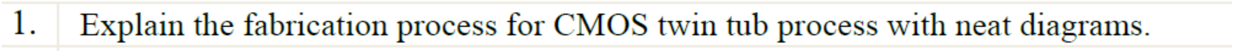


1.	Explain the fabrication process for CMOS twin tub process with neat diagrams.

In twin tub process both p-well and n-well for NMOS and PMOS transistors respectively are formed on the same substrate. The main advantage of this process is that the threshold voltage, body effect parameter and the transconductance can be optimized separately. The starting material for this process is p+ substrate with epitaxially grown p-layer which is also called as epilayer. The process steps of twin-tub process are shown in Figure 1.

The process starts with a p-substrate surfaced with a lightly doped p-epitaxial layer.

Step 1 : A thin layer of SiO_2 is deposited which will serve as the pad oxide.

Step 2 : A thicker sacrificial silicon nitride layer is deposited by chemical vapour deposition.

Step 3 : A plasma etching process is used to create trenches used for insulating the devices.

Step 4 : The trenches are filled with SiO_2 which is called as the field oxide.

Step 5 : To provide flat surface chemical mechanical planarization is performed and also sacrificial nitride and pad oxide is removed.

Step 6 : The p-well mask is used to expose only the p-well areas, after this implant and annealing sequence is applied to adjust the well doping. This is followed by second implant step to adjust the threshold NMOS transistor.

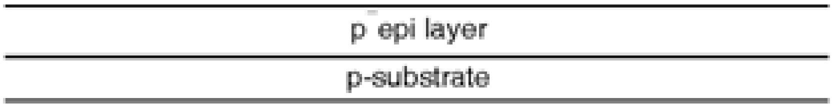
Step 7 : The n-well mask is used to expose only the n-well areas, after this implant and annealing sequence is applied to adjust the well doping. This is followed by a second implant step to adjust the threshold voltage of PMOS transistor.

Step 8 : A thin layer of gate oxide and polysilicon is chemically deposited and patterned with the help of polysilicon mask.

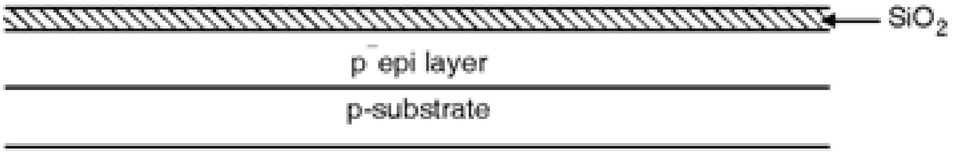
Step 9 : Ion implantation to dope the source and drain regions of the PMOS (p $^{+}$) and NMOS (n^{+}) transistors is used this will also form n^{+} polysilicon gate and p^{+} polysilicon gate for NMOS and PMOS transistors respectively.

Step 10 : Then the oxide or nitride spacers are formed by chemical vapour deposition (CVD).

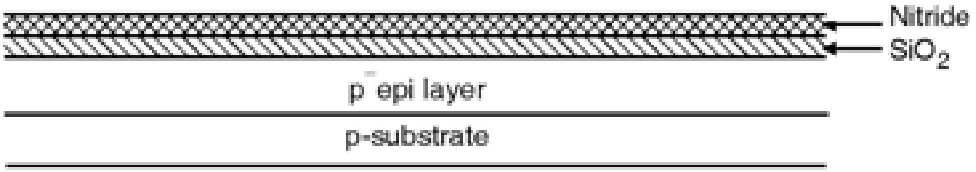
Step 11 : In this step contact or holes are etched, metal is deposited and patterned. After the deposition of last metal layer final passivation or overglass is deposited for protection.



