DEPARTMENT OF ELECTRONIC AND ELECTRICAL ENGINEERING

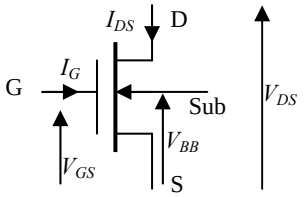
Spring Semester 2007-2008 (2 hours)

Answers to Introduction to VLSI Design/VLSI Design Questions 1…4

1. a. *Show how the expression for I_{DS} for a n-type MOSFET in the Ohmic region (shown in equation 1) can be derived.*

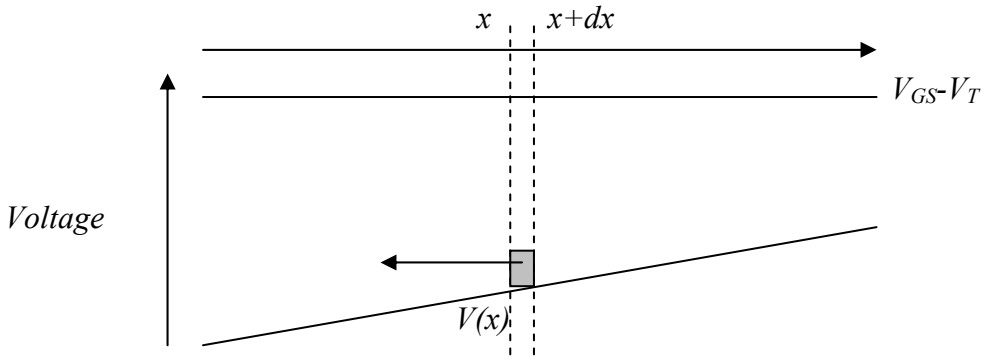
$$I_{DS} = \frac{\mu_E \cdot C_{OX} \cdot W}{L} \cdot \left(V_{GS} - V_T - \frac{V_{DS}}{2} \right) \cdot V_{DS} \qquad (1)$$

Ensure that your answer includes a diagram of a FET showing how the voltages and current are conventionally defined. Additionally, you should identify the condition that results in Ohmic behaviour.



Definitions

Consider the channel of an *n*-FET, with the distance along the channel from the source being *x* and the voltage (relative to the source) being *V(x)*.



Charge is developed at the interface between the channel and the gate insulator when the voltage across the insulator is greater than V_T and this can be simply modelled as a parallel plate capacitor where the voltage across the capacitor is V_{GS}-V_T-$V(x)$ where V_T is the threshold voltage, V_{GS} is the voltage on the gate relative to the source. The charge/per unit length of the channel at distance *x* from the source is:

$$Q/\text{unit length} = \frac{\varepsilon_0 \varepsilon_r W}{t_{OX}} \left(V_{GS} - V_T - V(x) \right)$$

Where $\varepsilon_0\varepsilon_r$ is the permittivity of the gate insulator, W is the width of the channel and t_{OX} is the thickness of the insulator.

The charge is mobile and moves under the influence of the electric field in the channel. Such that:

$$v(x) = \mu_E E(x) = \mu_E \frac{d}{dx} V(x)$$

Where μ_E is the mobility of the electrons and $v(x)$ is the velocity of the charge.

The current in the channel can be defined as:

$$I_{DS} = Q/\text{unit length} \cdot v(x) = \frac{\varepsilon_0 \varepsilon_r W}{t_{OX}} \left(V_{GS} - V_T - V(x)\right) \mu_E \frac{d}{dx} V(x)$$

and by integrating across the length of and voltages across the channel, we can resolve this in terms of the terminal voltages:

$$\int_0^L I_{DS} dx = \frac{\varepsilon_0 \varepsilon_r \mu_E W}{t_{OX}} \int_0^{V_{DS}} \left(V_{GS} - V_T - V(x)\right) dV(x)$$

$$I_{DS} L = \frac{\varepsilon_0 \varepsilon_r \mu_E W}{t_{OX}} \left[\left(V_{GS} - V_T\right) V(x) - \frac{V(x)^2}{2} \right]_0^{V_{DS}}$$

$$I_{DS} = \frac{\varepsilon_0 \varepsilon_r \mu_E W}{t_{OX} L} \left(\left(V_{GS} - V_T\right) V_{DS} - \frac{{V_{DS}}^2}{2} \right) = \frac{\varepsilon_0 \varepsilon_r \mu_E W}{t_{OX} L} \left(\left(V_{GS} - V_T\right) - \frac{V_{DS}}{2} \right) V_{DS}$$

