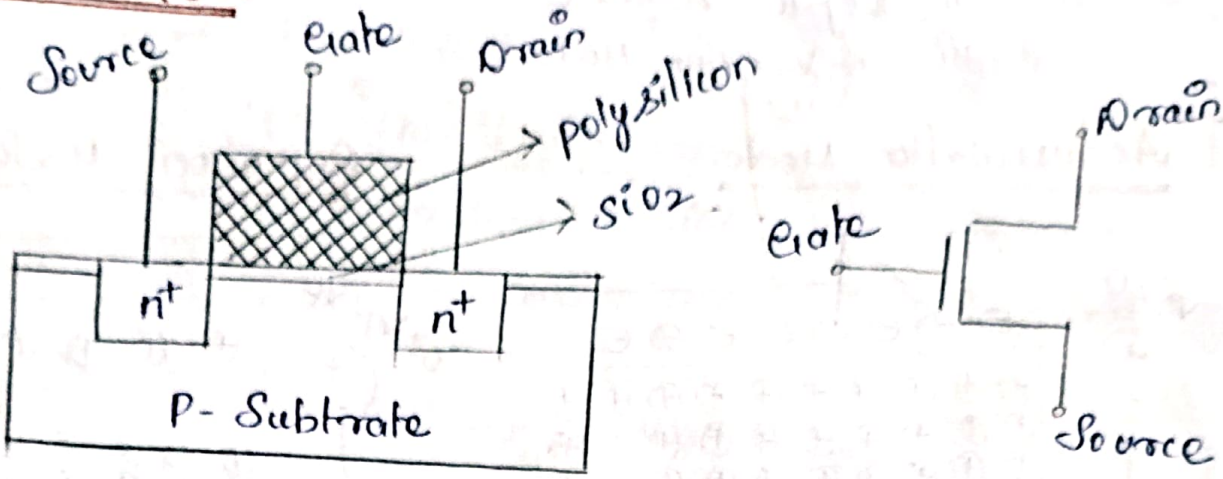
UNIT I - MOS Transistor Principle

VLSI Design

nMOS Transistor:-



→ n MOS transistor build on P type subtrate is moderate doping.

→ Source and drain are formed by heavily doped n-type impurities adjacent to gate.

→ SiO_2 is placed over the subtrate is b/w the source & drain

→ over SiO_2 a layer of poly silicon is formed over which gate terminal is taken.

Operation:-

→ Body is initially grounded. So, Pn junction is reverse bias.

→ If the gate is grounded no current is flow, transistor is off.

→ If the gate voltage is raised it creates an electric field that starts to attract free electrons towards the bottom of SiO_2 layer.

→ If the voltage is raised further the electrons out no. the holes and a thin region under the gate called the channel is inverted to act as an n-type semi-conductor.

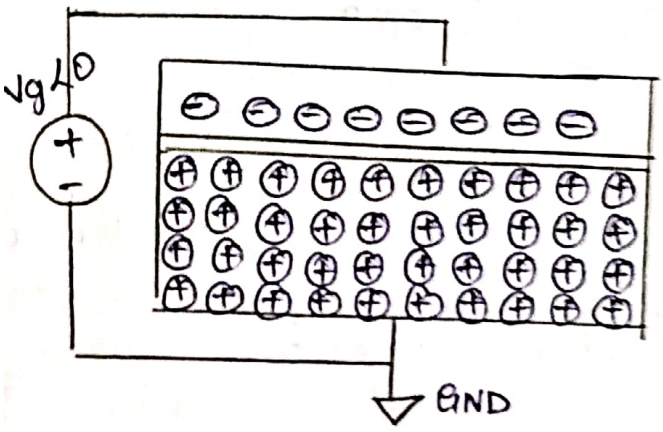
→ Hence, a conducting path of electrons carrier is formed from source to drain and current can flow. Now the transistor is ON.

Modes of operation :-

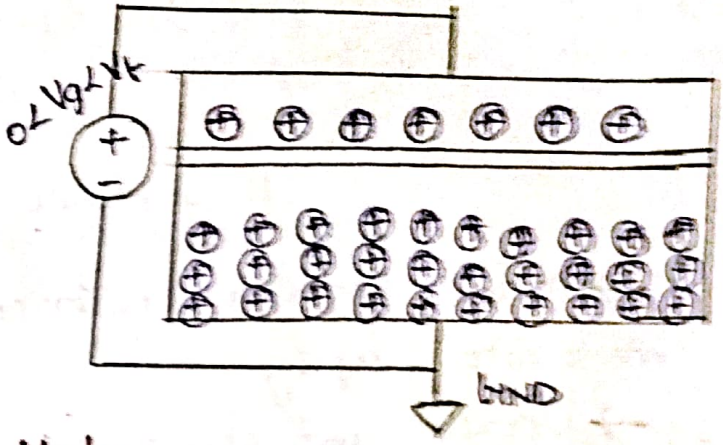
The following are the three types of Modes of operation of nMOS transistor.

