CMOS Digital VLSI Design
Prof. Sudeb Dasgupta
Department of Electronics and Communication Engineering
Indian Institute of Technology – Roorkee

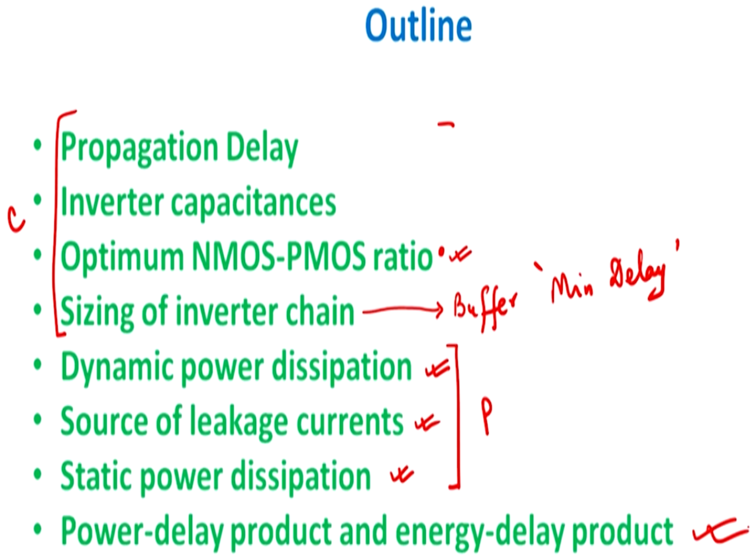
Module No # 02
Lecture No # 07
CMOS INVERTER BASICS – III

Welcome to this next module of the NPTEL online certification course for CMOS digital VLSI design and we start the second part of the inverter. In the first part we have seen the static characteristics of an inverter how do you draw a voltage transfer characteristics which is VTC and then how do you extract the switching threshold and the noise margins from the VTC. We have also understood that if these noise margins are available to you on which parameter is the noise margins and switching threshold are dependent.

What will be looking in this section on this slide for this module are the following we will be looking at the concept of propagation delay right. We will be looking at the inverter capacitances the main reason the motivation for this lecture is that at the end of the day when you are actually having a large chop right and you have primary impedance primary output available to you then you should know now much amount of delays is there between primary input and primary output across the critical path.

So that I could extract the total frequency or the maximum frequency of operation of the chip and therefore estimation of propagation delay in a chip is an important fact for that we require to have a knowledge of capacitances and resistances and therefore this is dedicated or this module is dedicated to find out the basic inverter capacitances on which factors do the depend.

(Refer Slide Time: 02:09)



Outline

- Propagation Delay
- Inverter capacitances
- Optimum NMOS-PMOS ratio ✓
- Sizing of inverter chain → Buffer 'Min Delay'
- Dynamic power dissipation ✓
- Source of leakage currents ✓
- Static power dissipation ✓
- Power-delay product and energy-delay product ✓

We also look into the optimal value of the NMOS to PMOS ratio means we have just now studied if you remember correctly just few at the previous lecture that you require the aspect ratio of PMOS should be about two to three times larger as compared to that of the NMOS's in order to have the symmetric VTC of the voltage transfer characteristics with VN approximately = VDD / 2 which is the natural coincidence or a very ideal value but in many of the cases I do not want the symmetric VTC to appear I want that it should be shifted either more towards VDD or less towards VDD.

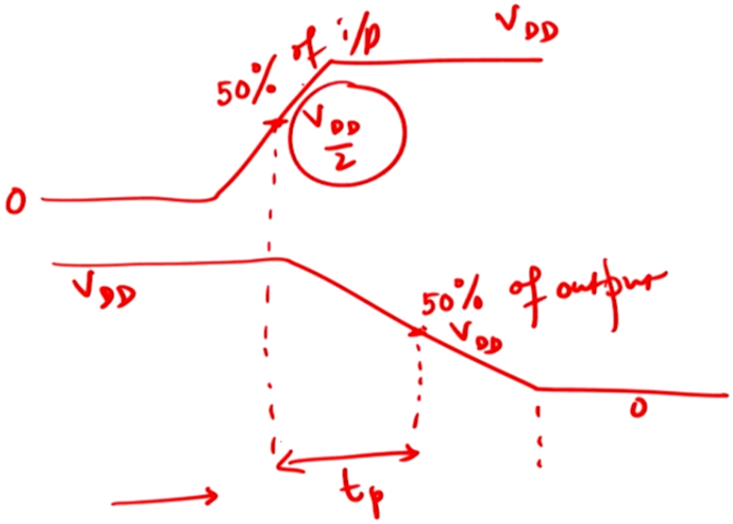
So I can make it VDD / 4 or I can make it 2 / 3 VDD and so on hence so forth so then and that stage what should my optimal ratio size of NMOS to PMOS that we are looking at this point then as I discussed you in the last part in the previous lecture that I generally have an inverter chain develop or I have a cascaded inverted chain which will take care of its activity as buffer right.

So we need to know the sizing of this buffer chain for this buffer chain right such that I get the minimum delay. So minimum delay condition have to found out so what we will be trying to do here is we will try to size the each inverter aspect ratio each inverter such that I get a minimum input and output we will be looking into the various understanding the power dissipation very important idea we will look at leakage currents and static.

So there are three parts of dynamic of three parts of power dissipation one is dynamic one is tactic basically a leakage one is short circuit another is leakage. So you will take all the three individually power dissipation and we will studying how does power dissipation look so the first four points which you see here are primarily finding out the value of C and this is primarily defined the value of power you get right.

And then we end our discussion with EDP and PDP this energy delay product and power delay product right. Let me come to you the basic fundamental principles of propagation delay you must have understood how do you find propagation delay is that suppose if I have an input transition taking place here.

(Refer Slide Time: 04:33)

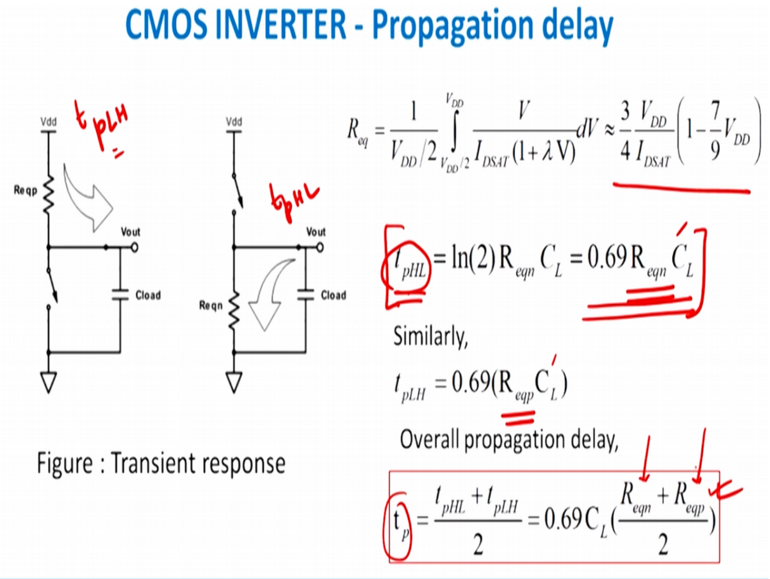


I have a input transition taking place here from 0 to 1 and the output is going from say 1 to 0 right now obviously if you see this is actually shifted in this direction because of the delay. So with that we defined like this as the difference between 50% of input rise so this is point where you have 50% of input rise to 50% of output fall this is 50% of output fall right and this 50% of input is this difference if you look very closely this difference which you see is referred to as my propagation delay 50% propagation delay right.

So I will have something like this difference is my TP propagation delay propagation delay from what to what? From high to low because your output was going from high to low right. So this is defined as propagation delay in this case TP 50% of the input rise to 50% of input fall is defined as my propagation delay. Now since it is 50% so I will so I can say safely assuming in voltage domain so it was 0 it is VDD right.

