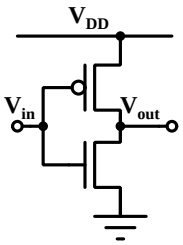
INDIAN INSTITUTE OF TECHNOLOGY BOMBAY
ELECTRICAL ENGINEERING DEPARTMENT

Solution to End Semester Examination

Sunday	**EE 671: VLSI Design**	Time: 13:30-16:30
Nov. 13, 2022	Autumn Semester 2022	Marks: 40

This question paper has 6 questions for a total of 6+12+10+4+5+3 = 40 marks. Quantitative answers should be accurate to 0.1%

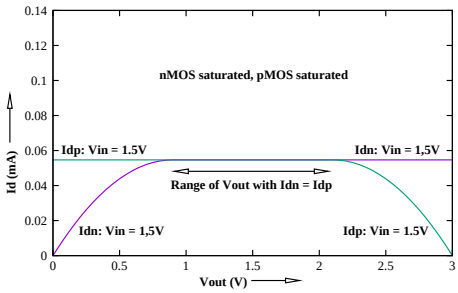
Q–1 Consider a CMOS inverter as shown below:



The n and p channel transistors have identical geometry with minimum channel length and width.
$K_n \equiv \mu_n C_{ox}(W/L)$, $K_p \equiv \mu_p C_{ox}(W/L)$ $\beta \equiv \mu_n/\mu_p$.

a) Find an expression for the input voltage V_{in} (in terms of V_{DD}, K_n, β and turn on voltages V_{Tn} and V_{Tp}) at which the current drawn from the supply is maximum. Derive an expression for this maximum current.
(Hint: You can use the graphical solution method to determine the input voltage at which the current drawn is maximum).

Soln. 1-a) Drawn current is maximum when both transistors are saturated.



Both transistor saturate simultaneously at a specific input voltage. If the input voltage is any lower, the nMOS current is less and limits the total current drawn. If it is higher, the pMOS current is less and limits the total current drawn.
Thus we first evaluate the input voltage for which both transistors are saturated.

Current through the two transistors must be equal. When both transistors are saturated, we get

$$\begin{aligned}
\frac{K_n}{2}(V_{in}-V_{Tn})^2 &= \frac{K_p}{2}(V_{DD}-V_{in}-V_{Tp})^2 \\
\text{So} \quad \frac{K_n}{K_p}=\beta &= \frac{(V_{DD}-V_{in}-V_{Tp})^2}{(V_{in}-V_{Tn})^2} \\
\sqrt{\beta}(V_{in}-V_{Tn}) &= V_{DD}-V_{in}-V_{Tp} \\
(1+\sqrt{\beta})V_{in} &= V_{DD}+\sqrt{\beta}V_{Tn}-V_{Tp} \\
\text{Therefore} \quad V_{in} &= \frac{V_{DD}+\sqrt{\beta}V_{Tn}-V_{Tp}}{1+\sqrt{\beta}}
\end{aligned}$$

This is the input voltage at which maximum current will be drawn by the circuit. The corresponding value of current is: $I_{Dmax} = \frac{K_n}{2}(V_{in}-V_{Tn})^2$.

$$\begin{aligned}
I_{Dmax} &= \frac{K_n}{2}\left(\frac{V_{DD}+\sqrt{\beta}V_{Tn}-V_{Tp}}{1+\sqrt{\beta}}-V_{Tn}\right)^2 \\
&= \frac{K_n}{2}\left(\frac{V_{DD}-V_{Tn}-V_{Tp}}{1+\sqrt{\beta}}\right)^2
\end{aligned}$$

– **[4]**

b) Evaluate the input voltage for which maximum current is drawn and the value of maximum current if $K_n = \mu_n C_{ox}(W/L) = 120\mu\text{A/V}^2$, $\beta \equiv \mu_n/\mu_p = 2.6$, $V_{DD} = 1.8\text{V},\ V_{Tn} = 0.4\text{V},\ V_{Tp} = 0.5\text{V}$.

Soln. 1-b) $V_{DD} = 1.8\text{V},\ V_{Tn} = 0.4\text{V},\ V_{Tp} = 0.5\text{V},\ \beta = 2.6$

$$V_{in} = \frac{1.8 + \sqrt{2.6} \times 0.4 - 0.5}{1 + \sqrt{2.6}} = 0.7445\text{V}$$

$$I_{Dmax} = \frac{K_n}{2}(V_{in} - V_{Tn})^2 = 60 \times (0.7445 - 0.4)^2 = 7.121\mu\text{A}$$

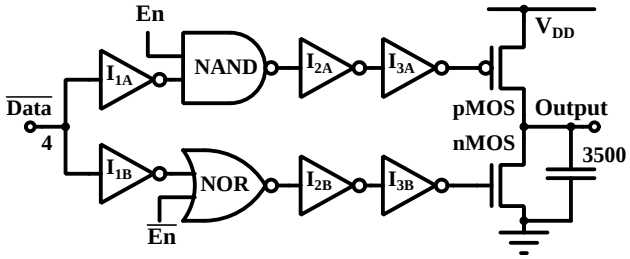
Just to confirm,

$$\begin{aligned} I_{Dmax} &= \frac{K_p}{2}(V_{DD} - V_{in} - V_{Tp})^2 \\ &= \frac{60}{2.6} \times (1.8 - 0.7445 - 0.5)^2 = 7.121\mu\text{A} \end{aligned}$$

– **[2]**

– **[Q1: 4+2 = 6 marks]**

Q–2 Consider an output driver using the NAND-NOR configuration for an I/O pad.



