

Design and Analysis of VLSI Subsystems
Dr. Madhav Rao
Department of Electronics and Communication Engineering
International Institute of Information Technology, Bangalore

Lecture - 54
Interconnects - RC delay, and Energy

(Refer Slide Time: 00:28)

Example: Find Delay with and without wire

1mm Wire,
 $C = 0.2 \text{ fF}/\mu\text{m}$
Width = $W = 0.125 \mu\text{m}$
thickness = $t = 0.22 \mu\text{m}$

$$R = \frac{\rho l}{t \times W}$$
$$= \frac{2.2 \mu\Omega \text{ cm} \times 1 \text{ mm}}{0.22 \mu\text{m} \times 0.125 \mu\text{m}}$$

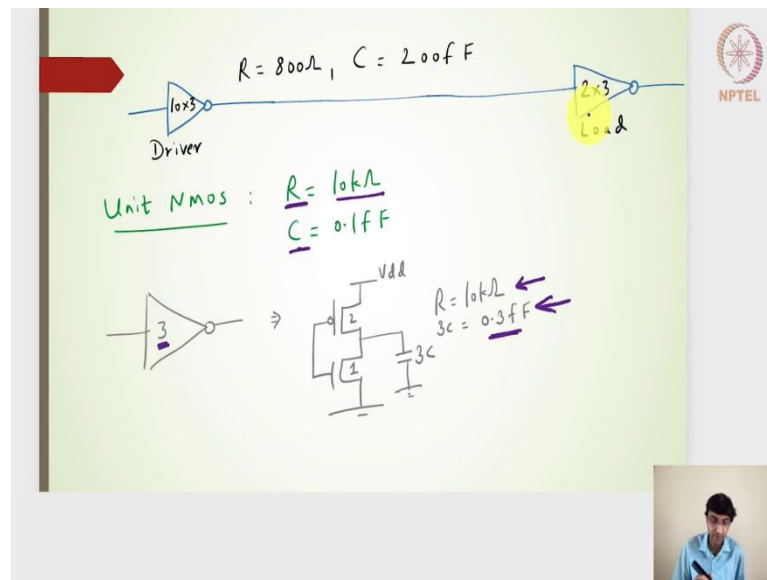
$R = 800 \Omega$
 $C = 0.2 \text{ fF}/\mu\text{m} \times 1000 \mu\text{m} = 200 \text{ fF}$

NPTEL

Hello students welcome to this lecture on the 2nd part of the Interconnects. In the last lecture we had seen one example where in a driver which is a part of the subsystem module 1 was connected to the another inverter or we can consider it to be a load. This being the load and we can consider this to be a subsystem module 2 and then this is the subsystem module 1 and it has been connected by a long 1mm length of wire and it has been characterized with a capacitance of $0.2 \text{ fF}/\mu\text{m}$ and with a width and thickness of this particular value of $0.125 \mu\text{m}$ and $0.22 \mu\text{m}$ respectively. We also calculated the resistance turned out to be 800Ω for this long length of wire and the capacitance turns out to be 200 fF .

We had also seen this particular driver the switching resistance and then the parasitic capacitance of the driver side and then the switching resistance and then the input capacitance as well as the parasitic capacitance on this particular inverter.

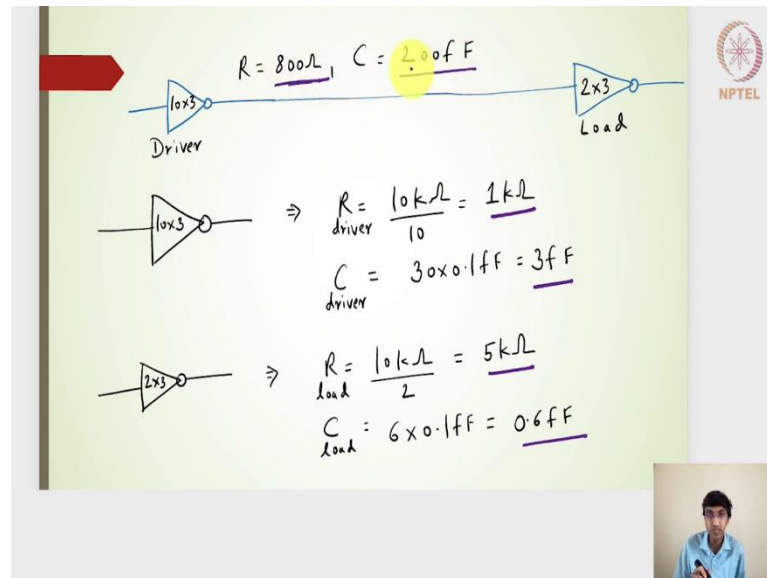
(Refer Slide Time: 01:38)