So I can safely assume that this point is effectively VDD / 2 right so if this is VDD this is 0 so this is VDD / 2. So I just need to find out at what point does VDD / 2 happen for input at what point does VDD/ 2 in the output find the difference in time domain and that will give me the propagation delay.

(Refer Slide Time: 06:14)



CMOS INVERTER - Propagation delay

$$R_{eq} = \frac{1}{V_{DD}/2}\int_{V_{DD}/2}^{V_{DD}} \frac{V}{I_{DSAT}(1+\lambda V)}dV \approx \frac{3}{4}\frac{V_{DD}}{I_{DSAT}}\left(1-\frac{7}{9}V_{DD}\right)$$

$$t_{pHL} = \ln(2)R_{eqn}C_L = 0.69R_{eqn}C_L'$$

Similarly,

$$t_{pLH} = 0.69(R_{eqp}C_L')$$

Overall propagation delay,

$$\boxed{t_p = \frac{t_{pHL}+t_{pLH}}{2} = 0.69C_L(\frac{R_{eqn}+R_{eqp}}{2})}$$

Figure : Transient response

So if you look here I get tp this is the formula which we use so remember it is V/ I = R very straight forward I get V / IDSAT into 1 + lambda V this is primarily because of CLM and we vary from VDD / 2 to VDD because that is the transition we are doing. So input 50% to up to VDD we are trying find out and R equivalent will be given by this formula actually please note there is a mistake here a lambda will be inserted after before VDD.

As you can see therefore higher the value of lambda which you see small be R equivalent higher the value of lambda implies that your CLM will have a such profile right. So rather than the ideal value of lambda = 0 if lambda is very high what happens is that it is almost like this slope which primarily means that this is VDS versus IDS then here even if you changes VDS by large amount your ID is not changing.

So del VDS, del IDS is approximately infinity so which means that is current source but as you make it more and more sloppy your resistance starts to fall down and this my exactly what is happening. So as you make a lambda higher and higher 1 – that goes on decreasing and as a result R equivalent goes on decreasing which means that the resistance offered by the device is actually reducing as you can see here this is the charging flow.

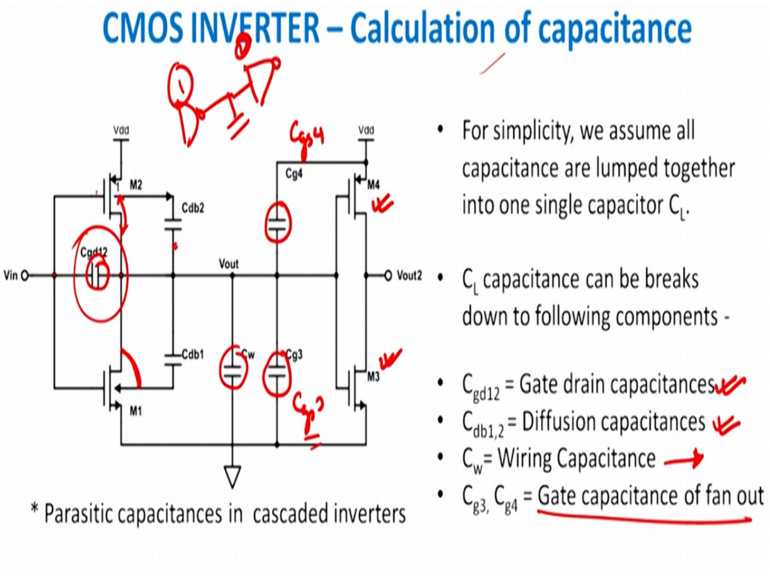
So I have got VDD charges to R equivalent P switch model charges C load and your NMOS was switched off in the second off cycle when when the NMOS is switched on it becomes R equivalent this was switched on switched off and this capacitance was discharging its current through the to the ground. With this knowledge to me therefore conclude that for this case therefore so I get this is my as my this thing so I defined tpHL.

tpHL is defined as the propagation delay for output to go from high to low right so please understand tpHL is basically the propagation delay for the output voltage to go from high to low right. So from generally we referred to this as high to low means 10 to 90 % we will also be going it this is the proximately goes to 0.69 LN of R equivalent to CL we got it from the first information of the first differential equation of RC common circuits.

Similarly tpLH which is so this is basically tp output is going from low to high so this low and high and this is tpHL high to low. So you see so tp so sorry tpHL high to low will have R equivalent N whereas tpLH which is this one will have R equivalent P as the value here in both the cases CL is the same. We define the overall propagation delay as the average of high to low and low to high and this come out to be R equivalent P + R equivalent N divided by 2 into 0.69 into CL where CL is the load capacitance.

We come to therefore so we therefore the previous slide we have if you know the value of RE equivalent N and R equivalent P you can calculate the total resistance of the average resistance seen by the device to the current flow we have just now therefore trying to find out how will your CL change or how what are the factors which influence the CL value the next slide we may get clear.

(Refer Slide Time: 09:55)



CMOS INVERTER – Calculation of capacitance

- For simplicity, we assume all capacitance are lumped together into one single capacitor C_L.
- C_L capacitance can be breaks down to following components -
- C_{gd12} = Gate drain capacitances
- $C_{db1,2}$ = Diffusion capacitances
- C_w = Wiring Capacitance
- C_{g3}, C_{g4} = Gate capacitance of fan out

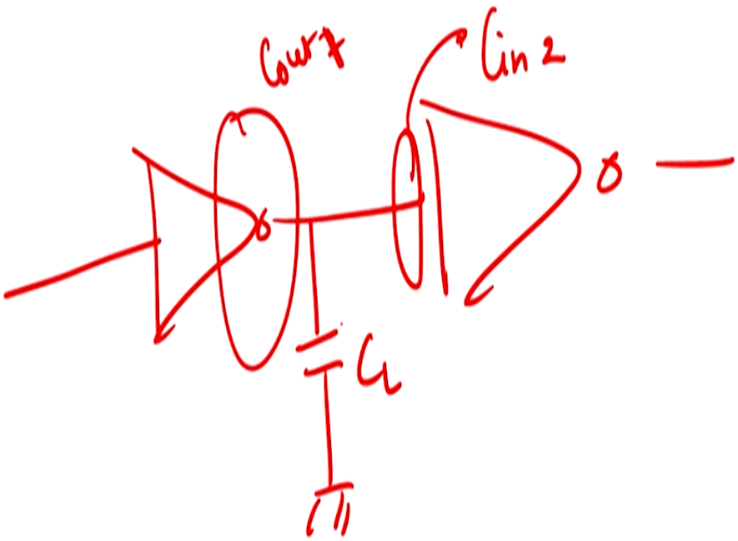
* Parasitic capacitances in cascaded inverters

If you look here I have one inverter here I have another inverter here so one inverter is driving another inverter here so there two inverter driving each other that is what you see in front of you and as you can see here that this as we discussed in our previous understanding Cdb2 is basically drain to bulk of the second transistor for M2 drain to bulk right. So this is drain to bulk sorry I will just make it clear to you this is drain to bulk this one right appears here drain to bulk of one is appearing here right we have gate to drain or gate to drain is basically gate to drain this one we also have the wiring capacitance Cw coming here.

And please understand this is the gate capacitance for four and gate capacitance for three this is what is it tell me this is basically gate to source. So Cgs so you can as well find out Cgs4 and this to be as Cgs3 gate to source right very important and as i discussed with you it consist of Cox and C depletion when we were discussing earlier stage but primarily C oxide is there available to you and we have to be very careful about it so adding.

