DEPARTMENT OF ELECTRONIC AND ELECTRICAL ENGINEERING

Answers to EEE310/6036 (Introduction to VLSI Design/VLSI Design)

1. *The schematic in **Figure 1** is the pull-up network for a CMOS circuit.*

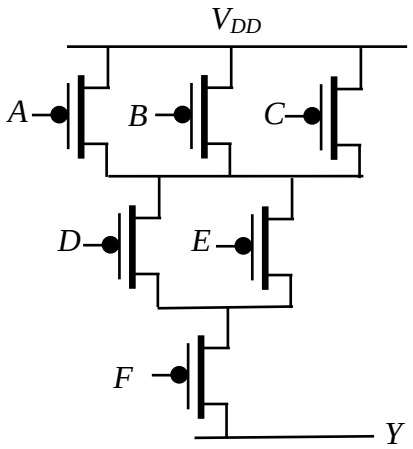
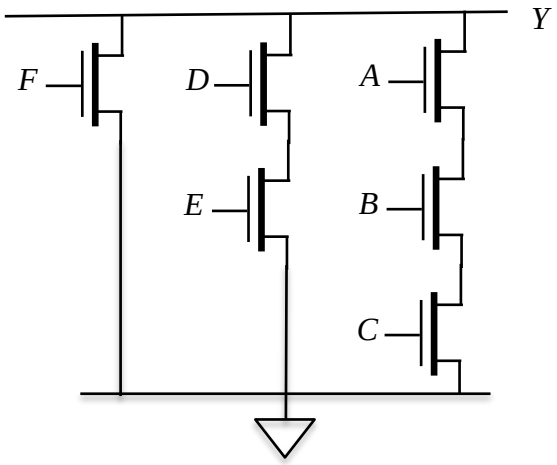


Figure 1: Pull-up Network

a. *Draw the corresponding pull-down network*



(5)

b. *Write down the function, Y, in terms of the inputs A, B, C, D, E and F.*

$$Y = \overline{ABC + ED + F}$$

(5)

c. *Size the transistors (in both the pull-up and pull down network) assuming that this is a minimum-sized gate and that $\mu_H = 0.4\mu_E$.*

Looking at the nFETs:

A,B,C = 3

D,E = 2

F = 1

Looking at the pFETs

A,B,C,D,E,F = 3/0.4 = 7.5

(5)

d. *Is there any significance in the ordering of the transistors i.e. A, B, C closest to the supply voltage; F closest to the output?*

In terms of functionality – no: the ordering is irrelevant. However, there are other issues that are affected. The capacitance seen at the output *Y* as a consequence of the transistors will vary with the ordering of the pFETs and the ordering given is probably the best. **(5)**

2. **a.** *Explain the problem with interconnect in integrated circuits as technology shrinks.*

Answer should cover:

Increasing clock frequency – reduced critical paths

Local lines shrink with the technology and nearly keep pace with the clock.

Chip size remains the same and now global spanning wires are electrically much longer.

Additionally, distributed resistance on such wires will increase and capacitance may not reduce to compensate.

Consequence, the IC cannot be globally synchronous without steps being taken to counter that fact that it may take multiple clock cycles for a signal to cross the chip from one side to the other.

(4)

b. *What is the optimum way to drive a large capacitance? Show that the relevant expressions to describe how to drive a large capacitance, C_{load}, starting from a minimum sized inverter with input capacitance C_{in}, are as follows.*

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

$$k = e^1$$

where k and N have their usual meanings in this context

The problem is that a small inverter will have a relatively poor current drive and this will result in a large capacitance changing voltage slowly. Conversely, a large capacitor will drive the capacitance quickly but will have a large input capacitance resulting is a problem for whatever drives the inverter. The solution is to use a sequence of inverter, scaling up in size to meet both constraints.

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:

$$\frac{Kk^iC_{in}}{k^{i-1}\beta(V_{DD}-V_T)} = \frac{KkC_{in}}{\beta(V_{DD}-V_T)}$$

where *K* is a constant of proportionality. That is, independent of the position in the sequence: the delay of each stage is equal. Whilst this is not necessarily a condition for minimum delay it does mean that none of the N inverters dominates the delay. Thus, the overall delay through *N* stages is:

$$\frac{KkC_{in}(N-1)}{\beta(V_{DD}-V_T)} + \frac{KC_{load}}{k^{N-1}\beta(V_{DD}-V_T)}$$

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:

$$\frac{KkC_{in}N}{\beta(V_{DD} - V_T)}$$

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:

$$\frac{KkC_{in}}{\beta(V_{DD} - V_T)ln(k)} ln\left(\frac{C_{load}}{C_{in}}\right)$$