i. Accumulation Mode
ii. Depletion Mode
iii. Inversion Mode

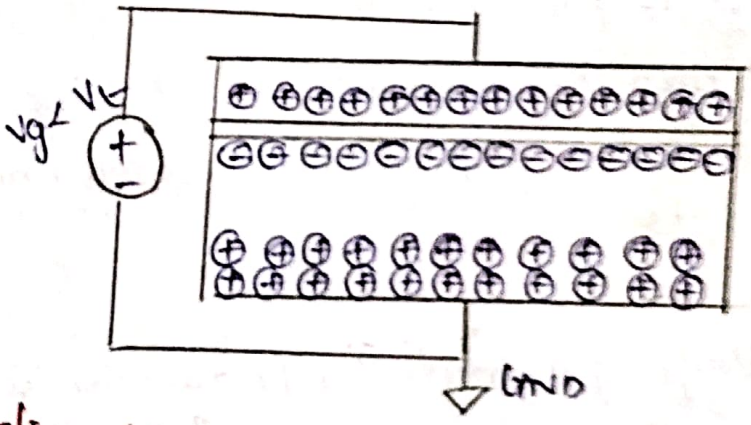
Accumulation Mode



Depletion Mode



Inversion Mode



Accumulation Mode [Vgs<0]:-

In this mode a small negative vtg is applied to the gate. Therefore, positively charged holes are accumulated beneath the gate.

Depletion Mode [0<Vg<Vt]:-

A low positive vtg is applied on the gate, resulting in some positive charge on the gate region the holes in the body are rippled from the region beneath the gate.

Inversion Mode [$V_g > V_t$]:-

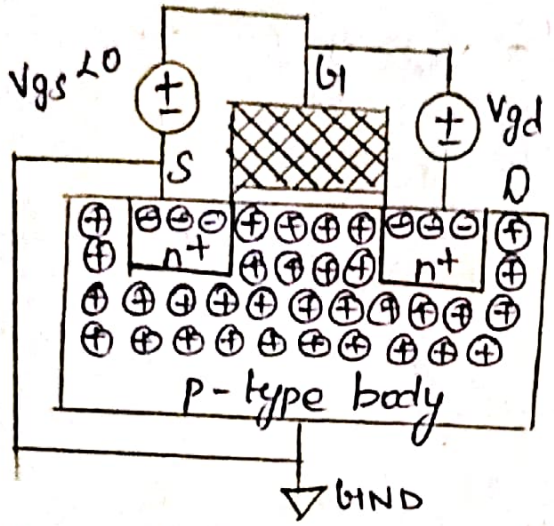
Higher positive gate voltage is applied which attracts more positive charge to the gate. As a result the minority carrier i.e electrons in the p-type substrate are attracted beneath the gate. This conductive layer of electrons in the p-type body called channel forms the inversion layer.

Operating Region of nMOS Transistor:-

There are three operating region.

1. cut-off region 2. Linear region 3. Saturation region

Cut-off Region:-

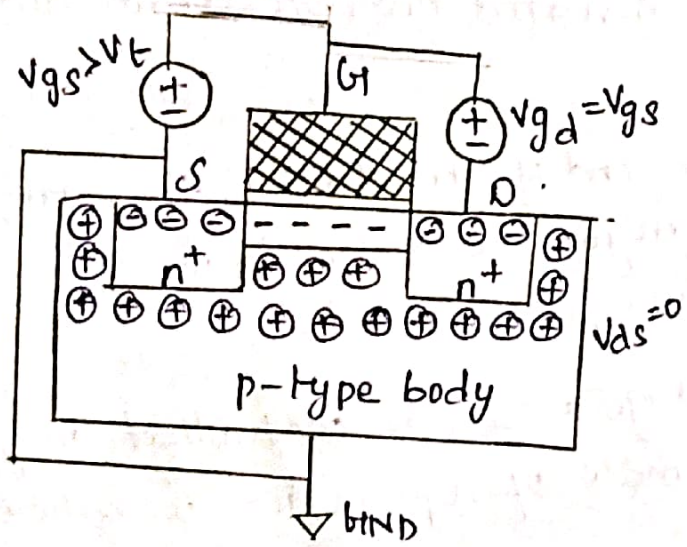


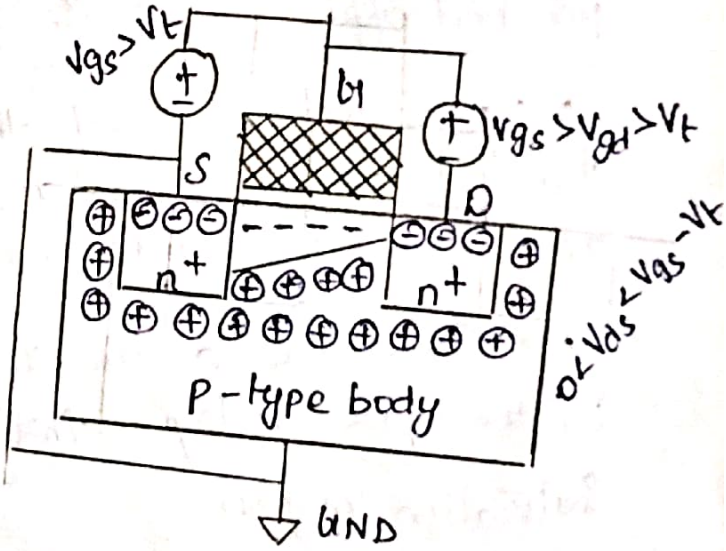
→ In this region $V_{gs} < V_T$.

