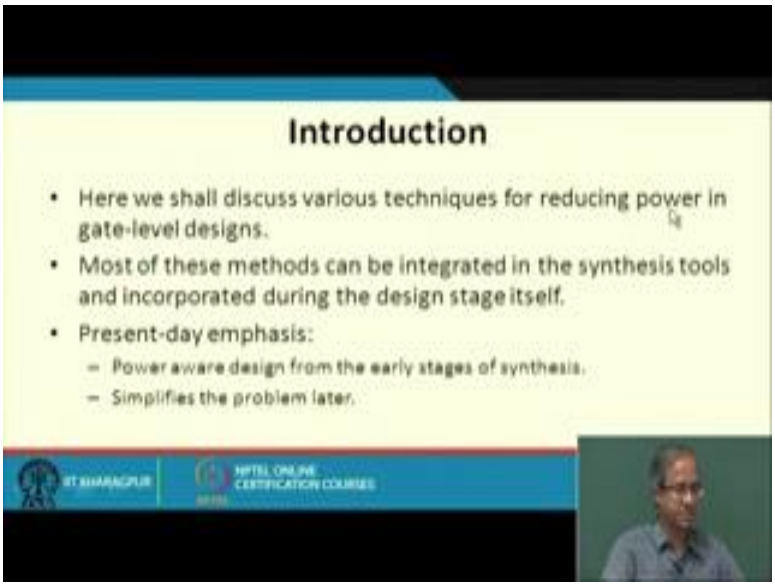
VLSI Physical Design
Prof. Indranil Sengupta
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture – 60
Gate Level Design for Low Power (Part 1)

So, in this lecture we shall be looking at some of the techniques to reduce power at the level of gates at the gate level. There are many interesting techniques which you can use by analyzing the signal activities that take place at the level of gates. So, we shall be looking at some of the quite commonly used approaches in this regard fine.

(Refer Slide Time: 00:43)



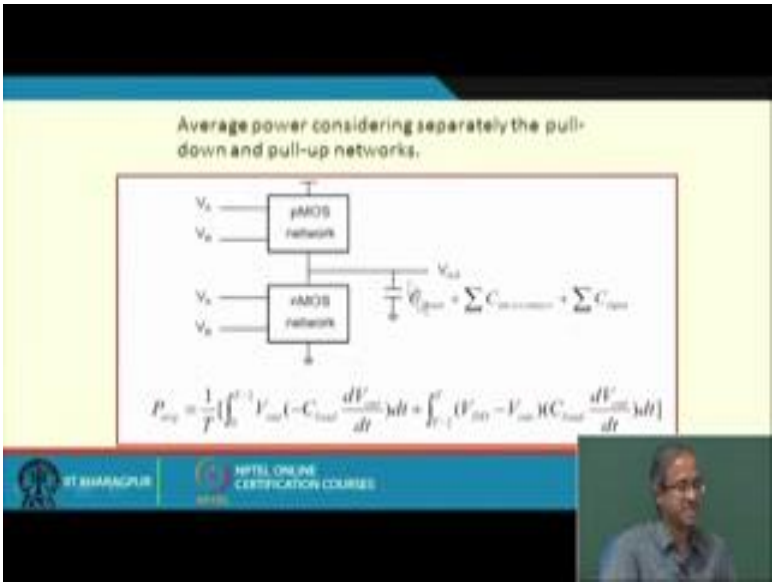
So, we shall look at as I said various techniques for reducing power; for designs which are expressed at the level of gates means here we may not be directly looking at the level of transistors where in some approaches of course, we shall be looking at, but in some of the approaches we shall be means creating an abstraction, we shall be looking only at the gates without going into the transistor level details.

So, even at the level of gates we can do some analysis, and you can do some restructuring thereby we can reduce the power consumption. Now the advantage of these gate level design techniques for reducing power is that, these methods can be quite easily integrated as part of the synthesis tools, like the cat tools which are there these can be integrated there itself and of course, one very important design principle that is being

followed is that, you try and incorporate power aware design philosophy from the very beginning, do not wait till your others things have been done only at the end you are you start thinking about power do not do that.

Even from the very high level starting from the behavior description itself, you keep at the back of your mind that your final circuits as to be power aware as to consume less power, unnecessarily power consumption should be stopped; these are the things that have to be kept in mind; this will simplify the problem later of course, yes.

(Refer Slide Time: 02:33)

