INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY
ELECTRICAL ENGINEERING DEPARTMENT

End Semester Examination

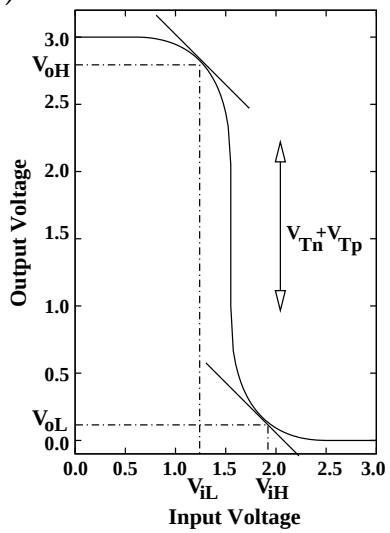
Tuesday	EE 671: VLSI Design	Time: 0930-1230
21-11-17	Autumn Semester 2017	Marks: 50

For all problems in this paper, use the simple transistor model with perfect saturation given at the end of the paper. Unless otherwise specified, take $V_{DD} = 1.8V$, $|V_{Tp}| = 0.45$V and $V_{Tn} = 0.35$V.

Quantitative answers must be accurate at least to 1%.
All intermediate calculations should be reported.

Q–1 **a)** The 'Low' and 'High' logic levels at the input and output are defined by the points on the transfer curve of an inverter where the slope is -1. Justify this choice.
How are static noise margins determined from the 'Low' and 'High' logic levels as defined above?

Soln. 1-a)



The transfer characteristics of a CMOS inverter are shown in the diagram on the left. For all points to the left of the first point and to the right of the second point where the slope $= -1$, the analog gain $\frac{\mathrm{d}V_{out}}{\mathrm{d}V_{in}}$ has an absolute value less than 1. Therefore, any noise or deviation at the input will be diminished at the output. On the other hand, for the region between these two, the gain is > 1. An input in these regions will be amplified at the output. For digital use, we want the output to be insensitive to the actual value of the input voltage, as long as it is within the specified range for the definition of a '0' or a '1'.

Thus points with a slope $= -1$ form natural boundaries for defining the highest voltage for '0' and lowest voltage for '1'. The input and output voltages at the defining points are termed (V_{iL}, V_{oH}) and (V_{iH}, V_{ol}). V_{ol} should be $< V_{iL}$, so that even if the output has some noise added to it, the next stage still treats it as a '0' at its input. Similarly, V_{oH} should be $> V_{iH}$, so that even in the presence of noise, a '1' at the output of one stage is still interpreted as a '1' at the input of the next. Therefore we define the static noise level at '0' as $V_{iL} - V_{ol}$ and the static noise level at '1' as $V_{oH} - V_{iH}$. Both should be positive and as large as possible for good noise immunity. – **[2]**

b) Consider a CMOS inverter. Assume that the geometries of the p and n channel transistors are so chosen that their conductance factors K_p and K_n are equal. Derive an expression for the input 'Low' and output 'High' values in terms of the supply voltage and absolute values of the turn on voltages.

Soln. 1-b) When the input is 'Low' and output is 'High', the n channel transistor will be in saturation and the p channel transistor will be in linear mode of operation. The absolute value of the gate to source voltage for the p channel transistor is

$V_{DD} - V_i$, while the absolute value of its drain to source voltage is $V_{DD} - V_o$ The current through the two transistor should be equal. Therefore,

$$\frac{K_n}{2}(V_i - V_{Tn})^2 = K_p \left((V_{DD} - V_i - V_{Tp})(V_{DD} - V_o) - \frac{1}{2}(V_{DD} - V_o)^2 \right)$$

We define $V_{DP} \equiv V_{DD} - V_o$. Since $K_n = K_p$, we can write,

$$\frac{1}{2}(V_i - V_{Tn})^2 = (V_{DD} - V_i - V_{Tp})V_{DP} - \frac{1}{2}V_{DP}^2$$

Which leads to the quadratic equation

$$\frac{1}{2}V_{DP}^2 - (V_{DD} - V_i - V_{Tp})V_{DP} + \frac{1}{2}(V_i - V_{Tn})^2 = 0$$

with solutions:

$$V_{DP} = (V_{DD} - V_i - V_{Tp}) \pm \sqrt{(V_{DD} - V_i - V_{Tp})^2 - (V_i - V_{Tn})^2}$$

Since the p channel transistor is in the linear regime, we must have $V_{DP} \equiv V_{DD} - V_o < V_{DD} - V_i - V_{Tp}$ and therefore, we must choose the negative sign in the equation. So

$$V_{DP} \equiv V_{DD} - V_o = (V_{DD} - V_i - V_{Tp}) - \sqrt{(V_{DD} - V_i - V_{Tp})^2 - (V_i - V_{Tn})^2}$$

$$\text{Therefore} \quad V_o = V_i + V_{Tp} + \sqrt{(V_{DD} - V_i - V_{Tp})^2 - (V_i - V_{Tn})^2}$$

$$\text{So} \quad V_o = V_i + V_{Tp} + \sqrt{(V_{DD} - V_{Tn} - V_{Tp})(V_{DD} - 2V_i + V_{Tn} - V_{Tp})}$$

Taking the derivative with respect to V_i and putting it equal to -1, we get

$$-1 = 1 - \sqrt{\frac{V_{DD} - V_{Tn} - V_{Tp}}{V_{DD} - 2V_i + V_{Tn} - V_{Tp})}}$$

$$\text{squaring, we get } 4 = \frac{V_{DD} - V_{Tn} - V_{Tp}}{V_{DD} - 2V_i + V_{Tn} - V_{Tp}}$$

$$\text{Therefore } 4V_{DD} - 8V_i + 4V_{Tn} - 4V_{Tp} = V_{DD} - V_{Tn} - V_{Tp}$$

Which gives

$$V_{iL} = \frac{3V_{DD} + 5V_{Tn} - 3V_{Tp}}{8}$$

Substituting for V_i in the equation

$$V_o = V_i + V_{Tp} + \sqrt{(V_{DD} - V_{Tn} - V_{Tp})(V_{DD} - 2V_i + V_{Tn} - V_{Tp})}$$

we get

$$V_{oH} = \frac{7V_{DD} + V_{Tn} + VTp}{8} = V_{DD} - \frac{V_{DD} - V_{Tn} - VTp}{8}$$

These are the values of the input Low and output High voltages. – **[4]**

c) For a CMOS inverter, $K_p = K_n = 100\mu\text{A/V}^2$. For what input voltage is the static current drawn by the inverter maximum? What is the value of this maximum static current?

