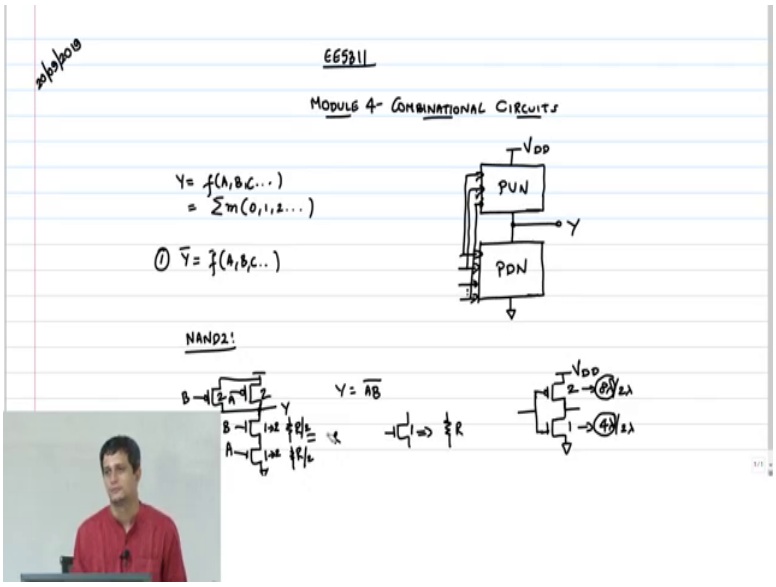
Digital IC Design
Prof. Janakiraman Viraraghavan
Department of Electrical Engineering
Indian Institute of Technology, Madras

Lecture – 35
Gate Sizing

(Refer Slide Time: 00:14)



Good morning everyone let us proceed with the discussion on combinational circuits right. So, yesterday we discussed how to construct any arbitrary Boolean function right. So, if you have Y equals f of A, B, C and so on right then you are going to assuming that this f is in a sum of products term, sum of min terms 0, 1, 2 and so on right.

Then you basically convert this step 1 you convert this Y to its complement function f bar of A, B, C and so on. And, then implement the SOP expression in the NMOS stack right. So,

you will get the pull down NMOS stack like this it is called pull down network the output Y for some specific combinations where the output goes to 0, Y will be 0, otherwise Y will be in it is high impedance state in this network right, then you construct the dual of this right.

The dual is basically converting series to parallel and complementing the inputs right, but of course, since these are all PMOS transistors you do not have to correct the complement of the input to the transistor just connect the same A, B, C and it should be fine ok. So, this is the pull up network to be reading and basically you have the same inputs going to all of this is where we started with this discussion and then we proceeded to sizing the transistors because we have to now fix the which of these transistors length is always fixed right.

So, therefore, I said that we consider the example of a NAND2 gate and if you had this implementation which is very obvious. So, I am not going through it in detail A, B, A and B output Y; Y equal to AB whole bar. Then the size of this NMOS transistor was to be chosen so that the net pull down strength of that NMOS stack in the worst case is as much as it is for a reference inverter right and what is our reference inverter? It is basically this guy V DD.

So, this is 1 and this is 2 and what is this 1 and 2 mean this is this could be for example, 4 lambda by 2 lambda, this could be 8 lambda by 2 lambda right. We are only talking about the width when we talk about 1 and 2 right. Ideally if area is a premium I should keep my area as low as possible.

So, I would have liked to keep the sizes like this 1, 1 even PMOS I would like to keep 1, but then we saw that because of mobility problems you have to double it in order to get equal rise and fall delay right. So, that we made 2, 2. So, I would have like to make this 1, 1, 2, 2 for the NAND as well, but the problem is that if a unit transistor width 1 offers a resistance of R, then a transistor of width alpha will offer a resistance of R by alpha right.

