The University of Michigan

Department of Electrical Engineering and Computer Science

EECS 427 Fall 2008

Quiz 1

Name:_____ UM ID:_____

For full credit, show the pertinent work leading to your answers

Problem #	Maximum Possible Points	Obtained Points
1	22	
2	12	
3	10	
4	16	
5	20	
6	20	
Total	100	

I have neither given nor received any unauthorized aid during this exam

Signed: _____

1.0 Problem 1 (22 Points) – Logical Effort & Tapering

Consider the circuit shown in Figure 1. Assume that the effective P/N ratios of all gates is 3 and a P/N ratio of 3 in an inverter results in equal rise and fall delays. The intrinsic delay of an inverter $p_{inv} = 1\tau$. Assume that the branch gates G and H are sized for equal stage effort with their corresponding on-path gate, D (and they have the same total output capacitance). Note that the 2-input NAND gates, C and D, have tapered NMOS stacks while all other gates are *NOT* tapered (including the 2-input NAND gate within the AND macro cell). Finally, note that the inverter and NAND gate inside the macro cell have a fixed size ratio relative to each other.

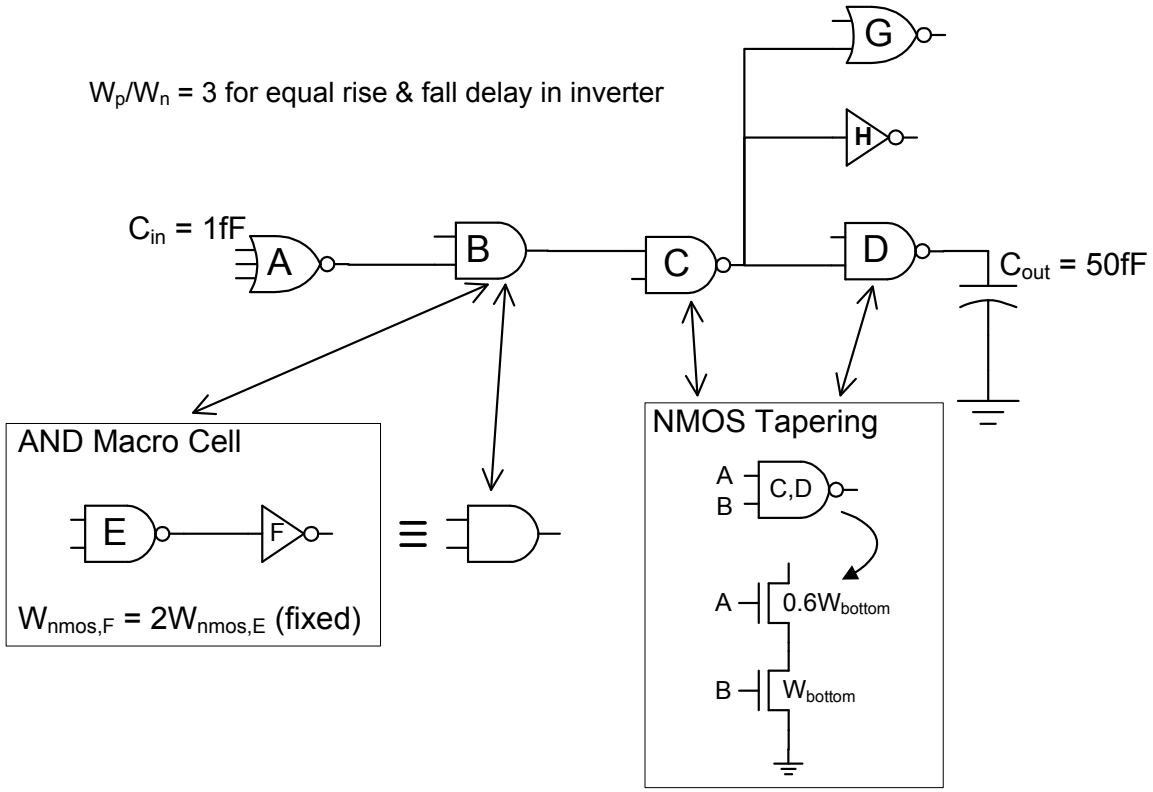


Figure 1. Circuit for Problem 1.0.

1.1 (10 Points) Fill in the table below, calculating the intrinsic delay, *p*, and the logical effort, *g*, for each gate. (Blank space is included on the following page for work.)

Table 1. Answer to Problem 1.1.

Gate Type	*p*	*g*
Inverter (H)	1τ	1
Nor2 (G)	2τ	7/4
Nand2 (D)	1.9τ	17/12
Nand2 (C)	1.9τ	23/20
And (B)	7τ	5/16
Nor3 (A)	3τ	10/4

(Work for Problem 1.1.)