Soln. 1-c) Both transistor will be on and in saturation for current to be maximum. When both are in saturation, their saturation currents must be equal. So

$$\frac{K_n}{2}(V_i - V_{Tn})^2 = \frac{K_p}{2}(V_{DD} - V_i - V_{Tp})^2$$

For $K_n = K_p$ this gives

$$V_i - V_{Tn} = V_{DD} - V_i - V_{Tp}, \quad \text{and therefore,} \quad V_i = \frac{V_{DD} - V_{Tp} + V_{Tn}}{2}$$

For the given constants,

$$V_i = \frac{1.8 - 0.45 + 0.35}{2} = 0.85\text{V}$$

At this voltage, the saturation current of the nMOS transistor is:

$$I_{max} = \frac{100 \times 10^{-6}}{2}(0.85 - 0.35)^2 = 50 \times 10^{-6}(0.25) = 12.5\mu\text{A}$$

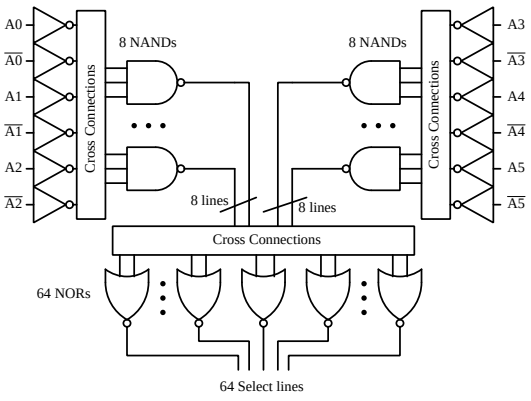
– **[2]**

d) Assume that the widths of the p channel transistors need to be twice the width of n channel transistors for matching their conductance factors K_p and K_n. The minimum width of an n channel transistor is 250 nm. Using thumb rules for scaling geometries, find the widths of the four transistors in a tri-stateable inverter so that it meets the same dynamic specifications as the minimum inverter with equal K_p and K_n values.

Soln. 1-d) The tri-stateable inverter has 2 n channel transistors in series and the 2 p channel transistors are also in series. In the inverter, the p channel transistor is twice as wide as the n channel transistor. Therefore in the tri-stateable inverter, the n width will be twice the n channel transistor width in the minimum inverter = 500 nm. The p channel width will be double this, = 1μm. – **[1]**

– **[Q1: 2+4+2+1 = 9 marks]**

Q–2



A two step decoder for 6 address lines, producing 64 select outputs is shown in the diagram on the left. Combinations of 3 lines out of $A0, \overline{A0}, A1, \overline{A1}, A2, \overline{A2}$ are fed to the first bank of 8 three-input NAND gates. Similarly, combinations of 3 lines out of $A3, \overline{A3}, A4, \overline{A4}, A5, \overline{A5}$ are fed to the other bank of 8 three-input NAND gates. 64 two-input NOR gates then accept one line each from the two banks of NANDs to produce 64 select lines.

The select lines are required to drive heavy loads equivalent to 512 minimum inverters each. Assume that the γ value representing the ratio of p channel widths to n channel widths in an inverter to produce equal rise and fall times is 2, and the parasitic inverter delay is 2.5.

– **P.T.O.**

a) What is the the parasitic delay and the logical effort of 3 input NAND gates and 2 input NOR gates? (Parasitic delay can be taken to be proportional to the total capacitive load at the output node for a gate providing equivalent drive to a minimum inverter). Report the results in a table.

Soln. 2-a) The three input NAND has 3 n channel transistors in series, so each will have a width of 3. Since $\gamma = 2$, the 3 p channel transistors which are in parallel, will have widths of 2. The two input NOR will have two n channel transistors in parallel, so each will have a geometry of 1. There are two p channel transistors in series and $\gamma = 2$, so each will have a width of 4. Geometries and consequently the logical effort and parasitic delay of the components in use will be as follows:

Gate	Width of n Trans.	Width of p Trans.	Total W per input	Total W at output	Logical Effort	Parasitic Delay
Inverter	1	2	3	3	1	2.5
3-input-NAND	3	2	5	9	5/3	7.5
2-input-NOR	1	4	5	6	5/3	5.0

– **[1]**

b) What is the optimum number of stages for minimum delay in this circuit, assuming that each of the input lines can drive 1 minimum sized inverter. Show the recommended configuration for the two step decoder with added inverters if necessary, so that overall delay is minimized.

Soln. 2-b) The input can drive one inverter while the output is required to drive 512 inverters, so $H = 512$.

The logical effort of NANDs as well as NORs is 5/3. Therefore $G = 5/3 \times 5/3 = 25.9$.

There are 6 input inverters which drive 8 three-input NANDs *i.e.* 24 inputs. Therefore the branching factor at the output at the first inverter is $24/6 = 4$. A total of 16 NANDs drive 64 two-input NORs. Therefore the branching factor at the output of NANDs is $64 \times 2/16 = 8$. Thus the overall branching factor is $4 \times 8 = 32$. Thus the overall path effort is $F = GBH = 512 \times 32 \times 25/9 = 45511.11$

We can find the asymptotic stage ratio ρ by solving the equation

$$p_{inv} + \rho(1 - \ln \rho) = 0 \quad \text{with } p_{inv} = 2.5$$

We define $f \equiv 2.5 + \rho(1 - \ln \rho)$. Then

$$f' = (1 - \ln \rho) + \rho\left(-\frac{1}{\rho}\right) = 1 - \ln \rho - 1 = -\ln \rho$$

Then starting with a guess value g for ρ, the next guess is given by Newton Raphson technique as

$$g - \frac{f}{f'} = g + \frac{2.5 + g(1 - \ln g)}{\ln g} = g + \frac{2.5 + g}{\ln g} - g = \frac{2.5 + g}{\ln g}$$

Starting with a guess value of 4, we iterate to get values for ρ as: 4.6888, 4.6524, 4.6523, 4.6523 ...
Thus, $\rho = 4.6523$. The optimum number of stages is then

$$N = \frac{\ln F}{\ln \rho} = \frac{\ln 45511.11}{\ln 4.6523} = 6.9767$$

Therefore the optimum number of stages is 7.

