UNIT 4

BASIC CIRCUIT DESIGN CONCEPTS

INTRODUCTION

We have already seen that MOS structures are formed by the super imposition of a number conducting ,insulating and transistor forming material. Now each of these layers have their own characteristics like capacitance and resistances. These fundamental components are required to estimate the performance of the system. These layers also have inductance characteristics that are important for I/O behaviour but are usually neglected for on chip devices.

The issues of prominence are

1.Resistance, capacitance and inductance calculations.

2.Delay estimations

3.Determination of conductor size for power and clock distribution

4.Power consumption

5.Charge sharing

6.Design margin

7.Reliabiltiy

8.Effects and extent of scaling

RESISTANCE ESTIMATION

The concept of sheet resistance is being used to know the resistive behavior of the layers that go into formation of the MOS device. Let us consider a uniform slab of conducting material of the following characteristics .

Resistivity- ρ

Width - W

Thickness - t

Length between faces – L as shown next

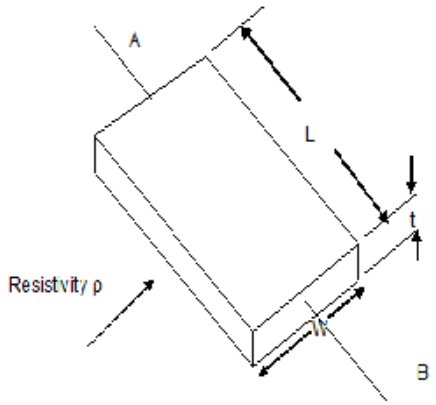

Figure 24:A slab of semiconductor

We know that the resistance is given by RAB= ρL/A Ω. The area of the slab considered above is given by A=Wt. There fore RAB= ρL/Wt Ω. If the slab is considered as a square then L=W. therefore RAB=ρ/t which is called as sheet resistance represented by Rs.The unit of sheet resistance is **ohm per square.** It is to be noted that Rs is independent of the area of the slab. Hence we can conclude that a 1um per side square has the same resistance as that of 1cm per side square of the same material.

The resistance of the different materials that go into making of the MOS device depend on the resistivity and the thickness of the material. For a diffusion layer the depth defines the thickness and the impurity defines the resistivity. The table of values for a 5u technology is listed below.5u technology means minimum line width is 5u and ƛ= 2.5u.The diffusion mentioned in the table is n diffusion, p diffusion values are 2.5 times of that of n. The table of standard sheet resistance value follows.

Layer	Rs per square
Metal	0.03
Diffusion n(for 2.5 times the n)	10 to 50
Silicide	2 to 4
Polysilicon	15 to 100
N transistor gate	10^4
P transistor gate	$2.5x\ 10^4$

SHEET RESISTANCE OF MOS TRANSISTORS

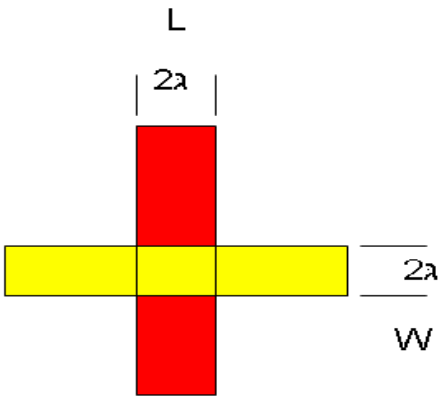

Figure 25 Min sized inverter

The N transistor above is formed by a 2λ wide poly and n diffusion. The L/W ratio is 1. Hence the transistor is a square, therefore the resistance R is 1sqxRs ohm/sq i.e. R=1x104. If L/W ratio is 4 then R = $4x10^4$. If it is a P transistor then for L/W =1,the value of R is $2.5x10^4$.

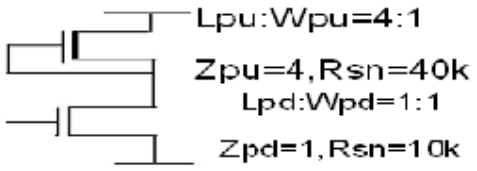

Figure 26 Nmos depletion inverter

Pull up to pull down ratio = 4.In this case when the nmos is on, both the devices are on simultaneously, Hence there is an on resistance Ron = 40+10 =50k. It is this resistance that leads the static power consumption which is the disadvantage of nmos depletion mode devices

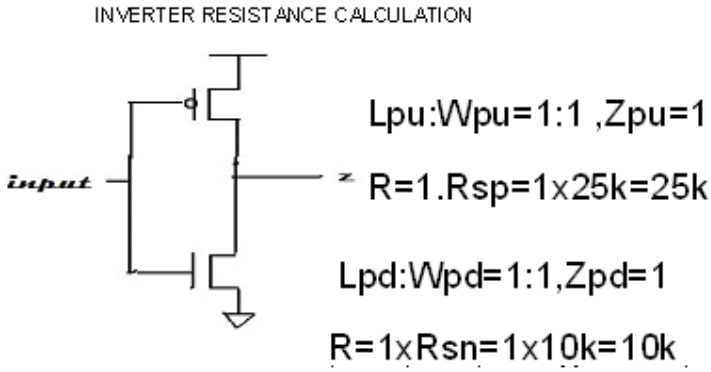

Figure 27: Cmos inverter

Since both the devices are not on simultaneously there is no static power dissipation