For a min-sized, tapered NAND2, we want $W_{eq,n} = 1$

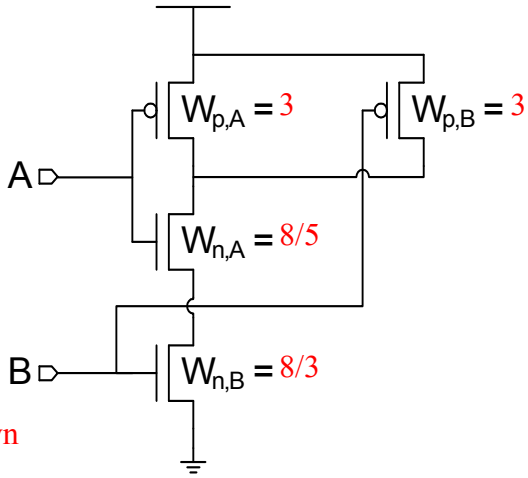
$$\Rightarrow \frac{1}{\frac{1}{W_{bottom}} + \frac{1}{0.6W_{bottom}}} = W_{eq,n}$$

$$\Rightarrow \frac{1}{\frac{1}{W_{bottom}} + \frac{1}{0.6W_{bottom}}} = 1$$

$$\Rightarrow \frac{1.6}{0.6W_{bottom}} = 1$$

$$\Rightarrow W_{bottom} = \frac{8}{3}$$

NAND2 (Tapered, Min-sized)



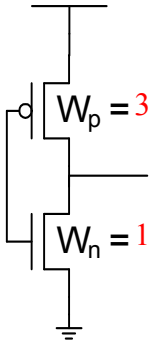
Given the min-sized NAND2 and INV values (shown on the right), we can calculate the p_{nand2} and the g_{nand2}.

$$p_{nand2} = \frac{3 + 3 + \frac{8}{5}}{3 + 1} = 1.9p_{inv}$$

$$g_{nand2,A} = \frac{3 + \frac{8}{5}}{4} = \frac{23}{20}$$

$$g_{nand2,B} = \frac{3 + \frac{8}{3}}{4} = \frac{17}{12}$$

INV (Min-sized)



For the AND macro cell, we perform a similar analysis. We begin with a minimum-sized inverter and then size the NAND gate based on the constraint that $W_{nmos,inv} = 2W_{nmos,nand}$. We then get the gate on the right. Now we can calculate p_{and} and g_{and}.

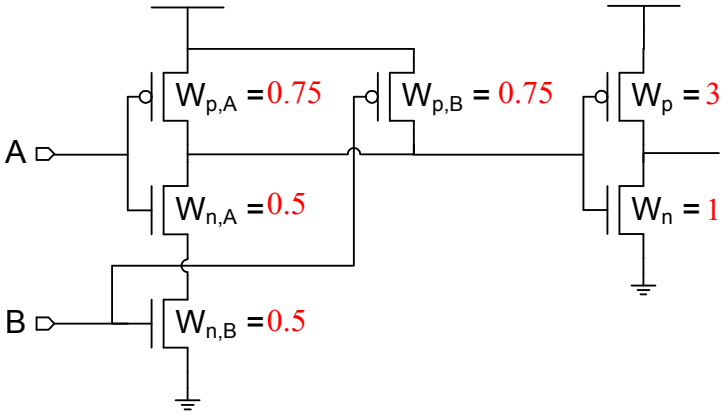
$$p_{and} = p_{nand2} + g\frac{C_{out,nand2}}{C_{in,nand2}} + p_{inv}$$

$$p_{and} = 2\tau + \frac{5}{4} \cdot \frac{C_{in,inv}}{C_{in,nand2}} + 1\tau$$

$$p_{and} = 3\tau + \frac{5}{4} \cdot \frac{4}{1.25} \cdot \tau = 7\tau$$

$$g_{and} = \frac{1.25}{4} = \frac{5}{16}$$

AND (B)



The rest of the cells are calculated as usual and their values are shown in the table on the previous page.

1.2 (4 Points) Find the path effort, F, for the path shown in Figure 1 and calculate the minimum delay using optimal gate sizing. Add extra inverters to the circuit to minimize the delay, if needed (refer to Table 2). Make sure the functionality of the circuit does not change.

Table 2. Table of ranges of path effort, F, and optimum stages, N.