Correspondingly, the actual stage ratio for 7 stages is

$$\hat{f} = F^{1/N} = 45511.11^{1/7} = 4.6286$$

The circuit as given has three stages – an inverter followed by a NAND, followed by a NOR. Therefore four inverters should be added to optimize delay.

It is clear that the sizes will be scaled up as we go from the input side to the final load. Therefore, it is preferable to have larger inverters rather than larger NANDS or NORS. This implies that we should try to put all four additional inverters at the end.

To check if this will run into problems, we start from the beginning.

Since the inputs can drive a single minimal inverter, the first inverter should have unit size. The stage effort of all stages should be $\hat{f} = 4.6286$. Therefore

$$gbh = 1 \times 4 \times \frac{C_{out}}{1} = 4.6286, \text{ So } C_{out} = 4.6286/4 = 1.1571$$

C_{out} of the inverter stage is C_{in} of the NAND stage. Each inverter represents transistor widths of 3 units. Therefore the NANDs can have a total transistor width of $3 \times 1.1571 = 3.4714$ at the input. This total width will be divided in the ratio 3:2 over the n and p channel input transistors. Thus the n channel transistor width will be $3 \times 3.4714/5 = 2.0828$, and the p channel transistor width will be $2 \times 3.4714/5 = 1.3886$. These widths are acceptable.

For the second stage, the NAND drives 8 NOR gates, and has a logical effort of 5/3. Thus

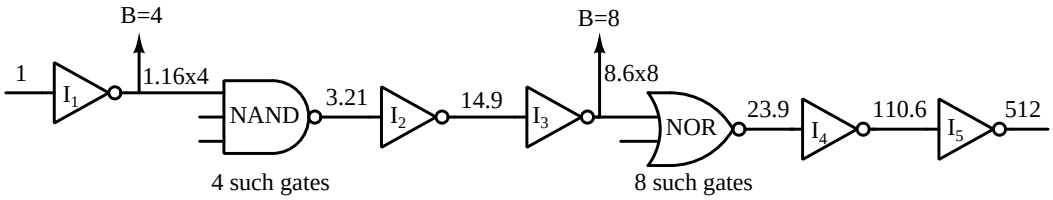
$$gbh = \frac{5}{3} \times 8 \times \frac{C_{out}}{1.1571} = 4.6286, \quad \text{So } C_{out} = \frac{4.6286 \times 3 \times 1.1571}{5 \times 8} = 0.4017$$

If we put NOR gates immediately following the NAND gates, the total input capacitance of each gate should be 0.4017 inverter units – or $3 \times 0.4017 = 1.2051$ transistor widths. This is too small to fit a NOR gate, which requires 5 transistor widths even while using minimum sized nMOS transistors for pull down.

Therefore we need to insert two inverters between the NAND and the NOR and do the 8 way branching to NORs *after* the inverters. (A single inverter will require that we change the NOR gate to AND – which will change the overall logical effort.).
So the final configuration should be – inverter, NAND, inverter, inverter, NOR, inverter, inverter for the seven stages. – **[3]** .

c) Compute the scale factors and absolute geometries for all transistors in the design.

Soln. 2-c)



Since the first inverter I_1 is to be driven by the inputs which can drive only one minimum inverter, its size should be minimum. This inverter drives 4 NAND gates.

$$gbh = 1 \times 4 \times \frac{C_{out}}{1} = 4.6286, \quad \text{So } C_{out} = \frac{4.6286}{4} = 1.1571$$

Therefore the input capacitance of the NAND gates should be equivalent to 1.1571 inverters.

The NAND gate is now followed by an inverter (I_2) without any branching.

$$gbh = \frac{5}{3} \times 1 \times \frac{C_{out}}{1.1571} = 4.6286, \quad \text{So } C_{out} = \frac{4.6286 \times 3 \times 1.1571}{5} = 3.2135$$

Therefore the input capacitance of I_2 should be 3.2135 units.

I_2 is followed by another inverter (I_3) without any branching at the output of I_2.

$$gbh = 1 \times 1 \times \frac{C_{out}}{3.2135} = 4.6286, \quad \text{So } C_{out} = 4.6286 \times 3.2135 = 14.8740$$

I_3 is required to drive 8 NORs. Then,

$$gbh = 1 \times 8 \times \frac{C_{out}}{14.8740} = 4.6286, \quad \text{So } C_{out} = \frac{4.6286 \times 14.8740}{8} = 8.6056$$

Thus the input capacitance of NORs is required to be 8.6056 inverter units. The NOR gate drives a single inverter (I_4).

$$gbh = \frac{5}{3} \times 1 \times \frac{C_{out}}{8.6056} = 4.6286, \quad \text{So } C_{out} = \frac{4.6286 \times 8.6056 \times 3}{5} = 23.899$$

So the input capacitance of I_4 is 23.899. This inverter drives the final inverter (I_5) without branching.

$$gbh = 1 \times 1 \times \frac{C_{out}}{23.899} = 4.6286, \quad \text{So } C_{out} = 4.6286 \times 23.899 = 110.62$$

Therefor the final inverter has input capacitance of 110.62. I_5 drives the final load.

$$gbh = 1 \times 1 \times \frac{C_{out}}{110.62} = 4.6286, \quad \text{So } C_{out} = 4.6286 \times 110.62 = 512$$

Which confirms that the final inverter will drive a load of 512.

