Answers to EEE310, 2001-02, Questions 1… 4

1.

Show that a simple expression for the dynamic power dissipation of a CMOS circuit due to switched capacitance is:

$$P_{sw-cap} = aC_{load}V_{DD}{}^2 f$$

Consider an input of capacitance C_{load} being driven at frequency f. Each time the input rises, the capacitor is charged to V_{DD} and the total charge on the capacitor will be $C_{load}V_{DD}$. When the input is driven low, this charge will discharge to ground. Hence in time $1/f$, a charge equal to $C_{load}V_{DD}$ is moved from V_{DD} to ground. Thus, the net current flow is $C_{load}V_{DD}f$ and the power dissipated is $C_{load}V_{DD}{}^2f$. Now, we can make the assumption that the input, as most inputs in a circuit would, does not switch at f but will switch with a lower probability, a. Thus, the effective power dissipation will be $aC_{load}V_{DD}{}^2f$.

Identify the other two major sources of power dissipation in a CMOS circuit.

The other two sources of power dissipation are:

- crowbar or short-circuit current which occurs because, during switching, both *n* and *p* FETs in a gate will be on momentarily so this is another frequency-dependent dissipation. Typically, with good design, this dissipation may only be 20% of the total and is usually minimised by keeping transitions short and matching input and output rise/fall times.

- leakage current which flows even when the transistors are in cut-off. This current exists even if the circuit is not switching and scales with the number of gates on and IC.

Estimate the power dissipation of the IC.

There are 5M 2-input gates and so there are 10M inputs. Each input consists of 1n and 1p transistor. Given that each gate is equivalent to a NAND gate, the gate consists of 2 stacked n FETs and 2 p FETs in parallel. The presumption is that the n transistors will be 2x minimum n width and the p FETs will be minimum p width (=2x n width due to difference in mobility). Hence, each input will look like a capacitor equivalent to area 0.18µm x (0.8µm+0.8µm).

The capacitance of each input will be $3.45\text{x}10^{-11}\text{x}0.18\ \text{x}10^{-6}\text{x}1.6\ \text{x}10^{-6}/10\ \text{x}10^{-9}$=0.994fF. There are 10M inputs and so the total gate capacitance is 9.94nF

Each gate output is connected to three drains (2p + 1n) and we assume that ½of the gate capacitance appears at the drain. There appears, therefore, to be a capacitor of width $0.5\text{x}2.4\text{x}10^{-6}$m and length $0.18\text{x}10^{-6}$m. So each output has capacitance 0.75fF. There are

5M gates and the total capacitance on each output will be 3.75nF. In truth, the effective capacitance will be different because the output and input capacitors are not grounded but connected and the capacitance at the drain varies with the state of the FETs. However, treating them as independent will give a 1st order estimate.

There is 1000m of interconnect and this is equivalent to a capacitor of area $1000mx(0.5+0.5+0.8+0.8)x10^{-6}m = 2.6x10^{-3}m^2$. The total capacitance of this interconnect is 208nF.

Thus, the total switched capacitance on the IC is 221.7nF.

Each node toggles with a probability of 0.125 and, hence, the average number of clock cycles required to charge and discharge a node is 16.

Consequently one over the time taken to charge and discharge this capacitance once on average is $af = 600x10^6/16=37.5x10^6$.

V_{DD}=1.8V and so $P_{sw\text{-}cap}=37.5x10^6x1.8^2x221.7x10^{-9}=29.94W$

$P_{total}=P_{sw\text{-}cap}/0.8 = 33.67W$.

2.

Discuss the factors that affect the manufacturing cost of a packaged ASIC and how these factors are combined to yield the overall cost.

The first part is bookwork. The cost of manufacturing splits down as follows:

The cost of a processed wafer (determined mainly from fab finance/running costs and wafer throughput – maybe 20000 per month) - W

The number of die per wafer (ignoring edge effect this will be the area of the wafer divided by the area of the IC) – A_W/A_{IC}

The systematic yield of the process – that is the proportion of wafers that are within normal process limits - Y_S

The yield of ICs due to random defects. There are a number of different yield models (Seeds, Murphy, ITRS) and it related defects / m^2 (maybe 1350) and A_{IC} to yield:

e.g. $Y_r = e^{-\sqrt{A_{IC} \cdot DefectDensity}}$

The working ICs will be packaged and this will be subject to further yields – dicing and attaching – Y_{att} and bonding Y_{bond} where this is raised to the power of the number of pins to be bonded (this might be related via Rents Rule to $kÖA_{IC}$). Additionally, there will be the cost of packaging which might be expressed simply as a multiplier of the number of

pins (neglecting the transitions between different package types as pin counts increase) – $k_{pin}\ddot{O}A_{IC}$. Based on this, a *simple* expression for cost might be:

$$CostPerPart = \frac{W}{Parts} + \frac{k_{pin}\sqrt{A_{IC}}}{Y_{att}Y_{bond}{}^{k\sqrt{A_{IC}}}}$$