N	F
1	0 - 5.83
2	5.84 - 22.3
3	22.4 - 82.2
4	82.3 - 300
5	301 - 1090
6	1091 - 3920
7	3921 - 14200

$$F = GBH \qquad G = \prod_i g_i \qquad B = \prod_i b_i \qquad H = \frac{C_{out}}{C_{in}} = \frac{50\,\text{fF}}{1\,\text{fF}} = 50$$

$$G = \frac{10}{4} \cdot \frac{5}{16} \cdot \frac{23}{20} \cdot \frac{17}{12} = \frac{1955}{1536} \approx 1.273$$

$$B = \frac{\frac{17}{3} + 4 + 7}{\frac{17}{3}} = \frac{50}{17} \approx 2.941$$

$$F = 1.273 \cdot 2.941 \cdot 50 = 187.174$$

From the table above, we find the optimal N for our F, $N = 4$

$$\hat{f} = 187.174^{\frac{1}{4}} = 3.699$$

$$\hat{d} = N\hat{f} + \sum_i p_i = 4 \cdot 3.699 + (3 + 7 + 1.9 + 1.9) \cdot p_{inv} = 28.60\tau$$

1.3 (4 Points) Compute the input capacitance of each of the gates A – D, as well as any inverters that you might have added, so that delay is minimum. Assume the gate capacitance of transistors, C_{ox}=1fF/μm^2 and the length of all the transistors to be 1μm.

$$C_{in,\,i} = \frac{g_i b_i C_{out,\,i}}{\hat{f}}$$

$$C_{in,\,D} = \frac{\frac{17}{12} \cdot 1 \cdot 50\,\text{fF}}{3.699} = 19.150\,\text{fF}$$

$$C_{in,\,C} = \frac{\frac{23}{20} \cdot \left(1 + \frac{4+7}{\frac{17}{3}}\right) \cdot 19.150\,\text{fF}}{3.699} = 17.512\,\text{fF}$$

$$C_{in,\,B} = \frac{\frac{5}{16} \cdot 1 \cdot 17.512\,\text{fF}}{3.699} = 1.480\,\text{fF}$$

$$C_{in,\,A} = \frac{\frac{10}{4} \cdot 1 \cdot 1.480\,\text{fF}}{3.699} = 1\,\text{fF}$$

1.4 (4 Points) Calculate the corresponding NMOS and PMOS transistor sizes for each of the gates A – D.