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

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

from this we can find that:

$\ln(k) = 1$

Therefore,

$k = e^1 = 2.71828$

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

(6)

c. *What, if any, are there any constraints on the value of N*

If, at some point, we have a signal with which we want the capacitor to be driven, then we must add an even number of buffers to this point to ensure that the signal applied to the capacitor is not inverted. **(2)**

d. *A minimum sized inverter has an input capacitance of C_{in} and the output resistance of the inverter can be simply modelled as:*

$$\frac{2}{\beta(V_{DD} - V_T)}$$

where the terms have their usual meanings. A load, C_{load}, equivalent to $400C_{in}$ is to be driven. The switching voltage for the inverters is $0.5V_{DD}$ and calculation shows you that the time taken for the output of a first order system, with a step input, to reach 0.5 of its asymptotic value is 0.7RC. Estimate the time (in terms of β, V_{DD}, and V_T) taken to drive the load capacitance from its resting state at V_{DD} or 0 to $0.5V_{DD}$. State any assumptions made.

$N = \ln(C_{load}/C_{in}) = \ln(400) = 5.99$ so round up to N=6 (even)

The delay of each stage can be estimated to be:

$$\frac{0.7 \times 2 \times C_{in}}{\beta(V_{DD} - V_T)}$$

Now, clearly, the various stages are driven by inputs that are not true steps (they

look like first order outputs – with the same time constant) and, indeed, it is not likely that the input to the first stage looks like a step. However, we can still use this simple approximation because the range of input voltages over which the output changes is still reasonably small. Consequently, the overall delay is:

$$\frac{0.7 \times 2 \times 6 \times C_{in}}{\beta(V_{DD} - V_T)} = \frac{8.4 \times C_{in}}{\beta(V_{DD} - V_T)} \quad \textbf{(8)}$$

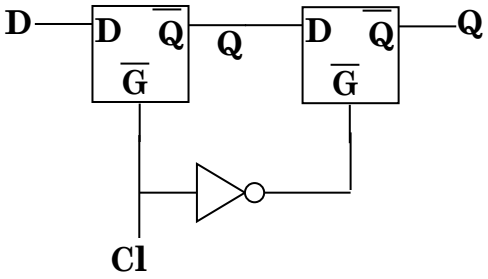
3. a. *Draw a transistor level schematic diagram for a master-slave D-type Flip Flip and explain its operation.*

A MS FF is made up from two latches whose truth tables are:

D	$\overline{G}$	Q_{new}
X	1	Q_{old}
0	0	$\overline{D}$

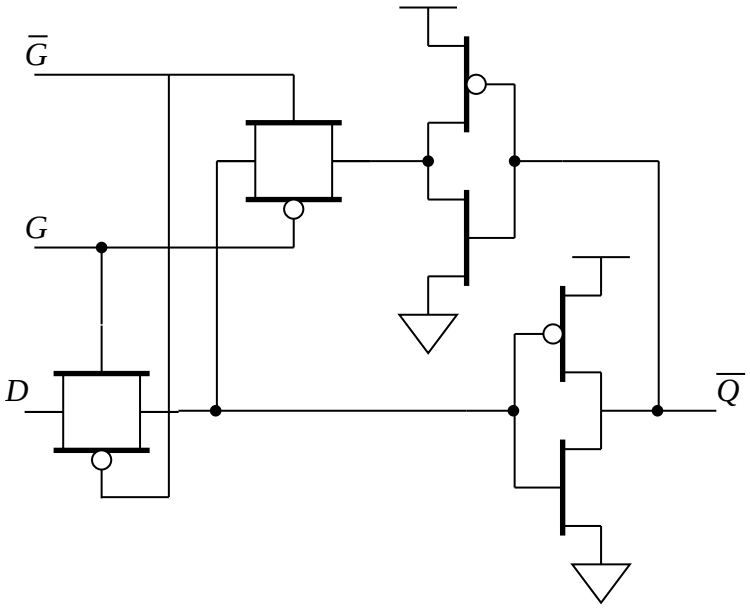
When the G input is 0, the latch is transparent allowing the inverted input to appear at the output after a small delay. When the G input is 1, the latch holds the value that was present at the 0-1 transition of the G input.

By cascading two latches driven by the clock and its inverse, as shown below.