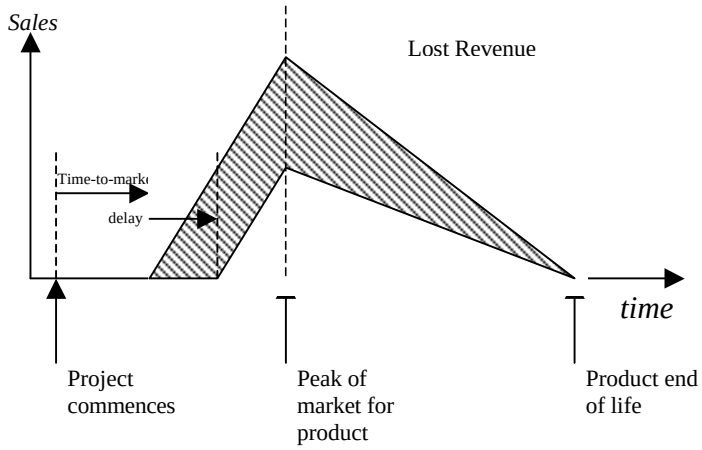
$$Parts = \frac{A_W}{A_{IC}} Y_s . e^{-\sqrt{A_{IC} \cdot DefectDensity}} Y_{att} Y_{bond}{}^{k\sqrt{A_{IC}}}$$

There will, of course, be a testing cost, and dependent on whether this is done in-house or is contracted-out this might incur an additional cost. However, assuming that it is done in-house, this might be factored into the cost of financing the fab (it is capital equipment after all) and will be reflected in *W*. However, what is certainly true is that unless testability is designed in from the outset, the cost of testing can be far in excess of what might have been required. In extreme cases, the IC may well be untestable and, hence, not marketable.

Time-to-market is a critical in the introduction of an ASIC to the market – explain, simply, why this is the case?

Time to market is critical because it affects overall revenue derived from selling an ASIC. This is based on the presumption that there is a point in time when the market for the ASIC will peak and this is influenced by a range of external factors (such as the readiness of the market, what other competitors are doing, etc).

It is assumed that sales will ramp up at a fixed rate and that, after the peak in the market, will decline until a point in time when the product is deemed to have reached the end of its life. Any delay in introduction of the product will result in lost sales as shown here and this equates directly to lost revenue against costs that are, essentially, fixed.



How is the semiconductor industry seeking to minimise the time-to-market and what factors make this difficult?

The semiconductor industry is addressing the time to market problem by a range of measures most of which hinge upon increased automation. The use of HDLs and RTL (behavioural?) synthesis promotes design at a highish level of abstraction, design migration, and design reuse. SoC and reusable blocks, common standards and 3rd party providers promotes the integration of large blocks within designs and rapid design of

complex systems. Design for Testability and the use of things such as scan testing and BIST ensures that manufactured devices can be tested (adequately?). However, as devices grow, the cost of designing these devices is expected to increase rapidly – even with these measures in place and the need to verify designs before manufacture is consuming a greater and greater proportion of the overall design time (50-60% currently). Additionally, design productivity itself is required to improve exponentially – even with the increased automation in place. This is one of the greatest challenges facing the design industry overt the next 5-10 years.

Find the volume of production at which it would be better to switch from FPGA to MGA and from MGA to CBIC and find the overall cost per ASIC at these volumes of production.

The design is 1M transistors and assuming that the project length is 1 year. For a CBIC each designer can output 91K transistors / year. Thus, 11 people are required to design the ASIC and the cost will be $5.5M (assuming $500K / designer/year – including salaries, on costs, CAD and support costs). For an MGA, the net effort will be 10/1.365 = 7.32 designer years and this will cost $3.66M. For the FPGA, the net effort would be 10/2.184 = 4.58 designer years and the design cost would be $2.29M.

The total cost of design and manufacturing per device will be:

(DesignCost + NRE) /volume + part cost (volume >= min order quantity)
(DesignCost + NRE + part cost*min order) / volume (volume < min order quantity)

Because there are different values based on quantity, the best idea is to make an assumption about the quantity, calculate a result and then check that the assumption was justified.