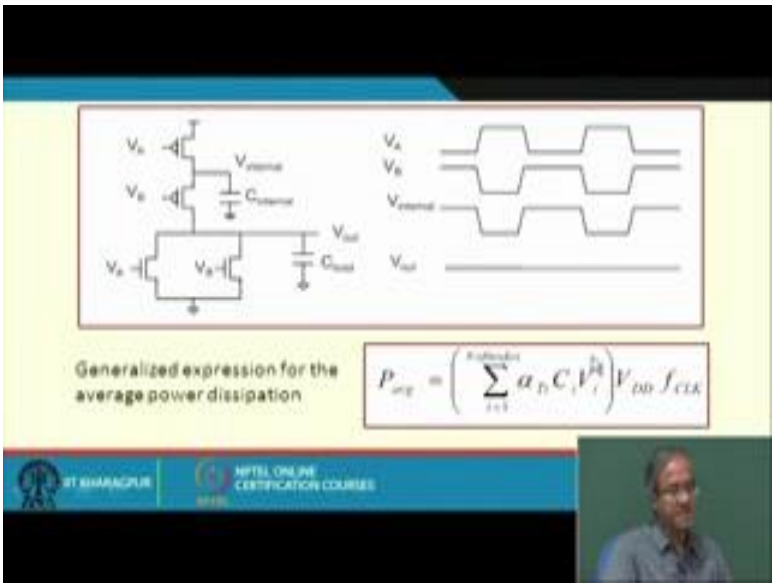


So, just recalling the power consumption in a cMOS network; so I have this pMOS cMOS this ground this V DD well. Here I am just showing that earlier I had shown only the load capacitance, but in practice this capacitance will be some of several things the sigma C input is the load capacitance, which is due to the driving of the gate of the other transistors. There can be the parasitic capacitance of interconnects this long interconnection line, and also the drain capacitance of the local transistors. So, sum total of all of these that will be the effective load capacitance, then the output will be driving right. And well this is a simplified equation, I am assuming that in every cycle which is 0 to capital T, the output is been switched from let us say 0 to V out and then from V out to 0 that means, 1 if charging and discharging.

So, I am showing the power consumption average power consumption 1 by 2 in two parts. I am saying that for half the cycles 0 to T by 2. So, the voltage of this node is V

out. So, this is discharging, C load dV out d t and in the other part you are charging. So, now, the voltage difference is V DD minus V out. So, if you take the sum total you will get the average power consumption over both the cycles charging and discharging.

(Refer Slide Time: 04:27)



And the other thing I want to also show with respect to this diagram is that you see you recall the average power dissipation expression.

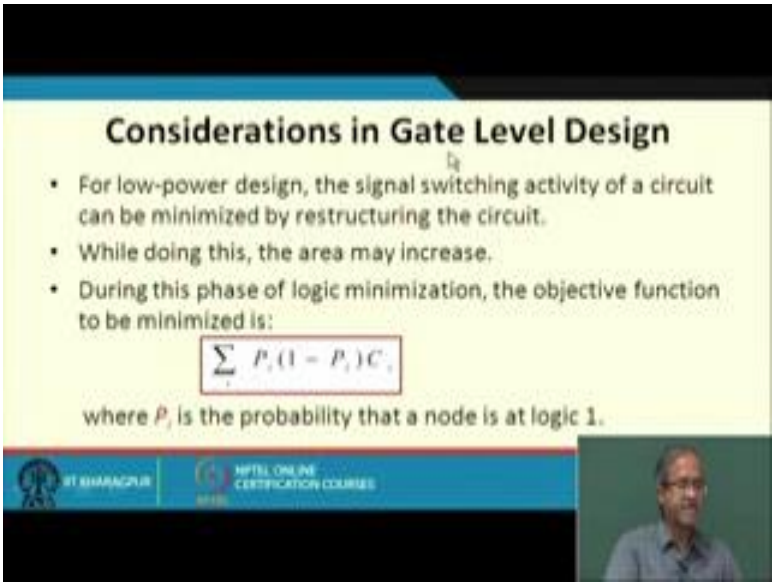
So, we showed an expression earlier, which looked like alpha C V DD square into f frequency of clock. But now we are refining this equation a little bit, we are saying is that well V DD into f clock is there of course, but here whenever we look at the charging and discharging of the capacitor, see it is not only the output node see all the other internal nodes of this gate; like here there is another junction out here. So, here also there will be an internal capacitance from this drain to ground, this is also not very negligible this also has to be considered.

So, whenever the inputs are changing, it is not only this load capacitance which is charging and discharging, but also the C internal which is charging and discharging; like for this example this is a 2 input nor gate, let us say V A and V B the 2 inputs are changing like this, V A is going from 0 to high, V B is going from high to low; so V internal. So, whenever V A is conducting, so V DD will come here. So, whenever V A is conducting; that means, it is 0, so V DD will be connecting here, it is 0 it is connecting

here the other time it will be 0, right. So, V internal will also be going through a charging and discharging phase.

So, C internal will also go through charging discharging. So, here you have to consider this charging discharging for all the nodes, not only the output node right. So, this is activity of the different nodes; activity of this node, activity of this node the corresponding capacitance and the corresponding voltage this is V out this is V internal. So, this will be a more accurate calculation of the power, right.

(Refer Slide Time: 06:39)



Now, let us look at the gate level design considerations. So, at the level of gates we are mainly concentrating at the dynamic power consumption. So, we have seen that the main reason for dynamic power consumption and also the short circuits power is whenever the gate output is switching state 0 to 1, 1 to 0. So, some signal activity we have we want to estimate the amount of signal activity, that is possible in a circuits right. So, there is an objective function we shall be explaining this a little later, this objective function we shall basically come back to it; where C denote the capacitance of a node, P i denote the probability, this P does not denote the power, but it is the probability that the logic value of a node is at logic 1.