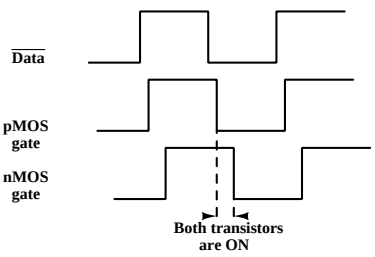
The final load is equivalent to 3500 template inverters.
The value of γ is 2.5.
The parasitic delay of inverters is 2.2.

(Hint: treat the output pMOS and nMOS transistors as logic gates by themselves and compute their logical efforts and parasitic delays by the usual method).

The combined input capacitance of the two branches should be equivalent to 4 template inverters.

a) Show with a timing diagram what will happen if delays through the two branches are substantially different.

Soln. 2-a) If the delays of the two paths are substantially mismatched, data edges will arrive at different times to the gates of pMOS and nMOS transistors.



If the pMOS path is faster, on a downward transition on $\overline{\text{Data}}$, the pMOS will turn on as soon as the faster transition has reached it. However, the downward transition will reach the nMOS later and as we can see from the wave forms, there will be an interval when the gate of pMOS is low, while the gate of nMOS is still high. During this interval, both transistors will be on. A similar problem will occur on the upward transition if the nMOS path is faster.

These are very large transistors and if they are simultaneously 'on', very high short circuit current will flow through the transistors, resulting in wastage of power. – **[1]**

b) Compute the logical effort and parasitic delays for the output pMOS and nMOS transistors treated as logic gates. While calculating the parasitic delay, notice that drains of *both* transistors are connected to the output.

Soln. 2-b) The pMOS transistor has a width of γ to supply the same current (during charge) as an inverter. Therefore its logical effort will be

$$g_{pMOS} = \frac{\gamma}{1+\gamma} = \frac{2.5}{3.5} = \frac{5}{7} = 0.714286$$

Similarly, the nMOS transistor has a width of 1 to supply the same current (during discharge) as an inverter. Therefore its logical effort will be

$$g_{nMOS} = \frac{1}{1+\gamma} = \frac{1}{3.5} = \frac{2}{7} = 0.285714$$

Notice that together the pMOS and nMOS transistors are equivalent to an inverter and their logical efforts add up to 1 as expected.

Both transistors see the pMOS width as well as the nMOS width as parasitic load at their output. Therefore their parasitic delay is $\propto 1+\gamma$. An inverter also sees a transistor width of $1+\gamma$ at its output. Thus, the parasitic delay of both transistors is $= p_{inv} = 2.2$. (This is not surprising – the parasitic delay of pMOS manifests itself during rise times, while the nMOS parasitic delay is seen during fall times).
Thus we have,

$$g_{pMOS} = \frac{5}{7} = 0.714286,\ p_{pMOS} = 2.2$$

$$g_{nMOS} = \frac{2}{7} = 0.285714,\ p_{nMOS} = 2.2$$

– **[1]**

c) Using logical effort, find the input capacitance and geometries of all transistors in the circuit for optimum and equal delay through the upper and lower paths. Report your results in a table like the following:

...

Transistor widths should be rounded off to one decimal place.

Soln. 2-c) The NAND gate has 2 nMOS transistors in series, so the width of nMOS transistors is twice the width of the nMOS in the template inverter. Its pMOS transistors are in parallel, each with a width of γ. Therefore the input capacitance of the NAND gate is $2+\gamma$, which leads to

$$g_{NAND} = \frac{2+\gamma}{1+\gamma} = \frac{4.5}{3.5} = \frac{9}{7} = 1.28571$$

Its parasitic delay is proportional to the total transistor width connected to the output node. This corresponds to the capacitance of one nMOS transistor of width 2 and 2 pMOS transistors with widths of γ each connected to the output. The template inverter has width of $1+\gamma$ connected to the output. Therefore,

$$\frac{p_{NAND}}{p_{inv}} = \frac{2+\gamma+\gamma}{1+\gamma} \quad \text{So } p_{NAND} = \frac{2+\gamma+\gamma}{1+\gamma}\ \ p_{inv} = 2 \times p_{inv} = 2 \times 2.2 = 4.4$$

The NOR gate has 2 n channel transistor in parallel, so each has a size of 1. The 2 pMOS transistors are in series, so their widths are 2γ each. The input capacitance is therefore $1+2\gamma$. Thus,

$$g_{NOR} = \frac{1+2\gamma}{1+\gamma} = \frac{6}{3.5} = \frac{12}{7} = 1.71429$$