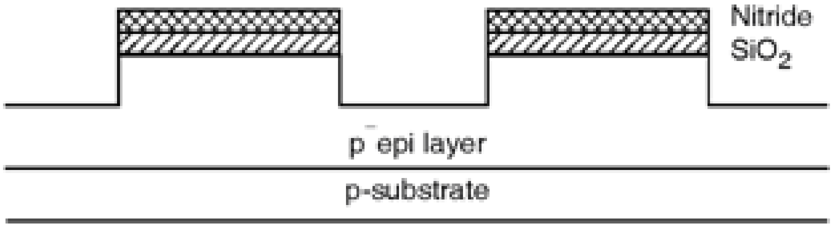
starting base material



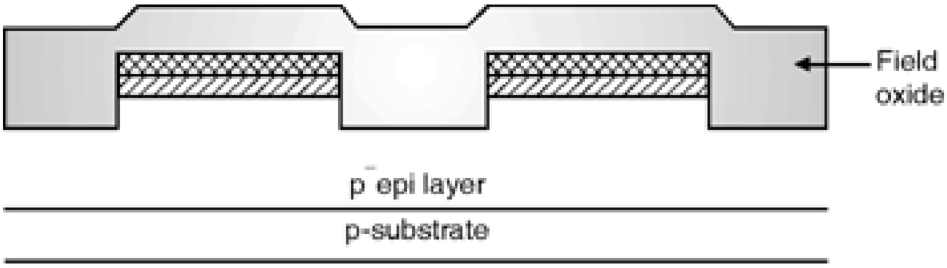
Step 1



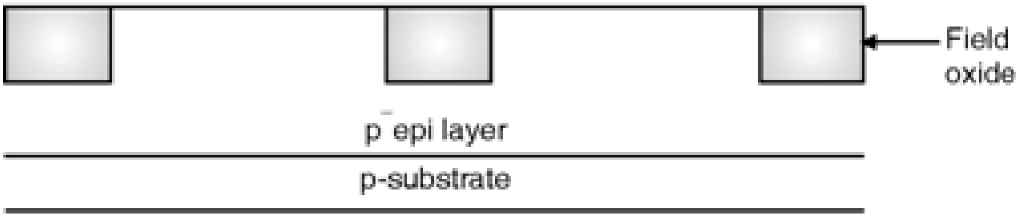
Step 2



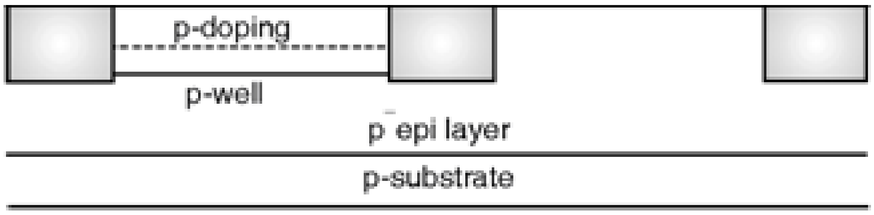
Step 3



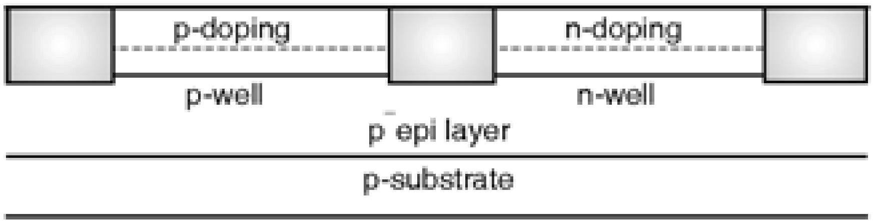
Step 4



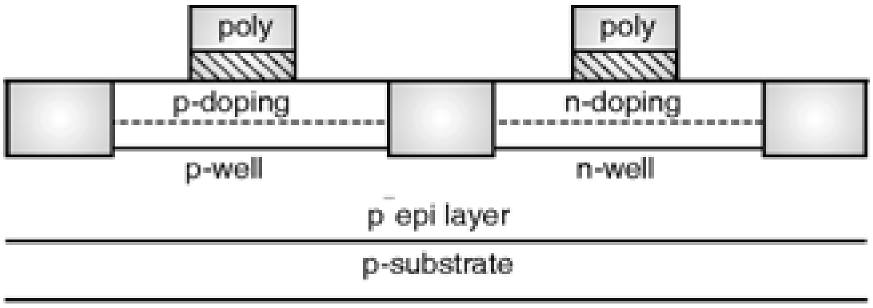
Step 5



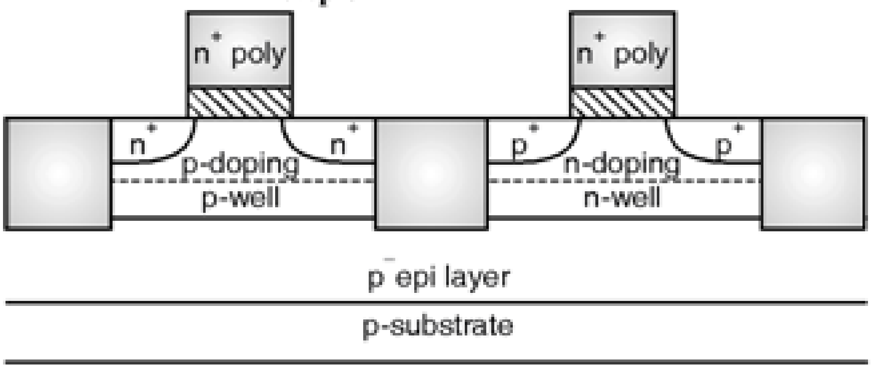
Step 6



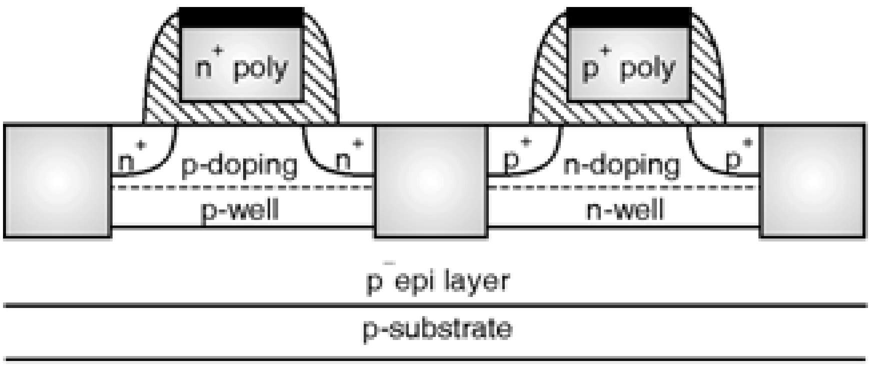
Step 7



Step 8



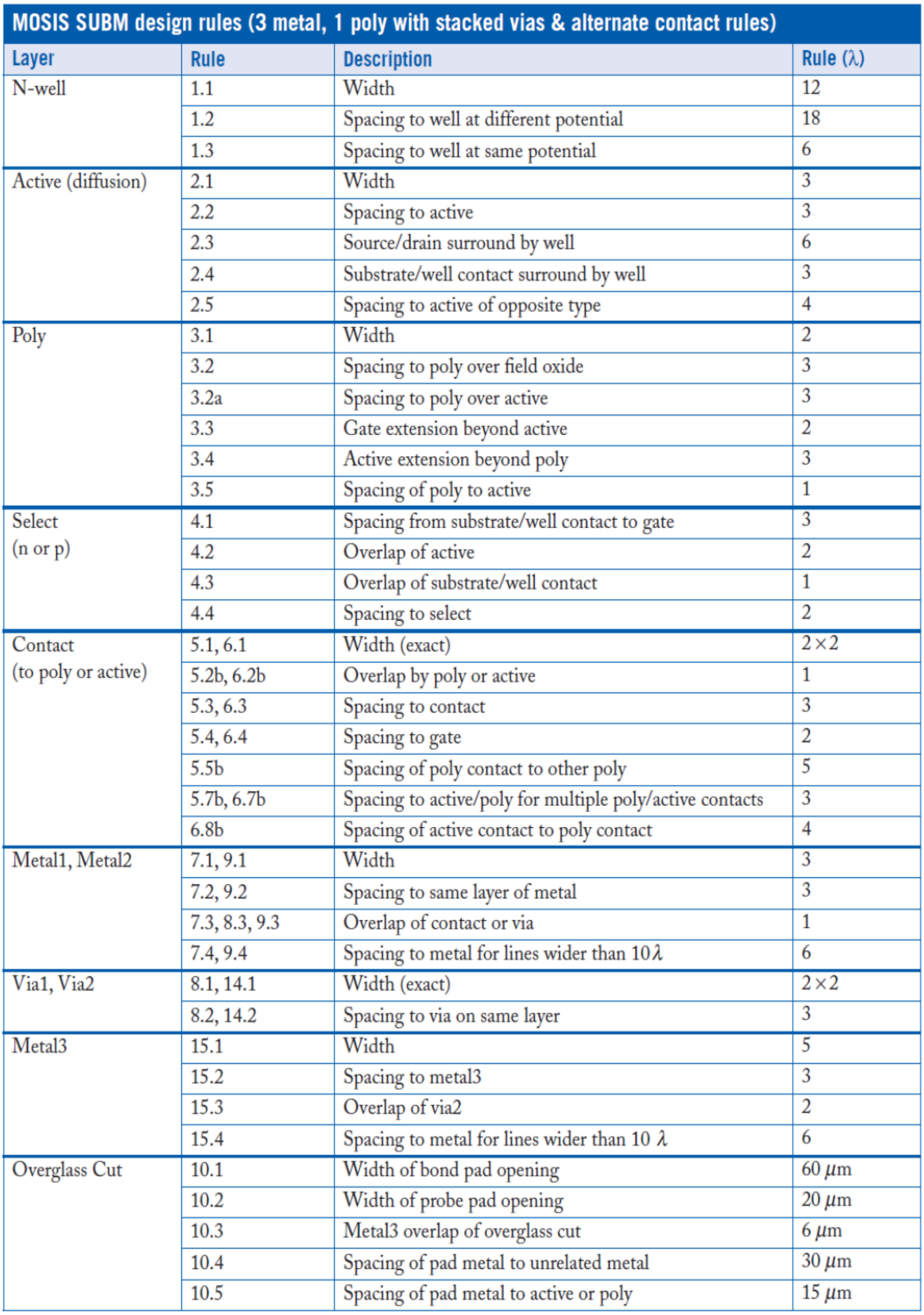
Step 9



Step 10

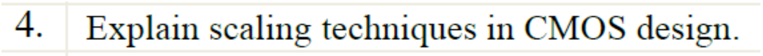
2. List the lambda based design rules.

LAMBA BASED LAYOUT RULES



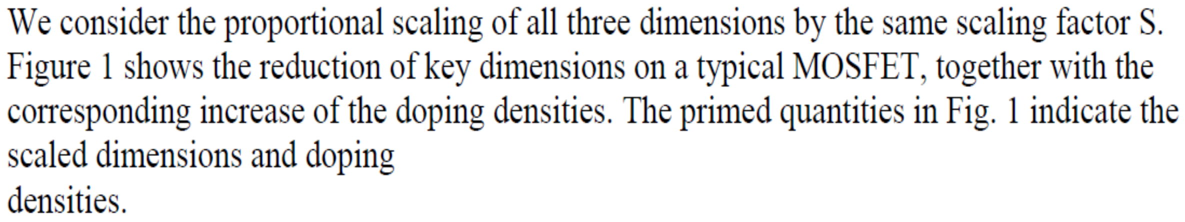
MOSIS SUBM design rules (3 metal, 1 poly with stacked vias & alternate contact rules)

Layer	Rule	Description	Rule (λ)
N-well	1.1	Width	12
	1.2	Spacing to well at different potential	18
	1.3	Spacing to well at same potential	6
Active (diffusion)	2.1	Width	3
	2.2	Spacing to active	3
	2.3	Source/drain surround by well	6
	2.4	Substrate/well contact surround by well	3
	2.5	Spacing to active of opposite type	4
Poly	3.1	Width	2
	3.2	Spacing to poly over field oxide	3
	3.2a	Spacing to poly over active	3
	3.3	Gate extension beyond active	2
	3.4	Active extension beyond poly	3
	3.5	Spacing of poly to active	1
Select (n or p)	4.1	Spacing from substrate/well contact to gate	3
	4.2	Overlap of active	2
	4.3	Overlap of substrate/well contact	1
	4.4	Spacing to select	2
Contact (to poly or active)	5.1, 6.1	Width (exact)	2 × 2
	5.2b, 6.2b	Overlap by poly or active	1