You see let me talk about the justification P i into 1 minus P i; you see P i is the probability that the node is at 1, and 1 minus P i is the probability that the node is at 0. Now at a node when there will be power consumption, when there will be a transition

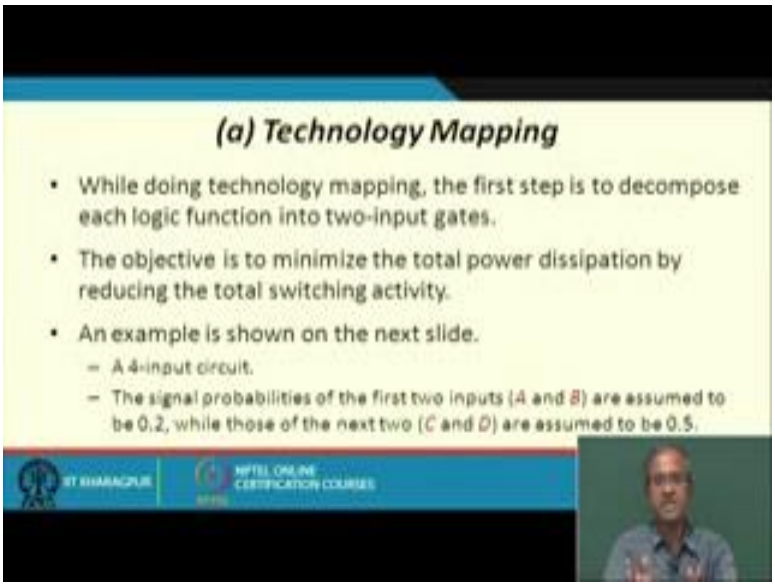
right. Actually when there is a transition that the value of the node is sometimes high, and sometimes low, then only there will be a transition. So, what will be the probability of that? Probability of high is P i and probability of low is 1 minus P i, so I take the product of these 2 pi into 1 minus P i. If I take the product it means I am saying that this is the probability that the logic value of the node will be at 1, and also will be at 0 which means there is a transition or vice versa 0 and 1 right.

That is why I take the product of P i and 1 minus P i that is the probability that there is a transition, multiplied by the capacitance value. So, if you take the sum total this will be the objective function that you want to minimize, because if I minimize this my dynamic power will also be minimum, fine.

(Refer Slide Time: 09:11)

So, there are many approaches which are used we shall be looking at some of this approaches one by one; technology mapping, phase assignment, pin swapping, Glitching power handling etcetera we shall be seeing these methods.

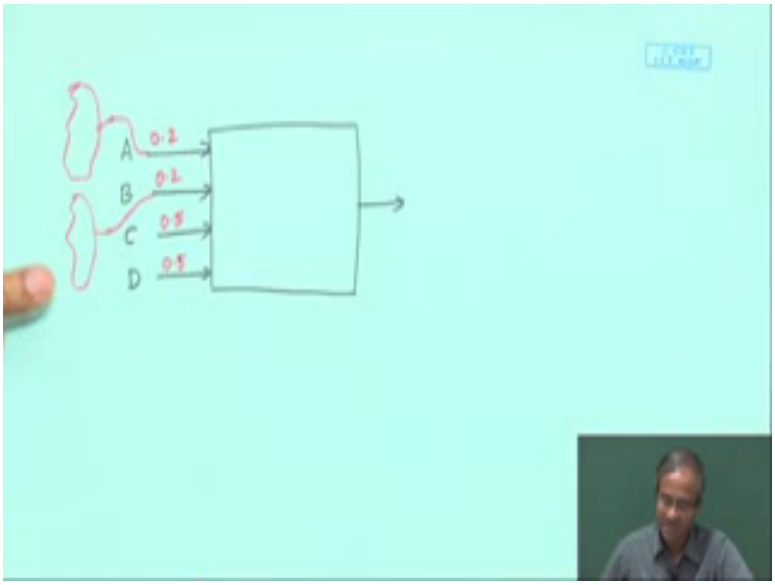
(Refer Slide Time: 09:26)



So, let us start with the technology mapping approach. See what is technology mapping? During logic synthesis; so what is logic synthesis? Given a specification I want to generate either a gate level or a cell level circuit. Technology mapping means I have a library where some standard cells are already present; may be two input gates, multiplexers this kind of things. So, when I synthesize my basic building blocks must be taken from that library only, this is called technology mapping. I generate a netlist of cells, where the cells are picked up from the technology library, this process is called technology mapping, fine.

So, here assuming that my technology library consists of 2 input gates which is quite true of course, few other type of gates are also there. So, here we are trying to define an objective function, where we are minimizing or trying to minimize the total switching activity; because we know that the power consumption is greatly dependent on the switching activity. So, we illustrate the process with the help of an example. So, we take a 4 input circuit with input A B C and D. See the idea is as follows.