The parasitic delay of the NOR gate is proportional to transistor widths connected to its output – which is two nMOS transistors of width 1 each and a pMOS transistor of width $2\gamma = 5$. Therefore,

$$\frac{p_{NOR}}{p_{inv}} = \frac{1+1+2\gamma}{1+\gamma} \quad \text{So } p_{NOR} = \frac{2+2\gamma}{1+\gamma} p_{inv} = 2 \times p_{inv} = 2 \times 2.2 = 4.4$$

For the upper path (with NAND), $G_A = 1 \times \frac{9}{7} \times 1 \times 1 \times \frac{5}{7} = \frac{45}{49}$.
There is no branching within the path, so $B_A = 1$.
The combined input capacitance of the two paths should be 4. Let us assume that C_{in} values of the two paths are $4r$ and $4(1-r)$ respectively.
Then for the upper path, $H_A = 3500/4r$. This leads to $F_A = \frac{45}{49} \times 1 \times \frac{3500}{4r}$. The path has 5 stages, so

$$\hat{f_A} = (F_A)^{1/5} = \left(\frac{45 \times 3500}{49 \times 4r}\right)^{1/5}$$

The total delay of the upper path is therefore:

$$5\hat{f_A} + 3p_{inv} + 2p_{inv} + p_{inv} = 5\left(\frac{45 \times 3500}{49 \times 4r}\right)^{1/5} + 6p_{inv}$$

Similarly, for the lower path $G_B = 1 \times \frac{12}{7} \times 1 \times 1 \times \frac{2}{7} = \frac{24}{49}$.
$B_B = 1$ again and $H_B = 3500/4(1-r)$. This leads to

$$F_B = \frac{24}{49} \times 1 \times \frac{3500}{4(1-r)}$$

This path also has 5 stages so

$$\hat{f_B} = (F_B)^{1/5} = \left(\frac{24 \times 3500}{49 \times 4(1-r)}\right)^{1/5}$$

So the total delay of the lower path is

$$5\hat{f_B} + 3p_{inv} + 2p_{inv} + p_{inv} = 5\left(\frac{24 \times 3500}{49 \times 4(1-r)}\right)^{1/5} + 6p_{inv}$$

Equating the delays, we find that parasitic delays are equal and cancel. Thus we get

$$5\hat{f_A} = 5\hat{f_B}, \text{ Therefore, } \hat{f_A} = \hat{f_B}, \text{ and so, } F_A = F_B$$

This leads to

$$\frac{45}{49} \times \frac{3500}{4r} = \frac{24}{49} \times \frac{3500}{4(1-r)} \qquad \text{so} \quad \frac{45}{r} = \frac{24}{1-r}$$

$$\text{Thus } 69r = 45 \quad \text{and so} \quad r = \frac{45}{69} = 0.652174$$

The input capacitance of the upper path is therefore $4r = 2.6087$ while the input capacitance of the lower path is $4(1-r) = 1.3913$.

$$\hat{f_A} = \left(\frac{45 \times 3500}{49 \times 4r}\right)^{1/5} = 4.1508$$

$$\text{Similarly } \hat{f_B} = \left(\frac{24 \times 3500}{49 \times 4(1-0.652174)}\right)^{1/5} = 4.1508$$

The two values come out equal as expected. We shall refer to these simply as $\hat{f}$ from now on.

We can now proceed to determine the sizes of all gates and transistors.

Upper Path:
Inverter I_{1A} :
$C_{in} = 4r = 4 \times 0.652174 = 2.6087$, Scale factor $= C_{in}/g = 2.6087$
n width $= 2.6087$, p width $= \gamma \times 2.6087 = 2.5 \times 2.6087 = 6.5217$

$$\hat{f} = 4.1508 = gbh = 1 \times 1 \times \frac{C_{out}}{C_{in}} \quad \text{So, } C_{out} = 4.1508 \times 2.6087 = 10.8282$$

NAND
$C_{in} = 10.8282$, Scale factor $= C_{in}/g = 10.8282 \times 7/9 = 8.42192$
n width $2 \times 8.42192 = 16.8438$, p width $= \gamma \times 8.42192 = 2.5 \times 8.42192 = 21.0548$

$$\hat{f} = 4.1508 = gbh = \frac{9}{7} \times 1 \times \frac{C_{out}}{C_{in}} \quad \text{So, } C_{out} = 10.8282 \times 4.1508 \times \frac{7}{9} = 34.9577$$

Inverter I_{2A} :
$C_{in} = 34.9577$, Scale factor $= C_{in}/g = 34.9577$
n width $= 34.9577$, p width $= \gamma \times 34.9577 = 2.5 \times 34.9577 = 87.3144$

$$\hat{f} = 4.1508 = gbh = 1 \times 1 \times \frac{C_{out}}{C_{in}} \quad \text{So, } C_{out} = 4.1508 \times 34.9577 = 145.103$$