	5.3, 6.3	Spacing to contact	3
	5.4, 6.4	Spacing to gate	2
	5.5b	Spacing of poly contact to other poly	5
	5.7b, 6.7b	Spacing to active/poly for multiple poly/active contacts	3
	6.8b	Spacing of active contact to poly contact	4
Metal1, Metal2	7.1, 9.1	Width	3
	7.2, 9.2	Spacing to same layer of metal	3
	7.3, 8.3, 9.3	Overlap of contact or via	1
	7.4, 9.4	Spacing to metal for lines wider than 10λ	6
Via1, Via2	8.1, 14.1	Width (exact)	2 × 2
	8.2, 14.2	Spacing to via on same layer	3
Metal3	15.1	Width	5
	15.2	Spacing to metal3	3
	15.3	Overlap of via2	2
	15.4	Spacing to metal for lines wider than 10 λ	6
Overglass Cut	10.1	Width of bond pad opening	60 μm
	10.2	Width of probe pad opening	20 μm
	10.3	Metal3 overlap of overglass cut	6 μm
	10.4	Spacing of pad metal to unrelated metal	30 μm
	10.5	Spacing of pad metal to active or poly	15 μm



4. Explain scaling techniques in CMOS design.

Scaling of MOS transistors is the systematic reduction of overall dimensions of the devices as allowed by the available technology, while preserving the geometric ratios found in the larger devices. The operational characteristics of the MOS transistor change with the reduction of its dimensions.



We consider the proportional scaling of all three dimensions by the same scaling factor S. Figure 1 shows the reduction of key dimensions on a typical MOSFET, together with the corresponding increase of the doping densities. The primed quantities in Fig. 1 indicate the scaled dimensions and doping
densities.