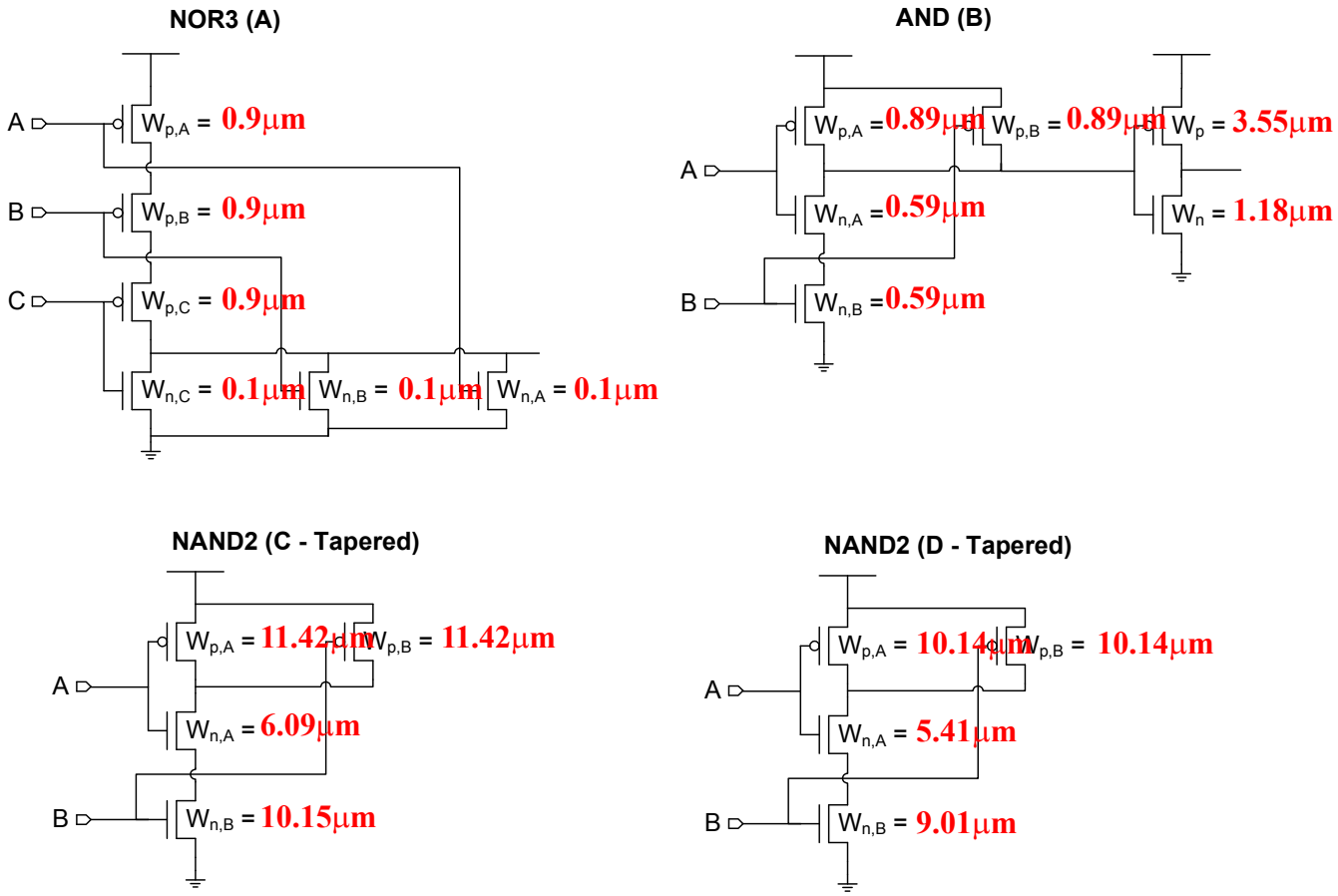


Figure 2. Answer to Problem 1.4.

2.0 Problem 2 (12 Points) – Layout

Given two primary inputs, *A* and *sel*, draw a stick diagram for the layout implementation for the tri-state inverter given by the truth table shown below in Table 3. Assume that the *sel* input is the latest arriving input and adjust your circuit and layout accordingly. Draw your diagram within Figure 3. Draw your transistor level circuit implementation below the layout.

Table 3. Truth table for Problem 2.0.

A	sel	Y
0	0	Z
0	1	V_{DD}
1	0	Z
1	1	0V

* Z represents a high impedance state
* Logic 0 represents 0V
* Logic 1 represents V_{DD}

Important: Continuous diffusion regions for nfets and pfets, as well as metal 1 rails must remain as shown in Figure 3. In other words, you cannot change the M1, diffusion, and poly that is already shown in Figure 3. To implement the tri-state inverter, *you can only connect poly, add M1 or add contacts*. Show your poly connections, added metal routes and contacts in Figure 3 and label your inputs *A*, *sel*, and output *Y*.

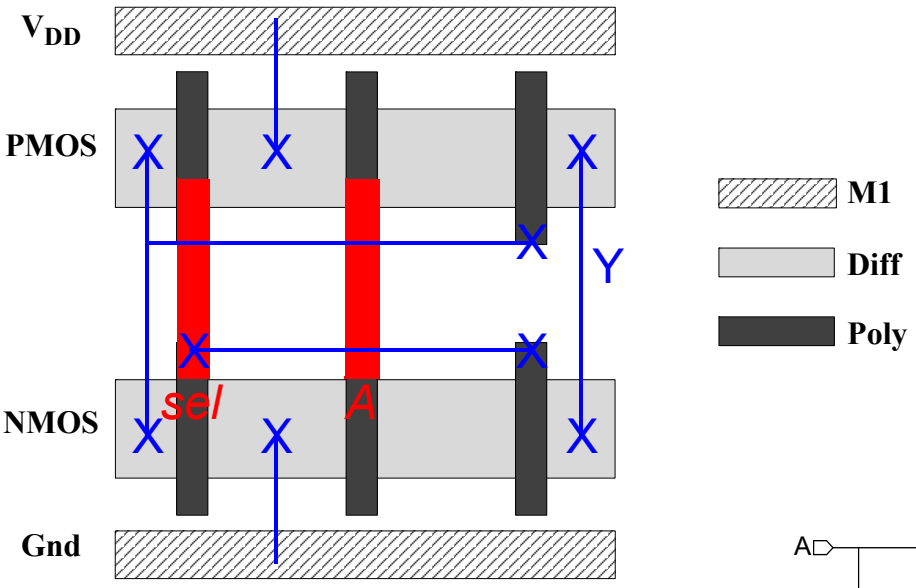
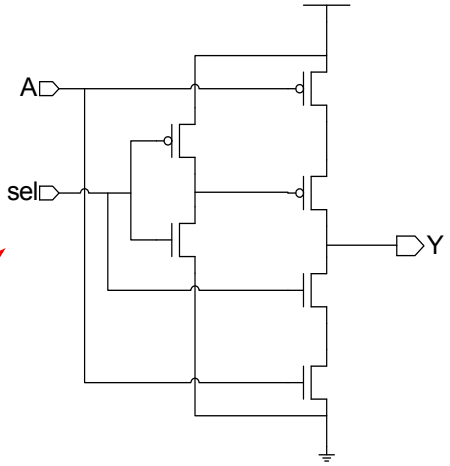


Figure 3. Incomplete layout for Problem 2.0.

Transistor-level circuit implementation:



3.0 Problem 3 (10 Points) – Antenna Rules

3.1 Consider the routing layout between three inverters g1, g2, and g3, as shown in Figure 4. Answer the following questions.

