UNIVERSITY OF CALIFORNIA
College of Engineering
Department of Electrical Engineering and Computer Sciences
Last modified on September 29, 2003 by Nuntachai Poobuapheun (nuntachp@eecs.berkeley.edu)

Borivoje Nikolic   Homework #4: Capacitance, Process Scaling, and Sizing   EECS 141

Problem #1: Adiabatic capacitor charging

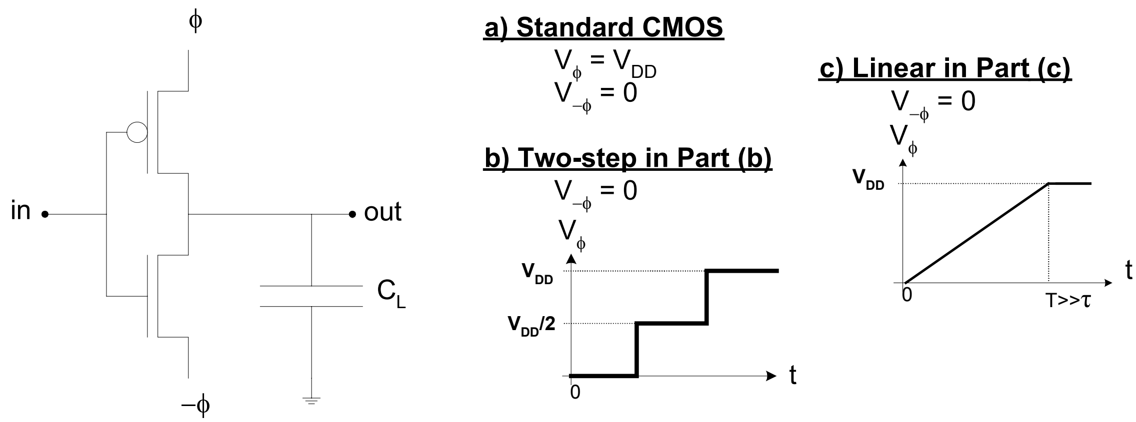

Figure 1a: Inverter with Variable Power Rails

Recall that the power consumption of an inverter has two principle components: static power dissipation and dynamic power dissipation. Adiabatic switching is one approach to mitigate these losses. This is accomplished by changing the voltage on the power and ground rails, instead of keeping them constant like in ordinary CMOS logic. Assume $C_L >> C_{INT}$ (ie. intrinsic capacitance is negligible).

a) Consider the transient response of a standard CMOS inverter as shown in Figure 1a (with constant rails at V_{DD} and 0) and external load C_L. Assuming sufficient time for the load capacitance to fully charge/discharge between transitions, how much energy is consumed charging the load for a pull-up (low to high) transition. How much energy is consumed for a pull-down (high to low) transition? After a pull-up transition, how much energy is stored on the capacitor? Why is this different from the total energy consumed to charge the capacitor?

$E_{L\to H} = CV_{DD}^2$
$E_{H\to L} = 0$
$E_{CAP} = ½\, CV_{DD}^2$ (the remaining $CV_{DD}^2 - ½\, CV_{DD}^2 = ½\, CV_{DD}^2$ is dissipated by the resistor)

b) Now, let's see what happens if the power is applied in two steps during a pull-up transition, as shown in Figure 1b. Assume that *in* = 0 and that there is plenty of time between the steps for the capacitor to charge up. How much energy does it take to charge the capacitor all the way up to V_{DD} using this two-step approach?
Derive a simple expression for N steps. Show that the power dissipation will be minimum when each step is equal to Vdd/N for N steps charging.

First, find the energy E_R dissipated by the resistor for each step.