The break-even volume between FPGA and MGA is when:

($2.29M + 0)/vol1 + 1000 = ($3.66M+50000 + 10*20000)/vol1

vol1 = (3660000-2290000+50000+10*20000)/1000 = 1620 parts

It is reasonable to assume that the break-even point will be well below the minimum order quantity for the MGA.

At this point the cost of manufacturing and designing each device is $2413.

The break-even volume between MGA and CBIC, assuming that the quantity is below the CBIC minimum order quantity, is when

($3.66M + 50000)/vol2 +10=($5.5M+200000 + 5*50000)/vol2

(5500000-3660000+200000-50000+250000)/10= vol2 = 224000.

and this is above the minimum order point. So …

($3.67M + 50000)/vol2 +10=($5.5M+200000)/vol2+5

(5500000-3660000+200000-50000)/5= vol2 = 398000.

At this point, the cost of manufacturing and designing each device is $19.32.

This assumes that the CBIC is above its minimum order quantity.

Calculate the best, overall cost per ASIC if the volume of production is 1 million devices.

The cost per device when the order quantity is 1M would be:
(5500000+200000)/1000000+5 = $10.7

3.

Derive an expression for the value of voltage at which the input and output value of an inverter are equal using the standard equations (and terms) that link V_{GS}, V_{DS}, and I_{DS}.

Equating drain currents and setting the gate and drain/source voltages to V_{IN} for the *n*FET and V_{IN}-V_{DD} for the *p*FET:

$$I_{DSN} = -I_{DSP} = \frac{b_N}{2} \cdot (V_{IN} - V_{TN})^2 = \frac{b_P}{2} \cdot (V_{IN} - V_{DD} - V_{TP})^2 = \frac{b_P}{2} \cdot (V_{TP} + V_{DD} - V_{IN})^2$$

From this (picking the appropriate root of the equation),

$$\sqrt{\frac{b_N}{b_P}} \cdot (V_{IN} - V_{TN}) = (V_{TP} + V_{DD} - V_{IN})$$

Hence:

$$V_{IN} \cdot \left(1 + \sqrt{\frac{b_N}{b_P}}\right) - \sqrt{\frac{b_N}{b_P}} \cdot V_{TN} = V_{DD} + V_{TP}$$

$$V_{IN} = \frac{V_{DD} + V_{TP} + \sqrt{\frac{b_N}{b_P}} \cdot V_{TN}}{\left(1 + \sqrt{\frac{b_N}{b_P}}\right)}$$

If we assume that V_{TP}=-V_{TN} and that the transistors are sized so that b_P=b_N then:

$$V_{IN} = \frac{V_{DD}}{2}$$

and, given that T1 and T2 have identical gains and operating conditions:

$$V_{OUT} = \frac{V_{DD}}{2}$$

What is metastability and what is its effect on a circuit?

Metastability is a problem that occurs in samplers when the data transitions are asynchronous with respect to the sampling clk or enable. During sampling, two inverters are connected one feeding-back to the other. If the data at the input at the instant that this

inverters are connected lies at some arbitrary voltage as it would be if the input was transiting at this point in time then the voltage at the output of the inverter pair will head exponentially towards the supply voltage dictated by whether the input voltage is above or below the voltage at which the inverter pair is perfectly balanced. If the time between sampling and observing the output voltage is too small then the output may be outside the logic high/low voltage margins with a finite probability – this is an upset.

The effect of metastability is characterised by $upsets/s = T_0 e^{-\frac{t_r}{t_c}} f_{clk} f_{data}$
Explain what the terms mean and from where they arise.

t_c is related to the 1/GBW of the fed-back inverter pair, t_r is the time between sampling and observation, T_0 is a constant related to the technology, circuit, and layout of the sampler. f_{clk} is the clock frequency whilst f_{data} is the effective frequency of the data input.

Explain how the effect of metastability can be reduced and how this is achieved in practice.

The main way of reducing the effect of metastability is to increase the observation time. A simple sampler is a flip-flop and putting two flip-flops in series increases the latency and the observation time. This seems odd but if the first flip-flop is in a metastable state from which the output voltage relaxes back to a stable state, the second flip-flop may well sample the metastable output at a value closer to its stable state after one clock cycle and the value continues to relax in the second flip-flop from its new value – so, effectively, the effect of adding another flip-flop is the same as waiting longer after sampling before observing.

An IC in a system simply samples an external 8-bit data bus to read data that is asynchronous to the clock of the IC. The clock frequency of the IC is 50MHz and the effective frequency of the data is 23MHz. T_0*=5ns, and* t_c*=1.5ns: Calculate the upsets per second.*

The probability of an upset/s/clk/data on each line is $p=5\text{x}10^{-9}e^{-20/1.5}=8.1\text{x}10^{-15}$ upsets/s/clk/data (The observation time is 20ns i.e. one clock cycle).