Alternative Computation

We could have started from the output end, once the configuration has been decided. In this case we begin with C_{out} values and evaluate C_{in} for each stage.
I_5 drives 512 inverters, so

$$\hat{f} = 4.6286 = 1 \times 1 \times \frac{512}{C_{in}}, \quad \text{So } C_{in} = \frac{512}{4.6286} = 110.62$$

I_4 is required to drive a load equivalent to 110.62 min. inverters, so

$$\hat{f} = 4.6286 = 1 \times 1 \times \frac{110.62}{C_{in}}, \quad \text{So } C_{in} = \frac{110.62}{4.6286} = 23.899$$

NOR Gate drives a load equivalent to 23.899 min. inverters. Therefore

$$\hat{f} = 4.6286 = \frac{5}{3} \times 1 \times \frac{23.899}{C_{in}}, \quad \text{So } C_{in} = \frac{23.899 \times 5}{3 \times 4.6286} = 8.6056$$

I_3 drives 8 NOR gates. Therefore,

$$\hat{f} = 4.6286 = 1 \times 8 \times \frac{8.6056}{C_{in}}, \quad \text{So } C_{in} = \frac{8 \times 8.6056}{4.6286} = 14.8740$$

I_2 just drives I_3. Therefore

$$\hat{f} = 4.6286 = 1 \times 1 \times \frac{14.8740}{C_{in}}, \quad \text{So } C_{in} = \frac{14.8740}{4.6286} = 3.2135$$

NAND drives I_2.

$$\hat{f} = 4.6286 = \frac{5}{3} \times 1 \times \frac{3.2135}{C_{in}}, \quad \text{So } C_{in} = \frac{3.2135 \times 5}{3 \times 4.6286} = 1.1571$$

I_1 drives 4 NANDs.

$$\hat{f} = 4.6286 = 1 \times 4 \times \frac{1.1571}{C_{in}}, \quad \text{So } C_{in} = \frac{4 \times 1.1571}{4.6286} = 1.0$$

So the input capacitance of $I_1 = 1$ as expected.

Calculation of geometries:

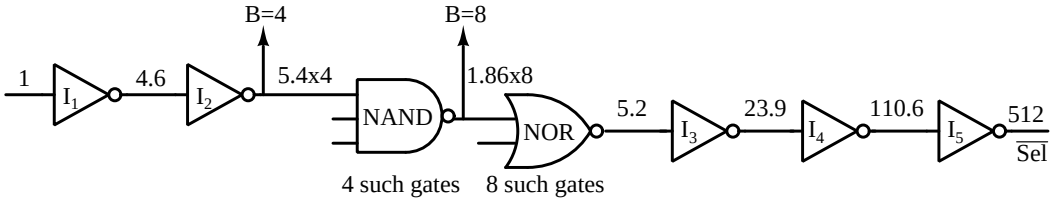
One inverter load is equivalent to 3 minimum transistor widths. We can work out transistor widths from input capacitances as follows:
I_1: $C_{in} = 1$ inv. $= 3$ transistor widths. This is a minimum inverter with n width $= 1$ and p width $=2$.
NAND: $C_{in} = 1.1571$ inv. $= 1.1571 \times 3 = 3.4713$ transistor widths. Therefore n width $= 2.08$, p width $= 1.39$.
I_2: $C_{in} = 3.2135$ inv. So n width $= 3.2135$, p width $= 6.427$
I_3: $C_{in} = 14.874$ inv. So n width $= 14.874$, p width $= 29.748$
NOR: $C_{in} = 8.6056$ inv. $= 25.82$ widths. Therefore n width $= 25.82 \times 1/5 = 5.163$ and p width $= 20.653$
I_4: $C_{in} = 23.899$, so n width $= 23.9$ and p width $= 47.8$
I_5: $C_{in} = 110.62$, so n width $= 110.62$, p width $= 221.24$.

These results are tabulated below:

	I_1	NAND	I_2	I_3	NOR	I_4	I_5
C_{in}	1	1.157	3.21	14.87	8.61	23.9	110.6
n width	1	2.08	3.21	14.87	5.16	23.9	110.6
p width	2	1.39	6.42	29.75	20.65	47.8	221.2

Alternative Configuration

To avoid the problem with heavy branching between the NAND and NOR stages, one can add another inverter *before* the NAND gates. This will make the NAND gate larger and enable it to drive 8 reasonable sized NOR gates. This configuration is shown below:



Notice that putting an extra inverter at the input still provides us with all address lines and their complements. Therefore at the NAND stage, we still have the required combinations to decode and the extra inversion does not matter. However, now there will be 3 inverters after NOR, so the output will be $\overline{Sel}$.

We can compute the scale factors and sizes for this configuration as:
Final stage loading is 512, which is driven by I_5. Therefore

$$\hat{f} = 4.6286 = 1 \times 1 \times \frac{512}{C_{in}}, \text{ So } C_{in} = \frac{512}{4.6286} = 110.62$$

I_4 is required to drive a load equivalent to 110.62 min. inverters, so

$$\hat{f} = 4.6286 = 1 \times 1 \times \frac{110.62}{C_{in}}, \text{ So } C_{in} = \frac{110.62}{4.6286} = 23.899$$

I_3 is required to drive a load equivalent to 23.899 min. inverters, so

$$\hat{f} = 4.6286 = 1 \times 1 \times \frac{23.899}{C_{in}}, \text{ So } C_{in} = \frac{23.899}{4.6286} = 5.1634$$

NOR Gate drives a load equivalent to 5.1634 min. inverters. Therefore

$$\hat{f} = 4.6286 = \frac{5}{3} \times 1 \times \frac{5.1634}{C_{in}}, \text{ So } C_{in} = \frac{5.1634 \times 5}{3 \times 4.6286} = 1.8592$$

The NAND gate drives 8 such NOR gates.

$$\hat{f} = 4.6286 = \frac{5}{3} \times 8 \times \frac{1.8592}{C_{in}}, \text{ So } C_{in} = \frac{1.8592 \times 5 \times 8}{3 \times 4.6286} = 5.3559$$

I_2 drives 4 such NAND gates. Therefore,

$$\hat{f} = 4.6286 = 1 \times 4 \times \frac{5.3559}{C_{in}}, \text{ So } C_{in} = \frac{4 \times 5.3559}{4.6286} = 4.6286$$

I_1 drives the inverter I_2.

$$\hat{f} = 4.6286 = 1 \times 4 \times \frac{4.6286}{C_{in}}, \text{ So } C_{in} = \frac{4.6286}{4.6286} = 1.0$$

So the input capacitance of I_1 is 1 as expected. From the input capacitances, we can compute transistor geometries as before: capacitance of 1 inverter is equivalent to 3 transistor widths.
For inverters, n width = C_{in}, p width = $2 \times C_{in}$
For the NAND, the width ratio between n and p is 3:2.
So n width = $C_{in} \times 3 \times 3/5$, p width = $C_{in} \times 3 \times 2/5$
For the NOR, the width ratio between n and p is 1:4.
n width = $C_{in} \times 3 \times 1/5$, p width = $C_{in} \times 3 \times 4/5$

	I_1	I_2	NAND	NOR	I_3	I_4	I_5
C_{in}	1	4.63	5.36	1.86	5.16	23.9	110.6
n width	1	4.63	9.64	1.12	5.16	23.9	110.6
p width	2	9.26	6.43	4.48	10.32	47.8	221.2

– **[6]**

d) Compute the delay of each stage and the total delay for the decoder.