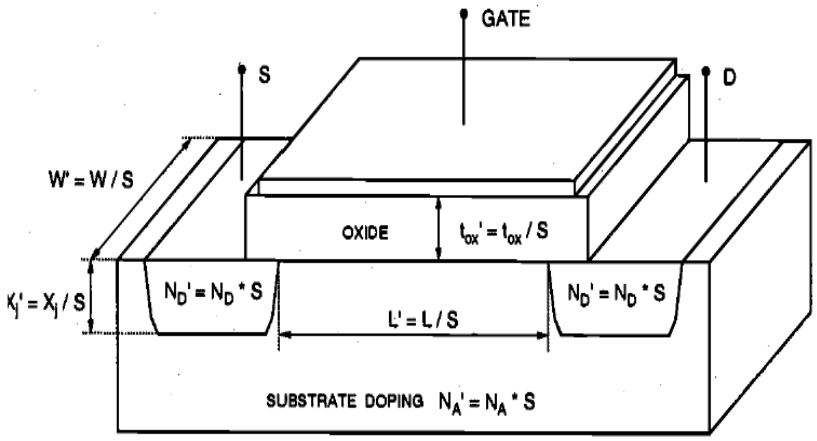


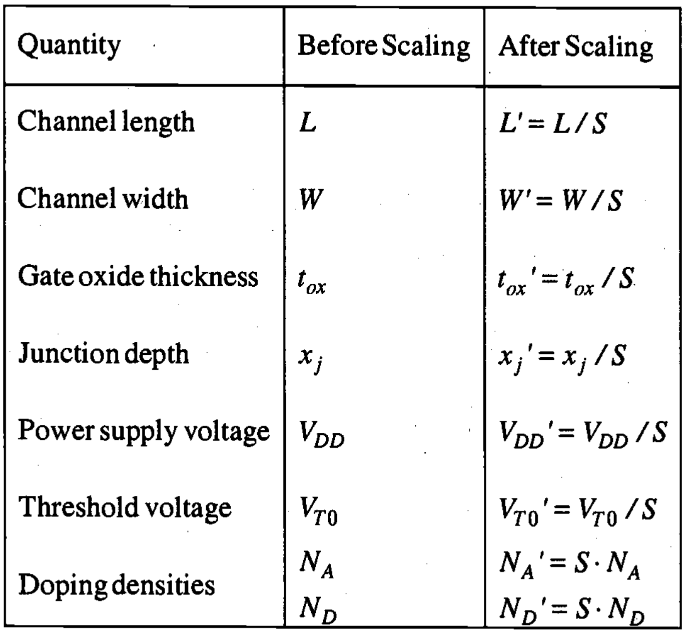
Figure 1. Scaling of a typical MOSFET by a scaling factor of *S*.

There are two basic types of size-reduction strategies: ***full scaling*** (also called constant-field scaling) and ***constant voltage scaling***. Both types of scaling have unique effects upon the operating characteristics of the MOS transistor.

Full Scaling (Constant-Field Scaling)

All potentials are scaled down proportionally by the same scaling factor say, S.

Table 1 lists the scaling factors for all significant dimensions, potentials, and doping densities of the MOS transistor. In order to maintain the field potential conditions, by Poisson’s equation, the doping densities must be increased by a factor S.



Quantity	Before Scaling	After Scaling
Channel length	L	$L' = L/S$
Channel width	W	$W' = W/S$
Gate oxide thickness	t_{ox}	$t_{ox}' = t_{ox}/S$
Junction depth	x_j	$x_j' = x_j/S$
Power supply voltage	V_{DD}	$V_{DD}' = V_{DD}/S$
Threshold voltage	V_{T0}	$V_{T0}' = V_{T0}/S$
Doping densities	N_A N_D	$N_A' = S \cdot N_A$ $N_D' = S \cdot N_D$

Table 1. Full scaling of MOSFET dimensions, potentials, and doping densities

Now let us consider the influence of full scaling described here upon the current-voltage characteristics of the MOS transistor. It will be assumed that the surface mobility μ is not significantly affected by the scaled doping density.

1. Gate oxide capacitance per unit area: The gate oxide capacitance per unit area, is changed as follows.

$$C_{ox}' = \frac{\varepsilon_{ox}}{t_{ox}'} = S \cdot \frac{\varepsilon_{ox}}{t_{ox}} = S \cdot C_{ox}$$

2. The aspect ratio *W/L* of the MOSFET will remain unchanged under scaling.

3. The transconductance parameter β or k_n will also be scaled by a factor of S. Let it be k_n'.

$$k_n' = S\beta.$$

4. The linear-mode drain current of the scaled MOSFET can now be found as:

$$I_D'(lin) = \frac{k_n'}{2} \cdot \left[2 \cdot (V_{GS}' - V_T') \cdot V_{DS}' - V_{DS}'^2\right]$$

$$= \frac{S \cdot k_n}{2} \cdot \frac{1}{S^2} \cdot \left[2 \cdot (V_{GS} - V_T) \cdot V_{DS} - V_{DS}^2\right] = \frac{I_D(lin)}{S}$$

5. Similarly, the saturation-mode drain current is also reduced by the same scaling factor.

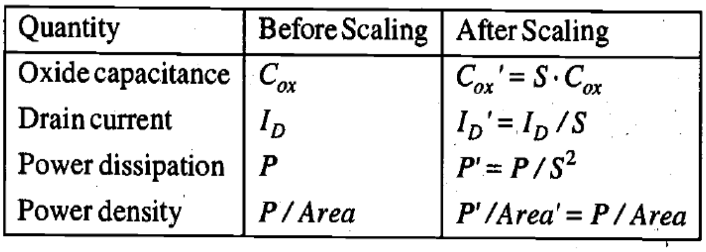
$$I_D'(sat) = \frac{k_n'}{2} \cdot (V_{GS}' - V_T')^2 = \frac{S \cdot k_n}{2} \cdot \frac{1}{S^2} \cdot (V_{GS} - V_T)^2 = \frac{I_D(sat)}{S}$$

6. The power dissipation of the transistor will be reduced by the factor S^2.

$$P' = I_D' \cdot V_{DS}' = \frac{1}{S^2} \cdot I_D \cdot V_{DS} = \frac{P}{S^2}$$

This significant reduction of the power dissipation is one of the most attractive features of full scaling.