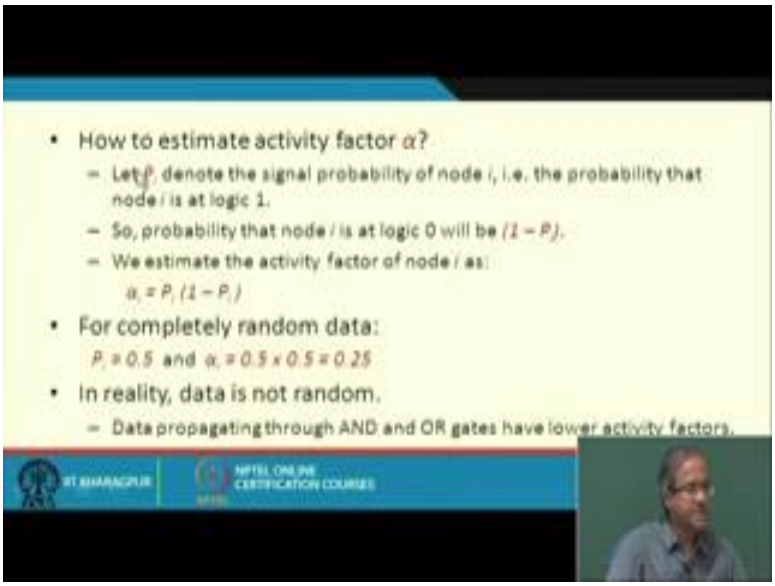
(Refer Slide Time: 11:06)



Let us say I have circuits there are 4 inputs A B C D. Now in this approach of course, there is an output, in this approach we are looking at the signal probabilities. So, I am assuming that the signal probability of the inputs A and B are 0.2 and 0.2 and C and D are 0.5 and 0.5 well. You can ask that how do I know these values? You see there can be some circuits from the output of which you are driving A, there can be some other circuit from where you are driving B, this circuits can determine that how frequently the output can be 1 output can be 0; there is a simple probability based calculation we shall be looking at it with respect to a block, that how this probability calculations are done.

So, 0.2 means the signal value at line A will be 1 with 20 percent probability will be 0 with 80 percent probability right. Similarly C and D will be equal 50 percent 50 percent, alright.

(Refer Slide Time: 12:25)

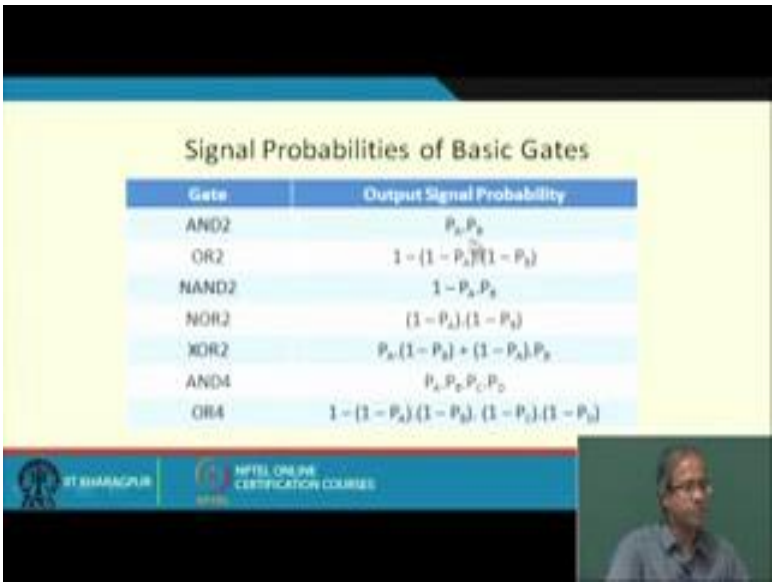


So, the estimation of the activity factor this is important. So, here we are saying P i will denote the signal probability and node i as we have just explained some time back, we use this notation P i to denote the signal probability of node i which means what is the probability that the particular line will be at logic 1.

So, the probability that the node is at logic 0 will be 1 minus P i, because the total probability has to be 1. So, activity factor from the consideration I explained just now sometime back, alpha I we defined as the product of P i and 1 minus P i, because there will be an activity whenever the node will be at 1 and also will be at 0, then only there will be a transition right. So, when the data which is coming is totally random, which means they are equally 0 or 1 the probabilities are equal. So, there is a 0.5 probability it is 1, and 0.5 probability it is 0 when the data is totally random.

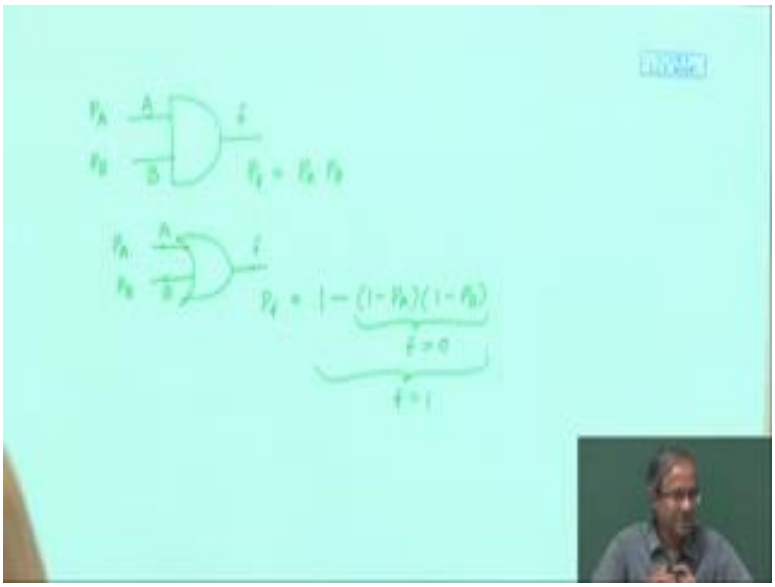
So, for totally random data P i will be 0.5 and hence alpha I will be 0.25, but in real circuits data will not be totally random, so the activity factors will typically will be much less. So, we shall be working out a simple example.