Soln. 2-d) The effort delay of each stage is 4.6286. The parasitic delay should be added to it to get the total stage delay.

The delay for each of the 5 inverter stages is $4.6286 + 2.5 = 5.1286$ units.
The delay of the NAND stage is $4.6286 + 7.5 = 12.1286$ units.
The delay of the NOR stages is $4.6286 + 5 = 9.6286$ units
Therefore the total delay is $5 \times 5.1286 + 12.1286 + 9.6286 = 47.4$ units.

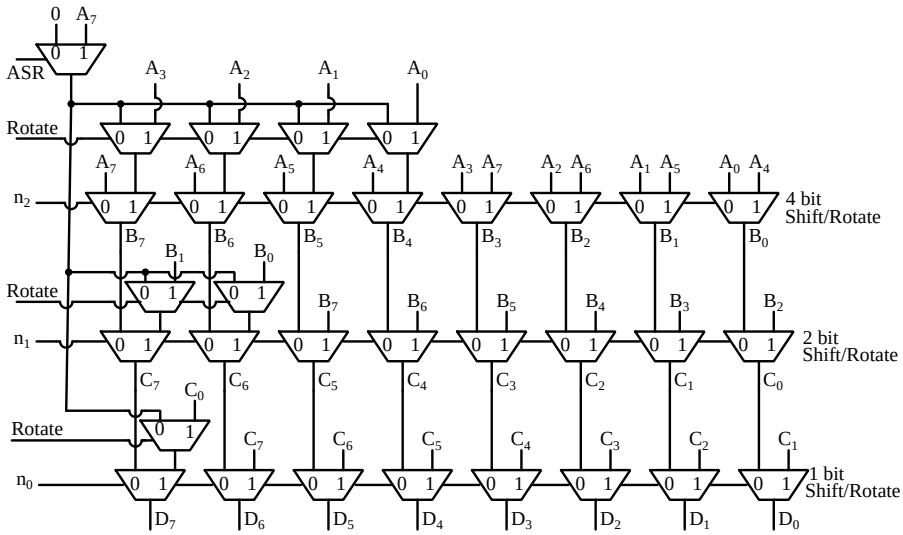
If the inverters were not added, the stage effort for each stage would have been $45511.11^{1/3} = 35.7031$.
The total delay in this case will be 35.7031+2.5+35.7031+7.5+35.7031+5 = 122.11 units.
(This is about 2.5 times the optimal delay!)

– **[Q2: 1+3+6+1 = 11 marks]**

Q–3 a) Show how we can construct a barrel shifter to carry out logical right shift, arithmetic right shift and rotate right, using only 2 input muxes.
Assume the operand to be 8 bit wide.

Soln. 3-a) The rotate operation can be performed by three rows of 2 input muxes. The top row selects between A_i and A_{i+4}, controlled by the top bit (n_2) of the shift amount. (Shift by 4 or 0 positions). Here $i + 4$ is modulo 8.

Similarly the second row selects between A_i and A_{i+2}, controlled by the middle bit of shift amount (n_1). The third row selects between A_i and A_{i+1}, controlled by the lowest bit (n_0) of the shift amount.



If a shift rather than rotate is required (Rotate = '0'), then top 4 bits of the first row, the top two bits of the second row and the top bit of the last row are fed 0. However, if an arithmetic shift is desired, than the value fed to top 4 bits of top row, top 2 bits of the middle row and the top bit of the bottom row is A_7 rather than 0.

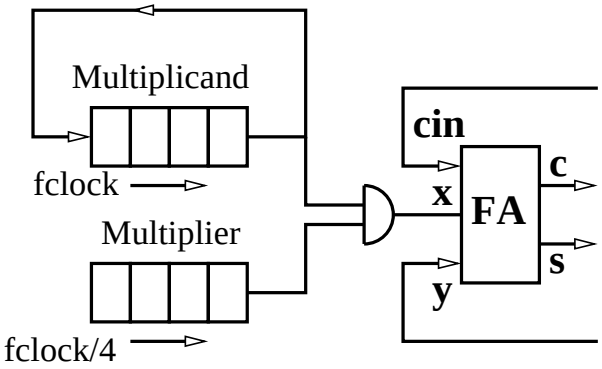
This is implemented by adding a 2 input mux which chooses between A_7 and 0 based on the control input 'ASR'. The selected value is then used by 4 additional muxes in the top row, 2 additional muxes in the middle row and 1 mux in the bottom row to replace the shifted data bits by this value. – **[5]**

b) Describe the operation of a bit serial multiplier, using a 4×4 multiplier as an example. Explain the operations which need to be carried out to take care of exceptions at the end of a row and show how these are implemented in hardware.

Soln. 3-b) The partial product generation as well as addition of partial products must be done serially in a bit serial multiplier. Each bit of the multiplier needs to be ANDed with each bit of the multiplicand to generate the partial products. This requires that all multiplicand bits be presented one after the other, every time a new bit from the multiplier is taken up. This can be managed by using a re-circulating shift register for the multiplicand, which is clocked at a rate which is m times faster than the clock to the multiplier shift register.

These partial products must be presented serially to one input of a full adder. The other input and C_{in} to the full adder have to be appropriately selected and timed to generate the correct product.

Consider a 4×4 bit serial multiplier.