(a) (4 Points) Which gate/gates (g1, g2, or g3) is/are the *most susceptible* to antenna violations?

Gate g2 is the most susceptible. Both g2 and g3 are susceptible to antenna violations but gate "g2" is more susceptible because it is exposed to excessive charge build up on the long M1 *and* M2 lines present.

(b) (3 Points) Which gate/gates (g1, g2, or g3) is/are the *least susceptible* to antenna violations?

Gate g1 is the least susceptible because source/drain nodes are less prone to punchthrough than their gate oxide counterparts (present in g2 and g3).

(c) (3 Points) Add one reversed biased diode to one of the existing metal nets to fix the antenna problem(s). Draw the diode within Figure 4. Explain below why you chose the location of your diode. (An example diode connection to M1 is shown in the bottom left corner of Figure 4.)

The diode was placed on the M1 net next to g2. When M1 is deposited, any charge buildup on the long M1 line with leak through the diode. Next, when M2 is deposited, the charge buildup should flow through the existing M1 lines into the diode once again.

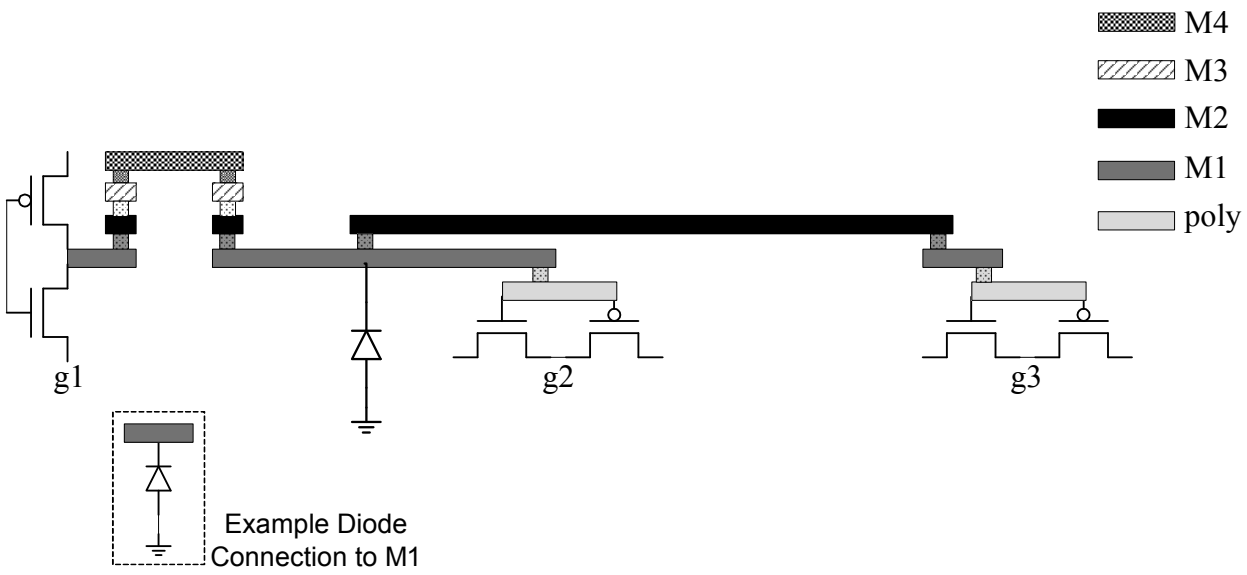


Figure 4. Circuit for Problem 3.0.

4.0 Problem 4 (16 Points) – Short Answer

Answer the following questions as briefly as possible. Use less than 4 sentences of explanation for each answer.

4.1 (2 Points) Explain why polysilicon was necessary for the development of a self-aligned process.

Polysilicon has a higher melting temperature (a higher thermal budget), so it can handle the high temperature anneal step needed after the ion implantation used to create the source/drain diffusion regions.

4.2 (2 Points) What is the approximate P/N ratio that provides equal rise and fall delays in an inverter for our 0.13µm process (used in the CAD assignments and discussed in lecture)?

2.6 +/- 0.1

4.3 (4 Points) List 2 current Resolution Enhancement Techniques (RETs) and in one sentence each, describe how they improve the lithography in modern processes.

Optical Proximity Correction --> Improves diffraction given a certain context which makes printed structures closer to the designed geometries.

Phase Shift Masks --> Uses optical interference constructively to improve photoresist exposure and, therefore, the optical resolution.

Immersion Lithography --> Increases refractive index to improve optical resolution.

Double Patterning --> Optimizes diffraction by only patterning half of geometries at a time.

4.4 (3 Points) Is transistor tapering more or less effective if the output capacitance of a gate is high? Explain.