The capacitance per unit area of the gate is:

$C_{OX} = \frac{\varepsilon_0 \varepsilon_r}{t_{OX}}$ and this gives rise to $I_{DS} = \frac{\mu_E \cdot C_{OX} \cdot W}{L} \cdot \left(V_{GS} - V_T - \frac{V_{DS}}{2} \right) \cdot V_{DS}$

(7)

b. *Describe what gives rise to the change from Ohmic to saturated behaviour for a MOSFET.*

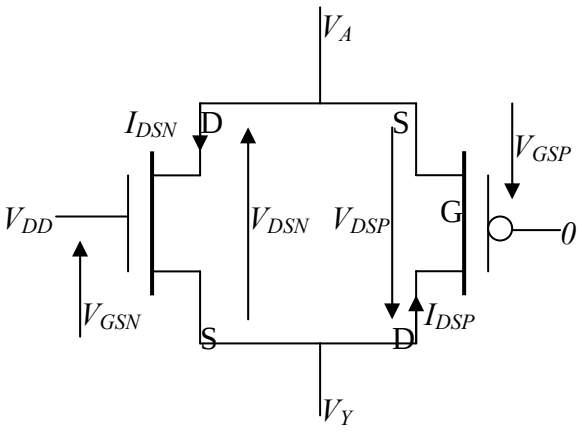
The mobile space charge, supporting conduction, only exists once sufficient carriers are drawn to the interface such that an excess of carriers (beyond the fixed charge in the depletion region) causes an inversion of the channel. This is when the voltage across the insulator is greater than V_T. At the drain end of the channel, the maximum voltage at which this occurs is V_{GS}-V_T. If the drain voltage is raised beyond this point, it threatens to stop the formation of the space charge. However, this does not happen but the channel pinches off (i.e. the channel is formed but reduces to almost nothing at the drain end, supporting a bigger voltage drop across the very end of the channel with the small available charge being driven by the larger electric fields – allowing current continuity along the channel. In this mode of operation, the current changes little with changes in V_{DS} i.e. it is saturated and depends only on V_{GS} (as long as the condition $V_{DS} > V_{GS}$-V_T continues to be met). **(7)**

c. *Show that the small-signal resistance of a MOSFET transmission gate is approximately as shown in Equation 2, for the case where the input and output voltages of the transmission gate are between V_T and V_{DD}-V_T. :*

$$r = \frac{1}{\beta(V_{DD} - 2V_T)} \qquad (2)$$

All the terms have their usual meaning and $\beta=\beta_P=\beta_N$ and $V_T=V_{TN}=-V_{TP}$.

Consider the following transmission gate where we assume that $V_A>V_Y$ without loss of generality



Based on the symmetry of the devices' operation, the terminals will be as defined in the diagram (that is D of the *n*-FET will be more positive than S and D on the *p*-FET will be more negative that S) and the voltages/currents will also be as defined.

Whilst V_A and V_Y are more than the threshold voltage away from V_{DD} and 0, respectively, the FETs will behave ohmically. Therefore:

for the *p*-FET

$$I_{DSP} = -\beta_P \cdot \left(V_{GSP} - V_{TP} - \frac{V_{DSP}}{2} \right) \cdot V_{DSP}$$

for the *n*-FET

$$I_{DSN} = \beta_N \cdot \left(V_{GSN} - V_{TN} - \frac{V_{DSN}}{2} \right) \cdot V_{DSN}$$

Substituting in for values from the above circuit gives:

$$I_{DSP} = -\beta_P \cdot \left((0 - V_A) - V_{TP} - \frac{(V_Y - V_A)}{2} \right) \cdot (V_Y - V_A) \qquad I_{DSN} = \beta_N \cdot \left((V_{DD} - V_Y) - V_{TN} - \frac{(V_A - V_Y)}{2} \right) \cdot (V_A - V_Y)$$

