EECS 151/251A Homework 7

Due Friday, November 11th, 2022 11:59PM

Instructions

Please read these instructions carefully before starting the homework.

Rounding

This homework involves decimal computations on a calculator. To simplify grading, please follow
the following rounding scheme.
1. Use **at least** 4 **digits** after the decimal points for all intermediate values.
2. Round all answers to the nearest hundredth (2 digits after the decimal point).

We will take answers within 2% of the instructor solution. (This gives a pretty generous mar-
gin for floating point errors. You should not have to worry about numerical precision at all.)

Notation

For a logic gate shown in the following format, x denotes the **input capacitance** of the gate. In
other words, the gate is sized such that the input capacitance of any single input is x.
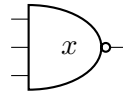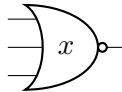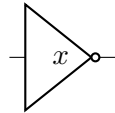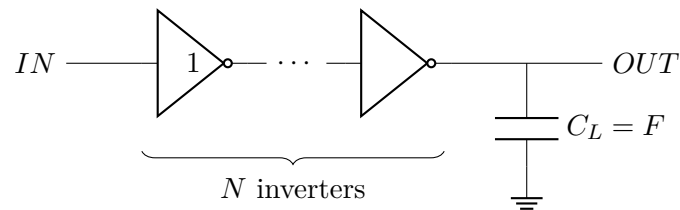
Problem 1: FO4 Delay

In this question, we explore the relationship between delay and the number of stages. We examine the notion of the FO4 delay and show experimentally why it is generally a good idea.

1.1 General Inverter Chain Delay

Now consider a general N-stage inverter chain. Assume $\tau_{inv} = 1$ and $\gamma = 1$. Recall that an inverter delay can be given as $t_{p,inv} = \gamma + FO$.

(a) Derive an expression for the minimum path delay in terms of the total fanout F and the number of stages N. $t_{delay}(N, F) =?$

(b) For the following values of F, find the optimal number of stages $\hat{N}$ and the optimal stage effort $\hat{f}$. Note: $\hat{N}$ should be an integer. Below is a table of potential N values that you might want to try. (An automated tool to explore all the values sounds like a great idea.)

For $F = 64$, $\hat{N} = $ _____, $\hat{f} = $ _____

For $F = 128$, $\hat{N} = $ _____, $\hat{f} = $ _____

For $F = 256$, $\hat{N} = $ _____, $\hat{f} = $ _____

For $F = 512$, $\hat{N} = $ _____, $\hat{f} = $ _____

For $F = 1024$, $\hat{N} = $ _____, $\hat{f} = $ _____

For $F = 2048$, $\hat{N} = $ _____, $\hat{f} = $ _____

F \ N	1	2	3	4	5	6	7	8
64								
128								
256								
512								
1024								
2048								

*This table is not graded

(c) Often times in circuit design, we may not be able to correctly model the delay expression $t_{delay}(F, N)$, due to noise, device physics, and other unpredictable variables; therefore, the number of stages that we pick, $\hat{N}$, may not always be the most optimal number of stages. In general, is it better to underestimate $\hat{N}$ or overestimate $\hat{N}$? In other words, is it better to have more stages or fewer stages than optimal? (*Hint*: Pick a fanout F_0 and plot the delay against N. Which side has a smaller slope?)

(d) If we increase the parameter γ, does the optimal fanout $\hat{f}$ increase, decrease or stay the same?

Solution:

(a) The stage effort for each stage should be the same $\hat{f} = \sqrt[N]{F}$. Then, $t_{delay}(N, F) = N(p + g\hat{f}) = N + N\sqrt[N]{F}$

(b) Plug in the given F and N values into the delay expression.

N / F	1	2	3	4	5	6	7	8
64	65	18	15	15.31	16.49	18	19.68	21.45
128	129	24.63	18.12	17.45	18.20	19.47	21	22.67
256	257	34	22.05	20	20.16	21.12	22.46	24
512	513	47.25	27	23.03	22.41	22.97	24.07	25.45
1024	1025	66	33.24	26.63	25	25.05	25.84	27.03
2048	2049	92.51	41.10	30.91	27.97	27.38	27.80	28.75

*This table is not graded

We can then pick out the N corresponding to the minimum delay for each row, and calculate $\hat{f} = \sqrt[N]{F}$

For $F = 64$, $\hat{N} = 3$, $\hat{f} = 4$

For $F = 128$, $\hat{N} = 4$, $\hat{f} = 3.36$

For $F = 256$, $\hat{N} = 4$, $\hat{f} = 4$

For $F = 512$, $\hat{N} = 5$, $\hat{f} = 3.48$

For $F = 1024$, $\hat{N} = 5$, $\hat{f} = 4$

For $F = 2048$, $\hat{N} = 6$, $\hat{f} = 3.56$

(c) It is better to have more stages. Looking at the table, the delay increases faster when we decrease the number of stages from the optimal point.