Transistor tapering is *less* effective if the output capacitance is large because one of the main benefits of tapering is lowering the total capacitance at the output node. Tapering can provide this lower capacitance by trading off lower junction capacitance at the output node for increased equivalent resistance in the tapered stack. Therefore, if the output load becomes large, then the amount of tapering needed to decrease the total output capacitance will increase the equivalent resistance dramatically, making tapering less effective.

4.5 (5 Points) – Use Figure 5 below to answer the next 2 questions about RET.

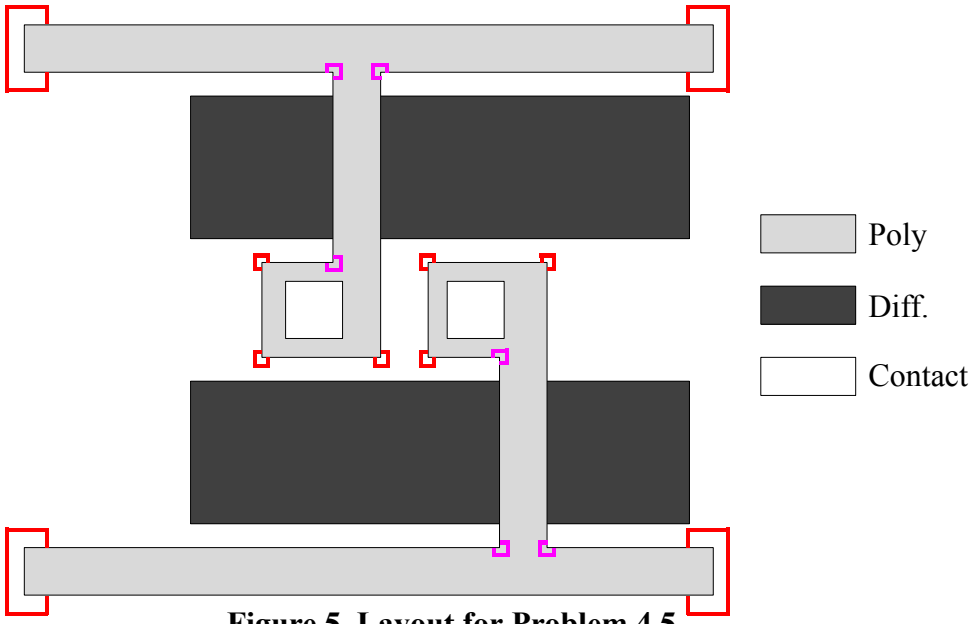


Figure 5. Layout for Problem 4.5.

(a) In Figure 5 above, indicate where the *polysilicon* mask will differ from the layout by drawing serifs, notches, and/or hammerheads in the correct locations.

Serifs and hammerheads are shown in red.

Notches are shown in magenta.

(b) Which design rules will be affected by the changes made to Figure 5 and how (increase or decrease, e.g., "metal-metal spacing increases", etc.)? Explain.

poly-poly spacing, poly-RX spacing, poly-CA spacing

5.0 Problem 5 (20 Points) – Adders

5.1 (9 Points) All inputs to a 16-bit, 4-stage carry-bypass adder, shown in Figure 6, are initially zeros. What new inputs (all of which arrive simultaneously) result in a worst-case delay to the sum? Give the specific input vectors for C_{in}, A<15:0>, and B<15:0>. In your answer, indicate which bit is the least significant bit (LSB). Explain briefly why this is the worst-case input. If necessary, use the delays t_{setup}, t_{c_gen}, t_{p_block}, t_{mux}, and t_{sum} in your explanation.
(NOTE: There may be multiple correct answers.)

Assume that the 5 main delay values (t_{setup}, t_{c_gen}, t_{p_block}, t_{mux}, and t_{sum}) are all within 2-3X of each other.

