University of California
College of Engineering
Department of Electrical Engineering
and Computer Sciences

E. Alon

Thursday, November 6, 2008
6:30-8:00pm

EECS 141: FALL 2008—MIDTERM 2

For all problems, you can assume that all transistors have a channel length of 100nm and the following parameters (unless otherwise mentioned):

NMOS:
V_{Tn} = 0.2V, μ_n = 400 cm^2/(V·s), C_{ox} = 1.125 μF/cm^2, v_{sat} = 1e7 cm/s, λ = 0
PMOS:
$|V_{Tp}|$ = 0.2V, μ_p = 200 cm^2/(V·s), C_{ox} = 1.125 μF/cm^2, v_{sat} = 1e7 cm/s, λ = 0

NAME	Last Solutions First

GRAD/UNDERGRAD	

Problem 1: _____/ 18

Problem 2: _____/ 30

Problem 3: _____/ 20

Total: _____/ 68

PROBLEM 1. (18 pts) SRAMs and Wires

In this problem, we will be looking at a 256 x 256 SRAM, where each 6T cell is 2µm wide and 1.5µm tall. All of the devices in the SRAM cell are minimum length (L = 0.1µm), and you can assume that both the NMOS access transistors and the NMOS pull-down transistors are 0.25µm wide, and that the PMOS transistors are 0.12µm wide. The bitline wires are 0.2µm wide. You can also assume that V_{DD} = 1.2V, $C_G = C_D$ = 1fF/µm, and that the wires have the following characteristics: C_{pp} = 100aF/µm^2, C_{fr} = 50aF/µm/edge, and $R_{sq,w}$ = 0.1Ω/□.

a) (6 pts) What is the total capacitance loading each bitline in this memory?

Drain cap/cell: $WC_D = 0.25\mu m \cdot 1fF/\mu m = 0.25fF$

Wire cap/cell: $WLC_{pp} + 2LC_{fr} = 0.2\mu m \cdot 1.5\mu m \cdot 0.1fF/\mu m^2 + 2 \cdot 1.5\mu m \cdot 0.05fF/\mu m$

$= 0.18fF$

Total cap: $256 \cdot (C_{drain} + C_{wire}) = 256 \cdot (0.25fF + 0.18fF)$

$= \boxed{110.08fF}$

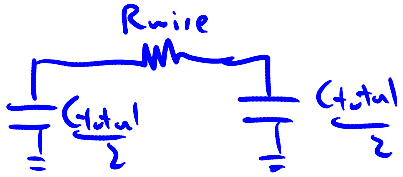
b) (7 pts) If we were reading to or writing from the SRAM array, what is the worst-case RC delay due to the bitline wire? You can assume that the input is a ramp (i.e., $t_p = \tau_{Elmore}$).

Already calculated wire cap itself - just need to notice that drain junction cap from transistors will slow the wire down too. So:

$$R_{wire} = R_{sq,w} \cdot \frac{L}{W_{wire}} = 0.1\Omega/\square \cdot \frac{256 \cdot 1.5\mu m}{0.2\mu m} = 192\Omega$$

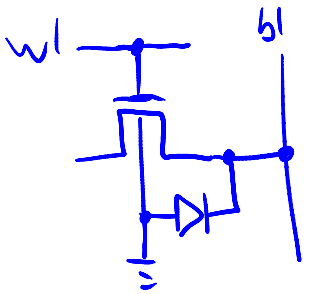
$$\boxed{t_{p,wire} = \frac{R_{wire} \cdot C_{wire}}{2} \approx 10.57ps}$$

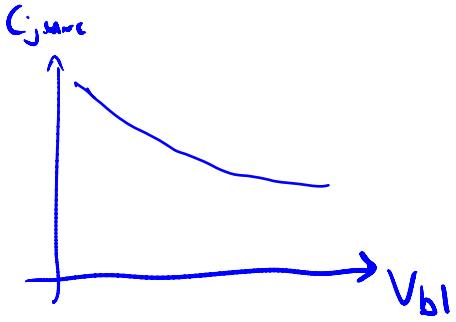
π model:



c) (5 pts) If the V_{DD} of the SRAM was raised to 1.3V, would the RC delay of the bitline wire increase, decrease, or stay the same? You don't need to do any calculations, but you do need to explain your answer.