Assuming that the gains are equal to β and that $V_{TN}=-V_{TP}=V_T$, and adding them in the direction of I_{DSN} we find:

$$I_{DS} = I_{DSN} - I_{DSP} = \beta \cdot \left(V_{DD} - V_Y - V_T - \frac{V_{AY}}{2} \right) \cdot V_{AY} + \beta \cdot \left(-V_A + V_T - \frac{V_{YA}}{2} \right) \cdot V_{YA}$$

This can be simplified because $V_{YA}=-V_{AY}$ and so:

$$I_{DS} = \beta \cdot \left(V_{DD} - V_Y - V_T - \frac{V_{AY}}{2} \right) \cdot V_{AY} - \beta \cdot \left(-V_A + V_T + \frac{V_{AY}}{2} \right) \cdot V_{AY}$$
$$= \beta \cdot \left(V_{DD} - V_Y - V_T - \frac{V_{AY}}{2} + V_A - V_T - \frac{V_{AY}}{2} \right) \cdot V_{AY}$$

Collecting terms:

$$I_{DS} = \beta \cdot (V_{DD} + (V_A - V_Y) - 2V_T - V_{AY}) \cdot V_{AY}$$
$$= \beta \cdot (V_{DD} + V_{AY} - 2V_T - V_{AY}) \cdot V_{AY}$$
$$= \beta \cdot (V_{DD} - 2V_T) \cdot V_{AY}$$

Consequently,

$$\frac{dI_{DS}}{dV_{AY}} = \beta(V_{DD} - 2V_T) = \frac{1}{r} \text{ and so } r = \frac{1}{\beta(V_{DD} - 2V_T)}$$

(10)

2.

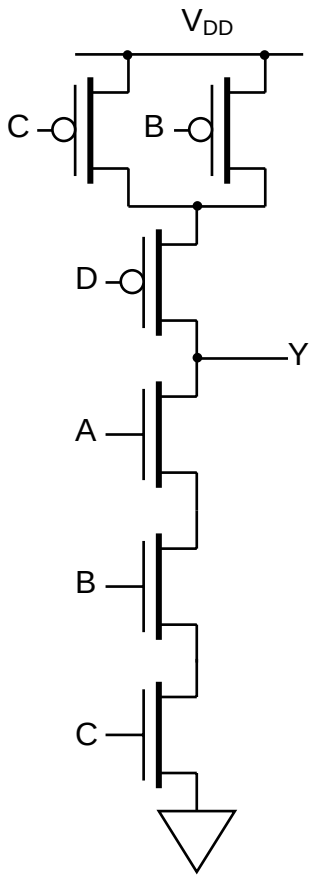


Figure 2: Incomplete Logic Circuit

*You are given an incomplete logic circuit, shown in **Figure 2** where the output Y is a function of four inputs: A, B, C, and D. You are told that to complete the circuit you need only to add transistors to the circuit without changing the way in which the existing transistors are connected. From this, you recognise that there are two possible ways in which the circuit could be completed.*