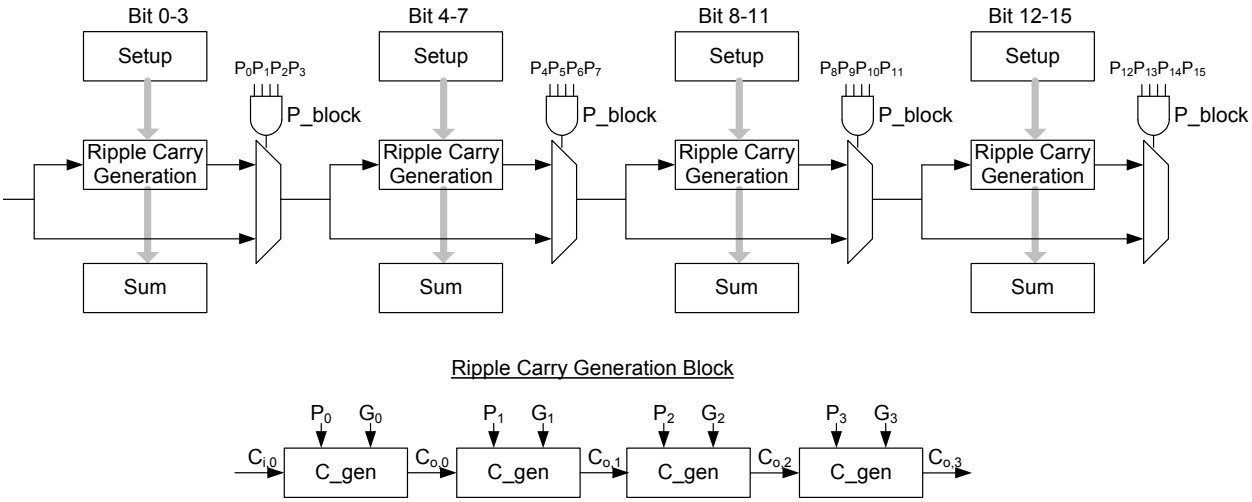


Figure 6. Carry bypass adder architecture used in Problem 5.0

Initial State:

C_{in} = 0

A = 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

B = 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

LSB

Next State = ? (next state results in worst-case delay; don't forget to mark the LSB of your vector)

C_{in} = 0

A<0:15> = 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 X*

B<0:15> = 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 X*

* Don't care

Explanation:

We need a generate in the first bit to start a ripple through the next three propagates. The rest of the bits need to be propagates to allow the original carry (from bit 0) to propagate all the way to bit 15.

$$t_{worst} = t_{setup} + 4t_{c\_gen} + 3t_{mux} + 3t_{c\_gen} + t_{sum}$$

5.2 (11 Points) Shown below are two subsequent input states to the same carry-bypass adder (from Figure 6). Determine the actual delay from the arrival of the new inputs to the final sum generation. Briefly show what the worst-case path is and explain why. Again, assume that all inputs arrive simultaneously. Express your answer as an equation utilizing t_{setup}, t_{c_gen}, t_{p_block}, t_{mux}, and t_{sum}. All LSB's (least significant bits) are on the left.

Initial input state:

$C_{in} = 0$

A = 1 1 0 0 0 0 1 0 1 0 0 1 1 0 0 1

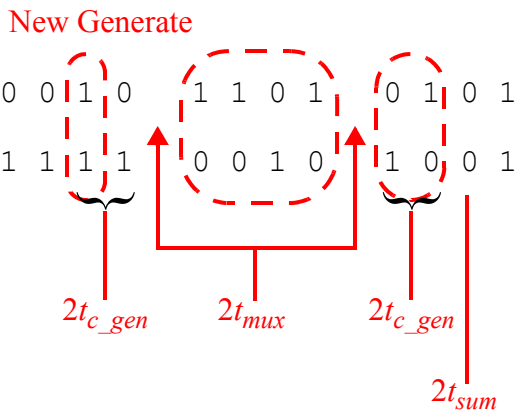
B = 0 1 0 1 1 1 0 1 0 0 0 0 0 1 0 1

Next input state:

$C_{in} = 0$

A = 1 1 1 1 0 0 1 0 1 1 0 1 0 1 0 1

B = 0 0 0 1 1 1 1 1 0 0 1 0 1 0 0 1



$t_{worst} = t_{setup} + 2t_{c\_gen} + 2t_{mux} + 2t_{c\_gen} + t_{sum}$

All other paths are shorter, so this is the worst path.