The resistance of non rectangular shapes is a little tedious to estimate. Hence it is easier to convert the irregular shape into regular rectangular or square blocks and then estimate the resistance. For example

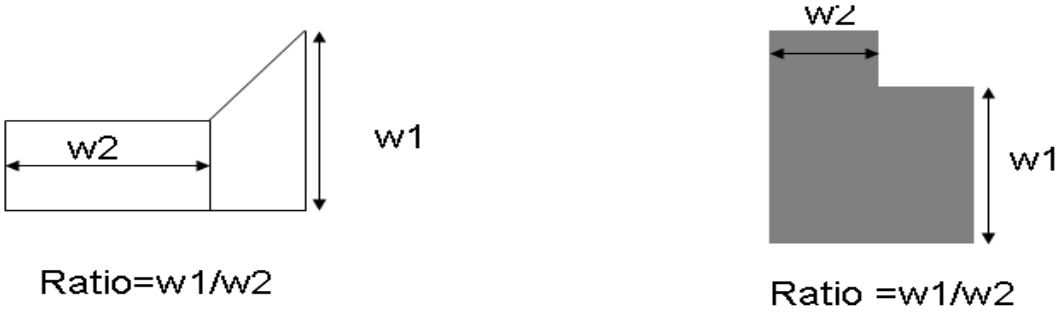

Figure 28:Irregular rectangular shapes

CONTACT AND VIA RESISTANCE

The contacts and the vias also have resistances that depend on the contacted materials and the area of contact. As the contact sizes are reduced for scaling ,the associated resistance increases. The resistances are reduced by making ohmic contacts which are also called loss less contacts. Currently the values of resistances vary from .25ohms to a few tens of ohms.

SILICIDES

The connecting lines that run from one circuit to the other have to be optimized. For this reason the width is reduced considerably. With the reduction is width the sheet resistance increases, increasing the RC delay component. With poly silicon the sheet resistance values vary from 15 to 100 ohm. This actually effects the extent of scaling down process. Polysilicon is being replaced with silicide. Silicide is obtained by depositing metal on polysilicon and then sintering it. Silicides give a sheet resistance of 2 to 4 ohm. The reduced sheet resistance makes silicides a very attractive replacement for poly silicon. But the extra processing steps is an offset to the advantage.

A Problem

A particular layer of MOS circuit has a resistivity ρ of 1 ohm –cm. The section is 55um long,5um wide and 1 um thick. Calculate the resistance and also find Rs

R= RsxL/W, Rs= ρ/t

Rs=1x10-2/1x10-6=10^4ohm

R= 104x55x10-6/5x106=110k

CAPACITANCE ESTIMATION

Parasitics capacitances are associated with the MOS device due to different layers that go into its formation. Interconnection capacitance can also be formed by the metal, diffusion and polysilicon (these are often called as runners) in addition with the transistor and conductor resistance. All these capacitances actually define the switching speed of the MOS device.

Understanding the source of parasitics and their variation becomes a very essential part of the design specially when system performance is measured in terms of the speed. The various capacitances that are associated with the CMOS device are

1.Gate capacitance - due to other inputs connected to output of the device

2.Diffusion capacitance - Drain regions connected to the output

3.Routing capacitance- due to connections between output and other inputs

The fabrication process illustrates that the conducting layers are apparently seperated from the substrate and other layers by the insulating layer leading to the formation of parallel capacitors. Since the silicon dioxide is the insulator knowing its thickness we can calculate the capacitance

$$C = \frac{\varepsilon_o \varepsilon_{ins} A}{D} \quad \text{farad}$$

Ɛεo= permittivity of free space-8.854x1014f/cm

εins= relative permitivity of sio2=4.0

D= thickness of the dioxide in cm

A = area of the plate in cm2

The gate to channel capacitance formed due to the sio2 separation is the most profound of the mentioned three types. It is directly connected to the input and the output. The other capacitance like the metal, poly can be evaluated against the substrate. The gate capacitance is therefore standardized so as to enable to move from one technology to the other conveniently.

The standard unit is denoted by □Cg. It represents the capacitance between gate to channel with W=L=min feature size. Here is a figure showing the different capacitances that add up to give the total gate capacitance

Cgd, Cgs = gate to channel capacitance lumped at the source and drain

Csb, Cdb = source and drain diffusion capacitance to substrate

Cgb = gate to bulk capacitance

Total gate capacitance Cg = Cgd+Cgs+Cgb

Since the standard gate capacitance has been defined, the other capacitances like polysilicon, metal, diffusion can be expressed in terms of the same standard units so that the total capacitance can be obtained by simply adding all the values. In order to express in standard values the following steps must be followed

1. Calculate the areas of area under consideration relative to that of standard gate i.e.4ƛ^2. (standard gate varies according to the technology)

2. Multiply the obtained area by relative capacitance values

   tabulated .