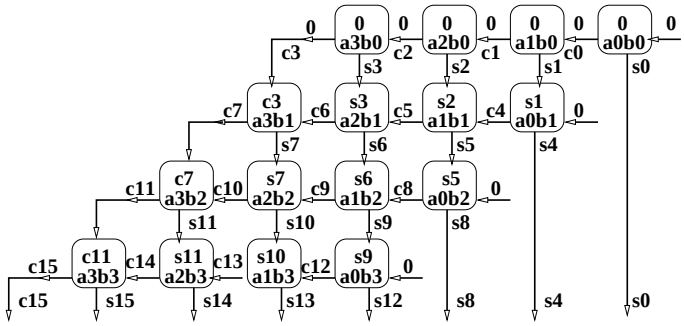


We need additions as follows: (terms in parentheses against each partial product are the times when these should be added).

			a3	a2	a1	a0
		×	b3	b2	b1	b0
			a3b0(3)	a2b0(2)	a1b0(1)	a0b0(0)
		a3b1(7)	a2b1(6)	a1b1(5)	a0b1(4)	
	a3b2(11)	a2b2(10)	a1b2(9)	a0b2(8)		
a3b3(15)	a2b3(14)	a1b3(13)	a0b3(12)			

for all additions, the earlier terms have to wait for 3 clock cycles before the later terms arrive. We can manage this by putting a 3 bit shift register at the sum output and presenting the delayed output at the 'y' input of the full adder. The carry output can be added immediately in the next clock, since it should always go to the next column to its left.

However just a 3 bit delay for the sum will not do as there has to be some exception handling at the end of each row of partial sums.

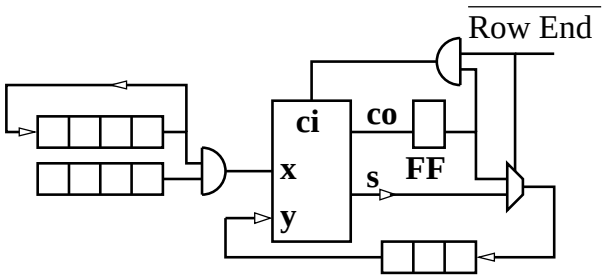


In the figure above, sum and carry terms are indexed by the clock interval in which these were generated. At clocks 0, 4, 8 and 12, carry input should be forced to 0. At clocks 7, 11 and 15, the adder y input should receive carry terms (c3, c7 and c11) instead of sum terms (s4, s8 and s12).

At clocks 0, 4, 8 and 12:

- Carry input should be forced to 0.
- The carry FF output (which is a 1 clock delayed version of cout) should be inserted in the 3-bit shift register.
  Thus C3 (which is always 0), C7 and C11 will emerge at clocks 7, 11 and 15 respectively.
- The sum terms should be taken out as result bits.

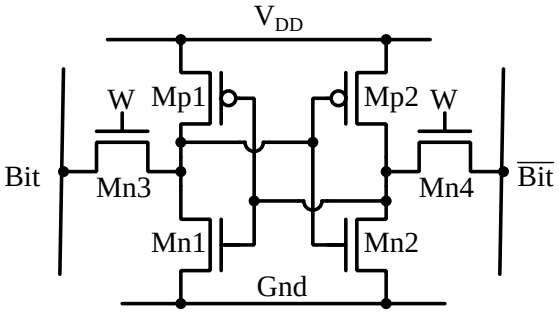
With this exception handling, the bit serial multiplier will be:



This circuit will do a 4x4 multiplication serially. – **[5]**

– **[Q3: 5+5 = 10 marks]**

Q–4



A six transistor static memory cell includes two cross-connected inverters and two N type access transistors (MN3 and MN4) connected to bit and $\overline{\text{bit}}$ lines. Gates of the access transistors are raised to V_{DD} by the word line W when the row containing the cell is selected.

a) Why do we use bit as well as $\overline{\text{bit}}$ lines? What would be the problem if we used only one of these?

Soln. 4-a) The change in voltage at the bit lines during the read operation is very small. This voltage cannot be sensed easily using single ended circuits. Differential circuits cancel out common mode noise and variations in pre-charge values of the bit line. Therefore we use bit and $\overline{\text{bit}}$ lines and a differential sense amplifier which brings up the small voltage differential to a rail to rail digital value.

If we can reliably read a smaller voltage difference, the time to discharge the bit / $\overline{\text{bit}}$ lines through the memory cell can be reduced, thus making the memory faster. – **[1]**

b) How is a "butterfly" diagram used for describing the behaviour of a cascade of two inverters? How is it used to find the stable and meta-stable equilibrium points of cross connected inverters? – **[2]**

c) The capacitance of the bit line is 2pF and it is initially charged to V_{DD}. Transistor MN1 is ON while MN2 is OFF. When the word line goes to V_{DD}, the bit line needs to be discharged to 1.6V for reliable reading by the sensing circuit. Assume that $K_n = 100\mu\text{A/V}^2$ for MN3.

i) If Mn1 is twice as wide as Mn3, find the voltage at the source of MN3 just as the discharge current starts flowing.

Soln. 4-c i) If the voltage at the source of Mn3 is V_s, its drain-source voltage as well as the gate-source voltage is $V_{DD} - V_s$. This means that Mn3 is in saturation. The gate-source voltage of Mn1 V_{DD}, while its drain voltage is small ($= V_s$). Therefore, this transistor is in the linear mode. The load transistor Mp1 is

OFF, because its gate is also at V_{DD}. Since Mn3 and Mn1 are in series, their drain currents must be equal. This gives:

$$\frac{K_{n3}}{2}(V_{DD} - V_s - V_{Tn})^2 = K_{n1}\left((V_{DD} - V_{Tn})V_s - \frac{1}{2}V_S^2\right)$$

Defining $V_1 \equiv V_{DD} - V_{Tn}$ and $\beta \equiv K_{n1}/K_{n3}$, we get

$$(V_1 - V_s)^2 = 2\beta\left(V_1 V_s - \frac{1}{2}V_s^2\right)$$

Which gives

$$V_1^2 + V_s^2 - 2V_1V_s = 2\beta V_1 V_s - \beta V_s^2$$

$$\text{Hence } (1+\beta)V_s^2 - 2(1+\beta)V_1V_s + V_1^2 = 0$$

Solving this quadratic equation, we get

$$V_s = \frac{2(1+\beta)V_1 \pm \sqrt{4(1+\beta)^2V_1^2 - 4(1+\beta)V_1^2}}{2(1+\beta)} = V_1 \pm \sqrt{V_1^2 - \frac{V_1^2}{1+\beta}}$$

Since Mn1 is in linear mode, its drain voltage should be less than $V_{DD} - V_{Tn} (\equiv V_1)$. Therefore,

$$V_s = V_1\left(1 - \sqrt{1 - \frac{1}{1+\beta}}\right) = V_1\left(1 - \sqrt{\frac{\beta}{1+\beta}}\right)$$

For $\beta = 2$, it evaluates to

$$V_s = (1.8 - 0.35)(1 - \sqrt{2/3}) = 0.2661\text{V}$$

– **[4]**

ii) Assuming that the current through MN3 remains constant at its initial value, find the time required to discharge the bit line to 1.6V.