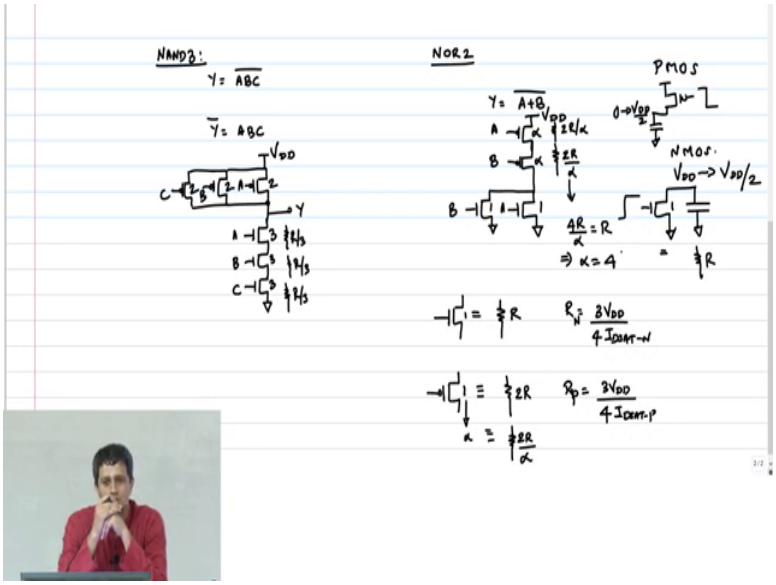
So, if this unit transistor now offers a resistance R and this is also R the net pull down resistance is 2R right this is equal to 2R, but we want the net pull down resistance to be R right. The reason is see now for NAND3 when I go to NAND3 I will have a 3 stack that will

become 3R, 4R and delay will keep going up. So, somewhere we have to normalize this delay.

The best thing to do is to by definition normalize it to the delay of the unit invert right; given that we are going to increase the size of the transistor anyway how much do I increase I have to standardize that. That we are saying is make the resistance worst case pull down resistance or pull up resistance equal to that of a reference CMOS inverter right.

And, therefore, this has to become 2 and this also has to become 2 in which case the resistance will be R by 2, R by 2, net resistance will become R ok.

(Refer Slide Time: 05:55)



So, now we will proceed to something like the NOR gate or let us do a NAND 3 quickly Y equal to ABC bar right. So, what do I do I am going to invert this function. I will get Y bar

equals ABC right. So, I am going to simply put three NMOS transistors in series A, B and C right. This again is A, B and C. The dual is make all series make it parallel right and correct the same inputs to those transistors.

So, now what should the size of the NMOS transistors be? 3R right because now there are three resistances if I make this 3, 3 and 3 then each will give me a resistance of R by 3, R by 3 and R by 3 right. What should be the PMOS size? 2, because the worst case pull a path for the PMOS is only one PMOS pulling up; if all the PMOS transistors pull up together it becomes a parallel combination and the resistance drops right. So, if the resistance drops there is no problem, delay is getting better right.

So, therefore, I can keep the PMOS transistors at 2 always irrespective of the number of inputs to the NAND gate ok. So, now let us look at the NOR gate, NOR2. Y equals A plus B whole bar. So, can you just draw the NMOS and PMOS stack first? Right. This is the circuit. So, what should be the NMOS transistor size now?.

1, because I have a parallel stack now worst case is only one NMOS transistor turns on, it is 1 ok. So, now if a unit NMOS transistor gave us a resistance of R ok, this is size 1 ok. Remember this formula was R was 3 V DD by 4 I DSAT ok. What would this transistor give us? PMOS transistor of same width it will give you 2R. Why because the mobility difference is there right.

So, if I look at the this is NMOS R R PMOS this is 3 V DD by 4 I DSAT-P, N. You take the ratio of the resistances assuming that the widths are the same it will eventually be the ratio of mobilities assuming all the other technology parameters are symmetric V TN equal to minus V T P, V DSAT equal to you know minus V DSAT P and all that then all those terms will cancel. You are free to work out this expression for a PMOS transistor charging.

So, how did we evaluate the resistance for an NMOS transistor we said that a capacitor was charged to V DD and this discharge to V DD by 2 right and this went up instantaneously. What is the equivalent for the PMOS transistor it is discharged initially and it has to charge to

V DD by 2 right. So, an identical scenario will occur there because the V DS or the PMOS will start from V DD and go exactly down to V DD by 2 right.