There are 8 lines in total and so the probability of an upset on any of the 8 lines should be $(1-(1-p)^8) =6.47\text{x}10^{-14}$ upsets/s/clk/data on the data bus.

So the total number of upsets/s at the stated clock and data frequencies should be $23\text{x}10^6\text{x}50\text{x}10^6\text{x}6.47\text{x}10^{-14}$ upsets/s = 74.5

It is noticed that this value is far too high. The designer also notes that 10000 such systems must be deployed and that only one failure every five days (due to metastability) across the systems deployed can be tolerated. What would the design of a metastable-resistant sampler be that would meet this requirement?

The sampler can consist of 1, 2, 3 … n flip-flops in series. There will be 74.5 u/s for 1 flip-flop or, across 10000 systems, there will be a total of 745000 u/s. With 2 flip-flops, the observation time is increased to 40ns and the probability of an upset on each line will be $5x10^{-9}e^{-40/1.5}=15.1x10^{-6}$ upsets/s/clk/data or 1.21 u/s across the deployed systems. With 3 flip-flops, the observation time is increased to 60ns and the probability of an upset on each line will be $5x10^{-9}e^{-60/1.5}=2.12x10^{-26}$ upsets/s or $1.95x10^{-6}$ u/s across the deployed systems. This is 1 upset every 511700 seconds or 1 upset every 5.9 days and this meets the specification.

4.
A logic gate is to be designed to implement the function:

$$Y = \overline{A.(B+C)}$$

Show how a standard CMOS circuit would be designed to implement this function. You should draw a schematic of the circuit and your answer must consider the sizing of transistors.

For the pull down network, we need to find a function that is the inverse of *Y* and we can only have non-inverted terms on the *rhs* or inverters will be needed on the input/output.

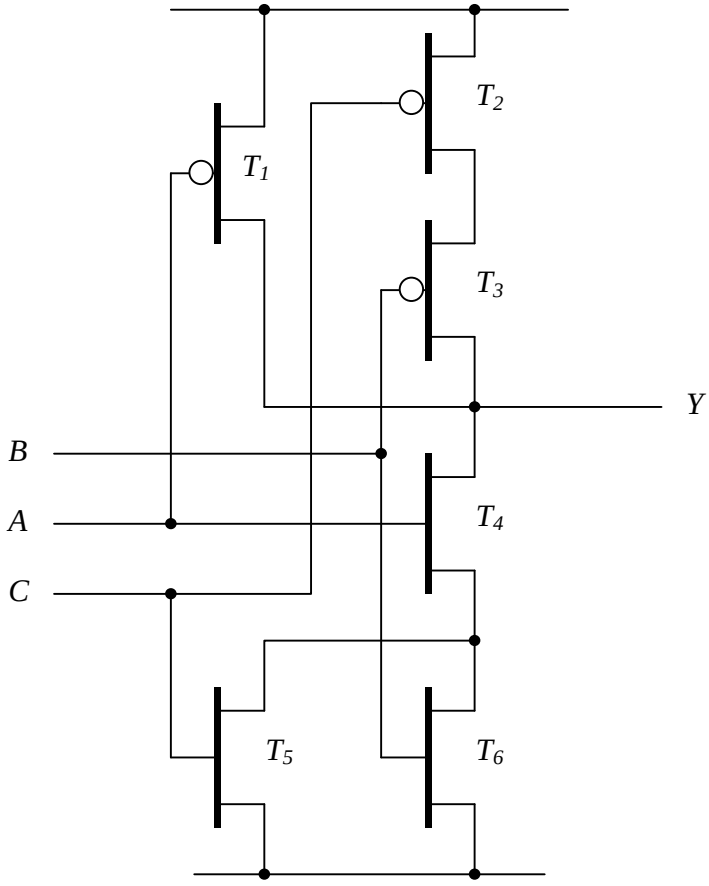
$$Y = \overline{A.(B+C)}$$
$$\overline{Y} = A.(B+C)$$

This implies a FET controlled by *A* in series with the parallel combination of FETs controlled by *B* and *C*.

For the pull-up network, we need to find the function formed by inverting the terms on the *rhs*. That is,

$$Y_{up} = \overline{A}.(\overline{B}+\overline{C})$$
$$Y_{up} = A+\overline{(\overline{B}+\overline{C})}$$
$$Y_{up} = A+B.C$$

That is a FET controlled by *A* in parallel with a series combination of FETs controlled by *B* and *C*.