i. *How do you know that the circuit is incomplete?*

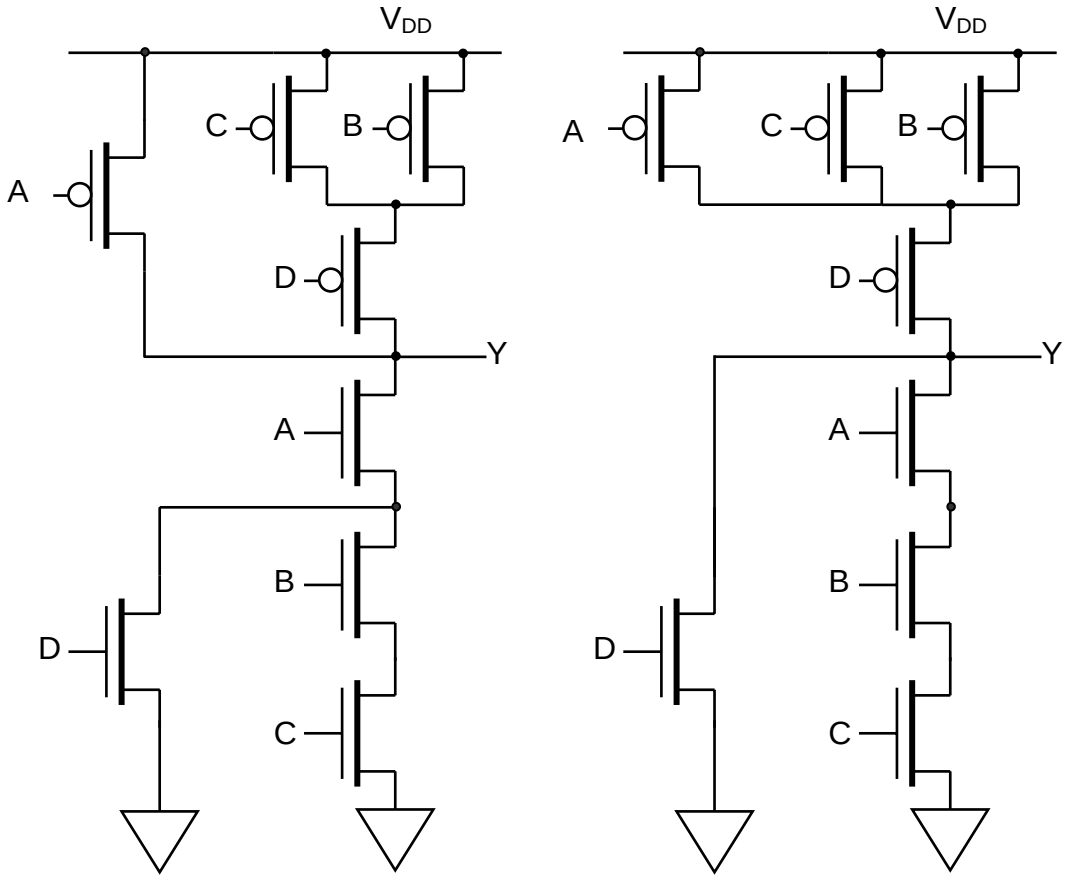
The pull-up and pull-down networks should have switches controlled by each and every input and the output should have a defined state for all combinations of inputs. This is not the case for the figure. **(2)**

ii. *Draw the two possible circuits that might represent the competed circuit.*

The pull-up and pull-down networks must be complementary – that is a series combination of switches in one network must have a parallel combination of switches controlled by the same inputs in the other network and vice versa. Looking at the figure, this means that C/B being in parallel in the PUN must imply C/B being in series in the PDN and this is the case. But D is in series with C//B in the PUN and this means that D must be in parallel with the B/C combination in the PDN – and this is not the case. Looking at the PDN, A is in series with B and C and so A must be in parallel with B//C in the PUN. However, there must be two choices based on the order in which we insert the two needed switches in the PU PD networks.

Option 1: If we insert a switch controlled by D in parallel with the B/C series combination in the PDN then the circuit must be completed by adding a switch controller by A in the PUN that is in parallel with the A/B//C network.

Option 2: If we insert a switch controlled by A in parallel with the B//C parallel combination in the PUN then the circuit must be completed by adding a switch controller by D in the PDN that is in parallel with the A/B/C network.



iii. *Write down the logical functions of the two possible circuits.* **(8)**

Option 1: $Y = \overline{A.(B.C + D)}$

Option 2: $Y = \overline{A.B.C + D}$ **(4)**

iv. *For any one of the completed circuits, size the transistors assuming that you want a minimum-sized gate (assume that the mobility of holes is half that of electrons).*