(d) The delay expression becomes $t_{delay} = N\gamma + N\sqrt[N]{F}$. Take $\gamma = 2$, then, $\hat{N} = 5$ for $F = 2048$. Which means that the optimal fanout $\hat{f} = \sqrt[N]{F}$ increases.

Problem 2: Branch Effort

(a) Consider the multi-stage path below. Find the optimal values of x, y, and z to minimize the delay from IN to OUT. Assume $\tau_{inv} = 10ps$, $\gamma = 1$, and $R_{eq,N} = R_{eq,P}$ for all gates.
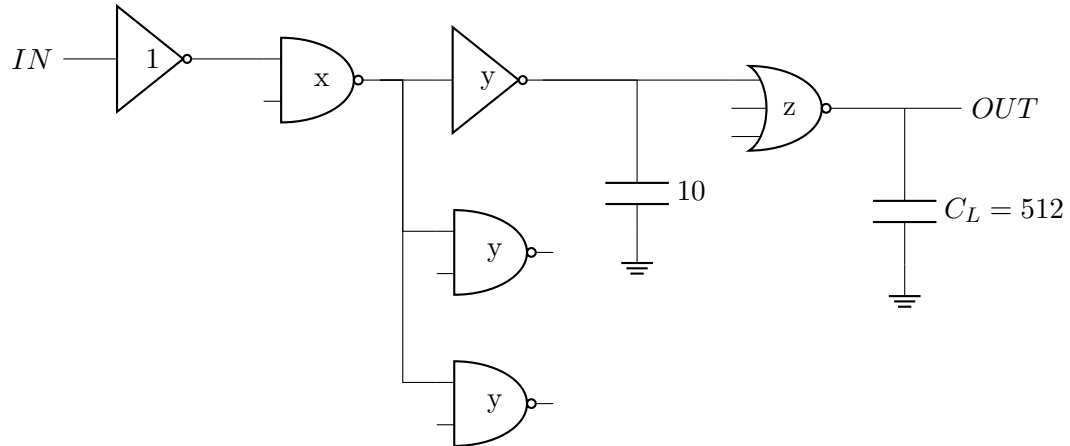
(b) What is the total delay of from IN to OUT? *Don't forget τ_{inv}!*

(c) If you were to add inverters to this path to reduce delay, where would you insert the inverters to keep the area as small as possible?

     A. Between 1 and x

     B. Between x and y

     C. Between y and z

     D. Between z and OUT

(d) We would like to add inverters between z and OUT. Using a FO4 estimate, how many inverters would you add? Assume output polarity does not matter. (*Hint:* FO4 doesn't mean the fanout for each stage should be 4. Think about what the fanout should be with respect to the logical effort.)

Solution:

(a) The delay expression is written as

$$\sum_{i=1}^{N} \tau_{inv}(p_i + g_i f_i)$$

The logical effort of an inverter is 1. The logical effort of a NAND2 gate is 3/2. The logical effort of a NOR3 gate is 4/2. Therefore, the delay expression becomes

$$t_{delay} = \tau_{inv}\left(1 \times x + \frac{3}{2} \times \frac{3y}{x} + 1 \times \frac{z+10}{y} + \frac{4}{2} \times \frac{512}{z} + \sum_{i=1}^{N} p_i\right)$$

We can then take the partial derivative of the delay with respect to x, y, z, and set the partial derivatives to 0 to get the optimal values.

$$\frac{\partial t_{delay}}{\partial x} = 1 - \frac{9}{2}\frac{y}{x^2} = 0$$

$$\frac{\partial t_{delay}}{\partial y} = \frac{9}{2}\frac{1}{x} - \frac{z+10}{y^2} = 0$$

$$\frac{\partial t_{delay}}{\partial z} = \frac{1}{y} - \frac{1024}{z^2} = 0$$

Solving the system of equations yields

$$x = 8.55$$
$$y = 16.25$$
$$z = 129.01$$

(b) The parasitic delay of an inverter is 1. The parasitic delay of a NAND2 gate is 2. The parasitic delay of a NOR3 gate is 3. The total delay is

$$\tau_{inv}\left(8.55 + \frac{3}{2} \cdot \frac{3 \cdot 16.25}{8.55^2} + \frac{129.01 + 10}{16.25^2} + \frac{4}{2} \cdot \frac{512}{129.01} + 1 + 2 + 1 + 3\right) = 250.14ps$$

(c) D. Between z and OUT. Inverters have the lowest logical effort; hence the highest strength to area ratio. It is always better to drive large capacitances with inverters.

