
  - Drain induced barrier lowering.
  - Gate induced drain leakage.

IT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, as the transistors becoming smaller, larger and larger number of transistor have been packed into a chip, the total contributions of static power dissipation are becoming more important. Because now in a chip we have of the order of a billion of transistor billions of transistors, and each of the transistor is consuming little bit of energy or power. So, the sum total if you take over all billions transistors it can become significant right.

So, I am not going to the detail, there can be several sources of these leakage currents, sub threshold leakage, reversed bias junctions, channel punch through oxide leakage etcetera. So, by doing some design manipulations you can control some of these. So, we shall see some of these techniques later.

So, with this we come to the end of this introductory lecture on low power design. In our next lectures we shall be moving in to some more details, about how we can actually control power dissipations by some design techniques.