Option 1: pA=2, pB=4, pC=4, pD=4, nA=2, nB=4, nC=4, nD=2

or

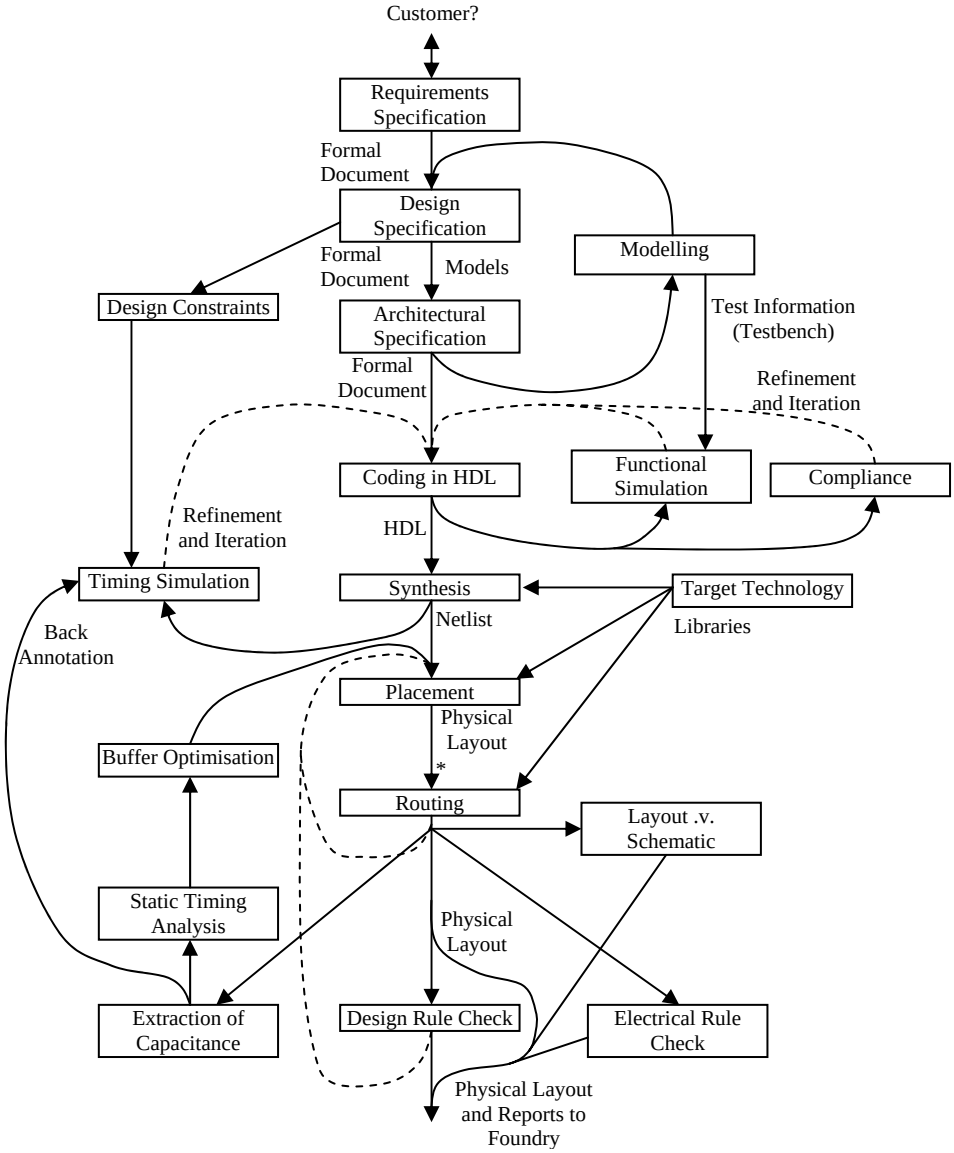
Option 2: pA=4, pB=4, pC=4, pD=4, nA=3, nB=3, nC=3, nD=1 **(4)**

v. *The substrate connections are not shown in **Figure 2**. Where would the substrates be connected for the n- and p-type MOSFETs?*

It is likely that all of the n and p transistors will, respectively, lie in a connected p and n region – so all the n transistors will share a common substrate voltage of 0V whilst all of the p transistors will share a common substrate voltage of V_{DD}. **(2)**

3. **a.** *Draw a flow diagram to represent a basic ASIC design flow, identifying the activity within each part of the design process.*

The flow diagram will be a subset of the following:



Major parts that need identifying are:

Specification phases identifying the role of modelling to determine the correctness of the specification and that the definition of the architecture is of primary importance.

Coding in HDL, linked with functional simulation to verify the correctness of the design (using information derived from the modelling phase).

Synthesis, using a library of cells from a target technology and set of constraints governing the process, coupled with timing simulation to ensure that the design is functional with estimates of delays added and meets the constraints.

Place and Route coupled with clock-tree (and test structure) insertion, extraction of more realistic delays, and timing simulation and static timing analysis/buffer insertion and optimisation to ensure that the laid-out design meets specifications.

Final checks, e.g. design rules, electrical rules, layout versus schematic to produce the foundry reports (without which the foundry will not fabricate) and production of design files for foundry. **(10)**

b. *What are the major differences between Cell-Based ICs, Structured ASICS, Masked Gate Arrays, and Field-Programmable Gate Arrays? How might you decide which technology to use in a particular design project?*

Answer should include:

Cell-Based IC – IC is designed using components drawn from a library of cells (which can be stitched together on the surface of the IC). Cells can be from NAND gates up to large macrocells (e.g. microprocessor, memory). High NRE costs because entire IC must be fabricated from scratch, highest performance, and smallest size for given functionality (mapping of design to cells is efficient). V.large designs possible.