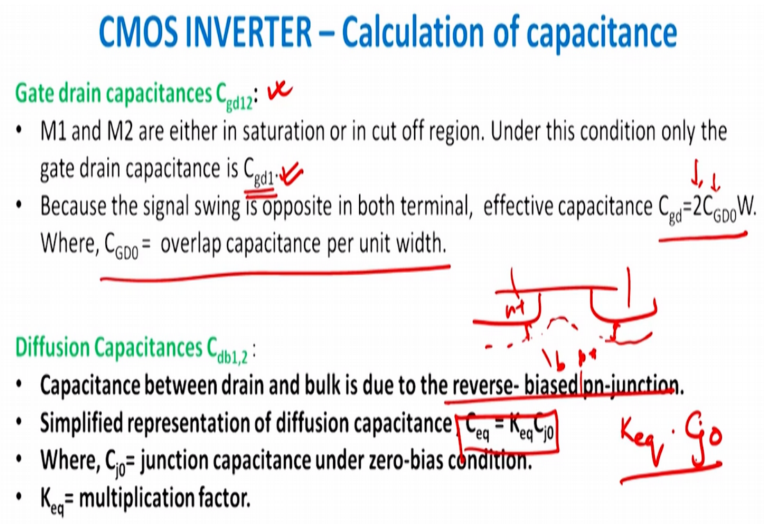
So as you can see this is gate drain diffusion we have a diffusion capacitance here Cw is the wiring capacitances primarily that capacitance is by virtue of the wire available to you and Cg3 and Cg4 are nothing but the gate capacitance of fan out. So this is the fan out here fan out is 1 here and therefore this is basically two devices M4 and M3 whose input capacitance is Cg. If you look very carefully the output capacitance of the first inverter also consist of the input capacitance of the second inverter like this is the first inverter or the second inverter.

(Refer Slide Time: 11:40)



The output capacitance of so if this I the first inverter this is the second inverter then the output capacitance so the first inverter which is CL here will consist of the output capacitance of first inverter and input capacitance of the second inverter right. So this will be in parallel and they will be added up and you will get a CL value here fine so I think I have made it clear that what is the composition of the these values of inverters.

(Refer Slide Time: 12:05)



CMOS INVERTER – Calculation of capacitance

Gate drain capacitances C_{gd12}:

- M1 and M2 are either in saturation or in cut off region. Under this condition only the gate drain capacitance is C_{gd1}.
- Because the signal swing is opposite in both terminal, effective capacitance $C_{gd}=2C_{GD0}W$. Where, C_{GD0} = overlap capacitance per unit width.

Diffusion Capacitances $C_{db1,2}$:

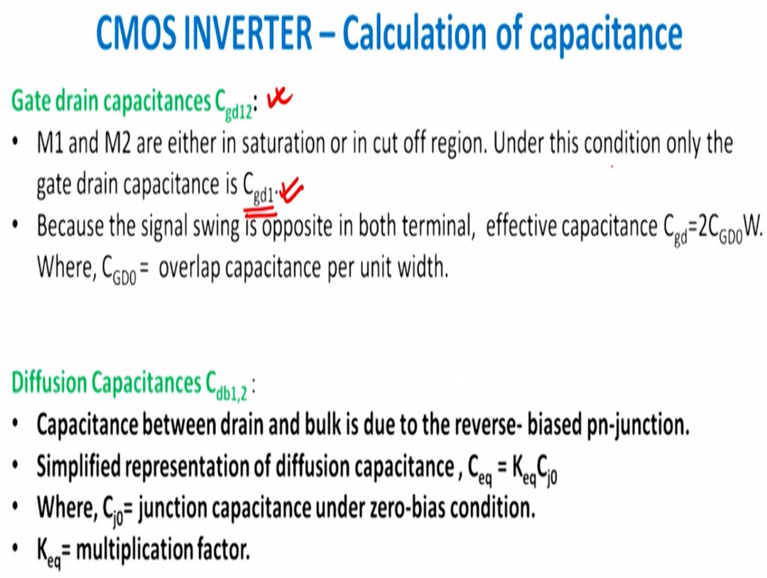
- Capacitance between drain and bulk is due to the reverse- biased pn-junction.
- Simplified representation of diffusion capacitance: $C_{eq} = K_{eq}C_{j0}$
- Where, C_{j0}= junction capacitance under zero-bias condition.
- K_{eq}= multiplication factor.

$K_{eq} \cdot C_{j0}$

Now what is the gate to drain? Gate to drain please see in the previous slide gate 2 drain is this one this is gate to drain so gate to drain we have got gate to drain here. So if you have gate to drain you will see that M1 and M2 the two transistors NMOS and PMOS will be either in saturation or in cut off. So under this condition only there will be Cgd1 right so there will be

Cgd1 only in this condition and there will be no other condition available to you as we discussed in our previous discussion that the gate to drain capacitance is Cgd1 right.

(Refer Slide Time: 12:48)



CMOS INVERTER – Calculation of capacitance

Gate drain capacitances C_{gd12}:

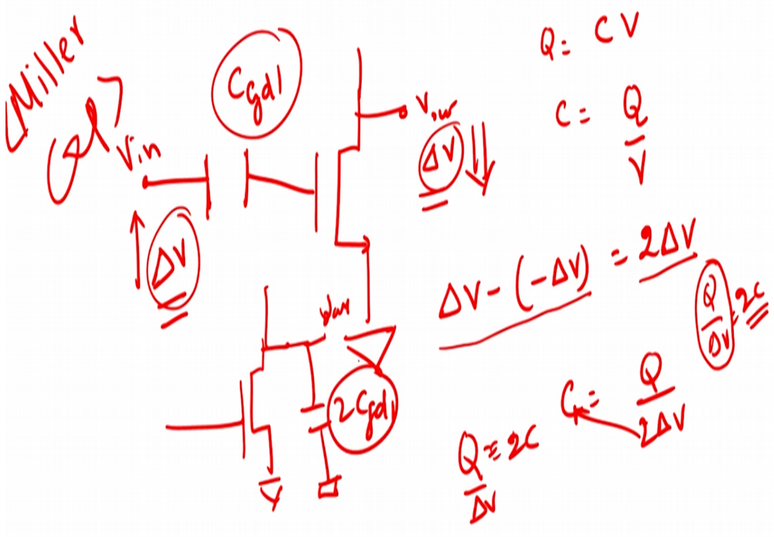
- M1 and M2 are either in saturation or in cut off region. Under this condition only the gate drain capacitance is C_{gd1}.
- Because the signal swing is opposite in both terminal, effective capacitance $C_{gd}=2C_{GD0}W$. Where, C_{GD0} = overlap capacitance per unit width.

Diffusion Capacitances $C_{db1,2}$:

- **Capacitance between drain and bulk is due to the reverse- biased pn-junction.**
- **Simplified representation of diffusion capacitance , $C_{eq} = K_{eq}C_{j0}$**
- **Where, C_{j0}= junction capacitance under zero-bias condition.**
- **K_{eq}= multiplication factor.**

So if we look at Cgd1 here I will explain to you what I am trying to say Cgd1 is gate to drain of 1 so this is drain gate to drain of 1 right this will appear when M1 and M2 are either in saturation or cut off because if there in cut off there will no channel form there and there will be therefore directly gate to drain will be come into picture and in saturation obviously you will have gate to drain available to you.

(Refer Slide Time: 13:14)



Let us see what happens in a general design issue so the design issue is let us suppose like this that I have got a may be NMOS here right let us suppose I have a Cgd1 and I have Vin here and I have V out coming out here so what I do let me see I have delta V change here the Cgd1 delta V change here then I will get a delta V down here. So if because this is a phase difference between gate and drain so if delta V increase there is a delta V drop here available to you because of this phase change.

Now what I do is if this means that this means that if you remember Q = CV or in other words if you mend to find out the capacitance C = Q / V which means that if you change on the input size by delta V you get a change delta. So what is the net change? Net change is delta V – of – delta V which is two times delta V with therefore C = Q / 2 times delta V right if you take this side and this side I get Q = therefore sorry Q / delta V = twice of C.