SUMMARY OF FULL/CONSTANT FIELD SCALING:

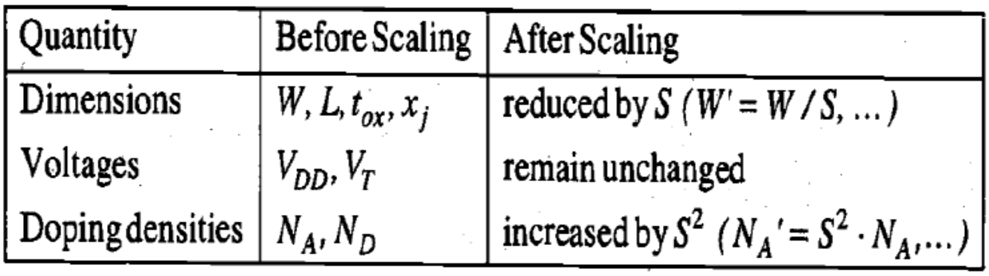


Quantity	Before Scaling	After Scaling
Oxide capacitance	C_{ox}	$C_{ox}' = S \cdot C_{ox}$
Drain current	I_D	$I_D' = I_D / S$
Power dissipation	P	$P' = P / S^2$
Power density	$P / Area$	$P' / Area' = P / Area$

While the full scaling strategy dictates that the power supply voltage and all terminal voltages be scaled down proportionally with the device dimensions, the scaling of voltages may not be very practical in case of circuits that may require certain voltage levels for all input and output voltages. For these reasons, constant-voltage scaling is usually preferred over full scaling.

Constant-Voltage Scaling: In constant-voltage scaling, all dimensions of the MOSFET are reduced by a factor of *S,* as in full scaling. But the power supply voltage and the terminal voltages remain unchanged.

Table 2 lists the scaling factors for all significant dimensions, potentials, and doping densities of the MOS transistor.



Quantity	Before Scaling	After Scaling
Dimensions	W, L, t_{ox}, x_j	reduced by S ($W' = W / S, \ldots$)
Voltages	V_{DD}, V_T	remain unchanged
Doping densities	N_A, N_D	increased by S^2 ($N_A' = S^2 \cdot N_A, \ldots$)

Table 2. Full scaling of MOSFET dimensions, potentials, and doping densities.

1. The doping densities must be increased by a factor of S^2 in order to preserve the charge-field relations.
2. The gate oxide capacitance per unit area *Cox* is increased by a factor of S, which means that the transconductance parameter β or k_n is also increased by S. Let it be k_n'

$$k_n' = S\beta$$

3. Since the terminal voltages remain unchanged, the linear mode drain current of the scaled
   MOSFET can be written as:

$$I_D'(lin) = \frac{k_n'}{2} \cdot \left[2 \cdot (V_{GS}' - V_T') \cdot V_{DS}' - V_{DS}'^2\right]$$

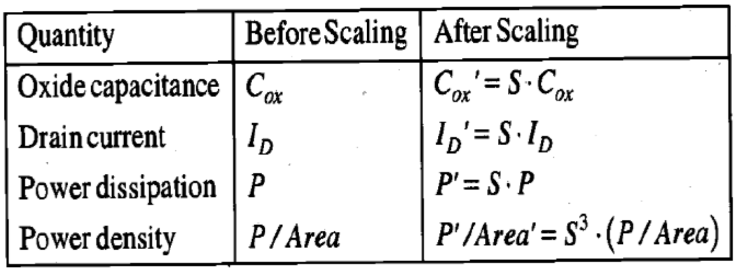
$$= \frac{S \cdot k_n}{2} \cdot \left[2 \cdot (V_{GS} - V_T) \cdot V_{DS} - V_{DS}^2\right] = S \cdot I_D(lin)$$

4. The saturation-mode drain current will be increased by a factor of S after constant voltage scaling. This means that the drain current *density* (current per unit area) is increased by a factor of S^3.
5. The power dissipation of the MOSFET increases by a factor of S.

$$P' = I_D' \cdot V_{DS}' = (S \cdot I_D) \cdot V_{DS} = S \cdot P$$

6. The power density (power dissipation per unit area) is found to increase by a factor of S^3 after constant-voltage scaling, with possible adverse effects on device reliability.

SUMMARY OF CONSTANT VOLTAGE SCALING



Quantity	Before Scaling	After Scaling
Oxide capacitance	C_{ox}	$C_{ox}' = S \cdot C_{ox}$
Drain current	I_D	$I_D' = S \cdot I_D$
Power dissipation	P	$P' = S \cdot P$
Power density	$P / Area$	$P' / Area' = S^3 \cdot (P / Area)$

5. Explain the RC equivalent model for CMOS inverter. Estimate the propagation delay using Elmore delay for an inverter driving m identical unit inverters if the driver width is 'w' times unit size.

RC DELAY MODEL

The RC delay model treats a transistor as a switch in series with a resistor. To get same rise and fall time in nMOS and pMOS, width of pMOS will be 2-3 times of nMOS. Hence equivalent circuit of MOS transistors will have:

- **Ideal switch + capacitance + resistance**
- **Unit nMOS has resistance R, capacitance C**
- **Unit pMOS has resistance 2R, capacitance C**

Capacitance is proportional to width and resistance is inversely proportional to width.