Inverter I_{3A} :
$C_{in} = 145.103$, Scale factor $= C_{in}/g = 145.103$
n width $= 145.103$, p width $= \gamma \times 145.103 = 2.5 \times 145.103 = 362.757$

$$\hat{f} = 4.1508 = gbh = 1 \times 1 \times \frac{C_{out}}{C_{in}} \quad \text{So, } C_{out} = 4.1508 \times 145.103 = 602.293$$

pMOS
$C_{in} = 602.293$, Scale factor $= C_{in}/g = 602.293 \times 7/5 = 843.210$
p width $= \gamma \times 843.210 = 2.5 \times 843.210 = 2108.03$

$$\hat{f} = 4.1508 = gbh = \frac{5}{7} \times 1 \times \frac{C_{out}}{C_{in}} \quad \text{So, } C_{out} = 602.293 \times 4.1508 \times \frac{7}{5} = 3500$$

This agrees with the specified final load.

Lower Path:
Inverter I_{1B} :
$C_{in} = 4(1-r) = 4 \times (1 - 0.652174) = 1.3913$, Scale factor $= C_{in}/g = 1.3913$
n width $= 1.3913$, p width $= \gamma \times 1.3913 = 2.5 \times 1.3913 = 3.47826$

$$\hat{f} = 4.1508 = gbh = 1 \times 1 \times \frac{C_{out}}{C_{in}} \quad \text{So, } C_{out} = 4.1508 \times 1.3913 = 5.77503$$

NOR
$C_{in} = 5.77503$, Scale factor $= C_{in}/g = 5.77503 \times 7/12 = 3.36877$
n width 3.36877, p width $= 2\gamma \times 3.36877 = 5 \times 3.36877 = 16.8438$

$$\hat{f} = 4.1508 = gbh = \frac{12}{7} \times 1 \times \frac{C_{out}}{C_{in}} \quad \text{So, } C_{out} = 5.77503 \times 4.1508 \times \frac{7}{12} = 13.9831$$

Inverter I_{2B} :
$C_{in} = 13.9831$, Scale factor $= C_{in}/g = 13.9831$
n width $= 13.9831$, p width $= \gamma \times 13.9831 = 2.5 \times 13.9831 = 34.9577$

$$\hat{f} = 4.1508 = gbh = 1 \times 1 \times \frac{C_{out}}{C_{in}} \quad \text{So, } C_{out} = 4.1508 \times 13.9831 = 58.0411$$

Inverter I_{3B} :
$C_{in} = 58.0411$, Scale factor $= C_{in}/g = 58.0411$
n width $= 58.0411$, p width $= \gamma \times 58.0411 = 2.5 \times 58.0411 = 145.103$

$$\hat{f} = 4.1508 = gbh = 1 \times 1 \times \frac{C_{out}}{C_{in}} \quad \text{So, } C_{out} = 4.1508 \times 58.0411 = 240.917$$

nMOS
$C_{in} = 240.917$, Scale factor $= C_{in}/g = 240.917 \times 7/2 = 843.210$
n width $= 843.210$

$$\hat{f} = 4.1508 = gbh = \frac{2}{7} \times 1 \times \frac{C_{out}}{C_{in}} \quad \text{So, } C_{out} = 240.917 \times 4.1508 \times \frac{7}{2} = 3500$$

This also agrees with the final load.
Notice that widths of the final pMOS driver and nMOS driver were independently computed through the upper and lower path. Still, the width of pMOS driver comes out to be $\gamma\times$ the width of nMOS driver ($= 2.5 \times 843.21 = 2108.03$).

The results can be summarized as:

Gate Label	Input capacitance	Logical effort	Parasitic delay	Scale Factor	nMOS width	pMOS width
I_{1A}	2.6087	1	2.2	2.6087	2.6	6.5
NAND	10.8282	1.28571	4.4	8.42192	16.8	21.1
I_{2A}	34.9577	1	2.2	34.9577	35.0	87.3
I_{3A}	145.103	1	2.2	145.103	145.1	362.8
pMOS	602.293	0.714286	2.2	843.210	-	2108.0
I_{1B}	1.3913	1	2.2	1.3913	1.4	3.5
NOR	5.77503	1.71429	4.4	3.36877	3.4	16.8
I_{2B}	13.9831	1	2.2	13.9831	14.0	35.0
I_{3B}	58.0411	1	2.2	58.0411	58.0	145.1
nMOS	240.917	0.285714	2.2	843.210	843.2	-