Which means that looking from the output side the capacitor was the Cgd1 from here has now will now become under this condition will now become what is known as you we will get 2Cgd1 this is what you will get this V out here. This is known as a Miller capacitances so how did you find a Miller capacitance? Miller capacitance again explain to you because this is quiet critical that as you increase the value of delta V here right your delta V where it will be increase in the output side because of phrase change which primarily means that for every therefore the change will be delta V – of – delta V which is + 2delta V C = Q / V

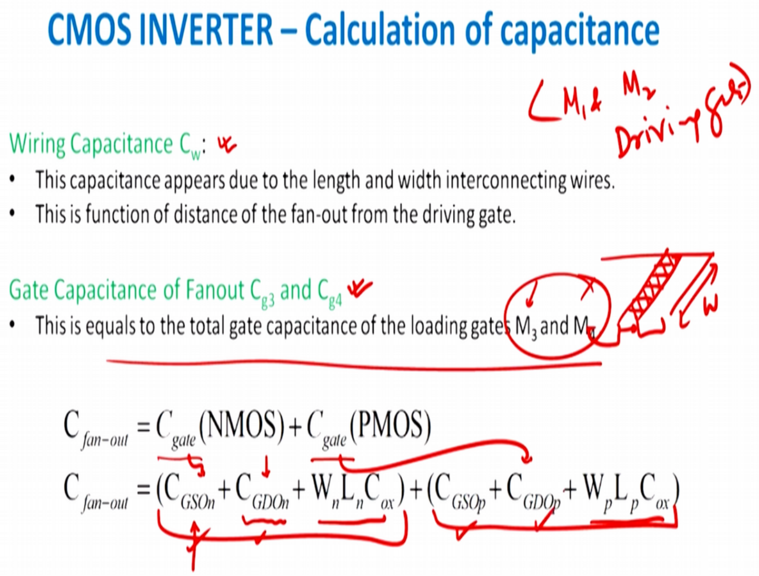
Now if you V changes by twice delta V so Q / 2 delta V take 2 on this side Q / delta V = twice C which means that this for the same change in the output side delta V is equivalent to say I have got capacitance 2C in the output side right 2 Cgd1 which means that input capacitance Cgd1 will appear to output side as to 2Cgd1 under such criteria right. So if you do not consider this under estimating the overall process parameter and that is the reason you see here that Cgd = 2Cgd 0 into w right why?

Because we have to multiply / W because this is pie unit width and therefore if you multiply by W I get the total Cgd and we multiply by 2 for the resource that we developed where Cgd0 is the overlap capacitance per unit width right which we discussed in our previous discussion there will be also remember drain to bulk because if you remember in our previous discussion this is bulk which is P type and this is N + type and this is P type there will be always a diffusion layer or there will be always a depletion layer here and here.

So there will be always a reverse bias Pn junction between the drain and the bulk and source and the bulk right. As a result there will be always a reverse bias pn junction this will result in the diffusion capacitance given by this quantity which is equals to K equivalent multiplied by Cj0 where Cj0 is the junction capacitance at 0 bias. So 0 bias means when you did not apply any bias just because of the doping condition differences there will be depletion region where a capacitance will be formed assuming the reverse bias more and more you end up having depletion thickness larger and larger.

So as a result Tox or a the thickness becomes larger capacitance as to fall down the capacitance start to decrease at a higher negative gate voltage.

(Refer Slide Time: 17:04)



CMOS INVERTER – Calculation of capacitance

(M1 & M2 Driving gate)

Wiring Capacitance C_w:

- This capacitance appears due to the length and width interconnecting wires.
- This is function of distance of the fan-out from the driving gate.

Gate Capacitance of Fanout C_{g3} and C_{g4}

- This is equals to the total gate capacitance of the loading gates M_3 and M_4

$$C_{fan-out} = C_{gate}(NMOS) + C_{gate}(PMOS)$$

$$C_{fan-out} = (C_{GSOn} + C_{GDOn} + W_n L_n C_{ox}) + (C_{GSOp} + C_{GDOp} + W_p L_p C_{ox})$$

Now K is equivalent is just an multiplication factor between 0 and 1 right now wiring capacitor as I discussed with you next is the wiring capacitance is primarily the capacitance which appears due to a finite width and length of the connected wires and it is a function of the distance which means that larger the length of the wire more will be cap loading capacitance of loading and lesser the length lesser will be the loading.

So higher is the cross sectional area lower will the loading so and hence so forth but primarily higher the length more will be loading available to you and therefore the capacitance loading will be much higher than this case as compared to the previous case. So that is the wiring capacitance which you see in front of you so we can do the last part which Cg3 and Cg4 and it is nothing but the total gate capacitance of the loading gate M3 and M4.

So we have driving gates M1 and M2 are referred to as driving gate since there driving the circuit and M3 and M4 at this is known as loading gates. So the loading gates M3 and M4 I have got an M1 and M2 referred to as driving gates right. So I have got driving gates and I have got loading gates right so C fan out is nothing but C gate of NMOS + C gate of PMOS because they are parallel right.

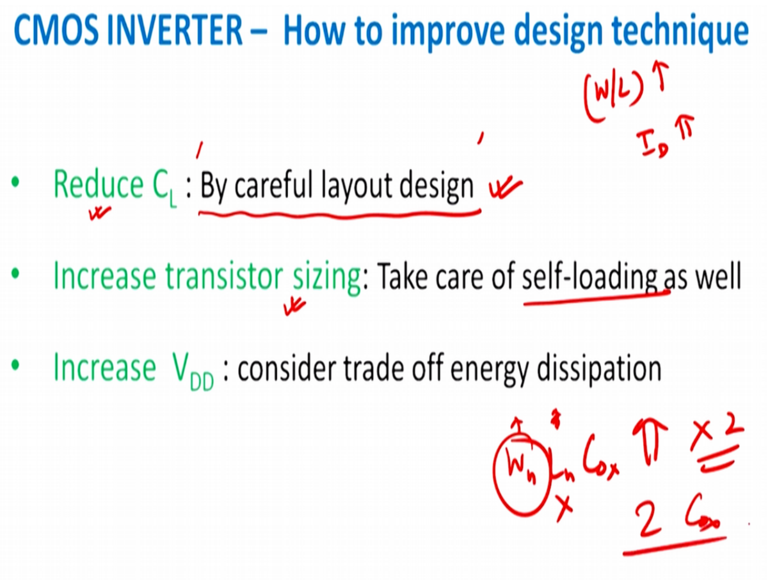
Now if you look very closely this is been broken up into this part and this is been broken up in this part right and let it see how its work out. If you remember Cox was oxide capacitance per unit area the area under the gate is nothing but W into L. L is the length of the device so if you take a cross section of the device this is your W and L was nothing but between these two point the distance here.

I will have a rectangular region and therefore W into L is nothing area under the gate and therefore this area is nothing but WL into LN into Cox. Same thing happens WP in LP into Cox CGDO means overlap capacitance gate to drain remember as you were assuming that there is not lateral diffusion so there will be self-aligned and therefore there will be no overlap but is this overlap you will have a CGD0n and you will have CSGSOn also available with you right.

Gate to source overlap gate to drain overlap for N type and you will have gate to source overlap for P type gate to drain to P type add all these together and you get the total inverter characteristics. So we will give you a numerical problem when we give an assignments to play with and then actually you can use this capacitive modeling to find out the total capacitance available to you when you had the cascade PR of the inverters available right.

Please make it a point of my understand that when you have a single device driving a simple capacitor and you have the cascade of inverters the formulation goes a bit different right and we will be discussing that as we move along that is the formulation will be different so the first question asked is how will you improve your speed very well very easy reduce your value of CL.