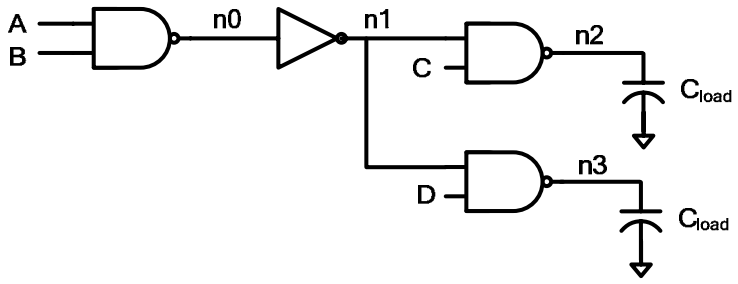
The RC delay would actually decrease because the effective capacitance from the drain junctions would decrease. Remember, C_D is a linearization of the actual non-linear junction cap; if V_{DD} is increased, the drain junctions become more reverse biased.





PROBLEM 2. (30 pts) Activity Factors and Sensitivity Analysis.

This problem will deal with the circuit shown below; throughout this problem you can assume that $C_D = 0$.



a) **(6 pts)** Assuming that all of the inputs A, B, C, and D have equal probability of being a 1 or a 0, what are the activity factors (i.e., $\alpha_{0\to1}$) at each of the nodes of the circuit (i.e., n0 – n3)?

n0: $P(n0=0) = P(A=1)\cdot P(B=1) = \frac{1}{4}$

$\alpha_{0\to1}(n0) = \frac{1}{4}\cdot\left(1-\frac{1}{4}\right) = \frac{3}{16}$

n1: inverter doesn't modify activity

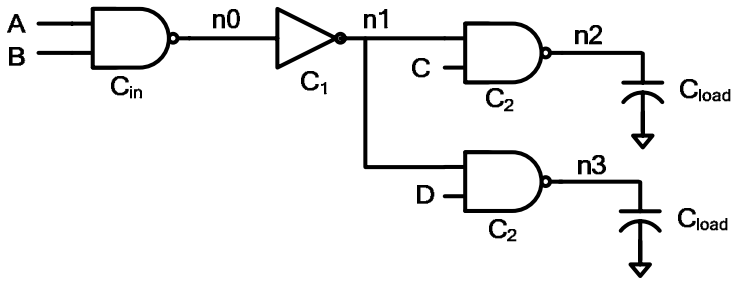
↳ $\alpha_{0\to1}(n1) = \frac{3}{16}$ Note: $P(n1) = P(n0=0) = \frac{1}{4}$

n2: $P(n2=0) = P(n1=1)\cdot P(C=1) = \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{8}$

$\alpha_{0\to1}(n2) = \frac{1}{8}\cdot\left(1-\frac{1}{8}\right) = \frac{7}{64}$

n3: Same as n2

Node	$\alpha_{0\to1}$
n0	3/16
n1	3/16
n2	7/64
n3	7/64



b) **(6 pts)** Assuming the circuit operates with a supply voltage V_{DD} and a clock frequency *f*, what is the total dynamic power consumed by this circuit as a function of C_{in}, C_1, C_2, and C_{load} (as labeled above)? Note that you should include the power dissipated by driving the A, B, C, and D inputs. If you aren't sure about your answers for part a), please assume that all of the activity factors are 1/8 (which is not the correct answer to part a).

$P_A: \alpha_{0\to1}(A) \cdot C_{in} V_{dd}^2 f = \frac{1}{4} C_{in} V_{dd}^2 f$

P_B: Same as $A \to \frac{1}{4} C_{in} V_{dd}^2 f$

$P_{n0}: \alpha_{0\to1}(n0) \cdot C_1 V_{dd}^2 f = \frac{3}{16} C_1 V_{dd}^2 f$

$P_{n1}: \alpha_{0\to1}(n1) \cdot 2 \cdot C_2 \cdot V_{dd}^2 f = \frac{3}{8} C_2 V_{dd}^2 f$