$E_{R,step1} = ½\ C(V_{DD}/2)^2 = CV_{DD}^2/8$
$E_{R,step2} = ½C(V_{DD}\text{-}V_{DD}/2)^2 = CV_{DD}^2/8$

$E_{total} = E_{R,step1} + E_{R,step2} + E_{cap} = ¾C\ V_{DD}^2$

In general, for N steps,

$E_{R,stepi} = ½\ C(V_{DD}/N)^2$
$E_R = \sum E_{R,stepi} = \sum ½\ C(V_{DD}/N)^2 = N/2\ CV_{DD}^2/N^2 = CV_{DD}^2/2N$
$E_{total} = E_R + E_{CAP} = CV_{DD}^2/2N + CV_{DD}^2/2$

In order to prove that the energy consumption will be minimal when using equal steps charging, we first have to write down an expression for total energy E_R dissipated by the resistor:

$E_{R,tot} = ½\ CV^2(V_1^2 + (V_2\text{-}V_1)^2 + (V_3\text{-}V_2)^2 + \ldots\ldots + (V_{DD}\text{-}V_{N\text{-}1})^2$

Then we take the partial derivative of E_R respective to each of the variables, then set all the derivatives to be zero:

$dE_R/dV_1 = 2(V_1) - 2(V_2\text{-}V_1) = 0 \Rightarrow \mathbf{V_2 = 2V_1}$
$dE_R/dV_2 = 2(V_2\ \text{-}V_1) - 2(V_3\text{-}V_2) = 0 \Rightarrow \mathbf{V_3 = 2V_2 – V_1 = 3V_1}$ (V_2=$2V_1$ from above)
$dE_R/dV_3 = 2(V_3 – V_2) - 2(V_4\text{-}V_3) = 0 \Rightarrow \mathbf{V_3 = 2V_3 – V_2 = 4V_1}$ (V_2 and V_3 from above)
…
…
…
$dE_R/dV_{N\text{-}1} = 2(V_{N\text{-}1} – V_{N\text{-}2}) - 2(V_{DD}\text{-}V_{N\text{-}1}) = 0 \Rightarrow \mathbf{V_{DD} = 2V_{N\text{-}1} – V_{N\text{-}1} = NV_1}$

so we have:

$$\mathbf{V_1 = V_{DD}/N}$$

From the above derivations, we can conclude that we have to increase the supply voltage by V_1=V_{DD}/N for each step in order to get the minimum energy consumption. This is equivalent to the N equal steps charging.

c) Consider now the linear increase of the supply voltage, illustrated in Figure 1c. Assume that T >> τ = RC and *in* = 0. How much energy does it take to fully charge the capacitor (to V_{DD})? Can you intuitively explain your result? Why is it no longer valid without the assumption that T >> τ? Verify your result using your result in (b).

Since τ >> T, the voltage drop across capacitor equals the supply voltage (V) at every point of the charging process. Hence, its charge is Q = CV.
The energy required to charge the capacitor fully is:

$$E = \int_0^\infty VIdt = \int_0^\infty V\frac{dQ}{dt}dt = \int_0^{V_{DD}} CVdV = \frac{1}{2}CV_{DD}^2$$

Since this energy equals the energy stored in the capacitor, **no energy is lost in the resistor!** This can be intuitively explained by the fact that the capacitor is charged so slowly that the capacitor voltage is always equal to the supply voltage and hence the voltage drop (and the current) across the resistor are 0.
If the assumption $T >> \tau$ doesn't hold, the voltage across the capacitor cannot follow the rapid increase of the supply voltage, and is thus smaller than the supply voltage. Hence, our assumption about zero voltage across the resistor is no longer valid.

The same result can be derived using the expression for N steps determined in (b), considering that, as the number of steps increases, the supply voltage characteristic approaches the linear increase of Figure 1c. For infinite number of steps, the characteristic is identical to that of the linear case and the formula derived in (b) gives:

$\lim(N\to\infty)\ E_{total} = CV_{DD}^2/2$

As an interesting aside, the charge could be saved somewhere else when the load capacitance is discharged. Since the charge was saved, a switch could be thrown and everything could run backwards to restore the charge to its original configuration. In this way, the computer consumes almost no energy, and this is the fundamental premise behind 'reversible computers'.