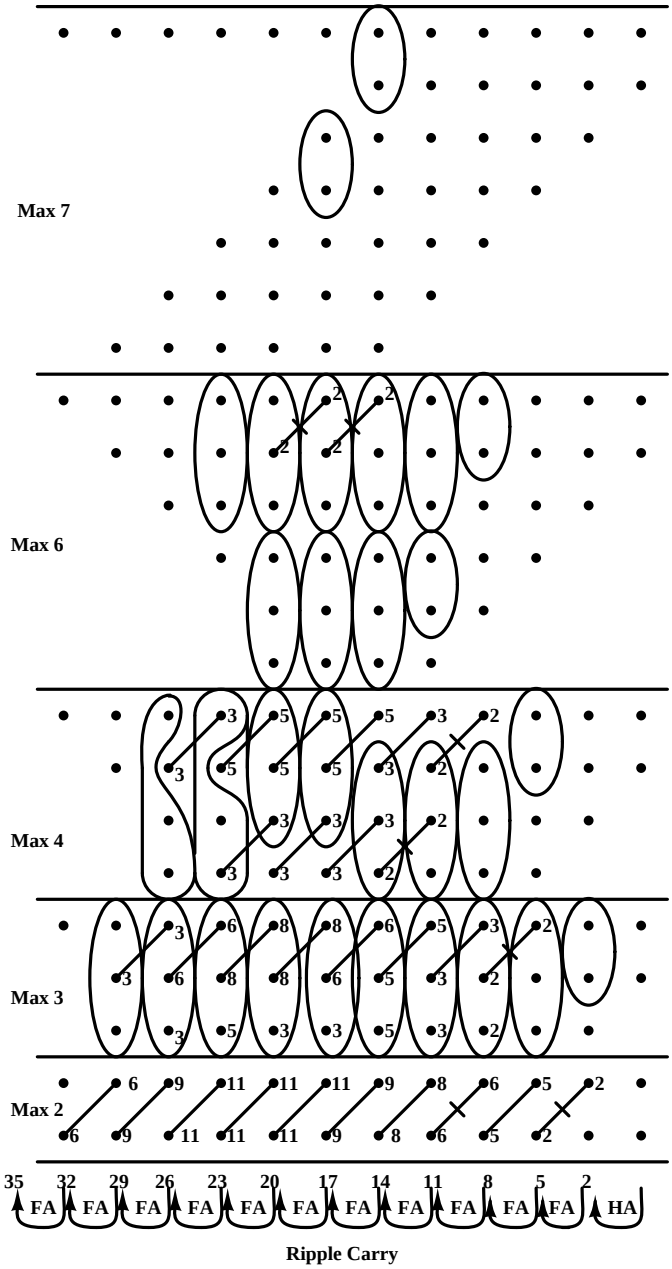
– **[10]**

– **[Q2: 1+1+10 = 12 marks]**

Q–3 Consider a Multiply and Accumulate (MAC) circuit using Dadda scheme for wire reduction which multiplies two 6 bit wide operands and adds a 12 bit operand to the product. The scheme is to be implemented using full adders which produce sum and carry outputs 3 units of time after the arrival of the latest input and half adders which produce sum and carry output 2 units of time after the arrival of the latest input.

a) Show the wire reduction scheme for the MAC circuit using a dot diagram.
Label each wire with the total delay accumulated by it as it passes through various adders in different stages. (You need not put a label on wires with zero delay). Wires fed to adders must be so selected that the total delay in the critical path for wire reduction is minimized.

Soln. 3-a) The dot diagram for the MAC circuit is given below. There is a maximum of 7 partial product bits in a column, so the reduction proceeds to 6, 4, 3 and 2 wires in 4 stages. The dots are labelled with the times when their data will be valid.



The latest dots to become valid at the end of wire reduction are at bits 6 to 9 at 11 units of time. All of these pass through 3 full adders and a half adder.

To optimize the critical path, we provide the fastest option to wires with highest delay. However, for every adder, we note the latest input bit included and if possible, consume other wires with the same delay in this adder because that will not change the time at which the output of this adder becomes valid. – **[8]**

b) What will be the time taken to complete the multiplication, inclusive of addition of the final two wires at each weight using a ripple carry adder made of the same full and half adders as the ones used for wire reduction.

Soln. 3-b) As can be seen from the dot diagram, the final addition will be complete at 35 units of time. This time is dominated by delays in the ripple carry adder. It is interesting to note that for every bit position, data on reduced wires becomes valid simultaneously or *before* the carry ripples to this bit. Therefore the critical path is through the ripple carry adder whose 11 full adders and a half adder account for 35 units of time.

(The following observations are not required in the answer to the question:)
This computation shows that a ripple carry adder is a poor choice for this multiplier. In fact the multiply and add operation takes no longer than the ripple carry addition of 12 bits with one half adder and 11 full adders.

For the 3 least significant bits, data on the reduced wires arrives at the same time as the rippling carry. Till this time, the ripple carry architecture does not incur any additional delay. For the remaining bits, data on reduced wires is ready much ahead of the ripple carry. Therefore, it is possible to break up the final adder into a 3 bit ripple carry adder and a faster adder architecture for the remaining bits.

An interesting follow up exercise is to use a 3 bit ripple carry adder for bits 0 to 2, an 8 bit logarithmic adder for bits 3 to 10 and a 1 bit ripple carry adder for the most significant bit. This is because logarithmic adders are most efficient when adding no of bits which are powers of 2. In any case, the final carry of a logarithmic adder arrives earlier than the intermediate carries – so the MSbit ripple carry addition will take place in parallel with the generation of middle bit carries and sums in the logarithmic adder.