(Refer Slide Time: 20:18)



CMOS INVERTER – How to improve design technique

- Reduce C_L : By careful layout design
- Increase transistor sizing: Take care of self-loading as well
- Increase V_{DD} : consider trade off energy dissipation

(W/L) ↑
I_D ⇑
W_n L_n C_ox ⇑ ×2
2 C_ox

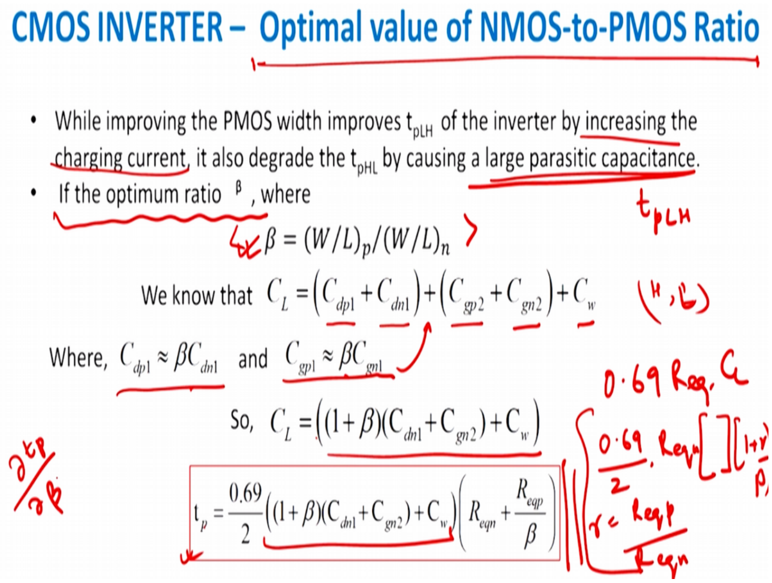
So what you do is reduce your CL right we will discuss this point do not worry about where it is coming from but reduce CL but somehow or other if you are able to reduce the total CL value by either making a overlap capacitance small or making your wiring length small or you can making the oxide thickness of the loading gates to be large Tox is large. So F silent ox by Tox for loading gate will be relatively low so and hence so forth I can take my I can reduce my CL and I can increase the frequency operation.

You can increase the transistor size why do you why do you want to increase the transistor size if you increase by for example W / L if you increase your current ID actually starts to becomes large. Where ID increases might for the same capacitance loading I will have require asset time to charge or discharge and therefore the speed will be larger but mind you as you increase the value by W / L you also increase the chances or in reality you also increase what is known as self-loading.

M3 you will see increase W / L we will also increase all W / L so when M3. M4 W/L increases then please understand WN into LN into Cox will also increase in quadratic fashion because you are increasing W as well L let us suppose or even fixing the value of L even if you double the value of WN you actually double the value of the overall oxide capacitance as a result your delay will be further reduced.

So you have to be very careful and you have to optimize the whole problem in order to achieve the best flow. The third point is increase VDD which is obviously to increase value of ID your performance in terms delay will be larger but then you pay the cost of higher power dissipation. Because VDD increasing means you are V into I into V is power. So the power will go on increasing drastically right so this is one of the areas which we look into.

(Refer Slide Time: 22:23)



CMOS INVERTER – Optimal value of NMOS-to-PMOS Ratio

- While improving the PMOS width improves t_{pLH} of the inverter by increasing the charging current, it also degrade the t_{pHL} by causing a large parasitic capacitance.
- If the optimum ratio β , where

$$\beta = (W/L)_p/(W/L)_n$$

We know that $C_L = \left(C_{dp1} + C_{dn1}\right) + \left(C_{gp2} + C_{gn2}\right) + C_w$

Where, $C_{dp1} \approx \beta C_{dn1}$ and $C_{gp1} \approx \beta C_{gn1}$

So, $C_L = \left((1+\beta)(C_{dn1} + C_{gn2}) + C_w\right)$

$$\boxed{t_p = \frac{0.69}{2}\left((1+\beta)(C_{dn1} + C_{gn2}) + C_w\right)\left(R_{eqn} + \frac{R_{eqp}}{\beta}\right)}$$

t_{pLH}

(H, L)

$0.69\, R_{eq}.\, C_L$

$\frac{0.69}{2}.\, R_{eqn}[\;][1+\frac{r}{\beta}$

$r = \frac{R_{eqp}}{R_{eqn}}$

$\frac{\partial t_p}{\partial \beta}$

Now let me come to the third for fourth topic the sub topic of this whole discussing and that is the basically the optical value of NMOS to PMOS ratio right optimal value of NMOS to PMOS ratio what do I mean by that? So you see if you improve the widths or if you increase the widths of PMOS W / L ratio of PMOS my tpLH low to high will obviously come down because it is so if your W / L ratios are higher right your resistance will be smaller and your therefore delays tpLH will be smaller by a previous discussion.

Also as I discussed with you any increase in the charging current so in both the cases the tpLH will be smaller. But it also degrades the tpHL right because of the large parasitic capacitance tpLH is what high to low right I agree with you high to low the PMOS will be cutoff right. Cutoff means what there is no channel available to you but you will still have gate to drain capacitance always available are you getting my point in our previous discussion.

So therefore if you go on increasing the value of W / L you end up also loading the pull down capacitance. So pull down capacitance will become higher so it becomes higher even on the loading capacitance right. So let us suppose I optimize it by a factor known as beta we assume beta to be equals to W / L of P divided by W / L of n right. Now you known as CL = diffusion capacitance of P1 + diffusion capacitance of N1 + gate capacitance of P2 and gate capacitance of N2 + the wiring capacitance which we have seen just now.

Now what we do is obviously CDp1 will be beta times Dn1 why it will be Dn1 because you have increase the size of PMOS beta times as compared to NMOS and therefore a capacitance will also be become double that of NMOS because it is totally depending on the profile parameter or on the on the device characteristics and Cgp1 will be also equal to bet times Cgnl right as a result if you up to out this and this back into this formula I get Cl = 1 + beta dn1+ gn cgn 2 + Cw right.

If you place this into 0.69R equivalent into CL formula I get this formula available to me fine if you take R equivalent outside I can safely write down to this to be 0.69 / 2 into R = right equivalent and then this whole thing right + 1 + R / beta well R = R equivalent P / R = n right R = P divided by R equivalent N you get in this case and this is 1 + R / beta right.

This is what you get finally for tp and this is the most optimized design which is available to you and for NMOS to PMOS ratio. The fundamental behavior therefore is that if now I have a tp relationship which you see in front of you and i have R values available to me here simply by deriving del tp del beta I can let you know the value of beta at which the tps will be minimum basic mathematics so I take tp del tp = 0 after doing some manipulation mathematical manipulation I get this to be true assuming that my wire capacitance Cw is very small capacitance as compared to CDn and CGn2 which is really the case.

There I get beta optimal optimum is actually equals to root of r now you see that is quiet interesting what I wanted to point out that the optimal ratio here what is R? R was basically equals to R equivalent if you go back to the previous slide R was equal to R equivalent P / R equivalent N right. So R equivalent P / R equivalent n right so root of r which means that if you make your R equivalent sorry R larger and larger your beta optimal will also increase what is R the ratio of the tool but ratio of the 2 will depend upon what factor aspect ratio so you have to make your W of p larger than the W of n agreed.

But please understand one important issue but when we were discussing of simple invertor and we were trying to fix our switching threshold in the middle of the whole VDD / 2 and assuming the threshold voltage are equal we say that we require only R beta was = R only remember it was only R because R was equals to 1 and we got VDD / 2 remember and that was the case when both pull up capability and pull down capability was exactly equal by making the widths of PMOS 2 to 3 times larger as compared to NMOS.