Problem #2: Process Scaling

A not very state-of-the-art embedded microprocessor from a company outside the valley consumes 0.72mW/MHz (excluding leakage power) when fabricated using a 0.18 μm process. With typical standard cells (gates), the area of the processor is 2 mm^2. Assume a 600MHz clock frequency, and 1.8V power supply. Its leakage power is 50μW. Assume short channel devices, but ignore second order effects like mobility degradation, series resistance, etc.

a) Power density is important for cooling the chip and packaging. Scale the circuit so, that the power density decreases to 150mW/mm^2 but the current density remains constant. What is the new frequency of the circuit?

Using Table 3.8 of the book, we get (using general scaling to satisfy both requirements):

$$\frac{(I/A)_{new}}{(I/A)_{prev}} = \frac{S^2}{U} = 1 \Rightarrow S^2 = U$$

$$\frac{(P/A)_{new}}{(P/A)_{prev}} = \frac{S^2}{U^2} = \frac{1}{U} \Rightarrow U = \frac{(P/A)_{prev}}{(P/A)_{new}} = \frac{(0.72mW/MHz)\times 600MHz/2mm^2}{150mW/mm^2} = 1.44$$

Hence, S can now be determined:

$$S = \sqrt{U} = 1.2$$

The frequency is inversely proportional to the delay, thus:

$$f_{new} = Sf_{prev} = 720MHz$$

b) Go now back to the original processor and calculate the scaling required for the circuit to dissipate 0.54mW/MHz. If there are many ways to do that, choose the one that gives maximum frequency without affecting the power density more than 20%. What is the die area of the new circuit?

We use general scaling to satisfy both requirements. For maximum frequency, we choose the maximum possible *S*, given the power requirements:

$$\frac{(P/A)_{new}}{(P/A)_{prev.}} = \frac{S^2}{U^2} = 1.20$$

$$\frac{(P/f)_{new}}{(P/f)_{prev}} = \frac{1}{SU^2} \Rightarrow \left(\frac{S^2}{U^2}\right)\frac{1}{S^3} = \frac{1.2}{S^3} = \frac{0.54mW/MHz}{0.72mW/MHz} \Rightarrow S = 1.17$$

The die area of the circuit is derived as follows:

$$A_{new} = \frac{A_{prev.}}{S^2} = \frac{2mm^2}{1.17^2} = 1.46mm^2$$

$$f_{new} = Sf_{prev} = 702MHz$$

c) If the threshold voltage in the 0.18μm process is 0.4V, what should be the threshold voltage in 0.13μm process with 1.3V supply voltage? Assuming 90mV/dec subthreshold slope, what would be the leakage power of the new processor?

Leakage Power = $V_{DD}I_{leak}$

$$\frac{V_{T,new}}{V_{T,prev.}} = \frac{1}{U} \Rightarrow V_{T,new} = 0.4V\left(\frac{1.3V}{1.8V}\right) = 0.289V$$

$$I_{leak,pre} = \frac{50\mu W}{1.8V} = 27.78\mu A$$

$$I_{leak,new} = I_{leak,prev}\left\{10^{(V_{T,prev}-V_{T,new})/(Subtreshold\_Slope)}\right\} = 27.78\mu A\left\{10^{(0.4-0.289)/0.09}\right\} = 475\mu A$$

Then we have the new leakage power:

$P_{leak,\,new}$ = (475μA)(1.3V) = 618μW

Problem #3: Propagation Delay and Energy

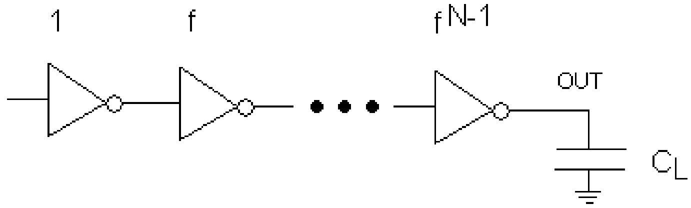

Figure 3a: Progressively Sized Inverter Chain

a) What is the delay of a minimum sized inverter driving another inverter *f* times its size? For the minimum sized inverter, assume input capacitance equal to C_{unit}, equivalent resistance through the NMOS or PMOS equal to R_{unit}, and intrinsic (self-loading) capacitance on the output also equal to C_{unit}. Assume that the capacitance and resistance values scale linearly with size. Your answer will be in terms of these parameters (no calculations!). Take the limit as f goes to 0 and call the result τ_{inv}.