3. This gives the value of the capacitance in the standard unit of

   capacitance □Cg.

Table 1:Relative value of Cg

layer	Relative value for5u technology
Gate to channel	1
Diffusion	0.25
Poly to sub	0.1
M1 to sub	0.075
M2 to sub	0.05
M2 to M1	0.1
M2 to poly	0.075

For a 5u technology the area of the minimum sized transistor is 5uX5u=25um2 ie λ=2.5u, hence,area of minimum sized transistor in lambda is 2 λX 2 λ = 4 λ2.Therefore for 2u or 1.2u or any other technology the area of a minimum sized transistor in lambda is 4 λ2. Lets solve a few problems to get to know the things better.

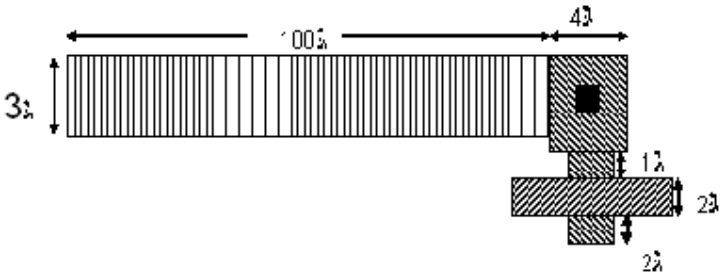

Figure 29 :Multilayered structure

The figure above shows the dimensions and the interaction of different layers, for evaluating the total capacitance resulting so.

Three capacitance to be evaluated metal Cm,polysilicon Cp and gate capacitance Cg

Area of metal = 100x3=300λ2

Relative area = 300/4=75

Cm=75Xrelative cap=75X0.075=5.625□Cg

Polysilicon capacitance Cp

Area of poly=(4x4+1x2+2X2)=22λ2

Relative area = 22λ2/4λ 2=5.5

Cp=5.5Xrelative cap=5.5x.1=0.55 □Cg

Gate capacitance Cg= 1□Cg because it is a min size gate

Ct=Cm+Cp+Cg=5.625+0.55+1=7.2 □Cg

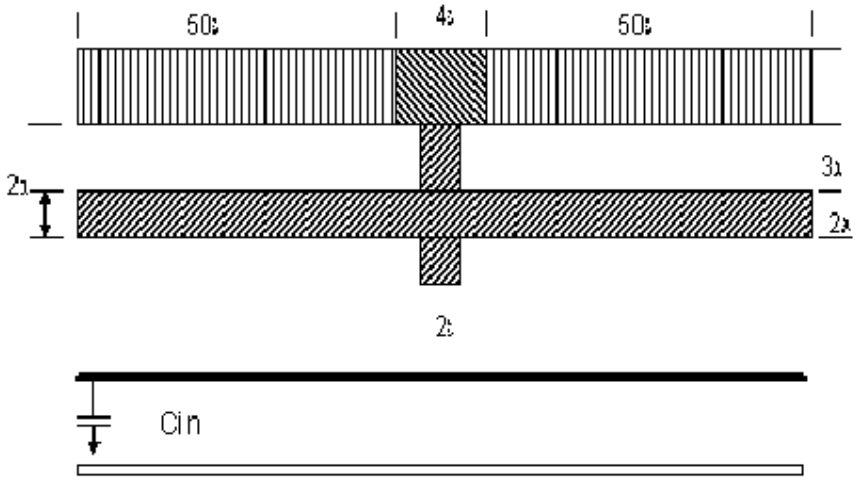

Figure 29:Mos structure

The input capacitance is made of three components metal capacitance Cm, poly capacitance Cp, gate capacitance Cg i.e Cin= Cm+Cg+Cp

Relative area of metal =(50x3)X2/4=300/4=75

Cm=75x0.075=5.625□Cg

Relative area of poly = (4x4+2x1+2x2)/4 =22/4 =5.5

Cp=5.5X0.1=0.55 □Cg

Cg=1 □Cg

Cin=7.175 □Cg

Cout = Cd+Cperi. Assuming Cperi to be negligible.

Cout = Cd.

Relative area of diffusion=51x2/4=102/4=25.5

Cd=25.5x0.25=6.25 □Cg.

The relative values are for the 5um technology

DELAY The concept of sheet resistance and standard unit capacitance can be used to calculate the delay. If we consider that a one feature size poly is charged by one feature size diffusion then the delay is Time constant 1T= Rs (n/p channel)x 1□Cg secs. This can be evaluated for any technology. The value of □Cg will vary with different technologies because of the variation in the minimum feature size.