(d) FO4 means that the stage effort for each stage should be 4. We can reverse engineering the size of each gate.

$1 \times \frac{x}{1} = 4 \implies x = 4$

$\frac{3}{2} \times \frac{3y}{x} = 4 \implies y = 3.56$

$1 \times \frac{z+10}{y} = 4 \implies z = 4.22$

The remaining path effort is $F_{remain} = \frac{4}{2} \times \frac{512}{z} = 242.53$. (Note that this is true because we're only adding inverters to the path, so the total logical effort of the remaining path is just the logical effort of the NOR3). Let the number of extra stages be N_{extra}. Then, $\sqrt[N_{extra}+1]{F_{remain}} \approx 4 \implies N_{Extra} = 3$. We should add 3 inverters to the end of the path.

Problem 3: Elmore Delay

3.1 Progressive Sizing

In this part, we will examine a sizing technique called stack tapering. Below are two different sizings of a NAND3 gate with the same total area. The drain capacitance is given as $C_d = 1.09 fF/um$ and the transistor resistance is given as $R_{on,1um} = 1.4k\Omega$. Assume $\gamma = 1$.
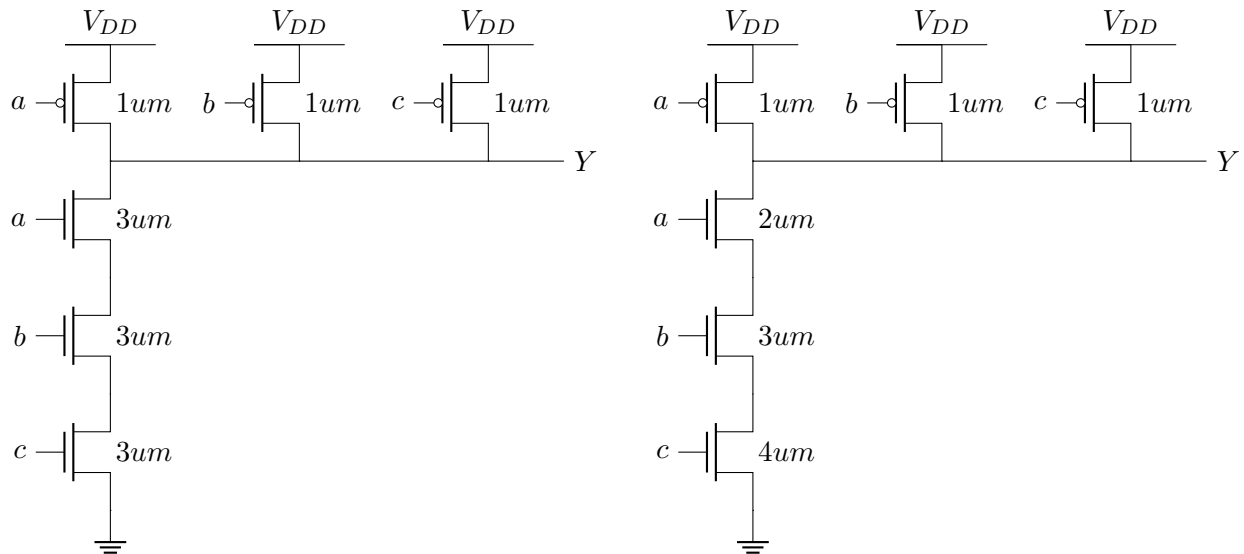

Figure 1: Conventional Sizing vs. Stack Tapering

(a) What is the worst-case input low-to-high transition for a conventionally sized NAND3 gate?

       A. $a = 0 \rightarrow 1, b = 1, c = 1$

       B. $a = 1, b = 0 \rightarrow 1, c = 1$

       C. $a = 1, b = 1, c = 0 \rightarrow 1$

(b) What is the worst-case input high-to-low transition for a conventionally sized NAND3 gate?