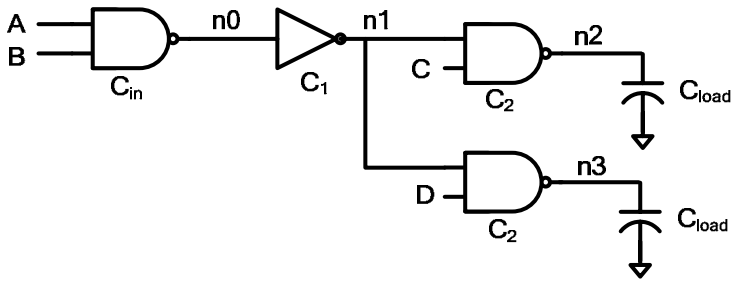
$P_C: \alpha_{0\to1}(C) \cdot C_2 \cdot V_{dd}^2 f = \frac{1}{4} C_2 V_{dd}^2 f$

P_D: Same as $C \to \frac{1}{4} C_2 V_{dd}^2 f$

$P_{n2}: \alpha_{0\to1}(n2) \cdot C_{load} \cdot V_{dd}^2 f = \frac{7}{64} C_{load} V_{dd}^2 f$

P_{n3}: Same as $n2 \to \frac{7}{64} C_{load} V_{dd}^2 f$

$$P_{total} = \left[\frac{1}{2} C_{in} + \frac{3}{16} C_1 + \frac{7}{8} C_2 + \frac{7}{32} C_{load}\right] V_{dd}^2 f$$



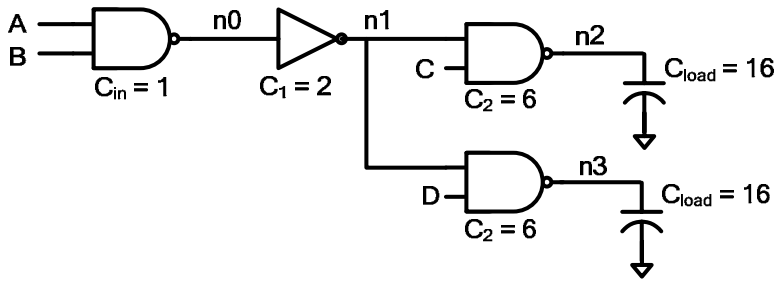
c) **(4 pts)** Assuming that A is the critical input (i.e., the last one to transition) and that the transistors are quadratic long-channel, what is the delay of the decoder (in units of t_{inv}) as a function of C_{in}, C_1, C_2, and C_{load}?

Long-channel devices $\rightarrow LE_{NAND2} = \frac{4}{3}$

$$t_p = t_{inv}\left[\frac{4}{3}\frac{C_1}{C_{in}} + \frac{2C_2}{C_1} + \frac{4}{3}\frac{C_{load}}{C_2} + 2\delta + \delta + 2\delta\right]$$

$\delta = 0$ since $C_0 = 0$, so:

$$\boxed{t_p = t_{inv}\left[\frac{4}{3}\frac{C_1}{C_{in}} + \frac{2C_2}{C_1} + \frac{4}{3}\frac{C_{load}}{C_2}\right]}$$



d) **(8 pts)** With the specific sizes shown above, what are the sensitivities S_{C1} and S_{C2}, where $S_x = (\partial\text{Power}/\partial x)/(\partial\text{Delay}/\partial x)$?

$$P_{tot} = \left[\frac{1}{2}C_{in} + \frac{3}{16}C_1 + \frac{7}{8}C_2 + \frac{7}{32}C_{load}\right] V_{dd}^2 f$$

$$t_p = \left[\frac{4}{3}\frac{C_1}{C_{in}} + \frac{2C_2}{C_1} + \frac{4}{3}\frac{C_{load}}{C_2}\right] t_{inv}$$

C_1: $\partial P/\partial C_1 = 3/16\ V_{dd}^2 f$

$$\frac{\partial t_p}{\partial C_1} = \left[\frac{4}{3}\frac{1}{C_{in}} - \frac{2C_2}{C_1^2}\right] t_{inv}$$

$$S_{C_1} = \frac{3/16}{\frac{4}{3} - \frac{12}{4}} \cdot \frac{V_{dd}^2 f}{t_{inv}} = -0.1125 \frac{V_{dd}^2 f}{t_{inv}}$$

C_2: $\partial P/\partial C_2 = \frac{7}{8} V_{dd}^2 f$

$$\frac{\partial t_p}{\partial C_2} = \left[\frac{2}{C_1} - \frac{4}{3}\frac{C_{load}}{C_2^2}\right] t_{inv}$$

$$S_{C_2} = \frac{7/8}{\frac{2}{2} - \frac{4}{3}\cdot\frac{16}{36}} \cdot \frac{V_{dd}^2 f}{t_{inv}} \approx 2.148 \frac{V_{dd}^2 f}{t_{inv}}$$

e) **(6 pts)** If you could change only one of the gate sizes (i.e., either C_1 or C_2), which one would you change, and in what direction? Be sure to explain your answer.