The D FF behaviour is exhibited. The LH (master) latch captures the input, D, and holds the inverse of this value at Q_m when Clk makes a 0 to 1 transition. The value is held at Q_m whilst Clk remains at 1. During this period, 0 is held at the G input of the RH (slave) latch and, thus, the inverse of Q_m is propagated to the output. Consequently, when Clk is high, the output, Q, is the value of D captured when Clk goes high. When Clk goes to 0, the master latch goes transparent, allowing the current value of D to propagate through the master latch. However, as Clk goes low, its inverse goes high. This traps the pre-existing value of Q_m in the slave latch. Hence the value of D is held at the output for a whole cycle of the Clk input.

The latch is constructed as follows:



The basic circuit is a bistable. When G is 0 the transmission gate connecting the input is off whilst the upper one, feeding the output of the upper inverter round to form the bistable circuit is on. When G goes to 1 the bistable loop is broken and

D propagates through the lower inverter to the output.

Care must be taken to ensure that the delay between *Clk* and its inverse is less than the delay from the *G* input of the master latch to the input of the slave latch. **(5)**

b. *What is metastability. Explain how it arises and what the consequences are?*

This problem arises when the input, *D*, is changing state at a critical point in time when *Clk* is rising. Essentially, this can cause the flip-flop to enter a metastable state (neither 1 nor 0). This indeterminate output value will relax to a defined state (near exponentially with time). In the worst case, consider what would happen if the *D* input is V_{switch} at the time when *Clk* goes high such that the voltage at the input to the forward inverter exactly balanced the pair of inverters i.e. $V_{in}=V_{out}$. When *Clk* goes high, the input *D* is disconnected and the bistable look is connected with this voltage, V_{switch} at the input and output of both inverters. In this case, the output would remain at this indeterminate state until noise unbalanced it slightly, at which point feedback would flip the output to a defined state. There is no telling how long this process may take – it could even take an infinite amount of time (with infinitesimal probability). The argument still holds if the inverters behave differently: there will be some critical voltage at which the loop is balanced in this way. Clearly, this is a pathological case but if this loop closes for any input voltage where V_{in} is above $V_{IL|max}$ and below $V_{IH|min}$ then the output will be indeterminate and will remain so until the output relaxes to within one of the defined voltage ranges. **(3)**

c. *Identify and explain the terms in the following expression, that models the number of upsets per second arising from metastability.*

$$upsets = T_0 e^{-t_r/t_c} f_{clk} f_{data}$$

t_r is the time allowed to *resolve* the sampled signal – that is, the time between the sampling event (clock edge) and the point in time that the output is observed.

t_c is a constant associated with the sampler and is basically 1/GBW of the sampling circuit at the point in time when the sampler is freezing the input value (this is the critical point in time when metastability occurs).

T_0 is a constant associated with the sampler and is related to technology and circuit design (the undefined voltage range and the rise times, for example). For a good circuit design and process, T_0 and t_c should be small and of the same order of magnitude.

f_{clk} is the clock frequency and f_{data} is the effective frequency at which the data changes **(2)**

d. *Metastability is likely to occur at the boundary between two clock domains – why is this the case?*

The reason for having separate clock domains on an IC is precisely because different clock frequencies are used within the separate domains. This may be unavoidable because, for example, a communication sub-system may be constrained by standard to use a particular frequency for operation but it may need to be connected to another part of the chip using an unrelated frequency. This combination is the worst case. Consider a pathological case where two clock frequencies are different by 1Hz. When data is being passed from one domain to the other, the data at the output of one domain (which will change state on the domain's rising clock edge) will be sampled at the input of the other domain by a

clock frequency that differs by 1Hz. Thus, as clock periods pass, the point in time at which data changes will drift through the clock period of the sampling clock such that it takes 1 second for the transition to drift by one clock period. Consequently, it will also slowly drift through the clock edge giving rise to metastablility. **(2)**

e. *How can the upsets per second at the boundary between two clock domains be reduced? What is a metastability-resistant sampler and how would it be constructed?*

The key to reducing upsets is to increase t_r the observation time. This can be done by:

a) designing FFs that have high GBW products and low values of T_0.
b) Using chains of FFs in series. t_r at the output of the first FF will be t_{clk} and so upsets might be relatively high. However, t_r at the output of the second FF will be $2t_{clk}$ and the upsets at this point will be exponentially smaller in number, relative to the output of the first FF. Similarly, the observation time after the third FF would be $3t_{clk}$. The reason why this is the case is that, at the end of the clock period, any upsets trapped in the first FF would be relaxing towards a defined logic state and it will be these relaxed values that are passed on to the second FF, etc. Consequently, using two FFs would be the same as waiting two times as long before observing. However, the difference is that throughput is kept high. Unfortunately, the down-side is that signals are delayed by the number of stages in the sampler and this must be reflected in the design/use of the signals. **(4)**