(Refer Slide Time: 14:06)



But before working out the example we look at how to compute the signal probability of the basic gates. See this equation may look a little complex, but the way they have been obtained are pretty simple consider a 2 input and gate.

(Refer Slide Time: 14:28)

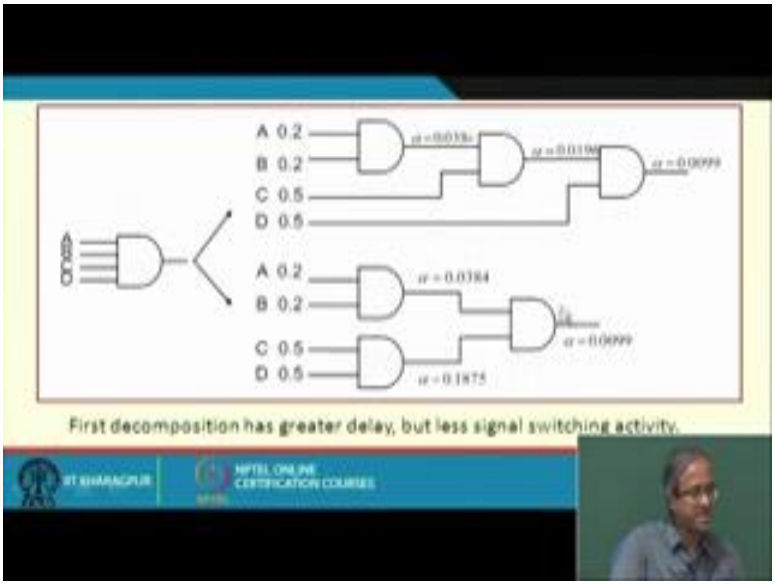


This is the probability that the input will be at 1. So, what will be the probability P f? The probability output will be at 1, for an and gate output will be 1 if both the inputs are 1, so it will be the product of P A and P B right. Similarly if we look at an OR gate same way 2 inputs the signal probabilities are P A and P B.

So, what will be P f here? See for an XOR gate, when will be the output be 1 while the rule is at least 1 of the inputs are 1; but it is difficult to write down probability that way, it is easier to write the probability that when the output will be 0. Let us see that we indirectly the output will be 0 if both the inputs are 0 right. So, what is both the input are 0 means 1 minus P A, 1 minus P B. So, 1 minus P A multiplied by 1 minus P B this is the probability that f will be 0 right, but we want the probability f is 1, if this is the probability of f is 0, then 1 minus of this will be the probability f will be 1, this whole thing will be probability f is 1. So, it follows like this right.

So, if you now look at the table you can justify the other ones; for a nand gate it will be just the reverse of and gate, the output of and gate will be 1 if both are 1 1, but nand gate will be 0. So, it will 1 minus of this. Nor gate this 1 minus will not be their; see XOR gate P A is 1, but P B is 0 or P A is 0 and P B is 1. For a 4 input and gate all the 4 are ones, 4 input or gate just an extension of these all 4 are 0, 1 minus of that, right.

(Refer Slide Time: 16:41)

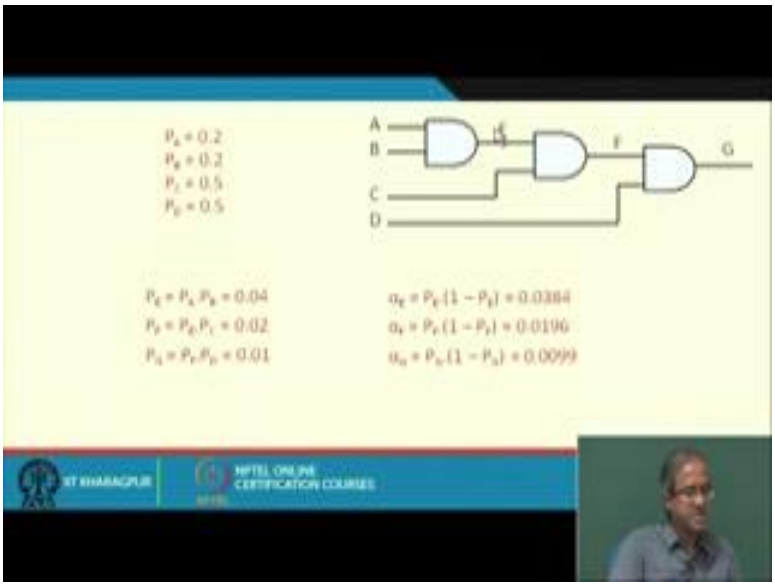


So, we shall be working out an example like this, I shall be showing you step by step evaluation. So, what we are basically saying is that, that I have a function which is a 4 input and, but my technology library as only 2 input. So, I can either partition my design like this or I can partition it like this. So, I do not know which 1 of them is better well; obviously, from the point of your delay, this one will be better because there are 2 levels

of gate delay, but here there are 3 levels of gate delay, delay will be more, but from power consumption point of view which one is better let us see.