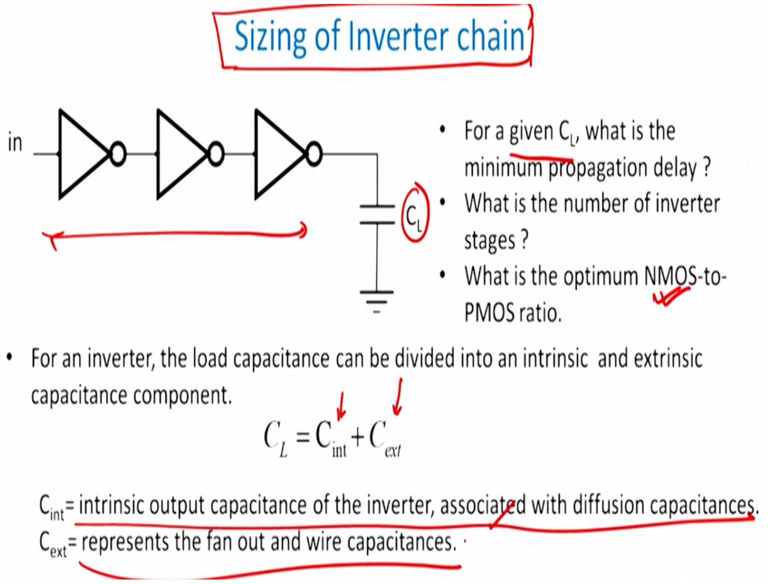
And therefore the pull up capability was almost equals to the pull down capability available to you and as a result it was perfectly symmetrical in natural and therefore beta = 1. Now what we were doing is we are telling no I do not want tpHL = tpLH that I want that my total delayed over all delay should be minimized for that I required this to be true bur root of R to be true. So root of R basically means if it is three times then the root of 3 will be available in this case right 1.7, 1.8 times right.

So if you look at the graph in front of you beta is basically as I discussed with you the Wp / Wn how because this is Wp / LLp, Wn / Ln, Ln Lp cancels out because they are equals so I get Wp / Ln equals to beta if you look at this graph this is propagation delay what is beta here right. So as the beta value goes on increasing right beta value goes on increasing become becoming what your PMOS is more and more stronger.

So what is happening is your low to high is falling down as I just told you but and it falls down like this right but as the beta value increases it high to low also as to fall down as you make the beta higher tpLH starts to falls down because PMOS is getting stronger but tpHL starts to rise up because your NMOS is getting weaker respect to NMOS so your pull down device is weaker now.

So I reach at optimum point somewhere here so this black line is nothing but summation as such of these two I reach out the point which is somewhere here right. So 2 point between 2.5 is a value of beta at which you get the minimum value of tp available to your right. So this is the optimal inverter equation which you see in front of you.

(Refer Slide Time: 29:08)



Sizing of Inverter chain

- For a given C_L, what is the minimum propagation delay ?
- What is the number of inverter stages ?
- What is the optimum NMOS-to-PMOS ratio.

- For an inverter, the load capacitance can be divided into an intrinsic and extrinsic capacitance component.

$$C_L = C_{int} + C_{ext}$$

C_{int}= intrinsic output capacitance of the inverter, associated with diffusion capacitances.
C_{ext}= represents the fan out and wire capacitances.

Now let us look how to size an inverter change a large number inverter are connected in series to each other. So how do I size my inverter so that I get a minimum delay right so for a what value for a given value of CL what is the minimum propagation delay between input and output and what is the number of stages at which I will get the minimum delay available to me and what is the optimum NMOS to PMOS ratio’s which is there.

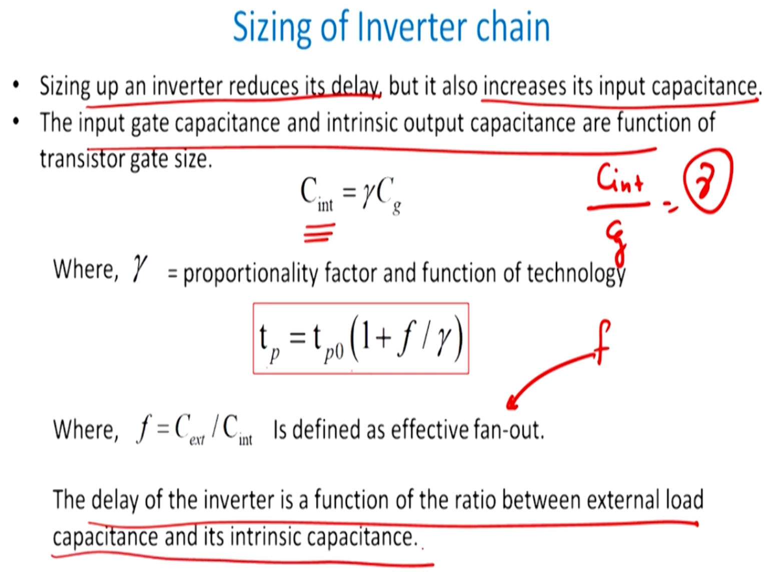
These three things we need we will be finding out over the next 5 to 10 minutes of work here now any load capacitance which you see here will be composed of intrinsic capacitances and extrinsic external capacitances. Intrinsic capacitances are what? Intrinsic capacitance is R which is self-loading capacitances because of the drain and diffusion and external capacitance is R by virtual wire by external sources by everything else you define right.

So I define intrinsic as output capacitance of inverter associated with diffusion capacitance and what is C external fan out and wire capacitances. So CG3 and CG4 + CW is external capacitance and CDB1 CGD1 and CGD2 are all diffusion capacitance right they are all intrinsic capacitances. So the propagation delay with tp will be = 0.69 into R equivalent into C internal + C external I think this is clear. So in that C I C internal I take it outside I get C external / C internal.

The delay of the each inverter will itself delayed by a factor equals to 0.969 equivalent by the thing so what I do is that replace this by tp0. So tp0 is what the delay of the inverter so what I am trying to see is that the chain of inverters right the chain of inverters if you look then this is CL this is in this CL is composed of C external in C internal and C external of C intrinsic of C external there will be a composition of C internal intrinsic. So intrinsic is what because of gate and diffusion because of drain and bulk there will be internal capacitances.

So the importance itself delayed by virtue of this without being even effected by the next stage gate capacitances. So as a result you get this as tp0 right 1 + C external / C internal.

(Refer Slide Time: 31:22)



Sizing of Inverter chain

- Sizing up an inverter reduces its delay, but it also increases its input capacitance.
- The input gate capacitance and intrinsic output capacitance are function of transistor gate size.

$$C_{int} = \gamma C_g$$

$\frac{C_{int}}{g} = \gamma$

Where, γ = proportionality factor and function of technology

$$\boxed{t_p = t_{p0}\left(1 + f/\gamma\right)}$$

f

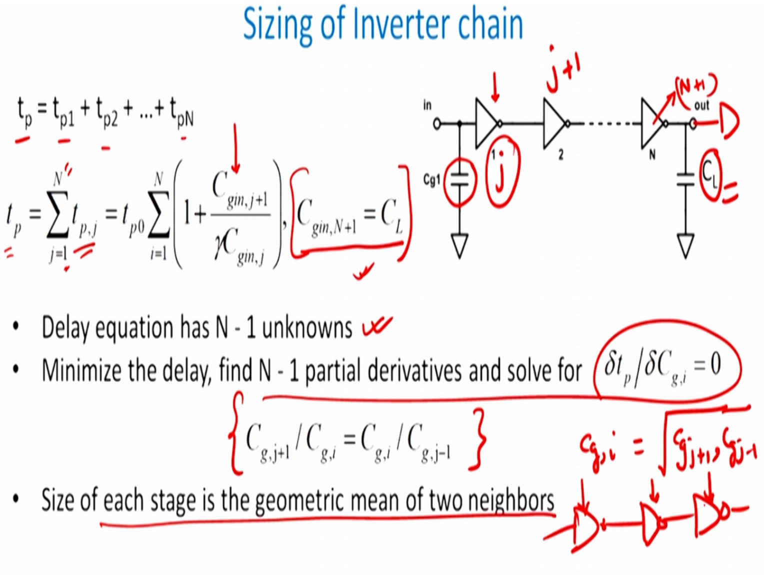
Where, $f = C_{ext}/C_{int}$ Is defined as effective fan-out.