f. *An 8 bit bus crosses a clock domain. The clock frequency on the receiving side of the boundary is 0.7GHz and the clock frequency on the sending side is 50MHz. Assuming random data, what would the sampler on the receiving side need to look like to ensure fewer than one upset every three years on data crossing the boundary. You can assume that both T_0 and t_c are equal to 0.1ns.*

$$upsets = T_0 e^{t_r/t_c} f_{clk} f_{data} = 0.1 \times 10^{-9} e^{-t_r/10^{-10}} \times 0.7 \times 10^9 \times 25 \times 10^6$$

$$upsets = 1.76 \times 10^6 e^{-t_r/10^{-10}}$$

So, if one FF is used then $t_r = 1/700\text{x}10^6$ and there will be 1.09 upsets per second.

If two FFs is used then $t_r = 2/700\text{x}10^6$ and there will be $6.833\text{x}10^{-7}$ upsets per second.

Treating this as the probability, p, that an upset will occur in a second. There are 8 lines in the data bus – each of which could be metastable (independently). Consequently, the probability of there being no error on any of the lines is $(1-p)^8$ and this is $1-8p$, approximately. Consequently, we can approximate the number of upsets/second to $8p = 5.4\text{x}10^{-6}$ or one upset every 182930 seconds – one every 2.11 days – this is not enough.

However, by adding another FF (3 in total), there will be $4.3\text{x}10^{-13}$ upsets/second, leading to $3.44\text{x}10^{-12}$ upsets/second on the bus or one upset every 9217 years. **(4)**

4. **a.** *For the circuit shown in **Figure 4**:*

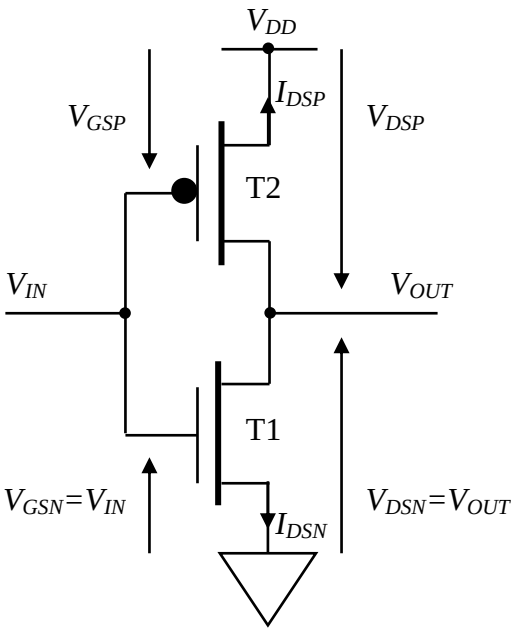


Figure 4: Inverter

show that the switching voltage (defined as the voltage when $V_{IN} = V_{OUT}$) occurs when:

$$V_{IN|switch} = \frac{V_{DD} - V_T\left(1 - \sqrt{\frac{b_N}{b_P}}\right)}{\left(1 + \sqrt{\frac{b_N}{b_P}}\right)}$$

and that when the gains of the transistors are equal, this is when $V_{IN} = V_{DD}/2$.

to solve this problem you may use the characteristics for n- and p-channel FETs, as follows:

n-FET *I/V* characteristic

$$\begin{aligned} I_{DSN} &= \frac{m_E \times C_{OX} \times W}{2L} \times \left(V_{GSN} - V_{TN}\right)^2 && (V_{TN} \le V_{GSN},\quad V_{DSN} \ge V_{GSN} - V_{TN}) \\ &= \frac{m_E \times C_{OX} \times W}{L} \times \left(V_{GSN} - V_{TN} - \frac{V_{DSN}}{2}\right) \times V_{DSN} && (V_{TN} \le V_{GSN},\quad V_{DSN} < V_{GSN} - V_{TN}) \\ &= 0 && (0 \le V_{GSN} < V_{TN}) \end{aligned}$$

p-FET *I/V* characteristic

$$\begin{aligned} I_{DSP} &= -\frac{\mu_H \cdot C_{OX} \cdot W}{2L} \cdot \left(V_{GSP} - V_{TP}\right)^2 && (V_{TP} \ge V_{GSP},\ V_{DSP} \le V_{GSP} - V_{TP}) \\ &= -\frac{\mu_H \cdot C_{OX} \cdot W}{L} \cdot \left(V_{GSP} - V_{TP} - \frac{V_{DSP}}{2}\right) \cdot V_{DSP} && (V_{TP} \ge V_{GSP},\ V_{DSP} > V_{GSP} - V_{TP}) \\ &= 0 && (0 \ge V_{GSP} > V_{TP}) \end{aligned}$$

$$\beta_P = \frac{\mu_H C_{OX} W}{L} \qquad \beta_N = \frac{\mu_E C_{OX} W}{L}$$

When $V_{IN} = V_{OUT}$ both T1 and T2 are in saturation. At this point:

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

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

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

Hence:

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

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

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

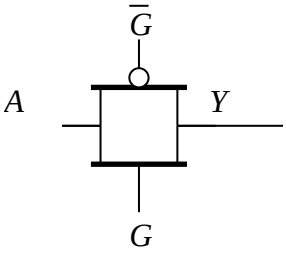
$$V_{IN} = \frac{V_{DD}}{2}$$

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

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

(6)

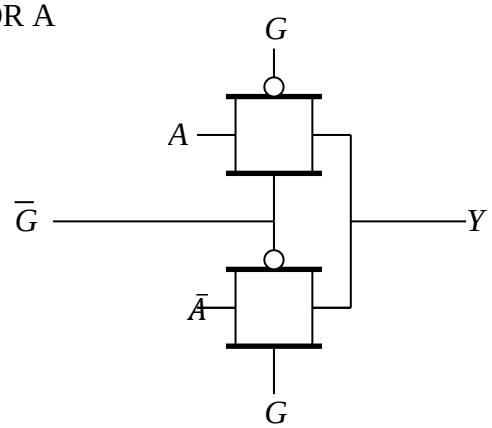
b. *Draw the schematic for a transmission gate.*



(2)

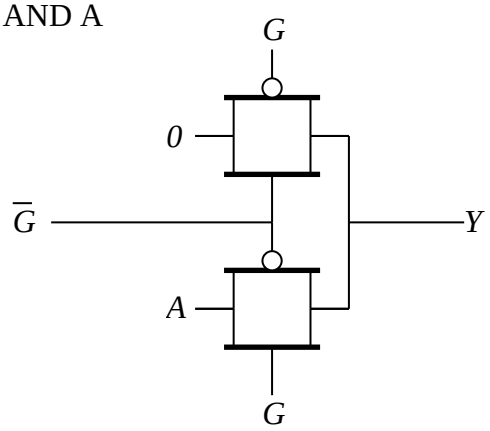
c. *Show how, using transmission gates and inverters only, the following logic functions can be implemented:*

i). Y = G XOR A



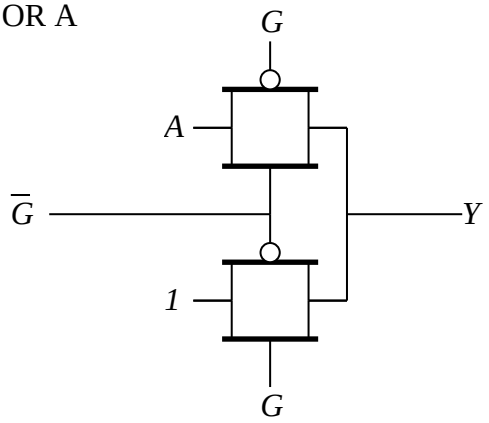
(2)

ii). Y = G AND A



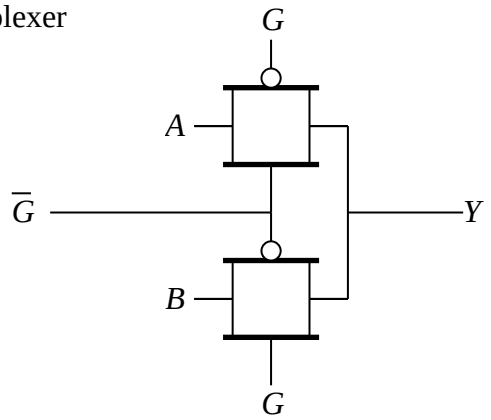
(2)

iii). Y = G OR A



(2)

iv). Multiplexer



(2)

d. **i).** *What are the limitations of using transmission gates to implement logic?*

Any current comes not directly from the power supply but comes from the output of any circuit driving the TGs. Consequently, this increases the output impedance and this makes it slower driving any subsequent stage. The longer the chains of TGs making up the logic, the greater the problem. **(2)**

ii). *How might you deal with these limitations?*

The simplest way (depends on implementing a more complex function) is to split the TGs up by adding inverters at various points to provide drive for the next part of the circuit. **(2)**