→ The source and drain have free electrons.

→ The junction b/w the substrate and source (or) drain are reverse biased. So, no current will flow. This mode of operation is called cut-off.

Linear Region:-





→ In this region $V_{gs} > V_T$ and inversion layer of electron called the channel connect the source and drain. The number of carrier and conductivity increases with the gate voltage.

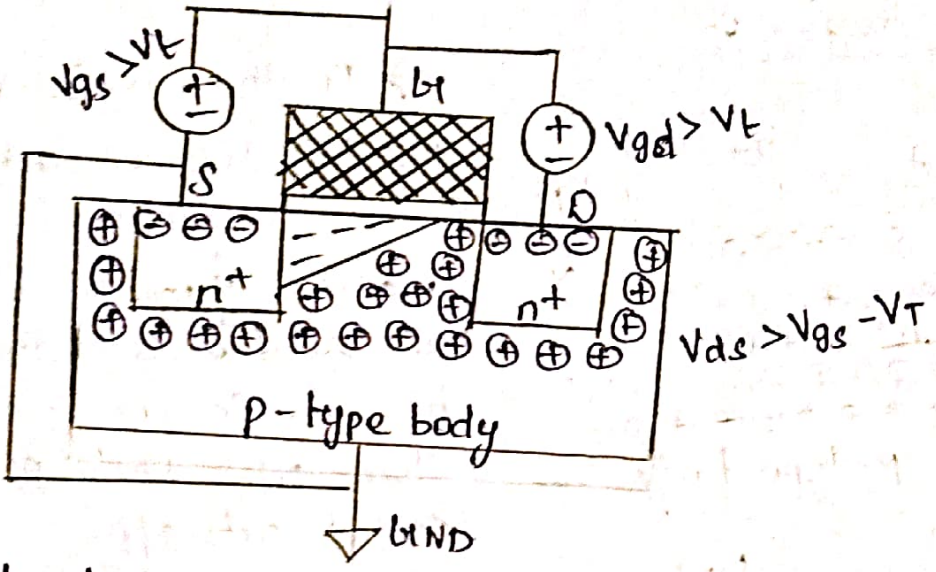
Therefore, $\boxed{V_{ds} = V_{gs} - V_{gd}}$

→ If $V_{ds} = 0$, then $V_{gs} = V_{gd}$.

→ There is a no Electric field tending to push the Electron from source to drain.

→ When a small positive potential V_{ds} is applied. the current I_{ds} flows through the channel from source to drain. This mode of operation is called linear (or) resistive (or) un-saturated (or) non-saturated region.

Saturation Region:-



→ If V_{ds} becomes sufficiently larger than $V_{gs} - V_T$. The channel is no longer inverted near the drain and becomes pinched off.

→ As Electrons reach the end of the channel, they are injected into the depletion region near the drain and accelerated toward the drain.

→ Above the drain voltage. I_{ds} is controlled only by the gate voltage. this mode of operation is called Saturation region.

② Electrical properties of CMOS circuit and device Modulae:-

The Electrical properties and device Modulae include the following characteristics.

1. Ideal IV characteristics
2. CV characteristics
3. Non-Ideal IV effects
4. DC transfer characteristics
5. Device Modelling.

(i) Ideal IV characteristics:-

→ The current voltage characteristics is known as IV characteristics.

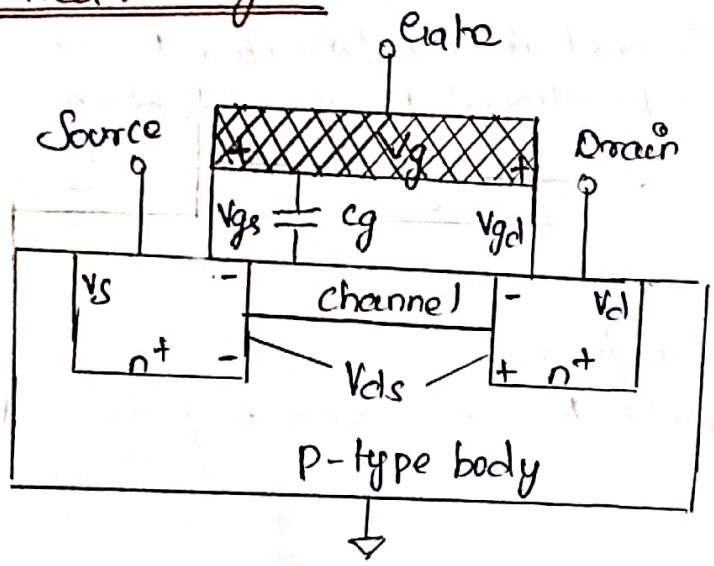
→ MOS transistor have three operating region.

