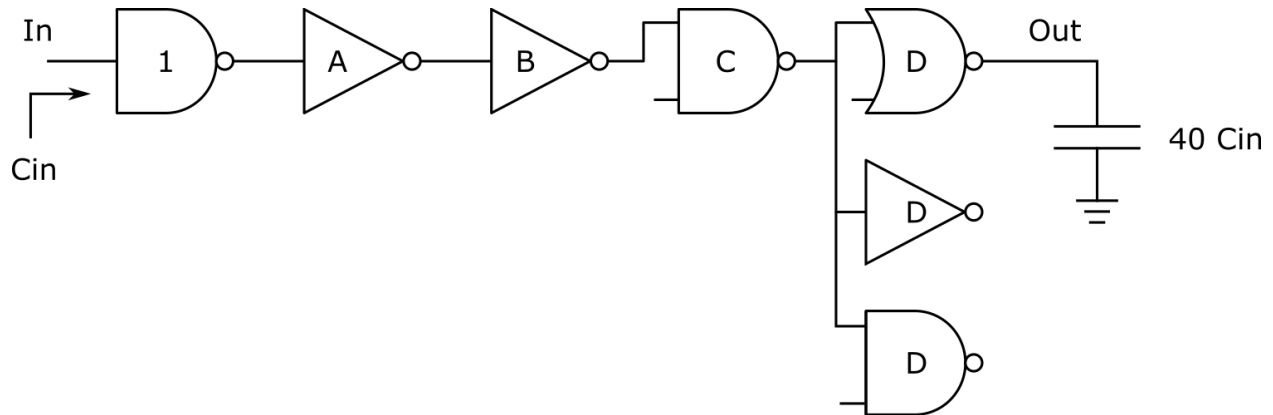


EECS 151/251A Homework 8

Due Friday, Nov 6th, 2020

Problem 1: Path delay with logical effort [10 pts]



For this problem, assume you have a reference inverter with $W_P = W_N = 1$. This technology has $\gamma = 1.5$

- (a) Determine the path effort from input to output, and optimal stage effort for the circuit shown above. Note that A, B, C, and D are the input capacitances of each gate.

Solution:

$$PathEffort = 405$$

$$OptimalStageEffort \approx 3.32$$

$$G = \frac{3}{2} * 1 * 1 * \frac{3}{2} * \frac{3}{2} = \frac{27}{8}$$

$$B = 1 * 1 * 1 * 1 * 3 = 3$$

$$F = 40$$

$$H = \frac{27}{8} * 3 * 40 = 81 * 5 = 405$$

$$SE = \sqrt[3]{405} \approx 3.32$$

- (b) Find the optimum A, B, C, and D in terms of C_{in} for minimum delay in the critical path (input to output).

Solution:

$$\begin{aligned}
 D &= 40.0 * \frac{3}{2} * 1/3.32 = 18.1 \\
 C &= 18.1 * \frac{3}{2} * 3/3.32 = 24.5 \\
 B &= 24.5 * 1 * 1/3.32 = 7.36 \\
 A &= 7.36 * 1 * 1/3.32 = 2.22 \\
 1 &= 2.22 * \frac{3}{2} * 1/3.32 \approx 1
 \end{aligned}$$

- (c) Consider the case where out output load is increased to $C_{out} = 60 * C_{in}$. To achieve the shortest path delay, is it better to add stages (two inverters in series) or to resize the current stages? If adding stages is the better option, where should they be inserted, and how should they be sized? If resizing the existing stages is better, what should the new sizing be?

Solution:

Resizing the existing stages is best.

$$A = 2.41, B = 8.68, C = 31.3, D = 25.0$$

Delay with resized stages:

$$t_{p,resized} = \tau_{inv} * (\gamma * (8) + (5 * 3.6)) = \tau_{inv} * 30.0$$

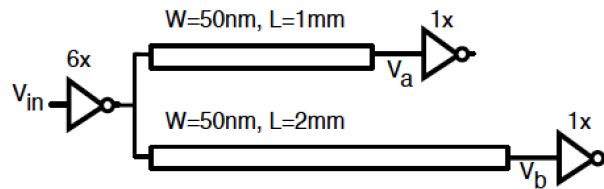
Delay with additional inverters at the output:

$$t_{p,w/inv} = \tau_{inv} * (\gamma * (10) + (4 * 3.2 + 3 * 1.7)) = \tau_{inv} * 32.9$$

Problem 2: Elmore Delay [8 pts]

For the following problem, $C_G = C_D = 2fF/\mu m$, the minimum sized (labeled as 1x in the picture) inverter has $L = 0.1\mu m$, $W_p = 2\mu m$, $W_n = 1\mu m$ and for this technology $R_{n,on} = 10k\Omega/sq.$ (i.e. the resistance of an NMOS with width W and length L is equal to $10k\Omega \frac{L}{W}$) and $R_{p,on} = 20k\Omega/sq.$ (i.e. the resistance of a PMOS with width W and length L is equal to $20k\Omega \frac{L}{W}$). Note that a 6x inverter has 6 times the width of a 1x inverter

For the wire, $R_{wire} = 0.1\Omega/sq.$, the parallel plate capacitance is $C_{pp} = 20aF/\mu m^2$ and the fringing capacitance per each side of wire is $C_{fr} = 14aF/\mu m$. The wire widths and lengths are shown in the picture.



- (a) Using the π wire model, draw the equivalent RC switch model. What is the propagation delay from a step at V_{in} to V_a and V_b ?