Notice that sensitivity of C_2 is not only larger in magnitude, its sign is positive. This means that if we decrease C_2, we get both lower power and lower delay. So, we should definitely decrease C_2.

PROBLEM 3. Scaling and Leakage (20 points)

For this problem you can ignore shoot-through current and assume that $C_G = C_D =$ 2fF/μm, and that the leakage current of the transistors is equal to $\frac{W}{L} I_0 e^{\left(\frac{-V_T}{1.5\times 25mV}\right)}$, where $I_{0,NMOS} = 4\mu A$ and $I_{0,PMOS} = 2\mu A$ are both independent of technology scaling.

a) **(6 pts)** What is the power dissipation of an inverter implemented in a 100nm technology with $V_{DD} = 1.2V$, $W_n = 10\mu m$, and $W_p = 20\mu m$ driving a capacitive load of 120fF? On average, the input to this inverter changes its state once every 25 cycles of a 2GHz clock.

Leakage for NMOS & PMOS equal:

$P_{leak} = \frac{10\mu m}{0.1\mu m} \cdot 4\mu A \cdot e^{(-200mV/1.5\cdot 25mV)} \cdot 1.2V$

$P_{leak} = 2.317\mu W$

Activity factor: 1 transition every 25 cycles → half of them are 0→1.

So $\alpha_{0\to 1} = 1/50$

$P_{dyn} = \frac{1}{50}\left[30\mu m \cdot (C_G + C_D) + 120fF\right](1.2V)^2 \cdot 2GHz$

$= 13.824\mu W$

$P_{total} = P_{leak} + P_{dyn}$

$\boxed{P_{total} \approx 16.14\mu W}$

b) **(6 pts)** If the same inverter is now scaled to a 50nm technology with full scaling (i.e., V_{DD} and V_T scaled along with the dimensions), what is the power consumption of the inverter in the new technology? You can assume that the capacitive load scales with the technology, but that the clock frequency is fixed.

$S = \frac{100nm}{50nm} = 2$ $\quad W_N = \frac{1}{2}W_p = \frac{10\mu m}{S} = 5\mu m$ $\quad C_{load} = 60fF$

$V_{DD} = \frac{1.2V}{S} = 0.6V$ $\quad V_T = \frac{200mV}{S} = 100mV$

$P_{dyn} = \frac{1}{50}\left[15\mu m \cdot (C_G + C_D) + 60fF\right] \cdot (0.6V)^2 \cdot 2GHz = 1.728\mu W$

$P_{leak} = \frac{5\mu m}{0.05\mu m} \cdot 4\mu A \cdot e^{(-100mV/1.5\cdot 25mV)} \cdot 0.6V = 16.68\mu W$

$\boxed{P_{total} \approx 18.4\mu W}$

c) **(8 pts)** Still keeping the clock frequency fixed in the 50nm technology, what is the power consumption of the inverter under fixed voltage scaling? Why is this power actually lower than the power of the inverter with full scaling?

$V_{DD} = 1.2V$

$V_T = 200mV$

$P_{dyn} = \frac{1}{50} \cdot 120ff \cdot (1.2V)^2 \cdot 2GHz = 6.912\mu W$

$P_{leak} = \frac{5\mu m}{.05\mu m} \cdot 4\mu A \cdot e^{(-200mV/1.5 \cdot 25mV)} \cdot 1.2V = 2.317\mu W$

$\boxed{P_{total} = 9.23\mu W}$

Power with fixed voltage scaling actually lower because leakage increased ~8x with full scaling. By bringing V_T back to 200mV (dropping leakage current by ~14x), overall power is lower (even though V_{DD} doubled).