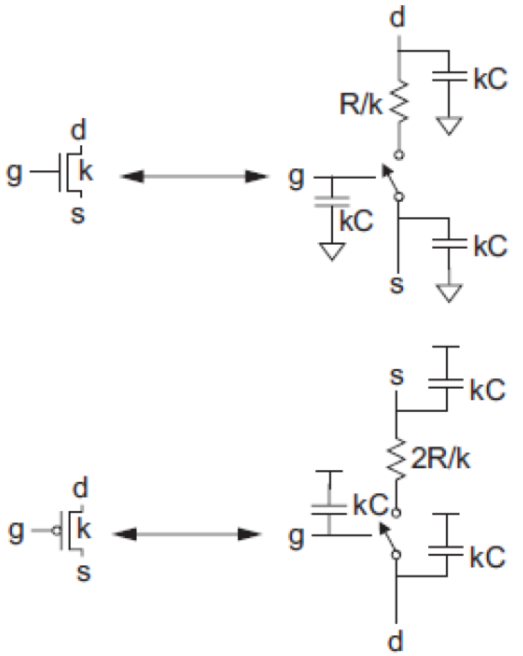
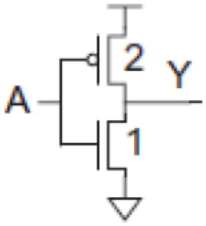


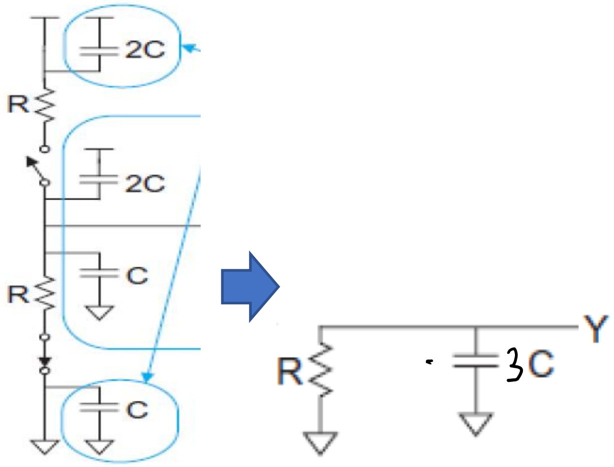
Figure 5: Equivalent circuits for transistors

Figure 5 shows equivalent RC circuit models for nMOS and pMOS transistors of width *k* with contacted diffusion on both source and drain. A transistor of *k* times unit width has capacitance *kC* and resistance is R/k. Diffusion capacitance depends on the size of the source/drain region. The pMOS transistor has approximately twice the resistance of the nMOS transistor because holes have lower mobility than electrons. For unit nMOS or pMOS, k=1.

Unit inverter has equivalent resistance of R for equal rise and fall times, hence for the pmos, in unit inverter transistor width is k=2 such that 2R/k = R for the pmos and for nmos it is k=1 such that R/k=R. Hence the pmos :nmos ratio is 2:1 as shown below.



Now the equivalent RC circuit for the inverter can be redrawn as below:



PROBLEM SOLUTION

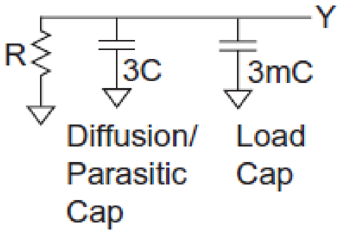


Figure 1

Figure 1 above shows an equivalent circuit for the falling transition. Each load inverter presents 3*C* units of gate capacitance, for a total of 3*mC*. The output node also sees a capacitance of 3*C* from the drain diffusions of the driving inverter. This capacitance is called *parasitic* because it is an undesired side-effect of the need to make the drain large enough to contact. The parasitic capacitance is independent of the load that the inverter is driving. Hence, the total capacitance is (3 + 3*m*)*C*. The resistance is *R*, so the Elmore delay is *tpd* = (3 + 3*m*)*RC*. The equivalent circuit for the rising transition gives the same results.

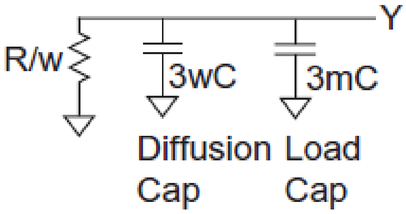


Figure 2

Figure 2 shows the equivalent circuit when driver is w times unit size. The driver transistors are *w* times as wide, so the effective resistance decreases by a factor of *w*. The diffusion capacitance increases by a factor of *w*. The Elmore delay is $tpd = ((3w + 3m)C)(R/w) = (3 + 3m/w)RC$.

6. Explain the transient response of a CMOS inverter and find its delay using Elmore delay model.

Transient response of a circuit is the solution of a differential equation describing the output voltage as a function of input voltage and time.

Delay is the time when the output reaches $V_{DD}/2$.

The differential equation is based on charging or discharging of the capacitances in the circuit. The circuit takes time to switch because the capacitance cannot change its voltage instantaneously. If capacitance *C* is charged with a current *I*, the voltage on the capacitor varies as:

$$I = C\frac{dV}{dt}$$

STEP RESPONSE OF INVERTER

Figure 1 shows an inverter *X*1 driving a load capacitance C_{out}. Suppose **a voltage step** from 0 to V_{DD} is applied to node *A* and we have to compute the propagation delay, *tpdf* , through *X*1, i.e., the delay from the input step until node *B* crosses $V_{DD}/2$.