Masked Gate Arrays – consist of sea of NAND gates (for example) in standard ICs on pre-diffused wafer. MGAs are fabricated as standard items. Design is synthesised down to NAND gate equivalents and the only steps needed to complete the IC are to add layers of metallization on top of pre-diffused wafer. Low NRE, quick turnaround, lower performance than CBIC. Not good with memory. Modest to Large designs possible.

Structured ASICS. Like MGA – pre-diffused with some layers of metallization added as standard. Contains sea of gates with some customisation possible and pre-defined blocks (e.g. memory, processor?). Design is synthesised down to instantiable gates and macro blocks used (or not used) as is (memory may allow different architectures to be used). Better performance than MGA quick turnaround, higher NRE costs than MGA. Large designs possible.

FPGA – programmable device. Contains sea of programmable blocks, memories, and possibly other functional units. Programmed by downloading stream of data (setting switches) or blowing fuses. Rapid turnaround, very-low NRE, worst performance in terms of size, speed, and power. Modest sized designs possible.

Ultimately, the choice is driven by basic constraints – will the technology give you the speed/power performance necessary and can the size of the design be accommodated in the technology. Thereafter, other financial considerations take over. Cost of development (measured by design, fabrication, risk) measured against the volume/price of the product. Against this must be weighed time to market and competition. **(5)**

c. *You work for a fabless IC company, and your company has to develop a single IC solution for a mass-market, mobile product that will encompass a PDA, media player, and personal communications (e.g. WiFi, bluetooth and GSM). Identify the factors that will influence the choice of technology and the fabricator of the IC.*

The issues that bear on the choice of technology are many and the answer should include some of the following (1 mark each):

High volume device and so the cost must be driven down but efficient technology such as CBIC with high NREs can be tolerated.

Single chip solution must have wide range of functionality in it so the design is likely to be large.

Mobile device so power efficiency is high on the list of constraints (clock gating, power down, etc.).

High performance required – such a device will have a significant DSP load (base-band processing for GSM/WiFi) data coding/decoding, etc.

Will have at least one processor and this processor/processor sub-system will need to accommodate a) standard operating system for PDA (e.g. Symbian, Linux) and protocol stack for comms aspect. Probably need a DSP device to support coding related activities and possible macro blocks to handle key hih performance aspects. Technology chosen must support the development of an end-to-end system without any gap (which would need plugging by the designer).

Single IC solution may need analogue section for comms (high frequencies) with defined SNR needed (in the presence of digital circuit).

Memory – likely to need some on-chip but will need most off-chip. **(5)**

4. a. i. *Describe the Synchronous Design Methodology and explain why it results in reliable designs.*

The SDM sets a framework in which synchronous circuit design become feasible. All FFs in a design are deemed to be clocked at the same point in time. The input to a FF, which is derived (possibly via combinational logic) from FF outputs elsewhere in the circuit. This approach means that, due to propagation delays though logic and wiring, the value of the input when a rising clock edge arrives will be the value set at the FF outputs (driving this input) at the last rising clock edge. Furthermore, this means that the values set at a FF output at one clock edge have the time equivalent to one clock cycle to propagate to and become properly set up at a FF input elsewhere in the circuit. The propagation delay along any possible path between two FFs (or FF and input/output) can, therefore, be analysed, against possible variation with PVT, to determine that the data at a FF input will always be valid when the next clock edge arrives. If the relative phase of all clock signals was not controlled in this way then data set at a FF output at one rising clock edge could propagate to another FF and be set up there before the same rising clock edge arrived at this FF. Under a different set of conditions, the same data might not arrive and be seen only at the next rising clock edge. This would mean that the circuit would behave differently under different conditions i.e. it would be unreliable. **(4)**

ii. *How does synchronous design relate to Register Transfer Level descriptions of circuits?*

An RTL statement within a clocked process (in VHDL for example) might be

```
Process (clk)
begin
   y <= a + b;
end process;
```

The value `y`, on the LHS, is stored in a register and is assigned after the rising edge of `clk` derived from values `a` and `b` on the RHS, which in this case pass through an adder (combinatorial logic). `a` and `b` will be stored in other registers and the values used to assign `y` at a given rising edge of `clk` will be values assigned to a and b at the previous rising edge. **(2)**

b. *What are the sources of variability in the performance/behaviour of an IC?*

Manufacturing is a variable process giving rise to significant differences in behaviour between two ICs manufactured at different times.

