Midterm Exam, Winter 2002, March 1

97.350: Digital Electronics

Department of Electronics, Carleton University

Instructor: Maitham Shams **Name:** _____

Aids: Open Book **ID:** _____

Q1	Q2	Q3	Q4	Q5	Total
/20	/20	/20	/20	/20	/100

- Write your name and ID number on all pages. Attempt all questions.
- If you are asked to make an assumption, then you must use it, but only for that particular question.
- You have *two hours* to write the exam. Approximate timing is given for each question. If you follow that, you should have 15 minutes at the end to review your answers.

Formula

$$
\begin{aligned}
C_{ox} &= \frac{\varepsilon_{ox}}{t_{ox}} \\
C_j &= \frac{C_{j0}}{(1 - V_D/\phi_0)^m} \\
K_{eq} &= \frac{-\phi_0^m}{(V_H - V_L)(1-m)}[(\phi_0 - V_H)^{1-m} - (\phi_0 - V_L)^{1-m}] \\
V_T &= V_{T0} + \gamma(\sqrt{|\, -2\phi_F + V_{SB}\,|} - \sqrt{|\, -2\phi_F\,|}) \\
I_D(\text{sat}) &= \frac{k'_n}{2}\frac{W}{L}(V_{GS} - V_T)^2(1 + \lambda V_{DS}) \\
I_D(\text{linear}) &= k'_n \frac{W}{L}\left((V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2}\right) \\
V_M &= \frac{r(V_{DD} - |\, V_{Tp}\,|) + V_{Tn}}{1 + r}, \quad r = \sqrt{\frac{k_p}{k_n}}
\end{aligned}
$$

Data

NMOS: $L = 1$ μm, V_{TO}=0.7 V, $k' = 20 \times 10^{-6}$ A/V^2, λ=0.07 V^{-1}, $C_{gdo} = C_{gso}$=0.4 fF/μm, C_j=0.3 fF/μm, C_{jsw}=0.8 fF/μm, t_{ox}=150E-10 m.

PMOS: $L = 1$ μm, V_{TO}=-0.7 V, $k' = 6 \times 10^{-6}$ A/V^2, λ=0.20 V^{-1}, $C_{gdo} = C_{gso}$=0.4 fF/μm, C_j=0.5 fF/μm, C_{jsw}=0.3 fF/μm, t_{ox}=150E-10 m.

General: m=0.5 (abrupt junction), ϕ_0=0.6 V, ε_{ox}=3.5 E-13 F/cm.

[1] (20 marks) [20 minutes] Implement a 2-input XOR gate in CMOS and size the transistors to match the delay of an inverter with PMOS-to-NMOS size ratio of 2 and NMOS width of 1 μm. Calculate the capacitance and energy consumption per switching due to each *input node* of this XOR gate. Supply voltage is 3 volts. *Ignore overlap and wiring capacitances.*

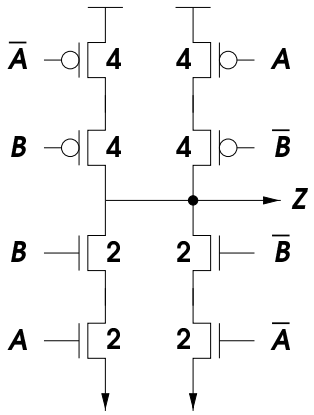
XOR function is given by

$$Z = A\bar{B} + \bar{A}B$$

which can be directly used to implement the pull-up network. The pull-down network can be implemented by inverting Z, which results in an XNOR function.

$$\bar{Z} = A\bar{B} \bar{+} \bar{A}B = AB + \bar{A}\bar{B}$$

Figure below shows the transistor-level schematic with proper sizing, i.e. widths of transistors in microns.



Each input, like A, is connected to an NMOS transistor of $\frac{W}{L} = \frac{2}{1}$ and to a PMOS transistor of $\frac{W}{L} = \frac{4}{1}$. Therefore, the total gate capacitance at each input node is given by

$$C_A = c_{ox}(WL)_{NMOS} + c_{ox}(WL)_{PMOS}$$

where

$$c_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} = \frac{3.5\ E-13\ \ (F/cm)}{150\ E-10\ \ (m)} = 2.33 \quad fF/\mu m$$

Hence, the total capacitance at each node is evaluated to

$$C_A = 2.33(2)(1) + 2.33(4)(1) = 13.98 \quad fF$$

Energy consumption per switching due to each input node is

$$E_A = C_A V_{DD}^2 = 13.98(3^2) = 0.125 \quad pJ$$

[2] (20 marks, 4 each) [20 minutes] Answer the following questions.