Estimate the switching time of the unloaded output of this gate.
(m_E=0.08m^2/Vs, m_H=0.04m^2/Vs, t_{OX}=10nm, W_{min}=1mm, L_{min}=0.25mm, e=3.45x10^{-11}F/m, V_{DD}=3.3V, V_{TN}=-V_{TP}=0.6V)

Assuming that b_N=2b_P and the minimum sized *n*FET can be modelled by a resistor R_o and that we want the worst case resistance between *Y* and any power supply to be R_o. T_1 can be 2x minimum width which will give it a resistance equal to R_o but $T_{2/3}$ are in series and must be 4x minimum width to give an overall resistance of R_o. Similarly, the resistance to ground will be set by T_4 in series with either T_5 or T_6 and so if all of the transistors are 2x minimum width then the worst case resistance will be R_o.

We can estimate R_o as:

$$R_o = \frac{2L_{\min}t_{OX}}{m_E e W_{\min}(V_{DD} - V_T)}$$

and given the minimum size of an *nFET*, the resistance between the output node and each supply is, worst case:
R_o=2x2.5x10^{-6}x10x10^{-9}/(0.08x3.45x10^{-11}x1x10^{-6}x(3.3-0.6))=671Ω.

C_{gate} for minimum sized transistor = 3.45x10^{-11}x1x10^{-6}x0.25x10^{-6}/10^{-8}=8.625x10^{-16}F.

We can estimate the capacitance at *Y* as being:
0.5$C_{gate|T3}$+0.5 $C_{gate|T1}$+0.5 $C_{gate|T4}$=0.5x(4+2+2)x8.625x10^{-16}F=3.45x10^{-15}F.

The rise/fall time will be ~2.2C_Y.R_o=2.2x3.45x10^{-15}x671=5.1ps.

A logic gate produces an output signal that must drive a large capacitance, C_{load}, but, obviously, a small gate will suffer large rise/fall times and delays. How can this delay be minimised.

The delay associated with driving a large load can be minimised by employing a sequence of inverters of increasing size. Consider a sequence of *N* inverters each one *k* times wider than the previous one. The objective is to find the values of *N* and *k* that minimise the overall delay from input to output.

The i^{th} inverter in the series drives an input capacitance of k^iC_{in} in the following gate and is of width $k^{i-1}W_{min}$. Consequently, delay of this stage is:

$$\frac{4k^iC_{in}}{k^{i-1}b(V_{DD} - V_T)} = \frac{4kC_{in}}{b(V_{DD} - V_T)}$$

That is, independent of the position in the sequence: the delay of each stage is equal. Thus, the overall delay through *N* stages is:

$$\frac{4kC_{in}(N-1)}{b(V_{DD} - V_T)} + \frac{4C_{load}}{k^{N-1}b(V_{DD} - V_T)}$$

where the second term represents the final stage that drives the load capacitance. We need to solve this for *k* and *N* and one other piece of information is needed. To keep the matching of the delay to its load constant, we can impose the boundary condition that:

$$C_{load} = k^N C_{in}$$

This also simplifies the expression for delay to:

$$\frac{4kC_{in}N}{b\left(V_{DD} - V_T\right)}$$

From the first of these expressions we can find that:

$$N = \frac{\ln\left(\frac{C_{load}}{C_{in}}\right)}{\ln(k)}$$

and the expression for delay becomes:

$$\frac{4kC_{in}}{b\left(V_{DD} - V_T\right)\ln(k)} \cdot \ln\left(\frac{C_{load}}{C_{in}}\right)$$

differentiating this expression *w.r.t. k* and setting equal to 0 yields:

$$\frac{4C_{in}}{b\left(V_{DD} - V_T\right)} \cdot \ln\left(\frac{C_{load}}{C_{in}}\right) \cdot \frac{1}{\ln(k)} - \frac{4C_{in}}{b\left(V_{DD} - V_T\right)} \cdot \ln\left(\frac{C_{load}}{C_{in}}\right) \cdot \frac{k}{\ln(k)^2} \cdot \frac{1}{k} = \frac{4C_{in}}{b\left(V_{DD} - V_T\right)} \cdot \ln\left(\frac{C_{load}}{C_{in}}\right) \cdot \left(\frac{1}{\ln(k)} - \frac{1}{\ln(k)^2}\right) = 0$$

from this we can find that:

$$\ln(k) = 1$$

$$k = e^1 = 2.71828$$

Therefore,

$$k = e^1 = 2.71828$$

$$N = \ln\left(\frac{C_{load}}{C_{in}}\right)$$