So, I have assumed that the signal probabilities are like this 0.2, 0.2, 0.5, 0.5, let us work out.

(Refer Slide Time: 17:31)

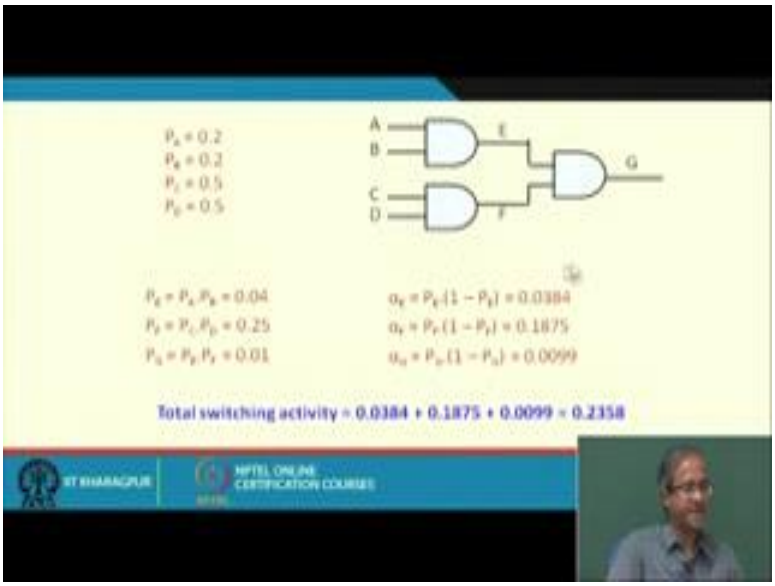


First let us look at this alternative, the first way of mapping where delay is 3 units. So, we are estimating power. So, what we have assumed? We have assumed that the input signal probabilities are these 0.2, 0.2, 0.5 and 0.5. So, we first calculate the probability values of the other lines just using those tables. So, what will be P E it will be product of P A and P B.

So, let see P will be is an and two input and gate product of P A and P B just multiple 0.04. P F will be product of P E and P C 0.04 into 0.5 it is 0.02, and what is P G? P F multiplied by P D 0.02 multiplied by 0.01. So, we have calculated the signal probabilities. Now let us calculate the alpha I values the signal activity values; now alpha by a definition is defined as P multiplied by 1 minus P. So, on this line E alpha will be P into 1 minus P, P is already calculated. So, if you just substitute it will be this alpha F similarly will be this and alpha G will be this.

So, for this circuit, the sum of the signal activities if you just add them up, so total switching activity will be 0.0679 right; for this kind of an alternative.

(Refer Slide Time: 19:31)

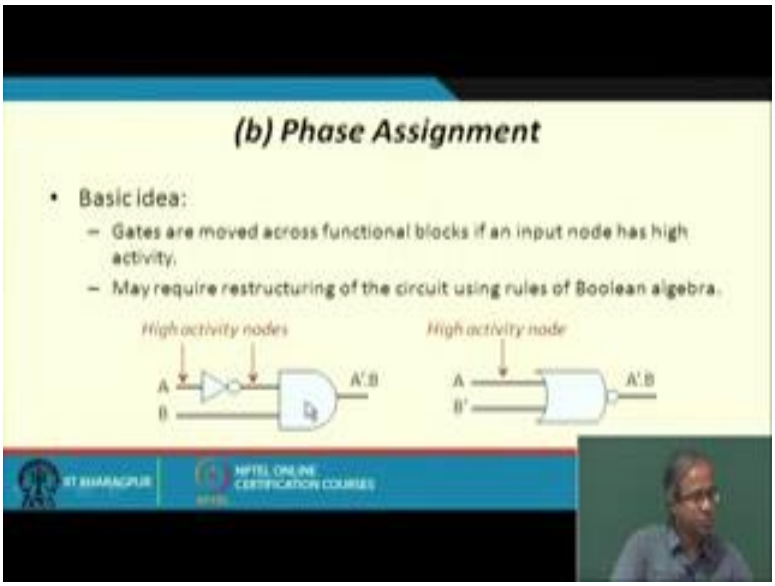


But now suppose we map the gate in this way, where the gates are in 2 level obviously, delay is less the gate will be faster, but with respect to power let us see, the same probability values. Now P values will be like this, P will be similarly P A P B 0 4, P F will be P C into P D 0.25, P G will be P into P F 0.01.

So, if you calculate now alpha E alpha F and alpha G, you will see the values are coming like this, but the value of alpha F is significantly higher right; and the total signal activity is also much larger in the earlier case it was about 0.06 something right, here it is 0.23. So, from the point of view of power dissipation, this is a very bad technology mapping.