2.a) *Explain* why is it possible to produce and market multi-million transistor IC chips using CMOS technology, but it is impossible with other technologies like Bipolar?

CMOS has much lower power consumption compared to Bipolar technology, especially static power. Hence, a multi-million CMOS chip produces far less heat. The amount of heat produced by millions of Bipolar transistors is not only difficult to remove, but also burns the chip itself.

2.b) *Explain* why Constant Electric Field (i.e. Full) Scaling of IC technologies is not feasible?

For Constant Electric Field Scaling, voltages and dimensions have to be scaled with the same factor, because $E = \frac{V}{d}$*. However, some voltage related parameters are physical constants and are not scalable, like built-in potential and band-gap potential. Thus, Full Scaling, although ideal, is not practical.*

2.c) *Explain* why an ideal Logic Gate must have an infinity input resistance and a zero output resistance?

The input resistance must be infinity so that logic gates do not drain any current and, hence, fan-out does not influence the output voltage levels (high and low). The output resistance, on the other hand, must be zero so that whether the output voltage level is high or low, there is always a perfect connection to V_{DD} *or GND. This maximizes output stability and noise immunity.*

2.d) *Explain* why Design Automation using software tools (CAD) has been successful with Digital Circuits but not with Analog Circuits?

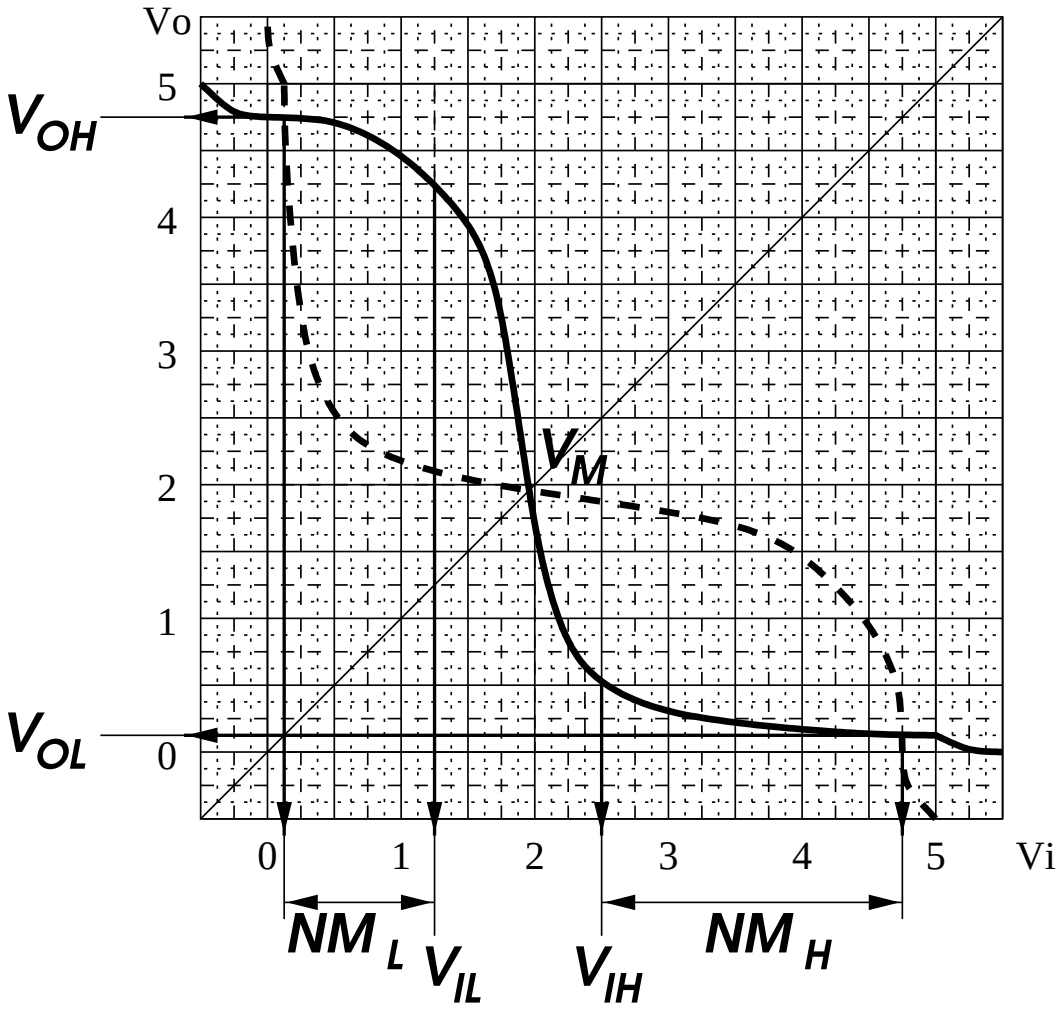
In digital circuits the idea of abstraction at different levels of device, circuits, logic gates, and system can be incorporated. Thus, the design procedure takes a hierarchical order. This is not as much possible in Analog circuits.