This was the switching resistance for a normal unit NMOS $10\text{K}\Omega$ and then the parasitic coming from or extracted from the unit NMOS transistor was around 0.1fF . For a 2:1 which is of the gate size of 3 inverter, we get the switching resistance as $10\text{K}\Omega$ and then the overall capacitance seen at the output node or the overall parasitic capacitance turns out to be 0.3fF .

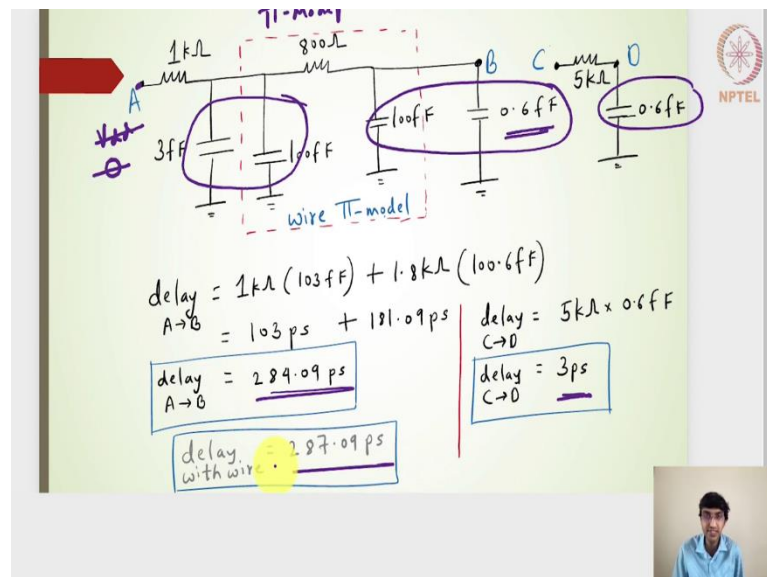
On the driver side we have the parasitic capacitance multiplied by 10 times and on the load side we will have multiplied by 2 times the capacitance of 0.3fF and then the switching resistance will get divided by 10 on the driver side and then on the load side it will get divided by 2.

(Refer Slide Time: 02:29)



This is what we have, $1\text{K}\Omega$ switching resistance and then 3fF is the total parasitic seen by this particular driver inverter and on the load side we have $5\text{K}\Omega$ and then 0.6fF on the load inverter with the interconnect switching resistance of 800Ω or rather resistance of 800Ω and then the capacitance of 200fF .

(Refer Slide Time: 02:58)



Now, if I actually put all those resistance and then the capacitance of the driver inverter of the interconnect and also the load inverter this is the circuit diagram which we get this is

the RC circuit diagram which we will get. The $1\text{K}\Omega$ is coming from the switching resistance of the driver inverter with that of the 3fF is the parasitic capacitance.

Now, that is the input A or rather this input A could be this is the switching resistance of the NMOS or the PMOS transistors based on the input to the inverter and then we have this 800Ω . We are using a π model here of the interconnect. The interconnect had a resistance of 800Ω and then 200fF . Out of the 200fF one part is going on this particular leg of the π model and another 100fF is going on to the another leg of the π model with the 800 ohm resistance is in the middle of this 100fF the 200fF . Lastly the 0.6fF is seen as an input capacitance of the load inverter and then I have taken the $5\text{K}\Omega$ as a switching resistance and then the 0.6fF as the parasitic capacitance is seen at the load inverter.

This particular point A it could be actually be V_{dd} or to the ground based on the input to that of the driver inverter. If the driver inverter input is 0 this A value will be V_{dd} and we have all the capacitance to be charged and then we should be able to find out the propagation delay for the rising one.

If the input to this particular driver inverter is actually 1 this particular A point will be connected to the ground supply that means, that all the capacitance should now start discharging to the ground.

I have left it very very generic in the sense that we need to estimate what is the propagation delay of rising or falling. Based on the input at the driver side we should be able to estimate that, but if I keep it generic then I should be able to find out the propagation delay it could be rising or the falling.