Solution:

- For the 1x inverter:

$$R_{N,1x} = R_{P,1x} = 10k\Omega \frac{0.1\mu m}{1\mu m} = 1k\Omega$$

$$C_{i,1x} = C_{o,1x} = (2\mu m + 1\mu m)2fF/\mu m = 6fF$$

- For the 6x inverter:

$$R_{N,6x} = R_{P,6x} = R_N/6 \approx 167\Omega$$

$$C_{o,6x} = 6 \cdot C_{o,1x} = 36fF$$

- For the top wire:

$$R_{w1} = 0.1\Omega \frac{1000\mu m}{0.05\mu m} = 2k\Omega$$

$$C_{w1} = 20aF/\mu m^2 \cdot 0.05\mu m \cdot 1000\mu m + 2 \cdot 14aF/\mu m \cdot 1000\mu m = 29fF$$

- For the bottom wire:

$$R_{w2} = 2 \cdot R_{w1} = 4k\Omega$$

$$C_{w2} = 20aF/\mu m^2 \cdot 0.05\mu m \cdot 2000\mu m + 2 \cdot 14aF/\mu m \cdot 2000\mu m = 58fF$$

The equivalent RC circuit is:

The delay from input to V_a is:

$$t_{p,a} = \ln 2 \cdot R_{w1} \cdot \left(C_{i,1x} + \frac{C_{w1}}{2} \right) + \ln 2 \cdot R_{p,6x} \cdot \left(C_{i,1x} + \frac{C_{w1}}{2} + C_{i,1x} + \frac{C_{w2}}{2} + \frac{C_{w1}}{2} + \frac{C_{w2}}{2} + C_{o,6x} \right)$$

$$\Rightarrow t_{p,a} = \ln 2 \cdot 62.38ps = 43.24ps$$

The delay from input to V_b is:

$$t_{p,b} = \ln 2 \cdot R_{w2} \cdot \left(C_{i,1x} + \frac{C_{w2}}{2} \right) + \ln 2 \cdot R_{p,6x} \cdot \left(C_{i,1x} + \frac{C_{w2}}{2} + C_{i,1x} + \frac{C_{w1}}{2} + \frac{C_{w2}}{2} + \frac{C_{w1}}{2} + C_{o,6x} \right)$$

$$\Rightarrow t_{p,b} = \ln 2 \cdot 162.55ps = 112.27ps$$

- (b) What is the skew (difference in arrival time between V_a and V_b)?

Solution:

$$t_{skew} = |t_{p,b} - t_{p,a}| = 69.03ps$$

Problem 3: Power, Energy, and Performance [12 pts]

151Laptops & Co. is producing a new line of computers tailored for specific workloads. They've come to you for advice on how to optimally configure their processors for each target workload.

The processor for each of these machines will be based on the company's newest RISC-V Core. This base core has been designed to run at a frequency of 3.8 GHz with a supply voltage of 1.0 V. The core has a total capacitance load of approximately 60nF. The design team has also informed you that by raising the supply voltage to 1.3V, the core can be overclocked to run at 4.4GHz. You can alternatively connect two cores together to build a multicore processor. The extra logic required to connect the cores adds 5nF of capacitance load. A characterization team has found that over a number of benchmarks, the average switching rate of gates never exceeds 10%. Use this worst case value in your calculations.

- a. How much power does the base core consume?

Solution:

$$P = \alpha CV^2 f = 0.1 * 60nF * (1V)^2 * 3.8GHz = 22.8W$$

- b. How much power does the single overclocked core configuration consume?

Solution:

$$P = \alpha CV^2 f = 0.1 * 60nF * (1.3V)^2 * 4.4GHz = 44.6W$$

- c. How much power does the multicore configuration consume?

Solution:

$$P = \alpha CV^2 f = 0.1 * 125nF * (1V)^2 * 3.8GHz = 47.5W$$

- d. The first custom machine targets graphic design and rendering workloads. The characterization team has benchmarked these workloads and found that 70% is parallel, and 30% is sequential. What configuration of the processor has the best throughput? What is its speedup over the base core?

Solution:

The dual core configuration has the best throughput. $DualCoreSpeedup = \frac{1}{0.3 + \frac{0.7}{2}} = 1.54$
 $OverclockedCoreSpeedup = 4.4GHz / 3.8GHz = 1.15$

- e. How efficient is each configuration: how much energy does each configuration consume to run the workload? You can leave this answer in terms of the time it takes the base core to run the workload t_b .

Solution:

$$E = t * P$$

$$BaseCoreEnergy = t_b * 22.8W$$

$$DualCoreEnergy = \frac{t_b}{1.54} * 47.5W = t_b * 30.8W$$

$$OverclockedCoreEnergy = \frac{t_b}{1.15} * 44.6W = t_b * 38.8$$

- f. Repeat (d) and (e) for a workload where only 20% the workload is parallel.

Solution:

The overclocked core configuration has the best throughput.

$$DualCoreSpeedup = \frac{1}{0.8 + \frac{0.2}{2}} = 1.11$$

$$OverclockedCoreSpeedup = 4.4GHz / 3.8GHz = 1.15$$

$$BaseCoreEnergy = t_b * 22.8W$$

$$DualCoreEnergy = \frac{t_b}{1.11} * 47.5W = t_b * 42.8W$$

$$OverclockedCoreEnergy = \frac{t_b}{1.15} * 44.6W = t_b * 38.8$$