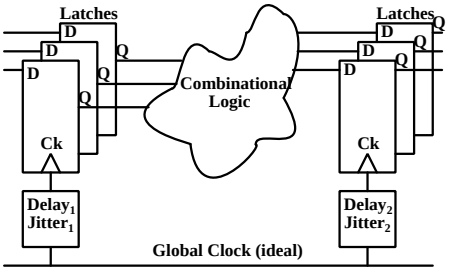
– **[2]**

– **[Q3: 8+2 = 10 marks]**

Q–4 A synchronous digital circuit uses a clock with period T.

a) What are the timing constraints due to clock skew, jitter, Clock to Q delays of the flipflops, their set up and hold times and the delay through the combinational logic between the flipflops?

Soln. 4-a) The figure below shows a typical stage in synchronous design with flipflops alternating with combinational logic. Two extreme cases needs to be considered to avoid malfunction:



1. Clock to the input set of flipflops is late, while the clock to the output set of flipflops is early - reducing the time available for processing data to a minimum and risking setup time violations.

2. Clock to the input set of flipflops is early, while the clock to the output set of flipflops is late, risking hold time violations.

Let the delay in the clock distribution tree to the input location be Δ_1, while that to the output location be Δ_2. The worst case value of $|\Delta_1 - \Delta_2|$ is the clock skew. Additionally, the arrival time of the clock at the same location fluctuates from cycle to cycle. Worst case value of this time dependent variation of the arrival time of the clock edge is called jitter. Jitter can be positive or negative.

Input Clock Late, Output Clock Early: Let the ideal clock (say at the clock source) tick at time 0. Then the latest time for clock arrival at the input flipflops is $\Delta_{1max} + J$ where J is the peak value of jitter.

Latest time for applying inputs to the combinational logic is: $\Delta_{1max}+J+\text{CktoQ}_{max}$, so the latest time at which data will reach D inputs of the output flipflops is $\Delta_{1max} + J + \text{CktoQ}_{max} + \text{LogicDelay}_{max}$.

This should be a setup time earlier than the arrival of the next clock at the output flipflops. Earliest time for clock arrival of the next clock at the output point is:

$T + \Delta_{2min} - J$. Thus we have the timing constraint:

$$\Delta_{1max} + J + \text{CktoQ}_{max} + \text{LogicDelay}_{max} \leq T + \Delta_{2min} - J - t_{setup}$$

This leads to

$$\begin{aligned} T &\geq \Delta_{1max} - \Delta_{2min} + 2J + \text{CktoQ}_{max} + \text{LogicDelay}_{max} + t_{setup} \\ \text{Or } T &\geq \text{skew}_{max} + \text{Jitter}_{peek-to-peek} + \text{CktoQ}_{max} + \text{LogicDelay}_{max} + t_{setup} \end{aligned}$$

Input Clock Early, Output Clock Late: In this case we are concerned that the data may change too early within the same clock period at the output flipflops and violate the hold condition.

Earliest arrival time for the clock at the input flipflops is: $\Delta_{1min} - J$. So the data gets applied to the combinational logic at $\Delta_{1min} - J + \text{CktoQ}_{min}$. Thus the earliest data change at the D inputs of flipflops will take place at $\Delta_{1min} - J + \text{CktoQ}_{min} + \text{LogicDelay}_{min}$.
This should occur a hold time later than the most delayed arrival of the current clock edge at the output flipflops. So

$$\Delta_{1min} - J + \text{CktoQ}_{min} + \text{LogicDelay}_{min} \geq 0 + \Delta_{2max} + J + t_{hold}$$

$$\text{Or} \qquad \text{LogicDelay}_{min} \geq \Delta_{2max} - \Delta_{1min} + 2J - \text{CktoQ}_{min} + t_{hold}$$

$$\text{Thus} \qquad \text{LogicDelay}_{min} \geq \text{skew}_{max} + \text{Jitter}_{peek-to-peek} - \text{CktoQ}_{min} + t_{hold}$$

– **[2]**

b) The worst case clock skew in the clock distribution circuit is 60 ps and the worst case peak to peak jitter is 25 ps. The circuit uses flipflops with set up and hold times of 30 ps, and with clock to Q delays of 40 ps. The worst case delay through the combinational logic is 180 ps.
What is the highest clock frequency at which we can operate this circuit? Is there a minimum limit on the delay through the combinational circuit in order to meet the hold specification? If so, how much is it?

Soln. 4-b) We have

$$T \geq \text{skew}_{max} + \text{Jitter}_{peek-to-peek} + \text{CktoQ}_{max} + \text{LogicDelay}_{max} + t_{setup}$$

Therefore the clock period should be greater than $60 + 25 + 40 + 180 + 30 = 335$ps. This corresponds to a clock frequency of 2.985 GHz or just under 3 GHz.