The delay from point A to point B is what we will try to estimate that. Using the Elmore delay method this is the resistance and then the capacitance these two will be in parallel and then we will have,

$$\begin{aligned} \text{delay}_{A \rightarrow B} &= 1\text{K}\Omega(103\text{fF}) + 1.8\text{k}\Omega(100.6\text{fF}) \\ &= 103\text{ps} + 181.09\text{ps} \\ \text{delay}_{A \rightarrow B} &= 284.09\text{ps} \end{aligned}$$

The delay of the C to D which is the next stage will be nothing but,

$$\text{delay}_{C \rightarrow D} = 5\text{K}\Omega \times 0.6\text{fF} = 3\text{ps}$$

If I put together given at the input to the driver, any changes in the driver inverter on the input side we will have the output at the point D which is nothing, but the output of the load inverter. Looking at the overall delay of these two stages including the inverter, including the π model of the inverter we actually get,

$$\text{delay}_{\text{with wire}} = 284.09\text{ps} + 3\text{ps} = 287.09\text{ps}$$

This is the delay with the interconnect or this is the delay with the wire.

(Refer Slide Time: 07:04)

The diagram illustrates an RC circuit model for two inverters connected by a wire. The driver inverter (A-B) has a switching resistance of $1\text{k}\Omega$ and a parasitic capacitance of 3fF . The load inverter (C-D) has a switching resistance of $5\text{k}\Omega$ and a parasitic capacitance of 0.6fF . A wire connects the output of the driver inverter (point B) to the input of the load inverter (point C). The total delay is calculated as the sum of the delay of the driver inverter and the delay of the load inverter.

Delay without wire = $1\text{k}\Omega \times 3.6\text{fF} + 5 \times 0.6\text{fF}$

Delay without wire = 6.6ps

Delay with wire = 287.09ps

Now, if I estimate delay without the wire that means, that the π model the R and then the capacitance which was divided into two parts and then we made the π model if that is not there what will be the delay of the two inverters, the driver inverter and then the load inverter? the driver inverter can be is represented in the form of the RC circuit.

This is the switching resistance of $1\text{K}\Omega$ and then 3fF and then this particular point is our B point where we will get the input capacitance of the load inverter. That will be 0.6fF and then the load inverters are switching resistance of 5k and then the 0.6fF as a parasitic capacitance. We will get,

$$\text{delay}_{\text{without wire}} = 1\text{K}\Omega \times 3.6\text{fF} + 5 \times 0.6\text{fF} = 6.6\text{ps}$$

Coming back to the present slide 287.09ps with the wire and without the wire the delay was 6.6ps. It turns out that the interconnects actually dominate in the overall performance or actually dominate in the estimation of the overall delay alright and then something has to be done to improve the performance or to reduce the delay.

(Refer Slide Time: 08:32)

Example to estimate switching energy

Estimate switching energy to send a bit of information from power rails.

$C_{\text{wire}} = 0.2\text{pF/mm}$

$E_{V_{\text{dd}}} = C_{\text{wire}} V_{\text{dd}}^2$

$E_{V_{\text{dd}}} = 0.2\text{pF/mm} \cdot 1^2$

$E_{V_{\text{dd}}} = 0.2\text{pJ/mm}$

The diagram illustrates a circuit where a V_{dd} source is connected to a wire with capacitance C_{wire} . The energy stored in the wire is shown as $\frac{1}{2} C_{\text{wire}} V_{\text{dd}}^2$.

In the last slide what we had seen the interconnects become very dominant in estimating the delay for a circuits for a simple circuit in this case where an inverter was connected to the another inverter and if I have a long interconnect of 1mm the 1mm line the 1mm copper wire turns out to be more dominant and gives us an overall delay of 287.09ps as compared to around 6.6ps it is a magnitude of 50 times more.

Now, here in this particular slide we want to estimate what is the energy. Normally for a circuit when we design there are three actual parameters, three aspects of the design one is what is the delay the second one is what is the energy or the power and then the third one what is the footprint.

Let us start thinking about what should be the energy that should be delivered by the power rails for this particular wire to communicate or to transfer a bit of information from one end of the wire to the another end and that particular energy we call it as the switching energy to send a bit of information from the power rails in the sense the V_{dd} rails. If I have a capacitance of the wire characterized for that particular interconnect as 0.2pf/mm.