So, what we are saying is that normally during technology mapping so far, we did not look at power; we simple looked at the delay how fast the circuit can work, but now there is an additional parameter, during technology mapping you can have another objective function, which will calculate the sum total of the signal activities, the activity factors alpha I; and tell you among multiple alternatives, which one is better with respect to power, also you know which one is better with respect to delay, then it is a tradeoff that you will have to decide that what is your objective fine.

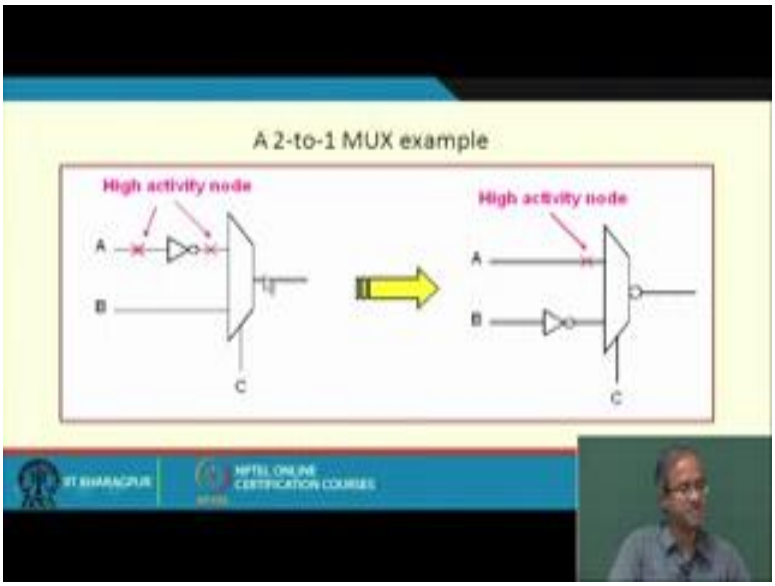
(Refer Slide Time: 21:12)



There is another technique which talks about something called phase assignment; here you can move a gate across functional blocks to reduce the activity. I am giving a very simple example, here we have a gate level circuit; just 2 gates, which computes the function A bar B, you see there is an inverter on input A. Let us assume that this line A is a high activity node where there is lot of signal activities, the value of P A is higher. So, because P A is higher, the output of this inverter will also have a higher value of p, because for every transition in the input there will be a transition in the output right.

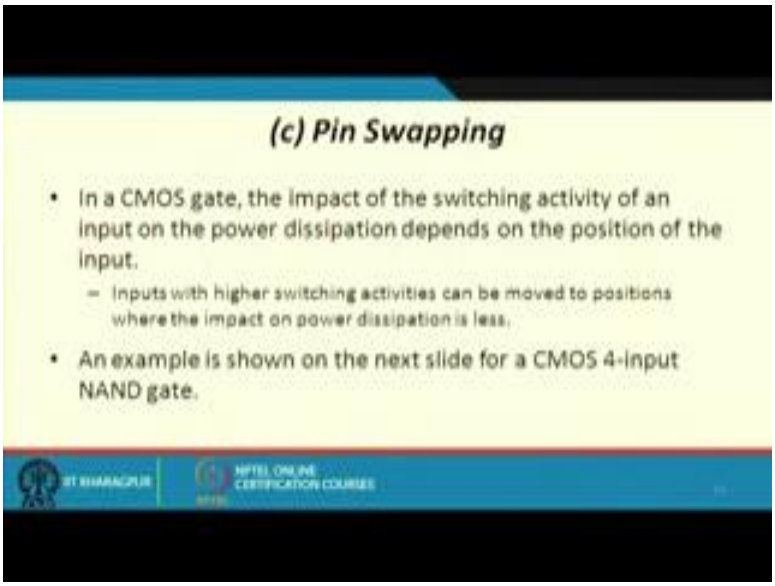
But suppose I make a technology mapping of this circuits like this So, that I modify this and gate into a nor gate and I change my input B from B to B bar, such that my output function remains the same right? But what I have achieved is that earlier there are 2 nodes that were high activity, but now there is a single node which is high activity right. So, naturally my power consumption will be reduced. So, we try to move gates around in sections of the circuit, where some higher activities are envisaged and you try to reduce the sum total of the these activities factor again alpha I values, by moving the gates around restructuring gates, using bullion algebra rules to change the functional wave like here we have modified an and gate into a nor gate just like that right, fine.