The delay for the inverter depends on its equivalent resistance and the capacitive load on its output, which is the sum of its intrinsic capacitance and the input capacitance of the next stage.

$\tau_p = \ln(2)R_1(C_{int,1}+ C_{in,2}) = 0.69\ R_{unit}\ (C_{unit}+ f^*C_{unit}) = 0.69R_{unit}C_{unit}(1+f)$

$\tau_{inv} = 0.69R_{unit}C_{unit}$

b) From part a), how much energy is consumed by the driving inverter after successive low to high (L→ H) and high to low (H→L) transitions, in terms of a supply voltage Vdd?

This should be a familiar result by now:

$E = C_{tot}V_{dd}^2 = C_{unit}(1+f)V_{dd}^2$

c) Consider the chain of N progressively sized inverters shown in Figure 3a (the first is minimum sized). If the output load C_L=40C_{unit}, what sizing factor f would minimize the total delay for a chain of N=5 inverters? Find the total delay of this chain in terms of τ_{inv}.

As discussed in lecture and the notes, minimum delay results when the inverters are sized so that each stage bears the same "effort", such that the delay is evenly distributed. Suppose the third inverter is g times larger than the second inverter. The second stage delay can be computed as

$\tau_{p2} = \ln(2)\ R_2\ (C_{int,2}+ C_{in,3}) = 0.69(R_{unit}/f\ )(f^*C_{unit}+ g^*f^*C_{unit}) = 0.69\ R_{unit}C_{unit}(1+g)$

where we have used the fact that the equivalent resistance of the second inverter is a factor of f smaller than that of the first, its intrinsic capacitance is a factor of f larger, and the input capacitance of the third inverter is g times that of the second. Comparing this result to that from part (a), we see the second stage delay is equal to the first if we use the same sizing ratio, or f=g.

Each stage delay is then just

$\tau_{pi} = 0.69\ R_{unit}C_{unit}(1+f) = (1+f)\ \tau_{inv}$

Continuing the inverter chain analysis, the output load C_L (=40C_{unit}) looks like an inverter with input capacitance 40 times the minimum sized one. Therefore, for a chain of five inverters, we require:

$f = (40)^{1/5} = 2.09$

This corresponds to progressive sizing {1, 2.09, 4.37, 9.15, 19.13}, such that the load capacitance is twice the input capacitance of the fifth inverter.

The total delay is then:

$\tau_{tot} = N\ (1+f)\ \tau_{inv} = 5\ (1+2.09)\ \tau_{inv} = 15.45\tau_{inv}$

d) Find the optimum number of inverters and sizing ratio for the output load specified in Part (c). Express the optimum delay again in terms of τ_{inv}. Considering your result for Part (b), do you think this inverter chain will consume more or less energy than a single inverter driving the output load?

For this problem its easiest just to plug in numbers for a few values (with $f=40^{1/N}$ and delay proportional to N(1+f) as in part (c)). The optimal number of inverters is then 3, with sizing ratio 3.42, corresponding to a delay $13.26\tau_{inv}$.

N	F	N(1+f)
1	40	41
2	6.32	14.64
3	3.42	13.26
4	2.41	14.06
5	2.09	15.45

If you play around with the size of the load, you will see that the **N which corresponds to f closest to 4** results in the smallest delay. That is why we say that a fanout of 4 (FO4) is typically best.

You can also directly find N that give f=4:

$$4 = (F)^{1/N} \Rightarrow N = \frac{\ln(F)}{\ln(4)} = 2.66$$

At this point, the easiest way to determine whether to use N=2 or N=3 is to compute the delay in both cases and compare the results.