The length of the wire is 1mm we can say that the overall capacitance of the wire is 0.2pF. What should be the energy that has been delivered by this V_{dd} ? and if I actually consider a simple model the L model which is R wire and then the C wire here. Alright I can actually consider different models π models or in T model, but this L model turns out to be very very simple to understand and that is why I have been used here in this particular slide.

If V_{dd} is connected to the resistance of the wire and then the capacitance or the energy that has been delivered by the V_{dd} is actually turns out to be $C_{wire}V_{dd}^2$, the energy delivered by the V_{dd} rail or the power rail. The energy consumed by the $C_{wire} = \frac{1}{2}C_{wire}V_{dd}^2$ the remaining $\frac{1}{2}C_{wire}V_{dd}^2$ will be dissipated as a heat from this resistance R_{wire} .

The energy that has been consumed by the C_{wire} will be half, the energy dissipated by the resistor will be $\frac{1}{2}C_{wire}V_{dd}^2$, but the overall energy that has been delivered by the V_{dd} , that this C_{wire} gets charged to V_{dd} will be nothing but $C_{wire}V_{dd}^2$ alright. The overall energy that has been delivered by the V_{dd} rail turns out to be,

$$E_{vdd} = C_{wire}V_{dd}^2$$

$$E_{vdd} = 0.2\text{fF}/\text{mm} \cdot 1^2$$

$$E_{vdd} = 0.2\text{pJ}/\text{mm}$$

If the length of the wire is 1mm, then we can say that the switching energy that has been delivered by the power rail is 0.2pJ, if it is 1mm of wire, if it is 2mm of wire it will be $0.2\text{pJ} \times 2$, hope this is clear moving ahead.

(Refer Slide Time: 12:23)

Example

Die of 20 mm × 20 mm running at 3 GHz in 65 nm process with pitch of 250 nm.
 → Half of the available wire tracks are used.
 Wire activity factor $\alpha = 0.1$ $C_w = 0.2 \text{ pF/mm}$

Number of wires = $\frac{20 \text{ mm}}{250 \text{ nm}} = 80 \times 10^3$

half of wire tracks = $N = \frac{80 \times 10^3}{2}$

$N = 40 \times 10^3$

$C_w = 40 \times 10^3 \times 0.2 \text{ pF/mm} \times 20 \text{ mm}$

$C_w = 160 \text{ nF}$

$\frac{N}{2} = 10\%$

NPTEL

Let us see an example here it is a very realistic example where we have a die where we have a chip of 20mm x 20mm size and we have a lot of wires here the pitching between the wires or rather I will say that the average pitching of the wires, pitching in the sense the distance between the or the spacing between the wires the individual wires or the neighboring wires is around 250 nanometers.

Overall this particular dimension is 20mm. The number of wires in this particular case will be nothing but 20 mm / 250 nm. I will get the total number of wires that is being present in this particular die and let us assume that we have only vertically placed wires we do not have horizontally placed wires at all.

This die is actually running at a frequency of 3Ghz in a 65 nanometer process technology process with a pitch of 250 nanometer that is something we have already seen and the another information is half of the available wire tracks are used. What it means is whatever is the number we get,

$$\text{Number of wires} = \frac{20 \text{ mm}}{250 \text{ nm}}$$

The number of wires if it is N only half of wire, half of this will be utilized the other half is not at all utilized. The other factor is the activity factor is given and then the $C_w = 0.2 \text{ pF/mm}$ has been characterized for the wires which are there in this, which are designed in this particular die. This wires may have connected to different transistors or a different

circuits, but we are ignoring the circuit at this point of time we will say that only the wires what is the overall energy that has been consumed.

Then the additional energy is actually consumed by the computational blocks because of the transistors or whatever is the design whether it is a NAND gate or an adder gate or a multiplier gate will be there from one end to the another end of the wire. That will be connected, but what we are saying is we are ignoring the computational blocks the logic blocks and we are only interested in how much is the energy that is consumed by the wires.

Let us look at this activity factor. The activity factor = 0.1, what it means is let us say that I have 10 wires 1, 2, 3, 4 and then 10th wire. The 1 to 10 wires out of this 10 wires at a particular time let us say that only 1 wire is where the logic block is sending the signal in the sense the logic block in this particular wire is now transitioning from 0 to 1 and then that has to be propagated to the other end of the wire.