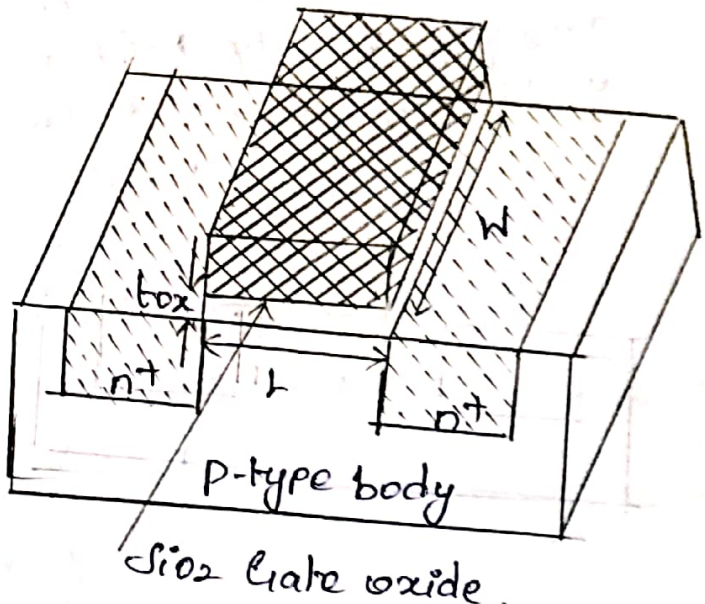
1. cut-off region 2. linear region 3. Saturation region.

Cut-off region:-

In this region gate to source voltage $V_{gs} < V_T$. So, no current flows from drain to source. $I_{ds} = 0$.

Linear region:-





→ In this region the gate attracts the carrier electrons to form a channel. the electron drift is proportional to the Electric drift b/w source and drain.

→ So, the current can be computed by knowing the amount of charge in the channel & the rate which moves.

$$Q_{channel} = C_g * (V_{gc} - V_t) \quad \text{—— (1)}$$

Where,

$C_g \rightarrow$ gate capacitance

$V_{gc} - V_t \rightarrow$ amount of V_{tg} attracting charge to the channel beyond the minimum required to invert from p to n.

The mean difference b/w the gate to channel potential

$$V_{gc} = V_{gs} - \frac{V_{ds}}{2} \quad \text{—— (2)}$$

→ let us consider the gate as a parallel plate capacitor

$$C = \frac{\varepsilon A}{d}$$

$$C_g = \frac{\varepsilon_0 \varepsilon_{ox} \cdot W_L}{t_{ox}} \Rightarrow C_g = C_{ox} \cdot W_L \quad \text{—— (3)}$$

where, $t_{ox} \rightarrow$ thickness of oxide

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \quad \text{—— (4)}$$

Sub (4) in (3),

$$C_g = \frac{\varepsilon_{ox}}{t_{ox}} \cdot W_L \quad \text{—— (5)}$$

where, $C_{ox} \rightarrow$ capacitance per unit area

→ Each carrier in the channel is accelerated to an average velocity which is proportional to the Electric field b/w source & drain.

$$V \propto E$$

$$V = \mu E$$

where, $\mu \rightarrow$ mobility of Electrons

$E \rightarrow$ Electric field b/w drain & source

$$E = \frac{V_{ds}}{L}$$

$$V = \frac{\mu V_{ds}}{L} \quad \text{—— (6)}$$

→ The time required for the carrier to cross the channel length is given by

$$T = \frac{L}{V} = \frac{\text{Length}}{\text{Velocity}}$$

→ The current b/w the source & drain is the total amount of charge in the channel divided by the time required to cross the channel.

$$I_{ds} = \frac{Q_{channel}}{T}$$

$$= \frac{C_g (V_{gc} - V_T)}{L/v}$$

$$= \frac{C_{ox} \cdot WL (V_{gs} - \frac{V_{ds}}{2} - V_T)}{L/v}$$

$$I_{ds} = C_{ox} \cdot Wv (V_{gs} - \frac{V_{ds}}{2} - V_T)$$

$$I_{ds} = \frac{C_{ox} \cdot W\mu V_{ds}}{L} (V_{gs} - \frac{V_{ds}}{2} - V_T)$$

$$\boxed{I_{ds} = \beta \cdot V_{ds} (V_{gs} - \frac{V_{ds}}{2} - V_T)}$$

Where, $\beta = \frac{C_{ox} \cdot W\mu}{L}$

Saturation region :-

For saturation region, $V_{ds} = V_{gs} - V_T$

$$I_{ds} = \beta (V_{gs} - \frac{V_{ds}}{2} - V_T) V_{ds}$$

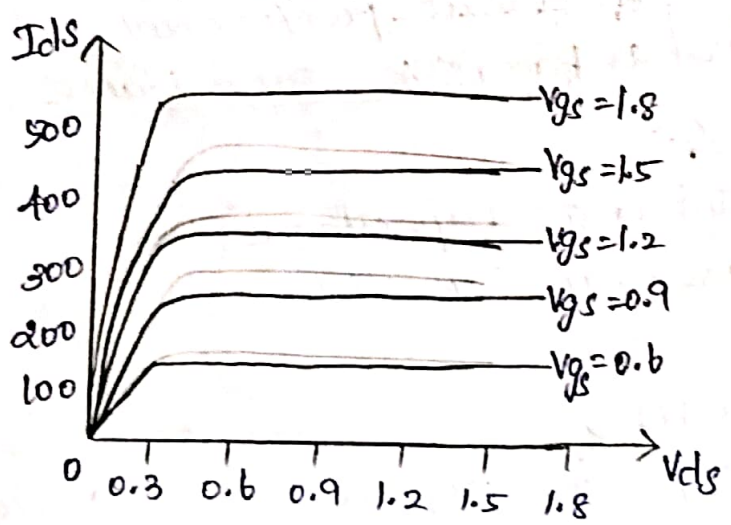
$$= \beta (V_{gs} - \frac{(V_{gs} - V_T)}{2} - V_T)(V_{gs} - V_T)$$