The delay of the inverter is a function of the ratio between external load capacitance and its intrinsic capacitance.

Sizing of the inverter as I discussed with you will reduce the delay agreed because you are actually making it much move faster because current capability is larger but the cost you are making for it is that it also increase the input capacitances. My input the input the gate capacitance is suppose I internal input capacitance and I have CG to be equals to gamma right this is basically a fraction and input gate capacitance is intrinsic gate capacitances are function of gate and we define that intrinsic gate capacitance is gamma times the gate capacitances.

Because there will be some proportionality factor which is there so what I do I just simply make it tp = Tp0 1 + f / gamma because gamma is c internal / CG and F = C external / C internal and therefore I get tp = tp01+f right. F is defined as the effective fan out right it is define as the effective fan out now therefore if you look at the delay therefore the delay for the inverter is the function between the ratio of the external load to its intrinsic capacitance which means that the delay of an inverter depends on the ratio of the external load capacitance to its internal capacitance that you can even find out at this stage also you see tp is depending on the value of C external to C 2 internal up to intrinsic and therefore I get ratio between external load and capacitance right.

(Refer Slide Time: 32:51)



Sizing of Inverter chain

$t_p = t_{p1} + t_{p2} + \ldots + t_{pN}$

$$t_p = \sum_{j=1}^{N} t_{p,j} = t_{p0}\sum_{i=1}^{N}\left(1+\frac{C_{gin,j+1}}{\gamma C_{gin,j}}\right), C_{gin,N+1} = C_L$$

- Delay equation has N - 1 unknowns
- Minimize the delay, find N - 1 partial derivatives and solve for $\delta t_p/\delta C_{g,i} = 0$

$$\{C_{g,j+1}/C_{g,i} = C_{g,i}/C_{g,j-1}\}$$

- Size of each stage is the geometric mean of two neighbors

$C_{g,i} = \sqrt{g_{j+1}, g_{j-1}}$

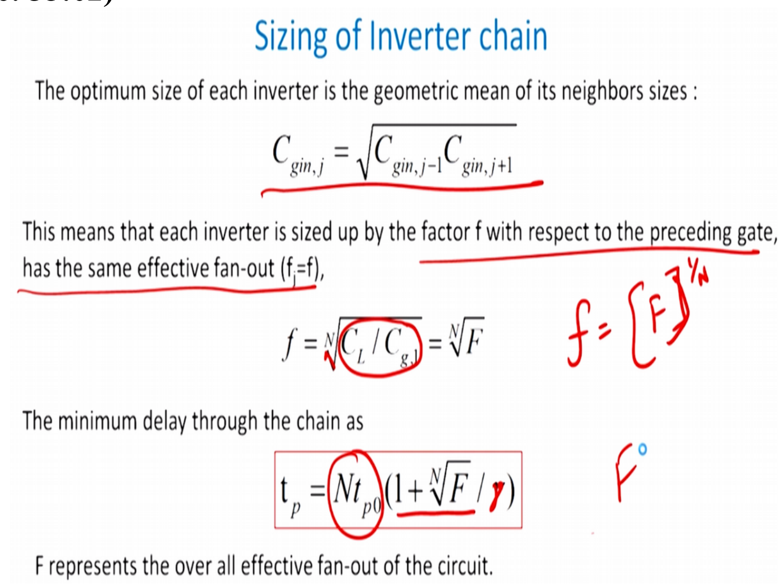
Now let us look at the sizing of the therefore sizing of the inverter chain so what I do I get tp = tp1 + tp2 + tp3 there are n number of gates available my driving gate as got input gate of Cg1 and last gate which is the final gate as the load capacitance of Cl. So going by the definition I get tp will be nothing but the some of tpj J moving from 1 to n so there are n gates each gate is giving tp1, tp2, tp3, tp4 you add of them you get the total tp.

Therefore breaking down is a tp0 is the propagation delay for the single gate which is that is constant independent of everything is depends only on the intrinsic gate it is depend upon the extrinsic gate therefore we can take it outside the summation side and therefore I get 1 + Cgin gate input gate capacitance of J = 1th 1 by j = 1th 1 so you can measure it for 1th 1 if this is j this will be j+1 right gate capacitance of this 1 divided by gamma times CJ n of what of this one.

So this right and I say that Cgin n+1 = Cn why? Because this is last is this one Nn+1 is the last stage which you see the output gate the input gate capacitance so if there would have been inverter here it is input capacitance would have been = Cl that is what I trying to say here. So there so there are N-1 equations there will be therefore n-1 partial derivatives to solve for this quantity.

So I get delta tp delta Cgi = 0 and then I would meet to minimize it if you solve for V I get this as the relationship right and this is quite an interesting relationship which you see. You can see therefore that if you try to find out I get Cgi right is nothing but square root of Cgj+1 comma Cgj - 1 which means that is size of each stage is the geometric mean of it is two neighbors that is very important that mean if I have one invertor here another inverter here I need to be very careful that this sizing here should be at least is the geometric mean of the this and this these two inverters right,

(Refer Slide Time: 35:02)



Sizing of Inverter chain

The optimum size of each inverter is the geometric mean of its neighbors sizes :

$$\underline{C_{gin,j} = \sqrt{C_{gin,j-1}C_{gin,j+1}}}$$

This means that each inverter is sized up by the factor f with respect to the preceding gate, has the same effective fan-out (f_j=f),

$$f = \sqrt[N]{C_L / C_{g,1}} = \sqrt[N]{F} \qquad f = [F]^{1/N}$$

The minimum delay through the chain as

$$\boxed{t_p = Nt_{p0}(1+\underline{\sqrt[N]{F}/\gamma})} \qquad F^{\circ}$$

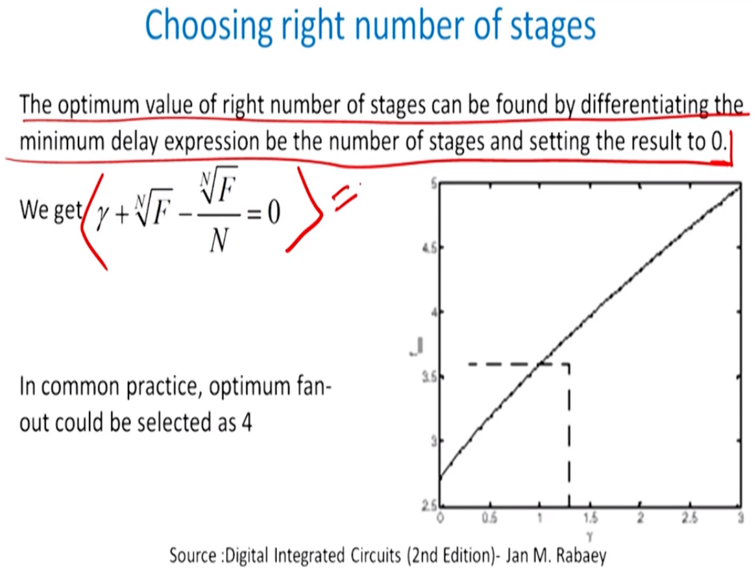
F represents the over all effective fan-out of the circuit.

You have to be very careful of that and that is what written as Cgin = Cg in j -1 into Cgin J+1 right. That means each inverter is sized by a factor of F with respect to its preceding gate right as a same effective fan out which means that if you therefore increase this sizing of individual inverter in the chain by a factor of F where I have already defined in the previous slide then you will able to minimize your delay under all conditions.

So we therefore define F to B = Cl / Cg1 square root of N of Nth product of that this is the capital F which is the effective fan out and square root of F so I get F = f = to the power 1 / N which we get from here. Therefore the minimum delay is what Ntpo why Ntpo because there are n number of stages so I multiple N with tpo delay of 1 that multiplied by 1 + root of NF / gamma right gamma is coming from previous slide right.

Where F represents the overall effective fan out the circuit capital F represent the overall effective fan out.