Soln. 4-c ii) MN3 is in saturation with $V_{GS} = 1.8 - 0.2661$ V. Therefore current through MN3, which discharges the bit line, is

$$\frac{K_{n3}}{2}(V_{GS} - V_{Tn})^2 = 50 \times 10^{-6}(1.8 - 0.2661 - 0.35)^2 = 70.08\mu\text{A}$$

The time taken to discharge the capacitor by (1.8-1.6) = 0.2 V will be $C\Delta V/I$ $= 2.0 \times 10^{-12} \times 0.2/70.08 \times 10^{-6}$ or 5.7075 ns. – **[1]**

d) Describe the sequence of operations during read cycle in a static RAM. How is it possible for the stored data to be destroyed during read if the RAM cell is not carefully designed?

Soln. 4-d) During the read cycle, the following actions take place.

1. The address is placed by the processor on the address lines.
2. The row and column address are decoded.
3. Bit and $\overline{\text{Bit}}$ lines are pre-charged.
4. The sense amplifier of the selected column is activated.

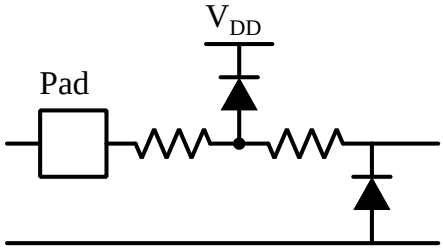
5. Word line for the selected row is pulled high.
6. After the time required to discharge one of the Bit and $\overline{\text{Bit}}$ lines, the Word line is brought low again.
7. The column amplifier determines which of the Bit and $\overline{\text{Bit}}$ lines is lower and outputs a '0' or a '1' accordingly.
8. This digital data is then buffered to the output pad.

As seen in the part above, during the read operation when we are trying to discharge the capacitance associated with the Bit or $\overline{\text{Bit}}$ lines through Mn3 and Mn1, the drain voltage of Mn1 rises by some amount. (It was 0.2661 V in the part above). This voltage is dependent on the value of β. If β is not large enough, this voltage rise can exceed V_Tn. If that happens, Mn2 will be turned ON, which will reduce the output voltage of the inverter Mn2-Mp2. This reduces the gate voltage on Mn1-Mp1 - which will further raise the voltage at the drain of Mn1. This positive feed back action can flip the memory cell such that Mn1 of OFF and Mn2 is ON. Thus the action of reading a cell could destroy the data stored in the cell. – **[2]**

– **[Q4: 1+2+4+1+2 = 10 marks]**

Q–5 a) At the input pads of a CMOS IC, we need to protect against high electrostatic voltages during handling and voltage excursions below ground and above the supply voltage during operation. What kind of device structures are used for protection against these hazards?

Soln. 5-a) Electrostatic voltage from human and machine handling can reach kilo volts in magnitudes. To protect thin oxides and junctions in the IC from these high voltages, special structures are incorporated on the chip.



A commonly used structure is the diode clamp shown in the figure on the left. A p+ diffusion in an n well acts a resistor as well as a diode to V_{DD}. Similarly, an n+ diffusion in a p well provides the resistor as well a diode to ground. The diode to V_{DD} limits any excursion about V_{DD} to within a diode drop and the diode to ground limits negative excursions.

However the IC may be exposed to handling before it is connected to a supply. In that case, high voltages will appear at the supply line through the upper diode. A clamp circuit has to be provided between V_{DD} and ground to prevent this from damaging the circuit. This clamp circuit is often a modified SCR (latchup) structure which occurs naturally in a CMOS circuit, or a field MOSFET, which used the field oxide instead of the gate oxide as the gate dielectric. This will turn on at voltages higher than V_{DD} and clamp the voltage on the supply line to its turn on voltage.

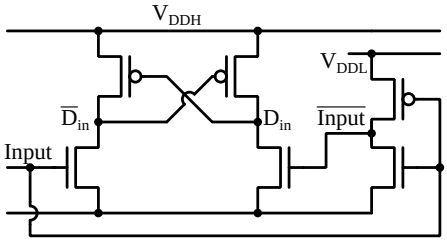
Since external voltages may transiently go below ground or above V_{DD}, it may forward bias the normally reverse biased junction of transistor source/drain with substrate or well. This injected current may turn on the latchup structures, which will damage the IC. To prevent this, all substrate and well regions around the transistors connected to external pads need to have guard bands of p+ regions in the p substrate and n+ regions around the n wells. These guard bands are

connected to ground and V_{DD} respectively through multiple contacts, so that they can collect the injected currents. – **[3]**

b) A chip designed for a supply voltage of 3.3V has to accept inputs from a source which provides logic levels of 0 to 1.8V. How can we use CVSL logic to translate the low swing logic at the input pads to the higher swing logic for use inside a chip? (Assume that a low voltage supply compatible with the low swing input is available on-chip).

Soln. 5-b) Since the input is low swing, the voltage corresponding to a '1' is only 1.8V. However, the internal CMOS circuits are operating at 3.3V. The PMOS transistor of an inverter will have its source connected to 3.3V and the gate connected to this input can only reach 1.8V. Thus the gate sees a negative bias of -3.3 V when the input is '0' and a negative bias of -1.5V while the input is '1'. (V_{Tp} is typically about half a volt). So the PMOS cannot be turned OFF by a '1' at the input. Then both the PMOS and NMOS transistors of a CMOS inverter will be ON when the input is '1' and static power will be wasted.

Thus we need a circuit in which PMOS is not driven by the input. We could have used a pseudo NMOS design, but that uses static power any way. Therefore a CVSL inverter, which does not waste static power and does not drive the PMOS by the input is an attractive choice.