Regardless of the number of inverters in the chain, the final load capacitance C_L has to be switched, contributing $C_L V_{dd}^2$ to the total energy. To minimize the total energy, we then would want to minimize the transistor contribution to the capacitance, which would mean using a single minimum sized inverter. As with many minimum energy solutions, this comes at a significant delay penalty.

Problem#4: Buffer Sizing

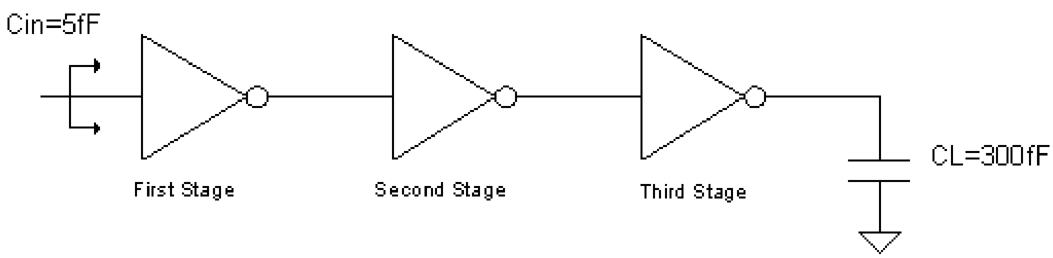

Figure 4a: Buffer Chain

In order to drive a large capacitance (C_L=300fF) from a minimum size gate (with input capacitance C_{in}= 5fF), you decided to introduce a two-stage buffer as shown in Fig. 4a. Assume that the propagation delay of a minimum size inverter is 20ps (for minimum size inverter delay in this case, there is only intrinsic (self-loading) capacitance on the output, which is equal to input capacitance –there is no other output loading). Also assume that the capacitance and resistance values scale linearly with size.

a) Determine the sizing of the two additional buffer stages that will minimize the propagation delay. What is the corresponding propagation delay?

This one is pretty straightforward:

$$F = \frac{300\,fF}{5\,fF} = 60$$

$$f = (60)^{1/3} = 3.915$$

Then we have:

$$C_{in,2} = 3.91 \times 5\,fF = 19.57\,fF$$

$$C_{in,3} = (3.915)^2 \times 5\,fF = 76.63\,fF$$

The total delay is:

$$t_p = N t_{p0}(1+f) = 3 \times 20\,ps \times (1 + 3.915) = 295\,ps$$

b) Given a supply voltage of 2.5V and activity factor of 0.5, what is the average energy-delay product of the circuit in part (a)?

The average energy per transition is:

$$E = \alpha\left(\sum C\right)V_{DD}^2 = 0.5 \times \left[2 \times \left(5fF + 19.57fF + 76.63fF\right) + 300fF\right] \times \left(2.5V\right)^2 = 1.57pJ$$

Note that the factor 2 in the equation above come from the fact that there are both input (gate) and output (self-loading) capacitances for each inverter.

So the average energy-delay product is:

$$EDP = 1.57pJ \times 295ps = 463 \times 10^{-24} J \cdot s$$

Please note that the 5fF input capacitance of the first inverter might be excluded from the energy calculation since it actually loads the preceding stage.

c) Determine the sizing of the two buffer stages that will minimize the average energy per transition while maintaining the propagation delay within 10% of the minimum value from part (a). For simplicity, assume that sizes are increasing in geometric fashion (1, f, f^2). This means that effective fanout of the first two stages (Cin2/Cin and Cin3/Cin2) is f, while the effective fanout of the last stage is equal to CL/Cin3. What is the new average energy-delay product?

In order to get the minimum energy-delay product, we want to make f to be as low as possible. This can be done to make all the buffers to be minimum sized and then f=1; however, we also have a delay constrain which limit the minimum acceptable value of f.