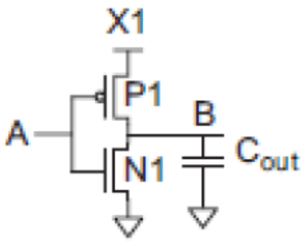
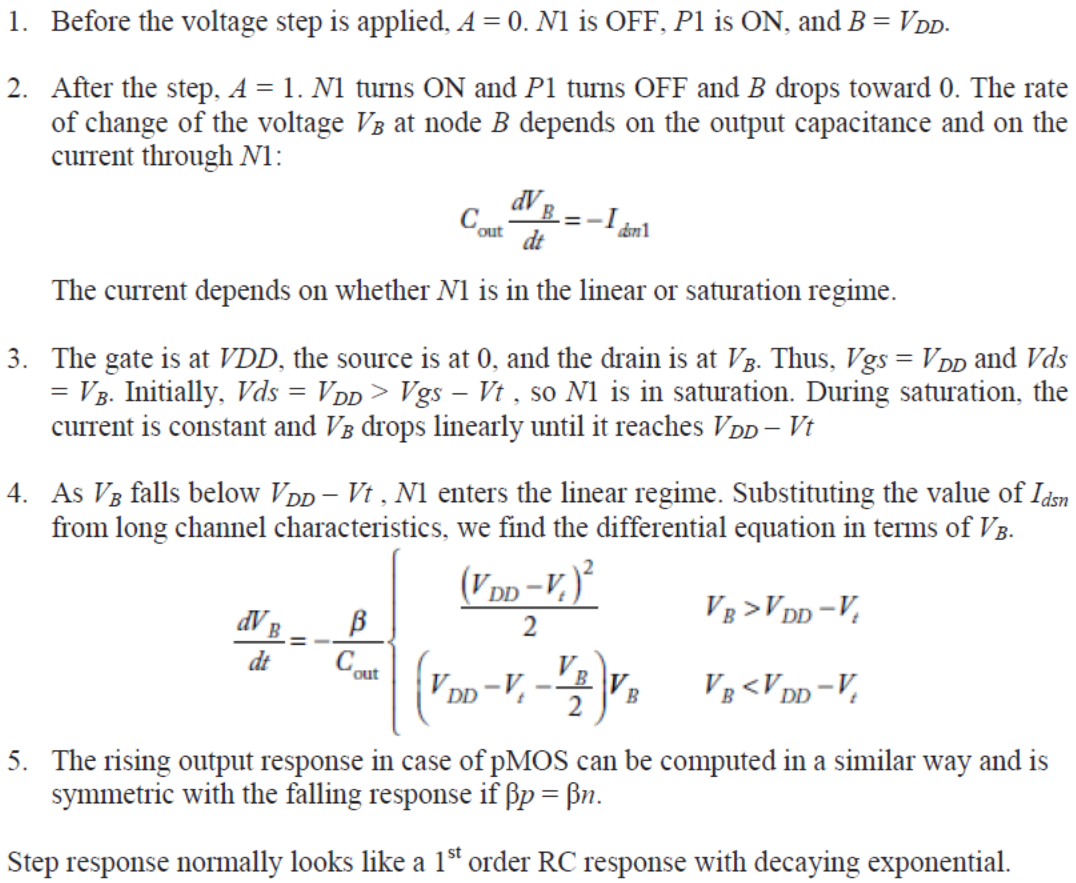


Figure 1: Inverter delay calculation using load capacitance



1. Before the voltage step is applied, $A = 0$. $N1$ is OFF, $P1$ is ON, and $B = V_{DD}$.

2. After the step, $A = 1$. $N1$ turns ON and $P1$ turns OFF and B drops toward 0. The rate of change of the voltage V_B at node B depends on the output capacitance and on the current through $N1$:

$$C_{\text{out}}\frac{dV_B}{dt} = -I_{dsn1}$$

   The current depends on whether $N1$ is in the linear or saturation regime.

3. The gate is at VDD, the source is at 0, and the drain is at V_B. Thus, $Vgs = V_{DD}$ and $Vds = V_B$. Initially, $Vds = V_{DD} > Vgs - Vt$, so $N1$ is in saturation. During saturation, the current is constant and V_B drops linearly until it reaches $V_{DD} - Vt$

4. As V_B falls below $V_{DD} - Vt$, $N1$ enters the linear regime. Substituting the value of I_{dsn} from long channel characteristics, we find the differential equation in terms of V_B.

$$\frac{dV_B}{dt} = -\frac{\beta}{C_{\text{out}}}\begin{cases} \dfrac{\left(V_{DD}-V_t\right)^2}{2} & V_B > V_{DD}-V_t \\ \left(V_{DD}-V_t-\dfrac{V_B}{2}\right)V_B & V_B < V_{DD}-V_t \end{cases}$$

5. The rising output response in case of pMOS can be computed in a similar way and is symmetric with the falling response if $\beta p = \beta n$.

Step response normally looks like a 1st order RC response with decaying exponential.

We know that the inverter can be modelled as a 1st order RC system.

Now, let us apply the RC model to estimate the step response of the first-order system shown in Figure 8.

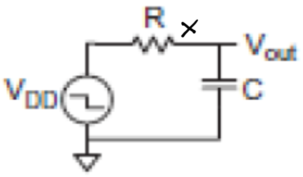


Figure 8: First order RC system

This system is a good model of an inverter sized for equal rise and fall delays. The system has a transfer function

$$H(s) = \frac{1}{1+sRC}$$ and the step response is $$V_{\text{out}}(t) = V_{DD}e^{-t/\tau}$$ where $H(s)= V_{out}(t)/V_{in}(t)$

where $\tau = RC$. The propagation delay, t_{pd} is the time at which V_{out} reaches $V_{DD}/2$, as shown in Figure 9. Putting the delay condition into the Vout equation and solving for t_{pd} we get

$$t_{pd} = RC\ln 2$$

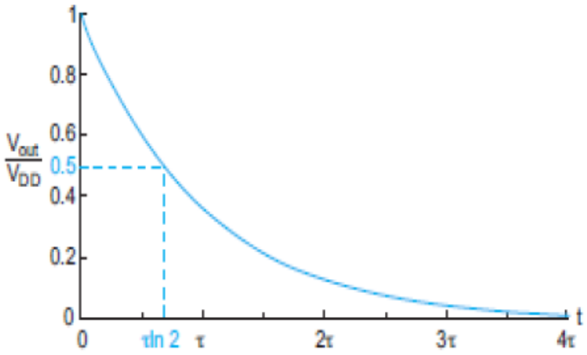


Figure 9: First-order step response

The factor of ln 2 = 0.69 is cumbersome. The effective resistance *R* is an empirical parameter anyway, so it is preferable to incorporate the factor of ln 2 to define a new effective resistance *R'* = *R* ln 2. Now the propagation delay is simply *R'*C. For the sake of convenience, we usually drop the prime symbols and just write t_{pd}=RC, where the effective resistance *R* is chosen to give the correct delay.