5u using n diffusion=104X0.01=0.1ns safe delay 0.03nsec

2um = 104x0.0032=0.064 nsecs safe delay 0.02nsec

1.2u= 104x0.0023 = 0.046nsecs safe delay =0.1nsec

These safe figures are essential in order to anticipate the output at the right time

INVERTER DELAYS

We have seen that the inverter is associated with pull up and pull down resistance values. Specially in nmos inverters. Hence the delay associated with the inverter will depend on whether it is being turned off or on. If we consider two inverters cascaded then the total delay will remain constant irrespective of the transitions. Nmos and Cmos inverter delays are shown next

NMOS INVERTER

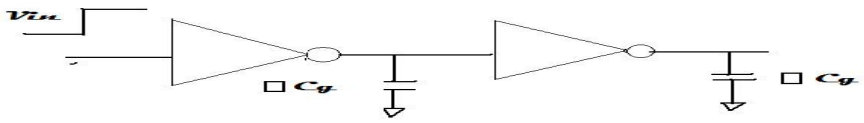

Figure 30: Cascaded nmos inverters

Let us consider the input to be high and hence the first inverter will pull it down. The pull down inverter is of minimum size nmos. Hence the delay is 1T. Second inverter will pull it up and it is 4 times larger, hence its delay is 4T.The total delay is 1T +4T= 5T. Hence for nmos the delay can be generalized as T=(1+Zpu/Zpd) T

CMOS INVERTER

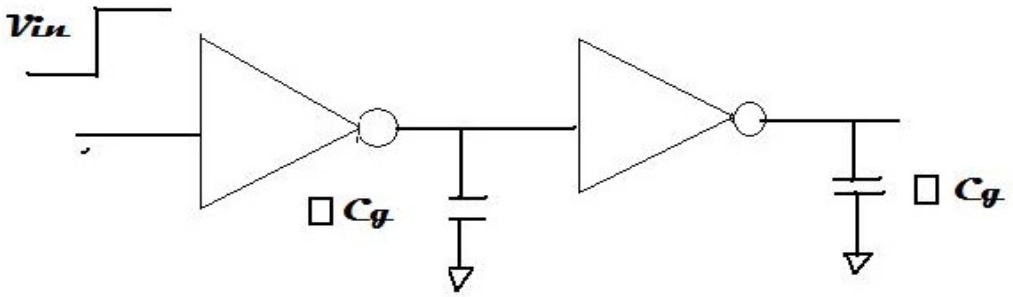

Figure 30 : Cascaded Cmos inverter

Let us consider the input to be high and hence the first inverter will pull it down. The nmos transistor has Rs = 10k and the capacitance is 2Cg. Hence the delay is 2T. Now the second inverter will pull it up, job done by the pmos. Pmos has sheet resistance of 25k i.e 2.5 times more, everything else remains same and hence delay is 5T. Total delay is 2T +5T = 7T . The capacitance here is double because the input is connected to the common poly, putting both the gate capacitance in parallel. The only factor to be considered is the resistance of the p gate which is increasing the delay. If want to reduce delay, we must reduce resistance. If we increase the width of p channel, resistance can be reduced but it increases the capacitance. Hence some trade off must be made to get the appropriate values.

FORMAL ESTIMATION OF DELAY

The inverter either charges or discharges the load capacitance CL. We could also estimate the delay by estimating the rise time and fall time theoritically.

Rise time estimation

Assuming that the p device is in saturation we have the current given by the equation Idsp=ßp(Vgs-|Vtp|)2/2

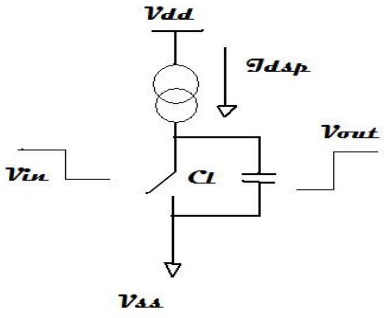

Figure 31 :Rise time estimation

The above current charges the capacitance and it has a constant value therefore the model can be written as shown in figure above. The output is the drop across the capacitance, given by

Vout =Idsp x t/CL

Substituting for Idsp we have Vout=ßp(Vgs-|Vtp|)2t/2CL. Therefore the equation for t=2CLVout/ßp(Vgs-|Vtp|).Let t=Tr and Vout=Vdd, therefore we have Tr = 2VddCL/ßp(Vgs-|Vtp|)2. If consider Vtp=0.2Vdd and Vgs=Vdd we have Tr = 3CL/ßpVdd

On similar basis the fall time can be also be written as Tf = 3CL/ßnVdd whose model can be written as shown next

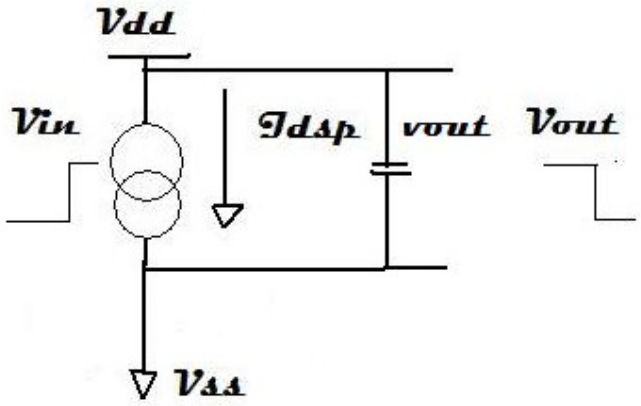

Figure 32 :Fall time estimation

DRIVING LARGE CAPACITIVE LOAD

The problem of driving large capacitive loads arises when signals must travel outside the chip. Usually it so happens that the capacitance outside the chip are higher. To reduce the delay these loads must be driven by low resistance. If we are using a cascade of inverter as drivers the pull and pull down resistances must be reduced. Low résistance means low L:W ratio. To reduce the ratio, W must be increased. Since L cannot be reduced to lesser than minimum we end up having a device which occupies a larger area. Larger area means the input capacitance increases and slows down the process more. The solution to this is to have N cascaded inverters with their sizes increasing, having the largest to drive the load capacitance. Therefore if we have 3 inverters,1st is smallest and third is biggest as shown next.