First of all, we have to write down an expression for the delay:

$$t_p = t_{p0}\left[\left(1 + \frac{C_{in,2}}{5fF}\right) + \left(1 + \frac{C_{in,3}}{C_{in,2}}\right) + \left(1 + \frac{300fF}{C_{in,3}}\right)\right]$$

Since $C_{in,2}/5fF = C_{in,3}/C_{in,2} = f$, we get:

$$t_p = 20ps\left[3 + 2f + \frac{300fF}{50fF \times f^2}\right] = 20ps\left[3 + 2f + \frac{60}{f^2}\right]$$

We want this delay to be within 10% of the minimum delay from part (a):

$t_{p,\min} \le t_p \le 1.1t_{p,\min} = 1.1 \times 295ps = 325ps$ (Note that $t_p < t_{p,min}$ is impossible)

Then we have:

$$295ps \le 20ps\left[3 + 2f + \frac{60}{f^2}\right] \le 325ps \Rightarrow 11.75 \le 2f + \frac{60}{f^2} \le 13.25$$

If you take the derivative of t_p respective to f, you will find out that t_p always increases as f decreases for $0 < f < 3.915$. This means we have to find f at the upper boundary of the equation above:

$$2f + \frac{60}{f^2} = 13.25$$

By either using a math solver or doing some iterations, we finally get:

$$f = 2.8$$

$$C_{in,2} = 2.8 \times 5fF = 14fF$$

$$C_{in,3} = (2.8)^2 \times 5fF = 39.2fF$$

The new average energy per transition is:

$$E = \alpha\left(\sum C\right)V_{DD}^2 = 0.5 \times \left[2 \times (5fF + 14fF + 39.2fF) + 300fF\right] \times (2.5V)^2 = 1.30pJ$$

So the new average energy-delay product is:

$$EDP = 1.30pJ \times 325ps = 423 \times 10^{-24} J \cdot s$$

d) Now go to the circuit in Fig.4b. Suppose that there is a 50fF wire capacitance between the second stage and the third (output) stage. What are the sizes of the two buffers in this case that will minimize the propagation delay?

The formula $f = (F)^{1/N}$ can not be used in this case since there is an interstage loading. To solve this problem we first have to write down an expression for the total delay.

$$t_p = t_{p0}\left[\left(1 + \frac{C_{in,2}}{5fF}\right) + \left(1 + \frac{50fF + C_{in,3}}{C_{in,2}}\right) + \left(1 + \frac{300fF}{C_{in,3}}\right)\right]$$

Take the partial derivatives of tp respective to $C_{in,2}$ and $C_{in,3}$ and set to zero:

$$\frac{\partial t_p}{\partial C_{in,2}} = 0 \Rightarrow \frac{1}{5fF} - \frac{50fF + C_{in,3}}{C_{in,2}^2} = 0$$

$$\frac{\partial t_p}{\partial C_{in,3}} = 0 \Rightarrow \frac{1}{C_{in,2}} - \frac{300fF}{C_{in,3}^2} = 0$$

Solve the two equations above using math solver or doing iterations, we get

$$C_{in,2} = 26.4fF$$

$$C_{in,3} = 88.9fF$$

This gives the total delay of:

$$t_p = 20ps\left[\left(1+\frac{26.4fF}{5fF}\right)+\left(1+\frac{50fF+88.9}{26.4fF}\right)+\left(1+\frac{300fF}{88.9fF}\right)\right]=339ps$$

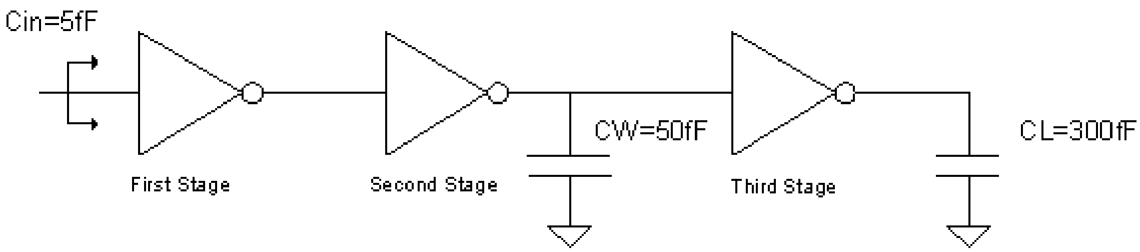

Figure 4b. Buffer Chain with Interstage Wire Capacitance