Additionally, we must have

$$\text{LogicDelay}_{min} \geq \text{skew}_{max} + \text{Jitter}_{peek-to-peek} - \text{CktoQ}_{min} + t_{hold}$$

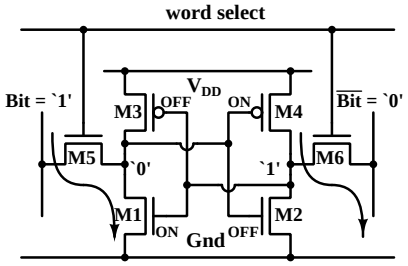
Therefore, we must have

$$\text{LogicDelay}_{min} \geq 60 + 25 - 40 + 30 = 75\text{ps}$$

Since this limit is > 0, we do have a minimum limit on Logic delay.
We must have LogicDelay $\geq$ 75ps. – **[2]**

– **[Q4: 2+2 = 4 marks]**

Q–5 Consider the 6 transistor static RAM shown below. Take the case when a '0' is stored in the cell (M1, M4 are ON, while M2, M3 are OFF).

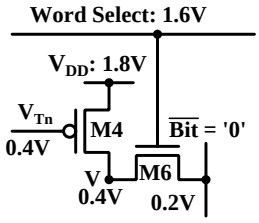


In order to write a '1' to this cell, the Bit line will be taken to logic '1', while the $\overline{\text{Bit}}$ line will be driven to logic '0', with the word select line at logic '1'. We want to compute the ratio of (W/L) values for M6 and M4, such that the voltage at the junction of M6 and M4 is brought down sufficiently to write a '1' to the cell reliably.

We make the simplifying assumption that M2 and M3 remain OFF during the whole operation and to prevent read upset, geometries of M1 and M5 have been so chosen that the voltage at the drain of M1 rises no higher than V_{Tn} when the Bit and Word Select lines are at '1'. Thus the gates of M4 and M2 are assumed to be at $V_{Tn} = 0.4$V.
The supply voltage V_{DD} is 1.8V, $V_{Tn} = |V_{Tp}| = 0.4$V.
The $\overline{\text{Bit}}$ line is at 0.2 V, the word select line is at 1.6V and the mobility ratio of n and p channel transistors is 2.4.

a) Find the ratio of (W/L) values for M4 and M6 such that the voltage at the junction of the two transistors is brought down to 0.4V.

Soln. 5-a) Because of the simplifying assumptions, the circuit to be analysed reduces to the one shown below.



For pMOS:

$$V_{GS} = 1.8 - 0.4 = 1.4\text{V}, \quad V_{DS} = 1.8 - 0.4 = 1.4\text{V}$$

For nMOS:

$$V_{GS} = 1.6 - 0.2 = 1.4\text{V}, \quad V_{DS} = 0.4 - 0.2 = 0.2\text{V}$$

For pMOS, $V_{DS} = V_{GS}$, so it is in saturation while the nMOS transistor is in linear mode.
Equating currents through the pMOS and nMOS transistors, we get

$$\frac{K_p}{2}(V_{GSp} - V_{Tp})^2 = K_n\left((V_{GSn} - V_{Tn})V_{DSn} - \frac{1}{2}V_{DSn}^2\right)$$

$$\text{Which gives:} \quad \frac{K_p}{2}(1.4 - 0.4)^2 = K_n(1.4 - 0.4)0.2 - \frac{1}{2}0.2^2$$

$$1 = \frac{2K_n}{K_p}(0.2 - 0.02) \qquad \text{So} \quad 1 = 0.36\frac{\mu_n}{\mu_p}\frac{(W/L)_n}{(W/L)_p} = 0.36 \times 2.4\frac{(W/L)_n}{(W/L)_p}$$

From this, we get

$$\frac{(W/L)_n}{(W/L)_p} = \frac{1}{0.36 \times 2.4} = 1.1574$$

– **[2]**

b) If we choose the minimum size for M6 as well as M4, what is the voltage to which the junction of M4 and M6 will be pulled down?

Soln. 5-b) If the n and p channel transistors have the same geometry and the voltage at the junction of M4 and M6 is V, we have:

$$V_{GSp} = 1.8 - 0.4 = 1.4\text{V}, \quad V_{DSp} = 1.8 - V, \quad V_{GSn} = 1.6 - 0.2 = 1.4\text{V}, \quad V_{DSn} = V - 0.2$$

Since the geometries are close to those computed in the part above, we expect V to be such that pMOS is saturated while nMOS is in linear mode. (This needs to be confirmed after calculating V). Equating currents, we get

$$\frac{K_p}{2}(1.4-0.4)^2 = K_n\left((1.4-0.4)(V-0.2)-\frac{1}{2}(V-0.2)^2\right)$$

Since geometries are the same, $2K_n/K_p = 2\mu_n/\mu_p = 2\times 2.4 = 4.8$. Let $x \equiv (V-0.2)$. Then