       A. $a = 1 \rightarrow 0, b = 1, c = 1$

       B. $a = 1, b = 1 \rightarrow 0, c = 1$

       C. $a = 1, b = 1, c = 1 \rightarrow 0$

We would now like to calculate the delay of the gate for the input transition

$$a = 1, b = 1, c = 0 \to 1$$

(c) First, draw out the equivalent circuit of the conventionally-sized NAND3 with $a = 1, b = 1, c = 1$ and fill in the values of each resistor and capacitor. Include all drain capacitors in your calculations. Assume all the capacitors are initially charged to VDD. Use $+inf$ to indicate open circuits. ($Hint$: Model off-transistors as open circuits.)
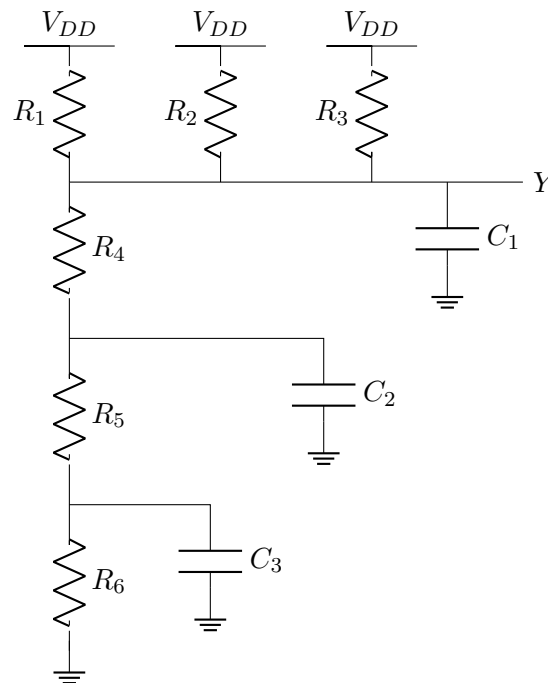

Figure 2: Equivalent circuit of conventionally-sized NAND3

(d) Calculate the Elmore delay from the ground node of R_6 to the output Y.

(e) Calculate the delay of the tapered NAND3 with the above input transition. What is the %-speedup of the tapered NAND3 over the conventional NAND3?

Solution:

(a) C. c is farthest away from the output node. So there is more internal capacitances that need to be discharged.

(b) C. Similarly, c is farthest away from the output node. There is more internal capacitances that need to be charged in this case.

(c) Use the given values to calculate R and C. $R_1 = R_2 = R_3 = +inf\,\Omega$, since they are off. $R_4 = R_5 = R_6 = 1400/3 = 466.67\Omega$. $C_1 = 3 \times 1um \times 1.09fF/um + 3um \times 1.09fF/um = 6.54fF$. $C_2 = C_3 = 2 \times 3um \times 1.09fF/um = 6.54fF$.

(d) The Elmore delay from the ground node of R_6 to Y is

$$ln(2)(R_6(C_3 + C_2 + C_1) + R_5(C_2 + C_1) + R_4(C_1)) = 12691fF \cdot \Omega = 12.69ps$$

(e) First, draw out the equivalent circuit for the tapered NAND gate and calculate the resistances and capacitances.
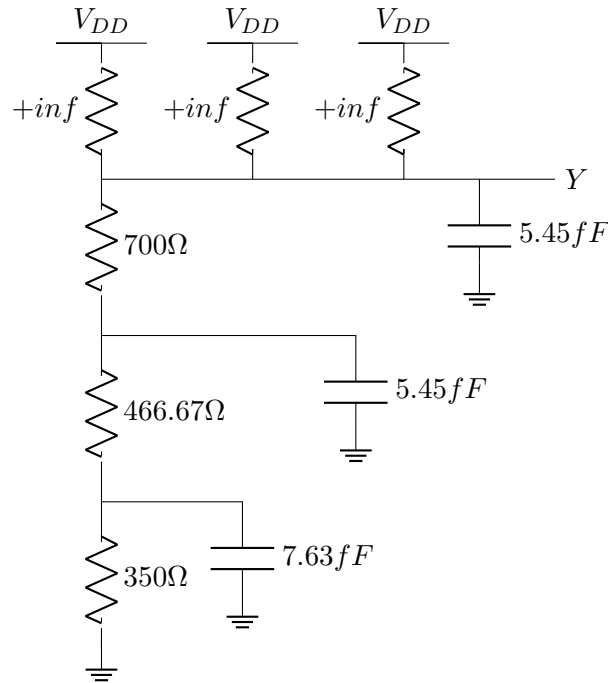
The delay is then $10.67ps$. The %-speedup is therefore $\frac{12.69 - 10.67}{12.69} = 15.92\%$