2.e) *Intuitively explain* whether the speed of a CMOS IC increases or decreases in a colder environment?

As temperature decreases the random motion of carries due to thermal velocity decreases. Therefore, more carries flow along the direction of the applied electric field. Thus, current increases and so does the speed of operation of the IC.

[3] (20 marks) [20 minutes]

The following figure shows VTC of an inverter. Units are volts. On the graph clearly *identify* V_{OH}, V_{OL}, V_{IH}, V_{IL}, NM_H, NM_L, and the switching threshold of the gate V_M. Then, state the numerical values for each.



$$
\begin{aligned}
VM &= 1.875\ V \\
VIL &= 1.25\ V \\
VIH &= 2.5\ V \\
VOL &= 0.125\ V \\
VOH &= 4.75\ V \\
NM_L &= VIL - VOL = 1.25 - 0.125 = 1.125\ V \\
NM_H &= VOH - VIH = 4.75 - 2.5 = 2.25\ V
\end{aligned}
$$

[4] (20 marks) [25 minutes] Design a CMOS inverter with a switching threshold of 2 volts to drive a load of 1 pF in an average propagation delay of 1 ns. The supply voltage is 3 volts. *Assumptions:* a) transistors operate in saturation during transitions, b) ignore the internal capacitances of the inverter, and c) the input to the inverter is an ideal pulse.

If we keep $L = 1\mu m$ (the minimum allowed device length by technology) for all transistors, then we only need to find the PMOS and NMOS transistor width, W_p and W_n, respectively.

Switching Threshold is given by

$$V_M = \frac{r(V_{DD} - \mid V_{Tp} \mid) + V_{Tn}}{1 + r}, \quad r = \sqrt{\frac{k_p}{k_n}}$$

Therefore,

$$\begin{aligned} 2 &= \frac{r(3 - 0.7) + 0.7}{1 + r} \\ r &= 4.33 \\ \frac{k_p}{k_n} &= 18.78 \end{aligned}$$

Since $k = k'(W/L)$,

$$\begin{aligned} 18.78 &= \frac{6}{20}\frac{(W/L)_p}{(W/L)_n} \\ \frac{W_p}{W_n} &= 62.6\mu\ m \end{aligned} \tag{1}$$

The propagation delay is the average of falling and rising delays.

$$\tau = \frac{1}{2}(\tau_r + \tau_f)$$

Where

$$\tau_r = \frac{CV_{DD}/2}{I_p}, \quad \text{and} \quad \tau_f = \frac{CV_{DD}/2}{I_n}$$

and, hence

$$\tau = \frac{1}{4}CV_{DD}\left(\frac{1}{I_p} + \frac{1}{I_n}\right) \tag{2}$$

To calculate I_n and I_p, we average current values at the beginning and end of the falling and rising transitions, respectively. Since the problem assumes that the transitions take place in saturation, we only need to use the saturation formulas.

$$\begin{aligned} I_p &= I_{pb} + I_{pe}; \quad \textit{where subscript } b \textit{ implies begin and } e \textit{ implies end} \\ I_{pb} &= \frac{k'_p}{2}\frac{W_p}{L}(V_{GS} - V_{Tp})^2(1 + \lambda V_{DS}); \quad \textit{where} \quad V_{DS} = -3\ V \end{aligned}$$