All other blocks are having a steady state voltage, there is no change in the signal, there is no change in the voltage level. At another instance we will have one more wire changing the signal at one end and that has to be communicated that has to be transferred to the other end of the wire.

The activity factor says that on an average the probability that the one end of the wire out of the 10 wires only 1 wire has to transfer the change in the transition from 0 to 1. In that sense the wire activity factor says that if I have 10 wires out of that only 1 wire needs to supply that amount of energy to transfer the signal to transfer the transition of 0 to 1 from one end to the other end, all other 9 wires does not have any transition.

The activity factor in that case it is given as 0.1 which says that 1 in 10 wires is likely to do the transition and that energy has to be supplied for that particular wire. What we say is from the $N/2$ wires whatever is that number only 10 percent or 0.1 of the $N/2$ wires needs that particular energy to be supplied needs that energy to be delivered by the V_{dd} rail.

That the proper transition from that one end of the wire will be reflected at the other end of the wire, hope this is clear. Of course, the wire capacitance of the wire is characterized to 0.2pF/mm. Let us try to evaluate and then reach out to what should be the energy that needs to be supplied by the V_{dd} .

$$\text{Number of wires} = \frac{20\text{mm}}{250\text{nm}} = 80 \times 10^3$$

$$\text{half of wire tracks} = N = 80 \times \frac{10^3}{2}$$

$$N = 40 \times 10^3$$

$$C_w = 40 \times 10^3 \times 0.2 \frac{\text{fF}}{\mu\text{m}} \times 20\text{mm} = 160\text{nF}$$

(Refer Slide Time: 18:26)

Energy delivered = $\alpha_w C_w V_{dd}^2 = 160\text{nF} \times 0.1 \times 1^2$

$E_{V_{dd}} = 16\text{nJ}$

Power delivered = $E_{V_{dd}} \times \text{frequency}$

$= 16\text{nJ} \times 3\text{GHz}$

Power = **48W**

Note: This is for one layer. Multiple layers will add more power dissipation by the power-rail.

1 clock
3GHz

Finally, the energy delivered is nothing but,

$$\text{energy delivered} = \alpha_w C_w V_{dd}^2 = 160\text{nF} \times 0.1 \times 1^2 = 16\text{nJ}$$

If the frequency is given, let us also try to estimate the power. The power delivered will be nothing but the energy divided by the time period of the clock signal. In that sense if the clock time period is not given, but the frequency is given, we can say that it could be multiplied by frequency and then we will get,

$$\text{Power delivered} = E_{V_{dd}} \times \text{frequency}$$

$$= 16\text{nJ} \times 3\text{GHz}$$

$$\text{Power} = 48\text{W}$$

The overall power turns out to be 48 watts just by the wires and then we have not even accounted for the computational blocks that are there in this particular die and we are also talking about only one level of the die there could be multiple dice which are stacked in a 3D manner and the overall power is actually it turns out to be very very huge and it is mostly.

In this particular situation where there are multiple wires it is all always the energy or the power is actually dominated by the wires and not by the computational blocks itself, that is what I have written here. Note this is one layer and multiple layers will add more power dissipation by the power rail only for the wire to transfer that particular logic from one end to the other end or whatever the transition that happens from 0 to 1 that has to be transferred to the other end of the wire and only for doing that the power that has been delivered by the V_{dd} rail turns out to be 48 watts which is kind of very very huge. This is the major problem and today's SOC designed the system on chip because there we see the high density of the wires that is kind of running around in the chip design and that has to be resolved.

The other important point here I wanted to make is why is this frequency coming into the picture or why is that clock time period coming into the picture? We have actually used the activity factor and activity factor it is basically a probability for the number of the amount of the or the number of the wires within a group of the wires where it is doing the transition, but for the power to be calculated we always consider it is nothing but over the average time period. Whatever energy we estimate and if I actually do the average of the energy over one particular time period that should give me the power.

The power is always with respect to the energy divided by the time. This particular time is nothing but we consider it for a one particular clock time period. Then normally for a chip design usually we will have a limited number of clocks to be at a primitive level what we are seeing is we will have one particular clock to be designed and we will measure the power with respect to this particular clock.

Now, that particular clock frequency is given as 3Ghz and hence we actually do estimate the power with respect to that particular clock frequency. Energy multiplied by that particular clock frequency is our 48W.