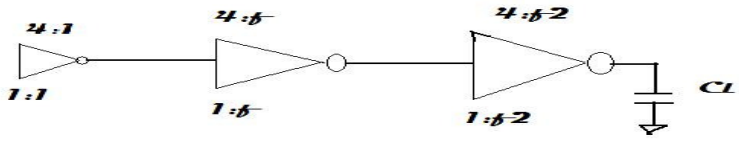

Figure 33:Cascaded inverters with varying widths

We see that the width is increasing by a factor of f towards the last stage. Now both f and N can be complementary. If f for each stage is large the number of stages N reduces but delay per stage increases. Therefore it becomes essential to optimize. Fix N and find the minimum value of f. For nmos inverters if the input transitions from 0 to 1 the delay is fT and if it transitions from 1 to 0 the delay is 4 fT. The delay for a nmos pair is 5 fT. For a cmos pair it will be 7fT

optimum value of f.

Assume y=CL/ ¤Cg = fN, therefore choice of values of N and f are interdependent. We find the value of f to minimize the delay, from the equation of y we have ln(y)=Nln(f) i.e N=ln(y)/ln(f). If delay per stage is 5fT for nmos, then for even number of stages the total delay is N/2 5fT=2.5fT. For cmos total delay is N/2 7fT = 3.5fT

Hence delay ά Nft=ln(y)/ln(f)ft. Delay can be minimized if chose the value of f to be equal to e which is the base of natural logarithms. It means that each stage is 2.7wider than its predecessor. If f=e then N= ln(y).The total delay is then given by

1.For N=even

td=2.5NeT for nmos, td=3.5NeT for cmos

2.For N=odd

transition from 0 to 1 transition from1 to 0

td=[2.5(N-1)+1]eT nmos td=[2.5(N-1)+4]eT

td=[3.59N-1)+2]eT cmos td=[3.5(N-1)+5]eT

for example

For N=5 which is odd we can calculate the delay fro vin=1 as td=[2.5(5-1)+1]eT =11e T

i.e. 1 +4+1+4+1 = 11eT

For vin =0 , td=[2.5(5-1)+4]eT = 14eT

4+1+4+1+4 = 14eT

SUPER BUFFER

The asymmetry of the inverters used to solve delay problems is clearly undesirable, this also leads to more delay problems, super buffer are a better solution. We have a inverting and non inverting variants of the super buffer. Such arrangements when used for 5u technology showed that they were capable of driving 2pf capacitance with 2nsec rise time.The figure shown next is the inverting variant.

I

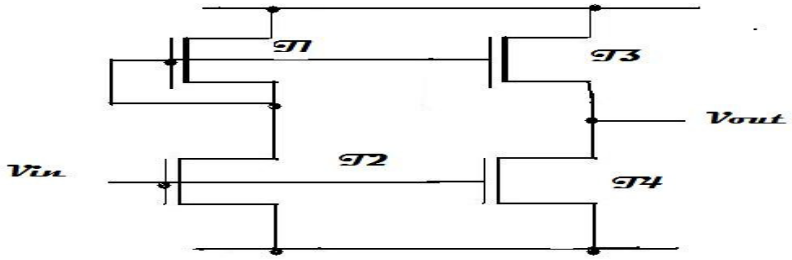

Figure 34:Inverterting buffer

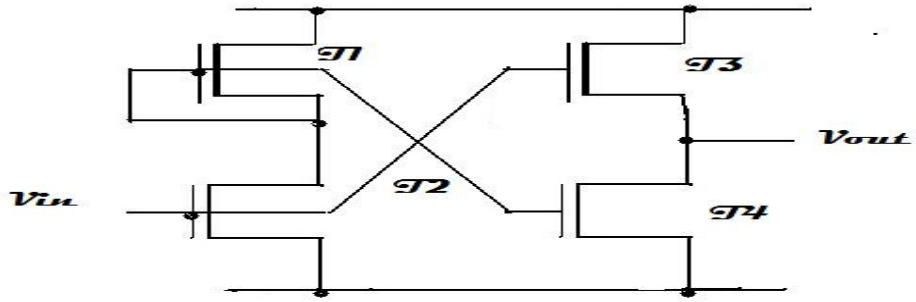

Figure 34:NonInverteing variant

BICMOS DRIVERS

The availability of bipolar devices enables us to use these as the output stage of inverter or any logic. Bipolar devices have high Tran con--ductance and they are able switch large currents with smaller input voltage swings. The time required to change the out by an amount equal to the input is given by $\Delta t=CL/gm$, Where gm is the device trans conductance. Δt will be a very small value because of the high gm. The transistor delay consists of two components Tin and TL. Tin the time required to charge the base of the transistor which is large. TL is smaller because the time take to charge capacitor is less by hfe which is the transistor gain a comparative graph shown below.