$$= \beta \frac{(2V_{gs} - V_{gs} + V_T - 2V_T)}{2} (V_{gs} - V_T)$$

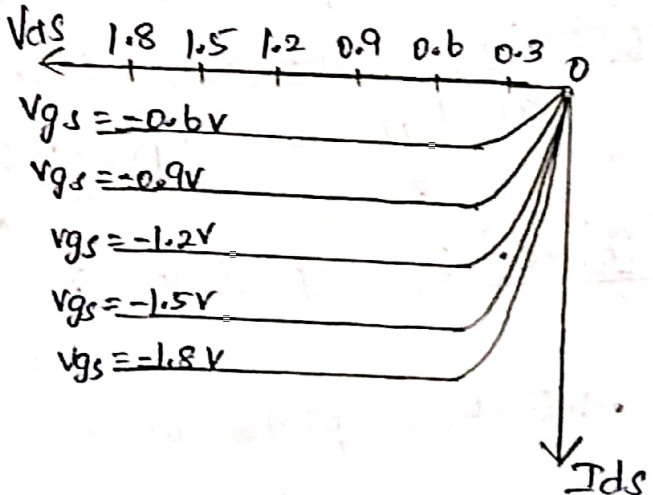
$$= \frac{\beta}{2} (V_{gs} - V_T)(V_{gs} - V_T)$$

$$\boxed{I_{ds} = \beta/2 (V_{gs} - V_T)^2}$$

For nMOS



For pMOS



(ii) C-V characteristics:-

→ Each terminal of MOS transistor has a capacitance with respect to the other terminal. This capacitance are non-linear and voltage dependent.

→ The capacitance of voltage characteristics has three different types of capacitance Model.

1. Simple MOS Capacitance Model
2. Detailed MOS capacitance Model
3. Detailed MOS Diffusion capacitance Model.

Simple MOS capacitance Model:

→ The gate of the MOS transistor is a good capacitance higher the gate capacitance to obtain high Ids.

→ The gate capacitor can be viewed as parallel plate capacitor with a gate on the top and the channel on the bottom between a thin oxide dielectric.

$$C_g = \frac{\varepsilon_{ox}}{t_{ox}} \cdot WL$$

$$C_g = C_{ox} \cdot WL$$

→ L specifies the minimum length which related to the speed and consumption of MOS transistor.

→ The advance Manufacturing process the channel length and oxide thickness are reduced by the same factor which is called Scalling factor.

→ In addition to the gate, the source and drain also add to the capacitance but it is not used by the operation of the device but it have an impact on circuit performance. Hence this capacitance is called as Parasitic capacitance.

Detailed MOS capacitance:-

It has two types (i) Intrensic capacitance
(ii) overlap capacitance

Intrensic capacitance:

$$C_g = C_0 = C_{ox} \cdot WL$$

→ Bottom plate of the capacitor depends on the Mode of

operation of the transistor as follow.

(i) **cut-off**: When the transistor is off. $V_{gs} > 0$ & $V_{gs} < V_T$.

Depletion region forms at the surface effectively moves the bottom plate downward from the oxide reducing the capacitance.

(ii) **Linear**: When $V_{gs} > V_T$.

→ Channel inverters and acts as a good conductive bottom plate $C_{gb} = 0$.

→ The channel is connected to the source & drain rather than body. At low values of V_{ds}, the channel charge shared between source & drain. $C_{gs} = C_{gd} = \frac{\varepsilon_0}{2}$

(iii) **Saturation**: When $V_{ds} > V_{gs} - V_T$.

Transistor saturates & channel pinches off. Because of pinch off the capacitance in saturation reduced to $C_{gs} = \frac{2}{3} C_0$ for an ideal transistor.

Approximation of Intrinsic MOS transistor:-

Parameter	cut-off	Linear	Saturation
C_{gb}	C_0	0	0
C_{gs}	0	$\frac{C_0}{2}$	$\frac{2}{3} C_0$
C_{gd}	0	$\frac{C_0}{2}$	0
$C_g = C_{gb} + C_{gs} + C_{gd}$	C_0	C_0	$\frac{2}{3} C_0$