(Refer Slide Time: 23:11)



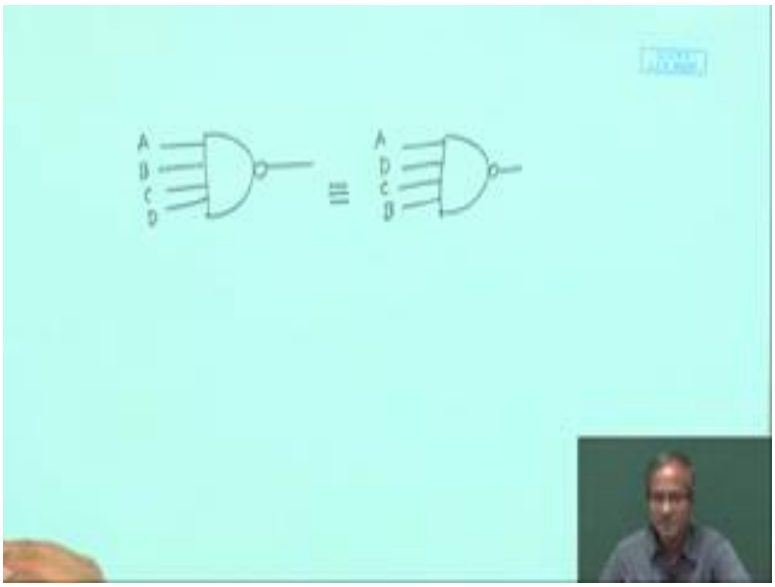
Let us take another example say a multiplexer; suppose I where 2 2 on multiplexer, where the 2 inputs A and B, A is inverted, and incidentally let us supports A is a high activity node. So, the output of this node gate will also be high activity. So, what I do I modify the circuits. So, I put the, I take out the inverter from here I put the inverter on line B and also I put an inverter on the output. So, functionally my multiplexer remains the same, because there will be two inversions. But what I have achieved instead of 2 high activity nodes, now I have a single high activity node; because output whatever it was a it remains same there is no change right. So, these are some simple examples which show that you can make some changes in your circuit netlist and also some functional blocks you can modify so as to reduce the signal activities, right.

(Refer Slide Time: 24:15)



Let us look at some more techniques; there is a technique called pin swapping, you see this is a interesting thing so far we have not considered this at all like what I am saying is something like this take a multi input gate, suppose I have a 4 input nand gate let say.

(Refer Slide Time: 24:46)

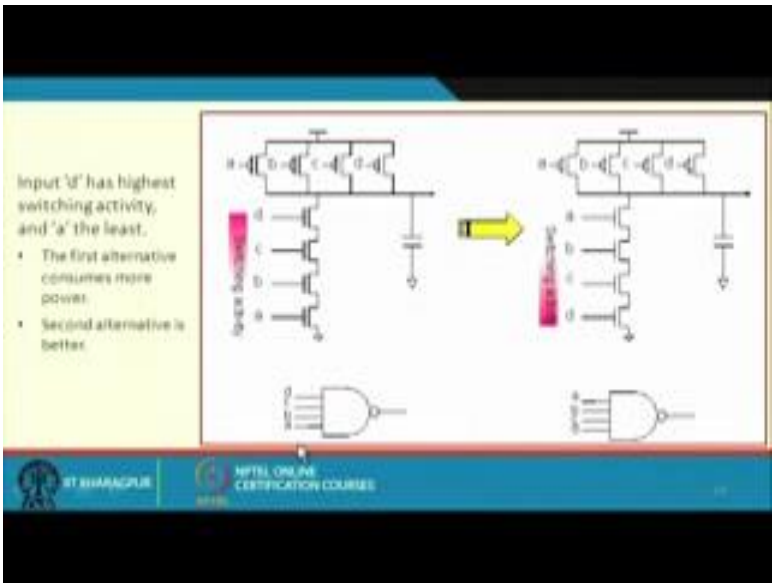


Normally when we have a multi input gate so, we connect ABCD like this, suppose we have another scenario same gate, but we connect the gates like say ADCB. Now because the nand function is commutative, where you can easily say that these two are equivalent, these two are equivalent functionally of course; but you may also be tempted

to think that well functionally they are equivalent, power dissipation wise also they should be equivalent because if there is a signal activity in one of these, there will also be signal activity in the other one and vice versa.

But in fact, in cMOS this does not always happen this is interesting. So, here what we says is that the impact of the switching activity of a gate input with respect to power dissipation, depends on the relative position of the input . So, now, we have another design flexibility or parameter, we can put the higher activity nodes to the inputs, whose impact on power dissipation is the least.

(Refer Slide Time: 26:18)



So, we are explaining this with the help of an example here; we take a 4 input nand function. So, the inputs we have connected are as follows d c b a, and this slide shows red means maximum activity, this lighter shade means less activity, which means the alpha value for d is the highest alpha value for a is the lowest.

Now, the observation here is follows; you see here we have only shown the load capacitance here, but as I had said there will be an capacitance with each of these intermediate nodes. So, if this node d switches very fast this is red, then due to some leakage current here or whenever this was on, this capacitance charging discharging will also be quite high alright. Because there will be a conducting path through here for some time. Similarly c is also quite frequently switching, so there will be times when both d and c will be conducting, and this capacitance will also be charging and discharging.

But if we put it the other way round, the least active input you put at the top. So, this transistor will be mostly non-conducting. So, even if the lowest transistor is very heavily switching, so the current switching and charging discharging impact will be the least. Because it is just down this track, lowest down these stack right. So, this is one design philosophy you can adopt, where looking at the signal activity of the input you can push the highest activity node to the lowest input like here a b c d will be a better input assignment, rather than d c b a.

So, with this we come to the end of this lecture, now in the next lecture we shall be looking at some more techniques at the level of gates by which you can control or reduce the power consumption in the circuit.

Thank you.