ELMORE DELAY

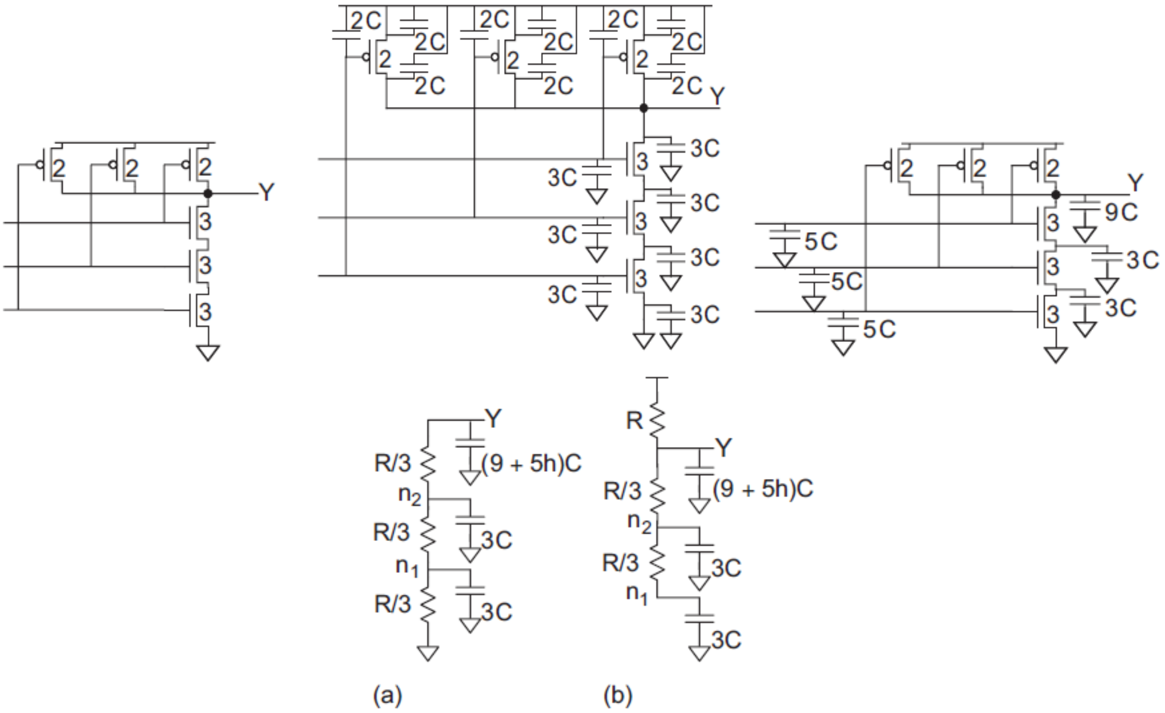
In general, most circuits of interest can be represented as an *RC tree*, i.e., an RC circuit with no loops. The root of the tree is the voltage source and the leaves are the capacitors at the ends of the branches. The Elmore delay model estimates the delay from a source switching to one of the leaf nodes changing as the sum over each node *i* of the capacitance *Ci* on the node, multiplied by the effective resistance *Ris* on the shared path from the source to the node and the leaf.

$$t_{pd} = \sum_i R_{is} C_i$$

Now applying Elmore model to the RC equivalent circuit in figure 8 above the propagation delay can be given as tpd= RC where R and C are the resistance at node x.

7.	Sketch a 3 input NAND and 3 input NOR gate and find their equivalent capacitances and resistances. Calculate the Elmore delays for the 3 input NAND gate.

3 Input NAND with transistor width values and its equivalent RC circuit



The NAND gate load presents 5 units of capacitance on a given input. For a fanout of h, figure (a) above shows the equivalent circuit including the load for the falling transition. Node $n1$ has capacitance $3C$ and resistance of $R/3$ to ground. Node $n2$ has capacitance $3C$ and resistance $(R/3 + R/3)$ to ground. Node Y has capacitance $(9 + 5h)C$ and resistance $(R/3 + R/3 + R/3)$ to ground. The Elmore delay for the falling output is the sum of these RC products, $tpdf = (3C)(R/3) + (3C)(R/3 + R/3) + ((9 + 5h)C)(R/3+ R/3 + R/3) = (12 + 5h)RC$. Figure (b) above shows the equivalent circuit for the falling transition. In the worst case, the two inner inputs are 1 and the outer input falls. Y is pulled up to VDD through a single pMOS transistor. The ON nMOS transistors contribute parasitic capacitance that slows the transition. Node Y has capacitance $(9 + 5h)C$ and resistance R to the VDD supply. Node $n2$ has capacitance $3C$. The relevant resistance is only R, not $(R + R/3)$, because the output is being charged only through R. This is what is meant by the resistance on the shared path from the source (VDD) to the node ($n2$) and the leaf (Y). Similarly, node $n1$ has capacitance $3C$ and resistance R. Hence, the Elmore delay for the rising output is $tpdr = (15 + 5h)RC$.

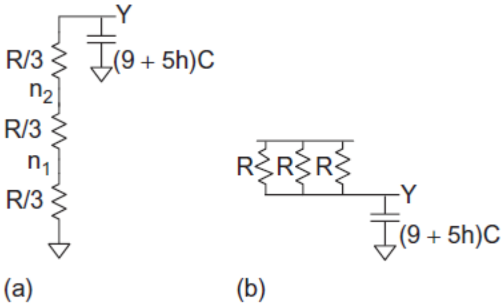
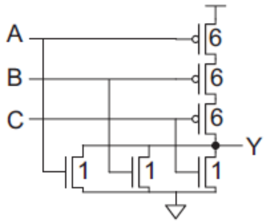


FIGURE [redacted] Equivalent circuits for contamination delay

The contamination delay is the fastest that the gate might switch. For the falling transition, the best case is that the bottom two nMOS transistors are already ON when the top one turns ON. In such a case, the diffusion capacitances on $n1$ and $n2$ have already been discharged and do not contribute to the delay. Figure (a) above shows the equivalent circuit and the delay is *tcdf* = (9 + 5*h*)*RC*. For the rising transition, the best case is that all three pMOS transistors turn on simultaneously. The nMOS transistors turn OFF, so *n*1 and *n*2 are not connected to the output and do not contribute to delay. The parallel transistors deliver three times as much current, as shown in Figure (b) above, so the delay is *tcdr* = (3 + (5/3)*h*)*RC*.

For the problem, h=0, hence, tpdf=12RC and tpdr=15RC, tcdf=9RC and tcdr=3RC.

3 INPUT NOR GATE with transistor widths


A
B
C
6
6
6
Y
1
1
1