Overlap capacitance:-

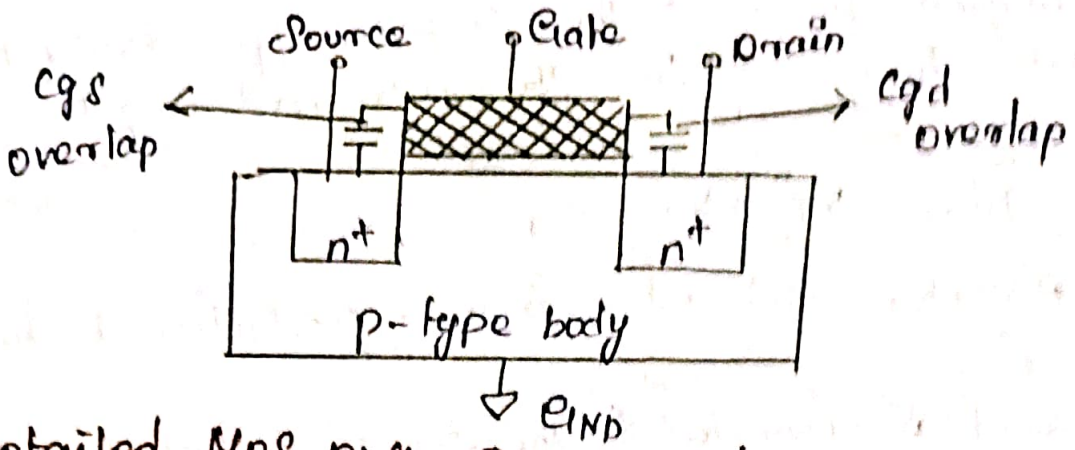
→ The gate overlap the source to drain by a small amount and it has fringing field with respect to source and drain, this leads to an addition overlap capacitance which is proportional to the width of the transistor.

$$C_{gs}(\text{overlap}) = C_{gso1} \cdot W$$

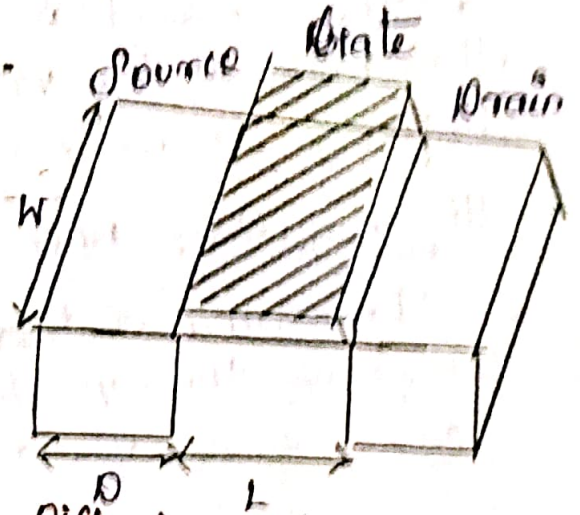
$$C_{gd}(\text{overlap}) = C_{gdo1} \cdot W$$

$C_{gs01} = C_{gd01} = 0.2 - 0.4\ fF/\mu m.$

[fF - femto farad, $1fF = 10^{-15} F$]



Detailed MOS Diffusion capacitance:-



Diffusion gateway region

$A_S = W.D \qquad P_S = 2W + 2D$

The total parasitic capacitance

$C_{sb} = A_S \cdot C_{jbs} + P_S \cdot C_{jbssw}$

→ The reverse bias pn junction b/w the source and the substrate contribute the parasitic capacitance and depends on its area (A) and the sidewall perimeter (P) of the source diffusion region.