6.0 Problem 6 (20 Points) – Sizing

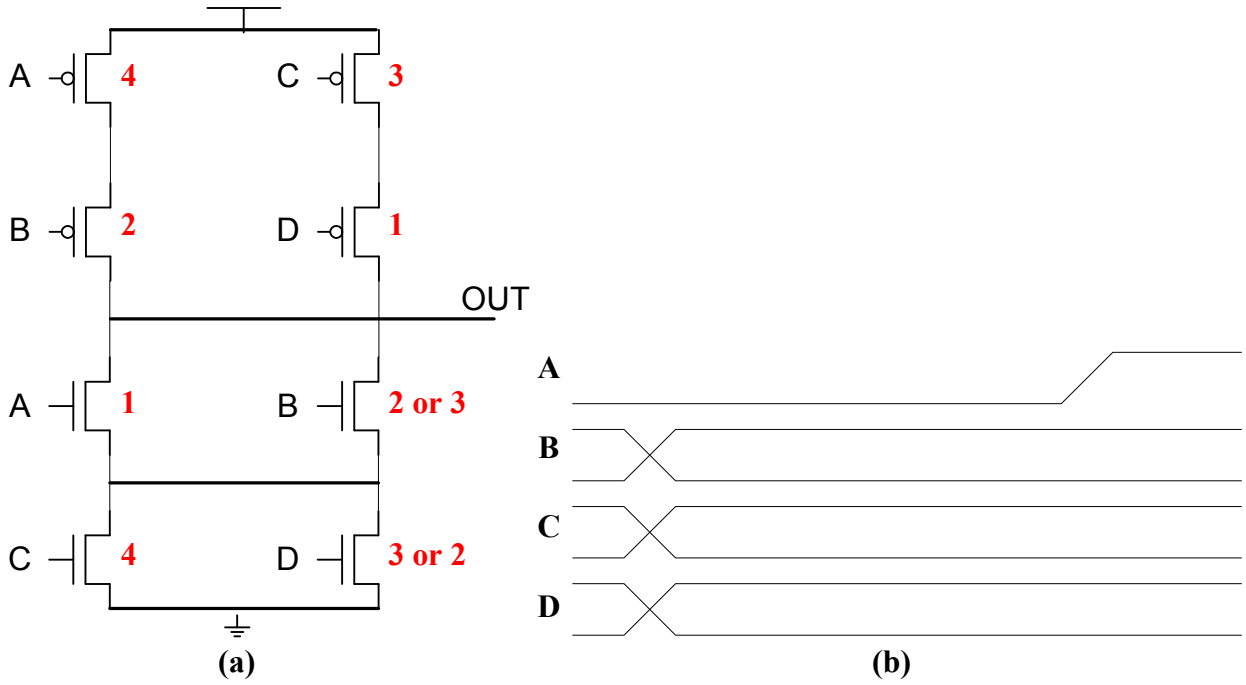


Figure 7. (a) Circuit for Problem 6.0.
(b) *Example* Scenario 1 – Input A rises long after other inputs.

6.1 (9 Points) In the gate above, all transistors are equally sized. The load capacitance on 'OUT' is small, all inputs are stable, and the output is high. Under these conditions, consider four scenarios. In each, *only one input switches* (either A, B, C or D), which causes 'OUT' to get pulled to ground (i.e., the other inputs are stable and their values are set such that 'OUT' is connected to ground when the critical input goes high). Order the scenarios from shortest input-to-output delay to the longest input-to-output delay. Consider the worst case for each if multiple delays are possible for a scenario.

Table 4. Answer to Problem 6.1.

Scenario	Rank*
A (Input 'A' rises)	2
B (Input 'B' rises)	1
C (Input 'C' rises)	4
D (Input 'D' rises)	3

* A rank of "1" represents the shortest input-to-output delay, while "4" represents the longest.

Scenario A --> discharges through 'C' only. Discharges 2 PMOS internal nodes.
Scenario B --> discharges through 'C' only. Discharges 1 PMOS internal node.
Scenario C --> discharges through 'A' only. Discharges 2 PMOS internal nodes and larger NMOS internal node.
Scenario D --> discharges through 'A' only. Discharges 1 PMOS internal node and larger NMOS internal node.
Compare the sizes of caps getting discharged to rank the scenarios.

6.2 (11 Points) Now, consider the scenario described in Problem 6.1 where input 'B' is the switching signal and all other inputs are not changing. Assign the transistors' sizes to minimize the fall delay for this scenario. For each of the transistors in the pull-up and pull-down networks, assign a width of 1-unit (the narrowest) to 4-units (the widest). All transistors have equal gate lengths. You should use each width of 1-, 2-, 3-, or 4-units exactly once each in the pull-up network, and exactly once each in the pull-down network. *Explain your choices* – in case of ambiguity, explain why one choice might be better than another or not. (Please write your sizings directly next to the transistors in the circuit diagram in Figure 7a).

For the PDN: We want 'A' smallest because it arrives early and will contribute to output loading. For the other gates, we want positive tapering for the fastest discharge. We use the table below to analyze the various cases (for worst case, only C *or* D will be on; we choose C to be on in this analysis, but the same analysis could be done using D).

For the PUN: Assign the narrowest sizes to the transistors loading the output capacitance.

Case	**(a)**	**(b)**	**(c)**	**(d)**	**(e)**
NMOS Size: B	1	1	2	2	3
NMOS Size: C	3	4	3	4	4
R	4/3 = 1.3	5/4 = 1.25	5/6 = 0.83	3/4 = 0.75	7/12 = 0.583
C_{int}	10	10	10	10	10
C_{out}	3	3	3	3	4

(a) $R*C_{out} = 4/3 * 3 = 4$

(b) $R*C_{out} = 5/4 * 3 = 3.75$

(c) $R*C_{out} = 5/6 * 3 = 2.5$

(d) $R*C_{out} = 3/4 * 3 = 2.25$

(e) $R*C_{out} = 7/12 * 4 = 2.33$

So case (d) has the lowest RC when there is no extrinsic output capacitance and it also presents a lower input loading than (e). However, since case (e) has the lowest *R*, then a large extrinsic output capacitance could cause case (e) to have the lowest delay.