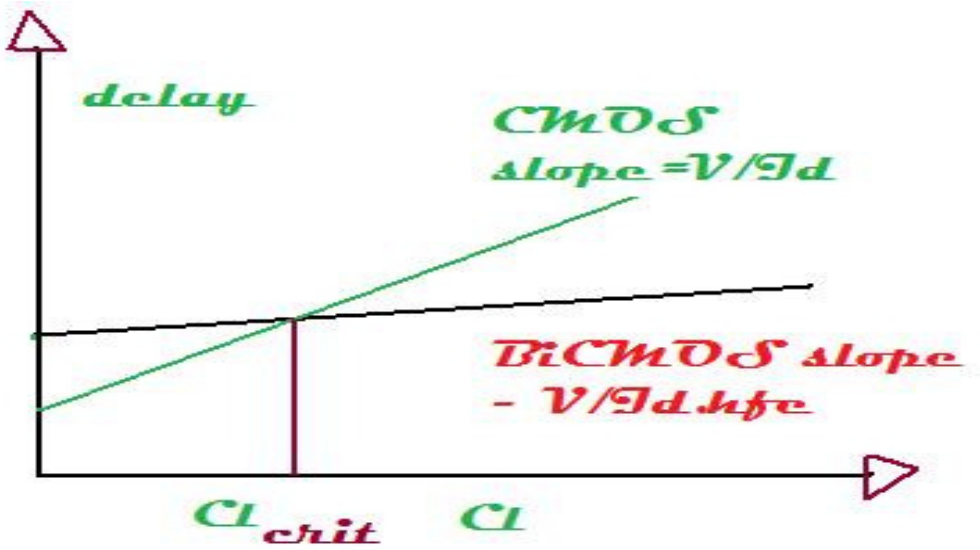

Figure 35

The collector resistance is another parameter that contributes to the delay.The graph shown below shows that for smaller load capacitance, the delay is manageable but for large capacitance, as Rc increases the delay increase drastically.

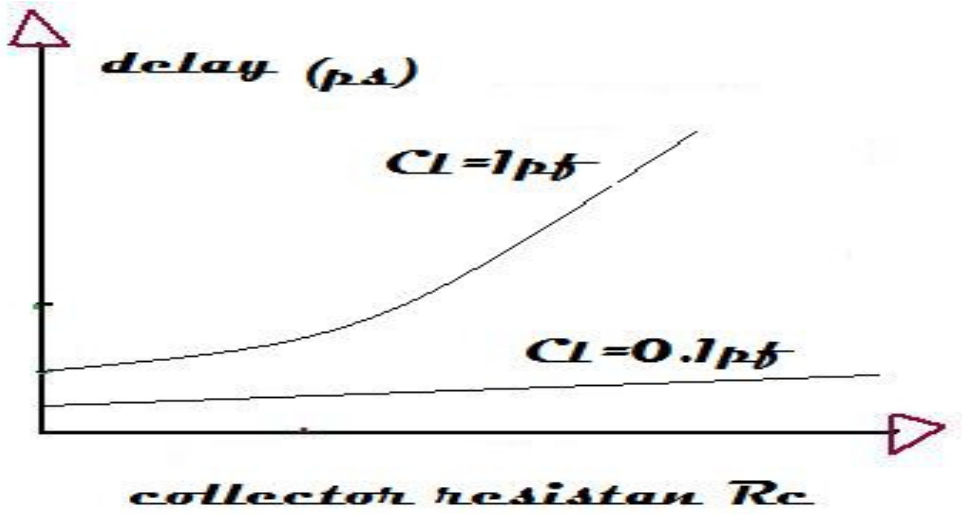

Figure 36

By taking certain care during fabrication reasonably good bipolar devices can be produced with large hfe, gm ,ß and small Rc. Therefore bipolar devices used in buffers and logic circuits give the designers a lot of scpoe and freedom .This is coming without having to do any changes with the cmos circuit.