Where,

C_{jbs} = junction capacitance per Area

C_{jbssw} = perimeter junction capacitance per length

$$C_{jbs} = C_j \left(1 + \frac{V_{sb}}{\psi_0}\right)^{-M_j}$$

Where,

C_j = junction capacitance at zero bias

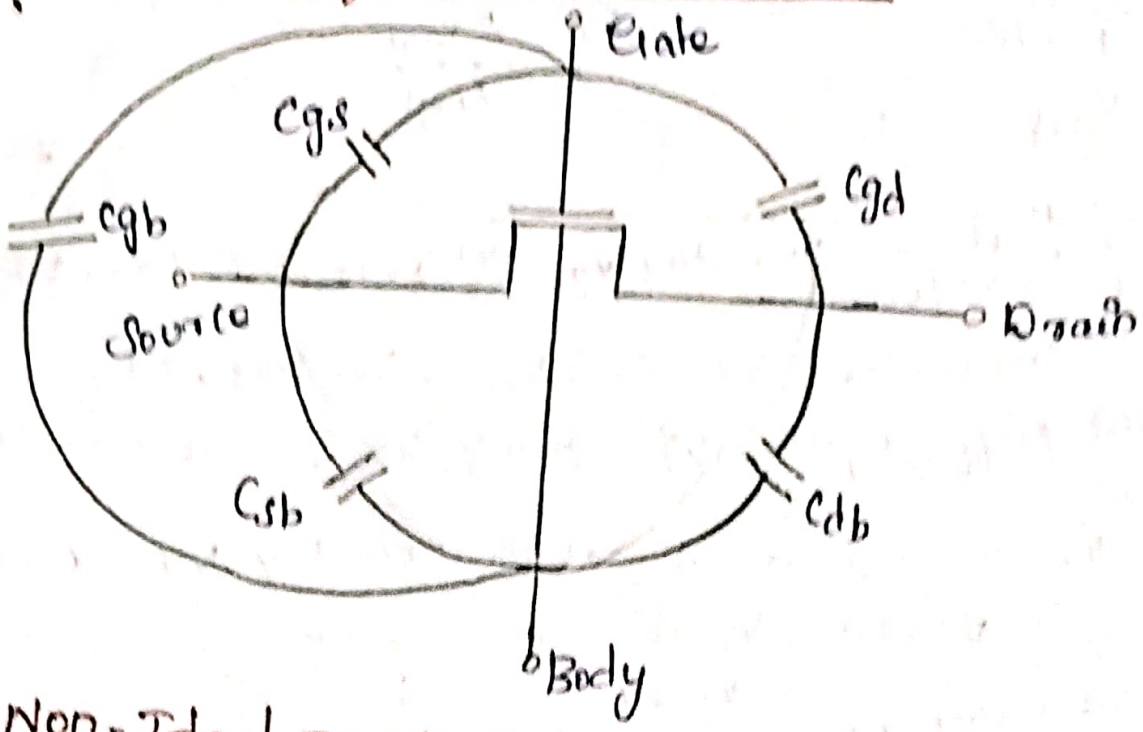
M_j = junction grading coefficient

V_{sb} = Voltage b/w source & bulk

$$\psi_0 = V_T \ln\left(\frac{N_A N_D}{n^2}\right)$$

The side wall capacitance $(C_{jbssw}) = \left(1 + \frac{V_{sb}}{\psi_0}\right)^{-M_{jssw}}$

Capacitance of an MOS Transistor:-



(iii) Non-Ideal IV effect:-

The non-ideal I-V effects include the following.

1. Velocity Saturation and Mobility Degradation.
2. Channel length Modulation
3. Body effect
4. Subthreshold condition
5. Junction Leakage.
6. Tunneling
7. Temperature Dependence
8. Geometry Dependence

Velocity Saturation

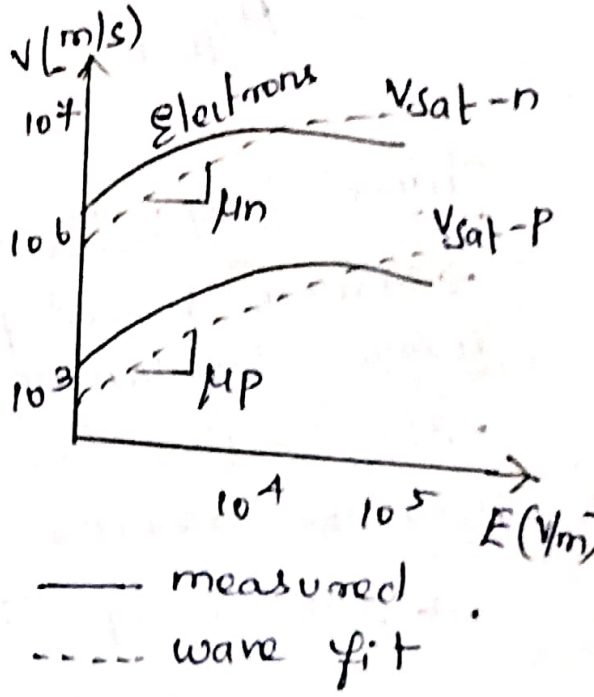
→ At high E_{lat}, carrier velocity rolls off.

→ carriers scatter of atoms in silicon lattice.

Velocity reaches V_{sat}

Electrons : 10^7 cm/s

Holes : 8×10^6 cm/s



$$v = \begin{cases} \dfrac{\mu_{eff} E}{1 + E/E_c} & E < E_c \\ v_{sat} & E \geq E_c \end{cases} \qquad E_c = \frac{2 v_{sat}}{\mu_{eff}}$$

Sub,

$$I_{ds} = \begin{cases} \dfrac{\mu_{eff}}{1 + V_{ds}/V_c} C_{ox} \dfrac{W}{L} \left(V_{gs} - \dfrac{V_{ds}}{2}\right) V_{ds}, & V_{ds} < V_{d,sat} \\ C_{ox} W (V_{gs} - V_{d,sat}) v_{sat}, & V_{ds} > V_{d,sat} \end{cases}$$

→ Note that μ_{eff} is also a function of V_{gs} due to mobility degradation.

→ Equility the two eqn above at the boundry $V_{ds} = V_{d,sat}$

$$V_{d,sat} = \frac{V_{gs} V_c}{V_{gs} + V_c}$$

→ Ideal transistor ON current Increases with V_{DD}^2

$$I_{ds} = \mu C_{ox} \cdot \frac{W}{L} \cdot \left(\frac{V_{gs} - V_t}{2}\right)^2 = \frac{\beta}{2} (V_{gs} - V_t)^2$$

→ Velocity saturated ON current increases with

$$I_{ds} = C_{ox} \cdot W (V_{gs} - V_t) V_{max}$$

Mobility Degradation:-

→ High overst efficiently reduces mobility.

— collisions with oxide Interface.

$$\mu_{eff-n} = \frac{540 \frac{cm^2}{V \cdot s}}{1 + \left(\dfrac{V_{gs} + V_t}{0.54 \frac{V}{nm} t_{ox}}\right)^{1.85}}$$