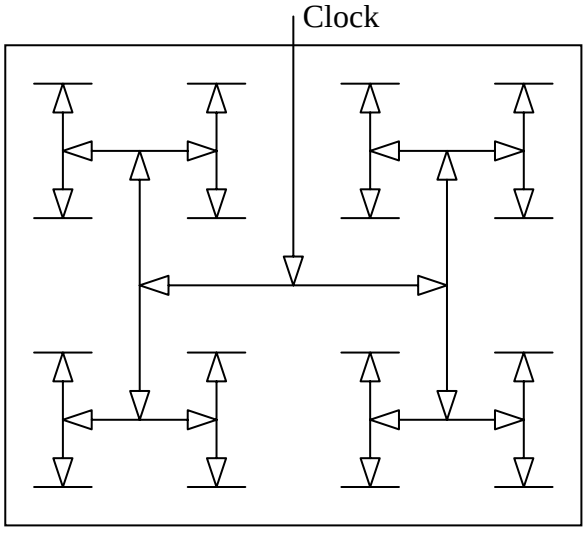
The performance of transistors is very temperature dependent and circuits slow down and the temperature increases.

A Circuit's performance is influenced by supply voltage such that as supply voltage falls, circuit slow down. Any circuit will be specified to work across a defined range of supply voltages and performance will change across this range.

Circuits must be designed such that they continue to work from worst to best case based on specification provided by the manufacturer. **(3)**

c. *How is the distribution of clock signals managed in a synchronous IC to ensure that the IC is clocked synchronously?*

The key issue is that clock edges should arrive everywhere across the surface of the IC at the same point in time (subject to some small skew that can be accounted for. Furthermore, the clock signal is the highest frequency on the IC and must drive a very large number of inputs to FFs. Both of these require a balanced clock tree. The clock is buffered and split, repeatedly in a symmetrical network such that the clock signals at the leaves of the tree are all pretty-much coincident and serve only a small, local region of the IC. The organisation shown here is not the only kind that might be acceptable.



(2)

d. **i.** *What particular problems exist for designers as a consequence of there being more than one clock domain in an IC?*

Data produced in one domain, in synchronism with the local clock will inevitably need to be passed to another domain and stored there under the control of the clock in the receiving domain. If the clocks are unrelated then this can give rise to metastability where the data is sampled in a FF in the receiving domain during a transition of the data (edges of the two sets of clocks become nearly co-incident). When the metastable data at the output is used in a logic circuit or sampled elsewhere it could be wrong and give rise to erroneous data. **(3)**

ii. *Is it important that designers should be able to design an IC with multiple clock domains? If so, why?*

Yes it is. Often, a large SoC has a number of domains and the clock frequency used in a block is mandated by the activity. So, a communication related block

may require a particular clock as determined by the communication standard and this implies a transition between this clock frequency and the frequencies used elsewhere on the device (e.g. within any processor/processing block

(2)

iii. *How do you make the transfer of signals between clock domains reliable?*

There are two basic ways.

With unrelated clocks metastability will occur but by increasing the time after sampling before the data is *observed* (i.e. used) you can reduce the risk that the data will still be in a metastable state at this time. The rate at which the signal relaxes back to a defined value is controlled by circuit and layout factors (e.g. bistable circuits with high effective gain-bandwith products will be better). Consequently, using samplers (FFs) that have been designed to minimise the risk is a good starting approach.

However, this may not be good enough. Consequently, trying to increase the observation time is often necessary. This can be achieved by sampling the data through a string of samplers (e.g. 2 or more). If the data at the front end is metastable, it will relax towards a defined state and as it passes through the delay line, this process will continue such that when the data reaches the end it is much more probable that it is in a defined state. This requires the system to be designed in the knowledge that data passing between domains will be delayed by a number of clock cycles and this might mean, in turn, that the time taken to read data across domains and then feed data back (requiring a metastable resistant sampler in the first domain) is relatively long.

(4)

NLS / MB