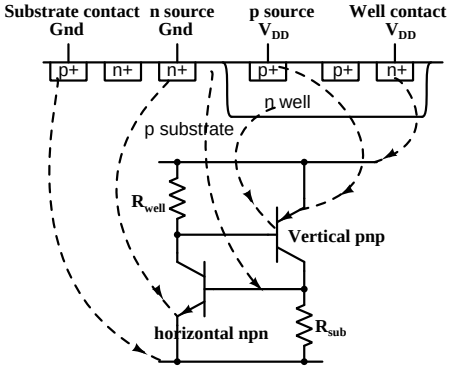
$$1 = 4.8(x-\frac{1}{2}x^2) \qquad \text{So} \quad 2.4x^2 - 4.8x + 1 = 0$$

$$2.4x^2 - 4.8x + 1 = 0 \quad \text{gives} \quad x = \frac{4.8 \pm \sqrt{4.8^2 - 4\times 2.4}}{4.8} = 0.2362\text{V}$$

Correspondingly, $V = x + 0.2 = 0.4362$V – **[3]**

– **[Q5: 2+3 = 5 marks]**

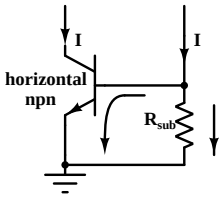
Q–6 The equivalent circuit of the latchup structure is given below:



Each transistor multiplies its base current by its β. This amplified current is then divided between the base current of the other transistor and the shunt resistor. Now the other transistor multiplies its base current by its β and feeds back this collector current to be divided between the shunt resistor and the base current of the first transistor.

Current drawn by the circuit increases uncontrollably when the closed loop current gain of this feedback circuit exceeds 1.

Consider only the npn transistor. To prevent latchup, we want to set the condition that the collector current of the npn transistor is $\leq$ the incoming collector current from the pnp transistor.



Assume that the intrinsic current gain of the parasitic bipolar transistors can be restricted to be ≤ 5.

The current equation of the base emitter junction is given by

$$I_{BE} = I_0(e^{qV/KT}-1) \quad \text{with } I_0 = 6\times 10^{-13}\text{A and} \quad KT/q = 26\text{mV}$$

a) Find the value of the shunt resistor, such that the product of current division ratio and current gain remains less than 1 for entering trigger current I of ≤ 1 mA.

Soln. 6) For the highest value of injected current (= 1 mA), the current through the diode should be no more than 1/5 mA. for this much current, the voltage drop across the diode is

$$\frac{KT}{q}\ln\left(\frac{I/5}{I_0}+1\right) = .026\ln\left(\frac{2.0\times 10^{-4}}{6\times 10^{-13}}+1\right) = 0.5102\text{V}$$

The resistor has this voltage across it and the current through it is 4/5 mA. So its resistance should be: $0.5102/(8\times 10^{-4}) = 637.8\ \Omega$. – **[2]**

b) How do we ensure that the value of the shunt resistor is low enough?

Soln. 6-b) In the part above, we have seen that latchup is prevented for trigger currents of up to 1 mA if the shunt resistor is $\approx 640\Omega$ or less. The shunt resistance is the well resistance to V_{DD} for pnp and the substrate resistance to ground for npn transistors.

To keep this value small, design rules require frequent metallic contacts to V_{DD} for the n well and to ground for the p well. Frequent metal contacts reduce the length of resistive path and keep the shunt resistance small. – **[1]**

– **[Q6: 2+1 = 3 marks]**

Paper Ends

Reference:

Transistor Model: You can use the following transistor model where required in this paper:

$K \equiv \mu C_{ox}(W/L)$ For $V_{GS} \leq V_T,\ I_{DS} = 0$

$$\text{For } V_{GS} \geq V_T \text{ and } V_{DS} \leq V_{GS} - V_T, \quad I_{DS} = K\left((V_{GS} - V_T)V_{DS} - \frac{1}{2}V_{DS}^2\right)$$

$$\text{For } V_{GS} \geq V_T \text{ and } V_{DS} \geq V_{GS} - V_T, \quad I_{DS} = \frac{K}{2}(V_{GS} - V_T)^2$$

Logical Effort

Logical effort for a gate is the ratio of its input capacitance and that of an inverter providing the same output drive.

To find the parasitic delay of a logic gate, we size it so that it gives the same output drive as the template inverter. Now the parasitic delay is **proportional to** total transistor width directly connected to the output node. Because of the proportionality, parasitic delay is often computed in relation to the template inverter.

Logical effort and parasitic delay are size independent.

For optimum delay, $\hat{f} = gbh$ should be the same for all stages. Here g is the logical effort, b is the branching factor, and h is the stage electrical effort.

$$b = \frac{\text{Total capacitance at the output}}{\text{on path capacitance at the output}} \qquad h = \frac{\text{on path capacitance at the output}}{\text{input capacitance}}$$

Newton Raphson iterative solution for a non-linear equation

To find the solution to a non-linear equation $f(x) = 0$ iteratively, we start with a guess solution g. Define $f'(x) \equiv \frac{\mathrm{d}f(x)}{\mathrm{d}x}$. A closer guess value is then given by

$$g_{next} = g - \frac{f(x)}{f'(x)} \quad \text{where } f(x) \text{ and } f'(x) \text{ are evaluated at } x = g.$$

We iterate over this computation till $f(x)$ is close enough to 0.