The circuit on the left receives a low swing input and inverts it using an inverter fed from the low voltage V_DD. Since the source of the pMOS of this inverter is at 1.8V, it will be turned off properly when the low swing input is '1'. The input and its complement so generated then drive a CVSL inverter as shown. The output provides a high swing version of the input data (and its complement). – **[2]**

c) How does the large output driver of an output pad also act as a protection device?

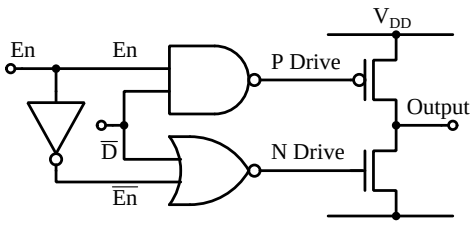
Soln. 5-c) The nMOS driver has a large n+ drain region in a p substrate connected to ground. This constitutes a diode to ground with the same polarity as the one shown for the input pad in part (a) above. The pMOS transistor has a large p+ drain region in an n well connected to V_{DD}. This constitutes a diode connected to supply with the same polarity as shown for the input pad above in part (a). Thus the output buffer itself acts as a protection device. – **[1]**

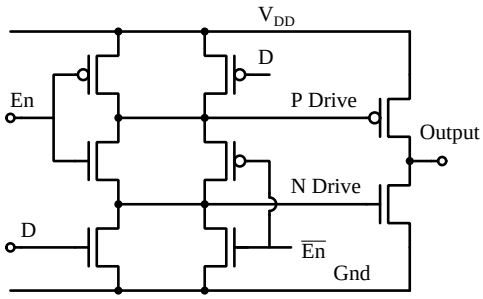
d) In a bidirectional pad, the output drivers need to be tri-stateable. Why is a NAND-NOR based driver structure preferred over 4 transistor tri-stateable inverters at the output?

Soln. 5-d) The tri-stateable inverter uses two pMOS transistors in series and two nMOS transistors in series. Output buffers need to have large widths in order to drive external loads. However, due to the series connection, these widths have to be doubled for both transistors. This requires a large area.

The NAND-NOR circuit shown below or its compact version can provide the ability to tri-state the output buffer, without needing a series transistor. As can be seen from the figure, when En = 0, the NAND output is forced to '1' which turns off the pMOS. At the same time, $\overline{\text{En}}$ is '1', which forces the NOR output to '0' and turns off the nMOS. Thus the output has both pMOS and nMOS turned off when

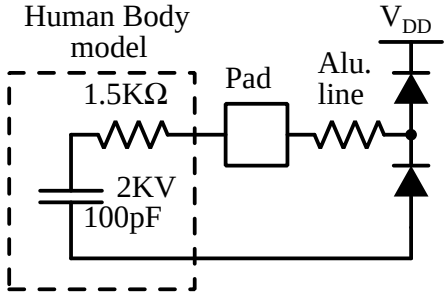
En=0. This accomplishes the tri-stating requirement.





The circuit on the right performs the same function as a Nand-Nor in a single 6 transistor circuit. – **[1]**

e) The human body model is used to emulate the electrostatic hazard to integrated circuits due to handling by human beings.



A 100 pF capacitor is charged to 2KV and discharged through a series resistor of 1.5KΩ and the IC input. Inside the IC, the bond pad size is 70μm ×70μm, which leads to an aluminium line which is 100μm long and 1μm wide. Aluminium thickness is 0.5μm in this layer. The aluminium line leads to a protection diode, whose breakdown voltage is negligible compared to 2KV, so the line may be considered to be terminated to ground.

Estimate the temperature rise in this line when the capacitor is discharged through the external 1.5KΩ resistor and this line to ground. The following assumptions may be made:
– The pad contributes negligibly to the resistance of the line.
– The entire mass of aluminium in the pad and line will be uniformly heated.
– No heat is lost to other structures in this short duration.
– The entire energy stored in the capacitor is used for heating the external resistor, pad and the line.

The relative density of aluminium is 2.7, its specific heat is 0.9J per gram per degree K and its resistivity is $2.7 \times 10^{-6}\Omega$Cm.

Soln. 5-e) Energy stored on the capacitor is $1/2 \times 100 \times 10^{-12} \times 4 \times 10^6 = 200\mu$ J.
Resistance of the Aluminium line is given by

$$R_{Al} = 2.7 \times 10^{-6} \times \frac{100 \times 10^{-4}}{1 \times 0.5 \times 10^{-8}} = 5.4\Omega$$

Since the current through the external resistor and the aluminium line is the same, the energy will divide in the voltage ratio. Therefore, energy dissipated inside the chip is $200 \times 5.4/1505.4\ \mu\text{J} = 0.71742\mu$J.
Total volume of Aluminium being heated is $(70 \times 70 + 100 \times 1) \times 0.5 \times 10^{-12}$ cm^3 = 2500×10^{-12} cm^3
Therefore mass of aluminium = $2.7 \times 25 \times 10^{-10}$g = 67.5×10^{-10}g
Temperature rise is given by

$$\Delta T = \frac{0.71742 \times 10^{-6}}{67.5 \times 10^{-10} \times 0.9} = 118.1^o\text{C}$$

– **[3]**

– **[Q5: 3+2+1+1+3 = 10 marks]**

Paper Ends

Reference

You can use the following transistor model for all questions in this paper:

$$\begin{aligned}
\text{For } V_{GS} \leq V_T, \quad & I_{DS} = \quad 0 \\
\text{For } V_{GS} \geq V_T \text{ and } V_{DS} \leq V_{GS} - V_T, \quad & I_{DS} = \quad K\left((V_{GS} - V_T)V_{DS} - \frac{1}{2}V_{DS}^2\right) \\
\text{For } V_{GS} \geq V_T \text{ and } V_{DS} \geq V_{GS} - V_T, \quad & I_{DS} = \quad \frac{K}{2}\left(V_{GS} - V_T\right)^2
\end{aligned}$$

$$\text{Here } K = \mu C_{ox}\,\frac{W}{L}$$

The asymptotic optimum stage effort for a chain of logic gates is given by:

$$p_{inv} + \rho(1 - \ln \rho) = 0$$