$$
\begin{aligned}
I_{pb} &= \frac{6\ \times 10^{-6}}{2}\frac{W_p}{L}(3-0.7)^2(1+0.2(3)) \\
I_{pe} &= \frac{k'_p}{2}\frac{W_p}{L}(V_{GS}-V_{Tp})^2(1+\lambda V_{DS}); \quad \textit{where} \quad V_{DS}=-1.5\ V \\
I_{pe} &= \frac{6\ \times 10^{-6}}{2}\frac{W_p}{L}(3-0.7)^2(1+0.2(1.5)) \\
I_p &= 23\ \frac{W_p}{L}\ \mu A
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
I_n &= I_{nb}+I_{ne}; \quad \textit{where subscript } b \textit{ implies begin and } e \textit{ implies end} \\
I_{nb} &= \frac{k'_n}{2}\frac{W_n}{L}(V_{GS}-V_{Tn})^2(1+\lambda V_{DS}); \quad \textit{where} \quad V_{DS}=3\ V \\
I_{nb} &= \frac{20\ \times 10^{-6}}{2}\frac{W_p}{L}(3-0.7)^2(1+0.07(3)) \\
I_{ne} &= \frac{k'_n}{2}\frac{W_n}{L}(V_{GS}-V_{Tn})^2(1+\lambda V_{DS}); \quad \textit{where} \quad V_{DS}=1.5\ V \\
I_{ne} &= \frac{20\ \times 10^{-6}}{2}\frac{W_n}{L}(3-0.7)^2(1+0.07(1.5)) \\
I_n &= 61.25\ \frac{W_n}{L}\ \mu A
\end{aligned}
$$

Substituting I_p and I_n in 2 yields

$$\frac{4}{3}\times 10^{-3} = \frac{1}{23W_p}+\frac{1}{61.25W_n} \tag{3}$$

Solving 1 and 3 together, we get

$$
\begin{aligned}
W_p &= 799\ \mu m \\
W_n &= 12.77\ \mu m
\end{aligned}
$$

To check the results, we may substitute back W_p and W_n to find

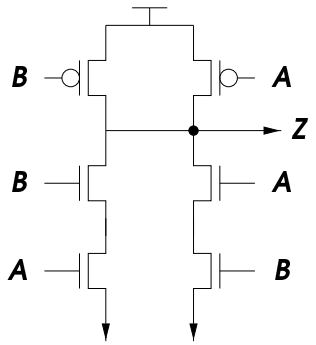
$$
\begin{aligned}
I_p &= 18.4\ mA \\
\tau_r &= 0.09\ ns \\
I_n &= 0.782\ mA \\
\tau_f &= 1.9\ ns
\end{aligned}
$$

and, thus, indeed $\tau = 1$ ns. The moral of this problem is that to shift V_M by only 0.5 V, we are using a huge PMOS transistor of 800μm width, that is, over 60 times larger than the NMOS. If we wanted to have $V_M = V_{DD}/2$, the PMOS transistor would have been around 30 μm wide.

[5] (20 marks) [20 minutes] Answer the following questions.

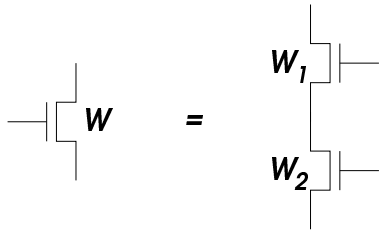
5.a) (8 marks) Draw the schematics of a symmetric CMOS NAND gate whose performance is independent of the arrival order of the inputs.

The following figure shows a CMOS NAND gate that does not differentiate between the two inputs and is symmetric with respect to their arrival order.



5.b) (12 marks) In the concept of *Equivalent Inverter*, two transistors of widths W_1 and W_2 connected in series are represented by a single transistor of width W. Derive the optimal sizes of W_1 and W_2 in terms of W, such that the overall area (and hence power consumption) of the series transistors is minimum.

This is similar to saying that you are looking for a resistor R for your circuit, but cannot find it. Instead, you have resistors of various sizes such that if you connect two of them R_1 and R_2 in parallel you can still end up with R. However, what values do you choose for R_1 and R_2 such that the total power consumption in the substitute circuit is minimum? If you solve this problem, you end up with the usual engineering practice of $R_1 = R_2 = 2R$. Nevertheless, you should be able to prove it scientifically. Here also we



can guess that the answer is $W_1 = W_2 = W$, but we need a formal solution. From the above figure we have

$$\frac{1}{W} = \frac{1}{W_1} + \frac{1}{W_2}$$

$$W = \frac{W_1 W_2}{W_1 + W_2} \tag{4}$$

To minimize area and power, one should minimize the total width of the series transistors W_T.

$$W_T = W_1 + W_2 \tag{5}$$

Replacing 5 into 4 to eliminate W_2 *gives*

$$W = \frac{W_1(W_T - W_1}{W_T}$$

$$W_T = \frac{W_1^2}{W_1 - W}$$

Differentiating the above with respect to W_1 *and equating to zero yields*

$$2WW_1 - W_1^2 = 0$$

Or,

$$W_1 = 2W$$

Similarly, one may show

$$W_2 = 2W$$