(Refer Slide Time: 36:11)



Choosing right number of stages

The optimum value of right number of stages can be found by differentiating the minimum delay expression be the number of stages and setting the result to 0.

We get $\gamma + \sqrt[N]{F} - \frac{\sqrt[N]{F}}{N} = 0$

In common practice, optimum fan-out could be selected as 4

Source :Digital Integrated Circuits (2nd Edition)- Jan M. Rabaey

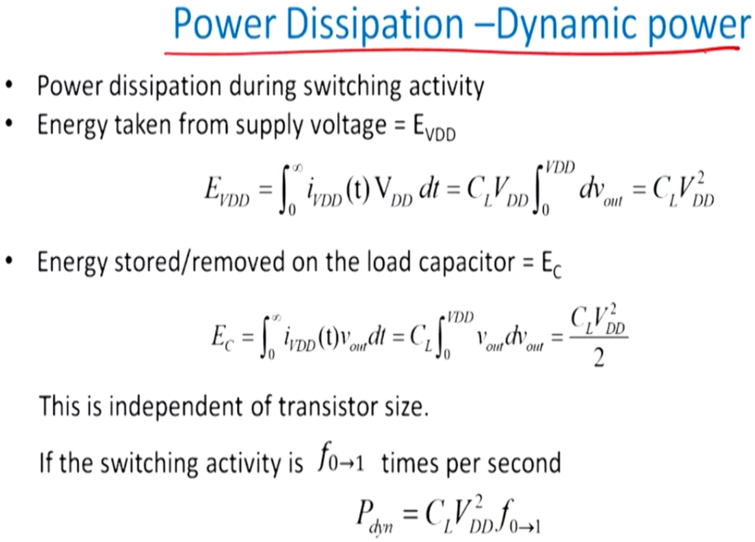
Now if you see carefully here carefully in this slide then as you increase value of N number of stages your Ntpo increases right Ntpo increases but at the same instant of time f to the power to the power say 1 / 2 it becomes 1 / 3, 1 /4 that starts to decreases right so you will have two competing effects at a time will come when they will cancel out each other and the tp will have minimum value.

The optimum therefore value of right number of stages can be found by differentiating the minimum delay expression to the number of stages and the result settling to 0 that is what I was telling you but del tp del N set to 0 and then C which was minimum right. So from that we get this expression that gamma + N of root F – nth root of F / N= 0. The common factors is that is that generally the optimum fan out selected is generally it goes to 4.

So this expression which you see in front of you will give you an idea about number of expression which you want or number of stages you want for delay chain to have a minimum value of gate voltage available to you. Before you move forward therefore let me recapitulate what we did till now we say what is the meaning of inverter delay? How does inverter delay depend upon the process parameters which is W / L if I have a single inverter driving a load and I have cascaded inverter driving a load how does the things change? What is the meaning of beta ratio? What is the primary meaning of optimized value of beta ratio right.

Then we came to the stage where we have inverter chain in this inverter chain we also found out that if I have inverter chain what is the minimum number of this inverters and I need to put in series so that I get the minimum delay we found out that N to be equals to 4. We also found out that the effective fan out should be minimized in order to achieve a minimum delay. So this was part which we have discussed there.

(Refer Slide Time: 38:21)



Power Dissipation –Dynamic power

- Power dissipation during switching activity
- Energy taken from supply voltage = E_{VDD}

$$E_{VDD} = \int_0^{\infty} i_{VDD}(\mathrm{t})\,\mathrm{V}_{DD}\,dt = C_L V_{DD}\int_0^{VDD} dv_{out} = C_L V_{DD}^2$$

- Energy stored/removed on the load capacitor = E_C

$$E_C = \int_0^{\infty} i_{VDD}(\mathrm{t}) v_{out} dt = C_L \int_0^{VDD} v_{out} dv_{out} = \frac{C_L V_{DD}^2}{2}$$

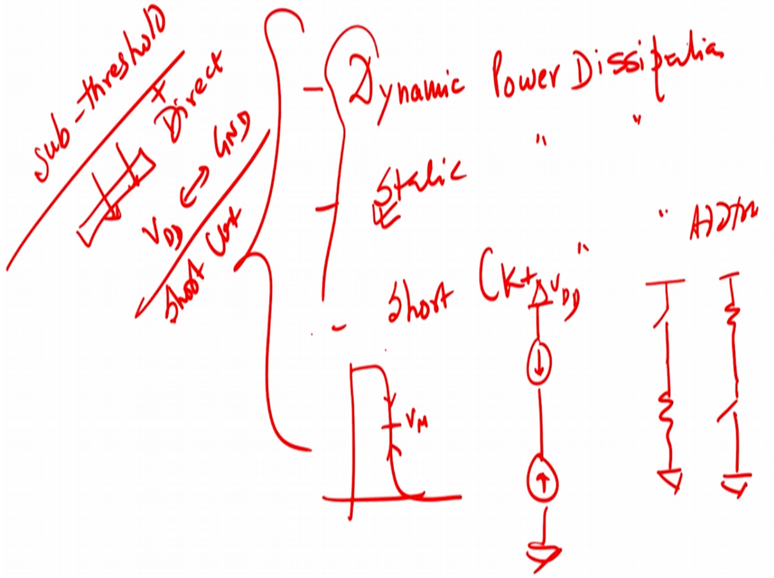
This is independent of transistor size.

If the switching activity is $f_{0\to 1}$ times per second

$$P_{dyn} = C_L V_{DD}^2 f_{0\to 1}$$

Now we come to the next section of our talk and that is basically your power dissipation right. In power dissipation we have three types of power which is generally available which is in a standard CMOS digital VLSI circuit.

(Refer Slide Time: 38:28)



We have what is known as a dynamic power dissipation right and we have a static power dissipation and we have a short circuit power dissipation. So that three types of power dissipation available to us dynamic, static and short circuit. Dynamic is the name suggest is when the name actually suggest the switching between 0 and 1 and so on and hence so forth that we define as the dynamic power dissipation and the inverter is switching between 1 and 0 right.

There active role of the load capacitor and the resistance offered by this NMOS and PMOS will come into picture we look at the short we look at the static one when the device is either in the off state then you will still have some leakage currents available to you right these are basically known as sub threshold leakage or you can also have some leakage to the oxide layer because of high electric field this is known as direct leakage right and you can have these if you add those 2 together they form about 5 o 10 % of the total power dissipation of the element to you.

We also have a short circuit dissipation please understand a very important issue remember there was a time when you did have something like this right you had a time something like this but there was also case when both for saturated was something like this and this was VDD which means during the switching threshold exactly at switching threshold both the devices NMOS and PMOS are in saturation and therefore for a very finite duration of time your VDD and ground are already in contact so there will be a short circuit short circuit available here fine.

So there will be short circuit which is available here and there will be short circuit current which is flowing from the device. Which means that which primarily means therefore is that and why is it very short I have discussed with you in previous short term when we were discussing VTC that if you look at this point this it the point I am talking about VM right. This is this moves so fast this movement is so fast across VM but this taste for a small duration of time but in that small duration VDD is connected to ground and as soon as what will happen is there is sudden peak in the current from VDD to ground.

And every time the switching takes place you will have always the short circuit current available to you. So if there are 50 switching's there will 50 short circuit currents which will be coming to you and as a result that will also add to the total power in the circuitry right with this knowledge with this idea I would like to stop here today and give you what we have discussed today is basically the inverter characteristics we have understood the static characteristics of inverter how to draw the VTC voltage transfer characteristics of an inverter.

How to extract noise margins from inverter we have also understood the basic building block of an inverter what is the propagation delay of inverter what is tpHL and tpLH how they are related to VDD and R equivalent and Cl how can I minimize the delay across an inverter chain what should be an sizing ratio available to you so that these can be modified to an proper extent with this I will just stop here and we will stop will stop next time with next differential of power dissipation thank you very much.