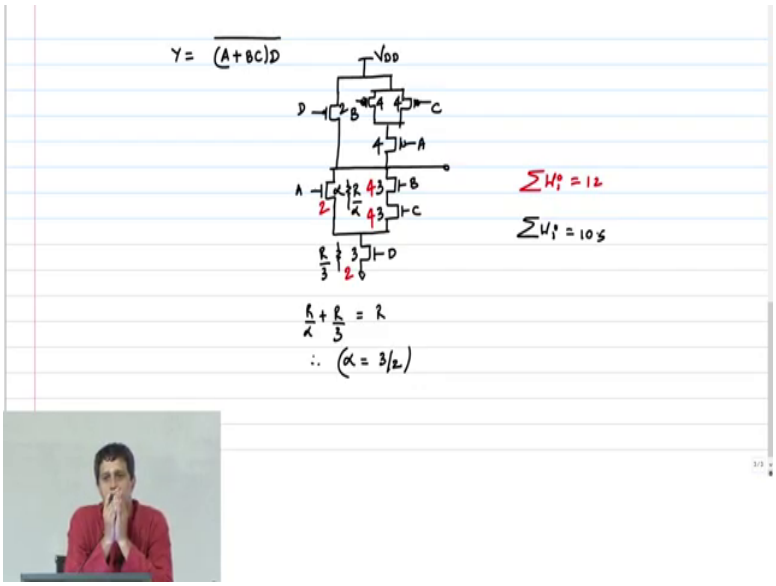
So, you can evaluate that just to tell you. So, this would give me a an equivalent resistance of R that is what I am saying for the PMOS this is a NMOS for the PMOS it is basically write like this right and the capacitor is going from 0 to V DD by because that is a definition of delay for rise time when the output goes from 0 to V DD by 2; fall time is when the output falls from V DD to V DD by 2 right. And, this has to drop instantaneously. So, you can evaluate this expression.

If the two technology parameters V DSAT, V TN and all that are symmetric then everything else will cancel out you will be left with only the ratio of mobilities right and therefore, a PMOS transistor with unit width is going to be a resistance of 2R. Now, if I scale this up the PMOS transistor by a factor of alpha. Let us say this one became alpha, then what would my resistance become? That is the key this is a equal resistance of 2R by alpha ok.

So, what I am going to now do is I am going to say that the width size of this PMOS transistors are all alpha. I want this resistance therefore, is 2R by alpha right the top transistor offers another 2R by alpha and therefore, the net resistance is what? 4R by alpha and I want this to be equal to R. The pull up resistance should also be equal to R because we have if the or you can just say that even the pull up resistance should be the same pull up resistance as offered by a reference inverter that we have already sized twice so, it is the same thing.

So, if I say this is R implies alpha equal to 4. So, from the inverter it is not surprising if you stack transistors you need to add p you know multiply the width by the stacking size right. So, unit inverter is 2; NOR2 is 4 NOR3 is 6 and so on ok. Any questions here?

(Refer Slide Time: 14:00)



So, let us look at maybe a slightly more involved example Y equals a let me see if I have A plus BC into D whole bar. So, can we implement this CMOS the first, then we size this is right. So, what are the sizes of for we tell me for the NMOS. So, first thing what is the worst case pull down path B, C, D. So, therefore, the sizes should be 3 3 3 right. So, 3 3 3 right

Now, what should the size of A, B? If B is 3 correct 3 by 2 cannot be less than 1, right that is a golden rule because the inverter has size 1, you cannot get less than that right. So, therefore, if this is alpha let us assume this is alpha then this resistance is R by alpha, this resistance is R by 3 and I want R by alpha plus R by 3 to be equal to R. Therefore, alpha will be equal to 3 by 2 ok.

Now, what about the option where this is made 4, 4; if both are 4 then what should DB? 2 correct. So, this will be 2. Now, if this is 2 then I can make A also 2 which one will you

choose? First one, why? Lowest width or lowest something else? Can you tell me what the summation of W i's these in both cases?.

This red one is how much? 12. Black one? 11 and a half, correct? 9 plus 1.5, 11.5 10 correct sorry 10.5. So, which one we will choose? The black one because that has lesser area right. So, in fact, this style of sizing where you size one gate one transistor higher than the other deliberately is also allowed, but that is basically it will come in when you are doing skew sizing ok.

Essentially what is happening there is we are preferring one input over the other input ok. So, I will come to that at a later point. But, for now pick the one that has least area and no other configuration is going to be useful right and of course, the point is between B, C and D if you cannot pick one input to be more critical than the other then you have to keep all of them the same size.

Otherwise what happens is depending on which input is switching you might get a different delay right, you do not want that right. Or the load that input is offering to the previous gate will be very different if you size it as 4, 4, 2 right, but if there is no reason to pick that input and specifically sacrifice the delay on other inputs then you do not adopt that in general ok. Any case the area is lesser in the equal assignment case.

So, what about PMOS? What should DB? 2 right? There is only one transistor path there no problem. Now, A and B it could go A and B or A and C right. So, B and C should be similar sizing and therefore, this has to be 4, 4 and 4 ok.