PROPAGATION DELAY

This is delay introduced when the logic signals have to pass through a chain of pass transistors. The transistors could pose a RC product delay and this increases drastically as the number of pass transistor in series increases.As seen from the figure the response at node V2 is given by CdV2/dt=(V1-V2)(V2-V3)/R For a long network we can write RCdv/dt =dv2/dx2, i.e delay ά x2,

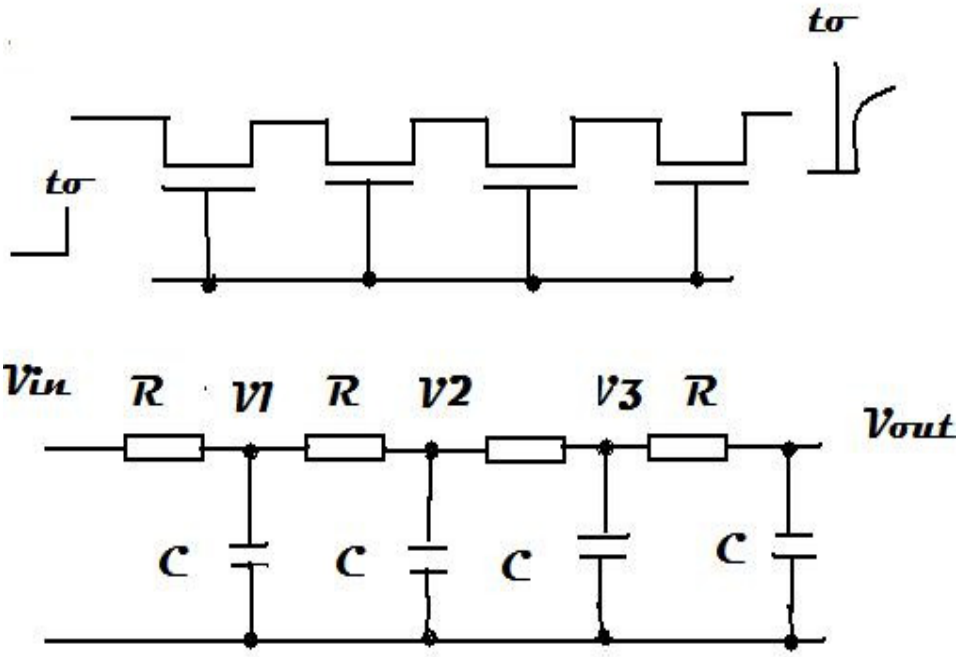

Figure 38

Lump all the R and C we have Rtotal=nrRs and C=nc□Cg where and hence delay =n^2rcT. The increases by the square of the number, hence restrict the number of stages to maximum 4 and for longer ones introduce buffers in between.

DESIGN OF LONG POLYSILICONS

The following points must be considered before going in for long wire.

1.The designer is also discouraged from designing long diffusion lines also because the capacitance is much larger

2.When it inevitable and long poly lines have to used the best way to reduce delay is use buffers in between. Buffers also reduce the noise sensitivity

OTHER SOURCES OF CAPACITANCE

Wiring capacitance

1.Fringing field

2.Interlayer capacitance

3.Peripheral capacitance

The capacitances together add upto as much capacitance as coming from the gate to source and hence the design must consider points to reduce them.The major of the wiring capacitance is coming from fringing field effects. Fringing capacitances is due to parallel fine metal lines running across the chip for power conection.The capacitance depends on the length l, thickness t and the distance d between the wire and the substrate. The accurate prediction is required for performance estimation.Hence Cw=Carea+Cff.

Interlayer capacitance is seen when different layers cross each and hence it is neglected for simole calculations. Such capacitance can be easily estimated for regular structures and helps in modeling the circuit better.

Peripheral capacitance is seen at the junction of two devices. The source and the drain n regions form junctions with the pwell (substrate) and p diffusion form with adjacent nwells leading to these side wall (peripheral) capacitance

The capacitances are profound when the devices are shrunk in sizes and hence must be considered. Now the total diffusion capacitance is Ctotal = Carea + Cperi

In order to reduce the side wall effects, the designers consider to use isolation regions of alternate impurity.

CHOICE OF LAYERS

1.Vdd and Vss lines must be distributed on metal lines except for some exception

2.Long lengths of poly must be avoided because they have large Rs,it is not suitable for routing Vdd or Vss lines.

3.Since the resistance effects of the transistors are much larger, hence wiring effects due to voltage dividers are not that profound

Capacitance must be accurately calculated for fast signal lines usually those using high Rs material. Diffusion areas must be carefully handled because they have larger capacitance to substrate.

With all the above inputs it is better to model wires as small capacitors which will give electrical guidelines for communication circuits.

PROBLEMS

1.A particular section of the layout includes a 3ƛ wide metal path which crosses a 2ƛ polysilicon path at right angles. Assuming that the layers are seperated by a 0.5 thick sio2,find the capacitance between the two.

Capacitance = ε0 εins A/D

Let the technology be 5um, ƛ=2.5um.

Area = 7.5umX5um=37.5um

C=4X8.854X10-12 x37.5/ 0.5 =2656pF

The value of C in standard units is

Relative area 6 ƛ2 /4 ƛ2 =1.5

C =1.5x0.075=0.1125□Cg

2 nd part of the problem

The polysilicon turns across a 4ג diffusion layer, find the gate to channel capacitance.

Area = 2 גx 4ג =8 ג2 Relative area= 8 ג2 / 4 ג2 =2

Relative capacitance for 5u=1

Total gate capacitance = 2□Cg

Gate to channel capacitance>metal

2. The two nmos transistors are cascaded to drive a load capacitance of 16□Cg as shown in figure ,Calculate the pair delay. What are the ratios of each transistors. f stray and wiring capacitance is to be considered then each inverter will have an additional capacitance at the output of 4 □Cg .Find the delay.

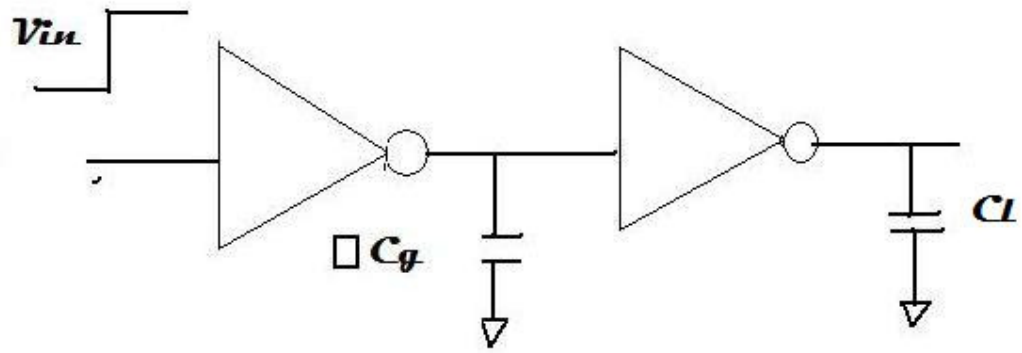

Figure 40

Lpu=16ג Wpu=2 ג Zpu=8

Lpd=2 ג Wpd=2 ג Zpd=1

Ratio of inverter 1 = 8:1

Lpu=2 ג Wpu=2 ג Zpu=1

Lpd =2 ג Wpd =8 ג Zpd=1/4

Ratio of inverter 2 = 1/1/4=4

Delay without strays

1T=Rsx1□Cg

Let the input transition from 1 to 0

Delay 1 = 8RsX□Cg=8T Delay 2=4Rs(□Cg +16 □Cg)=68T Total delay = 76T

Delay with strays

Delay 1 = 8RsX(□Cg+ 4□Cg) = 40T Delay 2= 4RsX(□Cg+ 4□Cg +16 □Cg)=84T

Total delay = 40+84=124T

If T = 0.1ns for 5u ie the delays are 7.6ns and 12.4ns

SCALING OF MOS DEVICES

The VLSI technology is in the process of evolution leading to reduction of the feature size and line widths. This process is called scaling down. The reduction in sizes has generally lead to better performance of the devices. There are certain limits on scaling and it becomes important to study the effect of scaling. The effect of scaling must be studied for certain parameters that effect the performance.

The parameters are as stated below

1.Minimum feature size

2.Number of gates on one chip

3.Power dissipation

4.Maximum operational frequency

5.Die size

6.Production cost .

These are also called as figures of merit

Many of the mentioned factors can be improved by shrinking the sizes of transistors, interconnects, separation between devices and also by adjusting the voltage and doping levels. Therefore it becomes essential for the designers to implement scaling and understand its effects on the performance

There are three types of scaling models used

1.Constant electric field scaling model

2.Constant voltage scaling model

3.Combined voltage and field model

The three models make use of two scaling factors 1/ß and 1/ά . 1/ß is chosen as the scaling factor for Vdd, gate oxide thickness D. 1/ ά is chosen as the scaling factor for all the linear dimensions like length, width etc. the figure next shows the dimensions and their scaling factors

The following are some simple derivations for scaling down the device parameters

1.Gate area Ag

Ag= L x W. Since L & W are scaled down by 1/ ά. Ag is scaled down by $1/ ά^2$

2.Gate capacitance per unit area

Co=εo/D, permittivity of sio2 cannot be scaled, hence Co can be scaled 1/1/ß=ß

3.Gate capacitance Cg

Cg=CoxA=CoxLxW. Therefore Cg can be scaled by ßx1/ άx1/ ά= ß/ $ά^2$

4.Parasitic capacitance

Cx=Ax/d, where Ax is the area of the depletion around the drain or source. d is the depletion width .Ax is scaled down by 1/ά2 and d is scaled by 1/ά. Hence Cx is scaled by

$1/ ά^2$ /1/ ά = 1/ ά

5.Carrier density in the channel Qon

Qon=Co.Vgs

Co is scaled by ß and Vgs is scaled by 1/ ß,hence Qo is scaled by ßx1/ß =1.

Channel resistance Ro

Ron = L/Wx1/Qoxμ, μ is mobility of charge carriers . Ro is scaled by1/ά/1/ άx1=1

Gate delay Td

Td is proportional to Ro and Cg

Td is scaled by 1x ß/$ά^2$ = ß/$ά^2$

Maximum operating frequency fo

fo=1/td,therefore it is scaled by 1/ ß/$ά^2$ = $ά^2$/ß

Saturation current

Idss= CoμW(Vgs-Vt)/2L, Co scale by ß and

voltages by 1/ ß, Idss is scaled by ß /$ß^2$= 1/ß

Current Density

J=Idss/A hence J is scaled by 1/ß/1/ά2 = $ά^2$ /ß