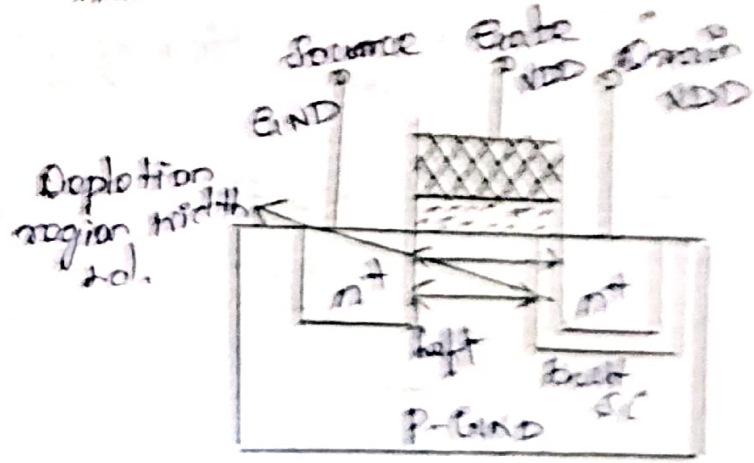
$$\mu_{eff-p} = \frac{185 \frac{cm^2}{V \cdot s}}{1 + \dfrac{|V_{gs} + 1.5 V_t|}{0.338 \frac{V}{nm} t_{ox}}}$$

Channel length Modulation:-

→ Reverse biased p-n jn forms a depletion region.

→ Region blw n and p with no carrier

→ Width of depletion L_d



region grows with reverse bias.

$$L_{eff} = L - L_d$$

→ Shorter eff gives more current

- I_{ds} increases with V_{ds}
- even in saturation.

$$I_{ds} = \beta/2 \, (V_{gs} - V_t)^2 (1 + \lambda V_{ds})$$

Where, λ → channel length Modulation coefficient

- Not feature size
- Empirically fit to I-V characteristics.

Body Effect:-

→ The potential diff. b/w source and body affect the threshold voltage. It is a fourth transistor terminal.

→ V_{sb} affect the charge required to invert the channel.

→ Increasing V_s or decreasing V_b increases V_T.

$$V_t = V_{to} + \gamma(\sqrt{\phi + V_{sb}} - \sqrt{\phi_s})$$

Where, ϕ_s = surface potential at threshold.

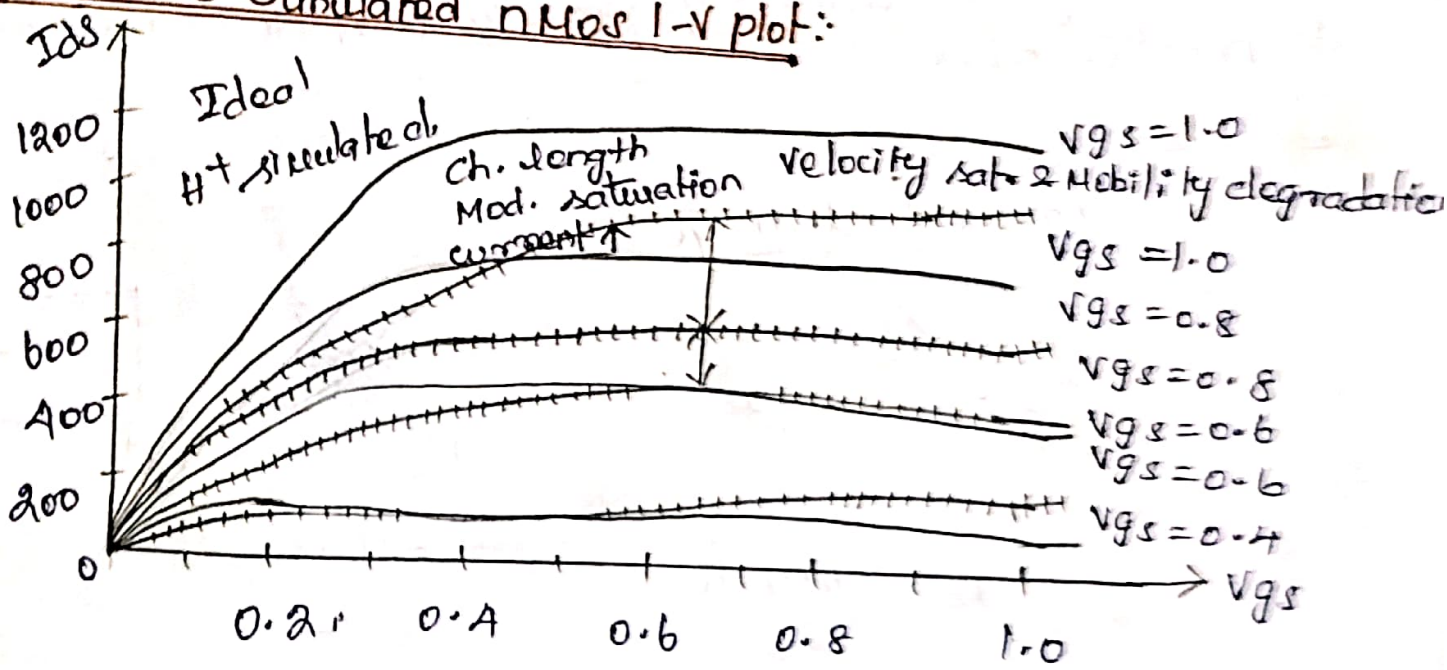
$$\phi_s = 2V_T \ln \frac{N_A}{N_i}$$

- Depends on doping level N_A.
- Intrinsic carrier concentration n_i.

Where, γ = body effect coefficient

$$\gamma = \frac{t_{ox}}{\varepsilon_{ox}} \sqrt{2q\varepsilon_{si} N_A} = \frac{\sqrt{2q\varepsilon_{si} N_A}}{C_{ox}}$$

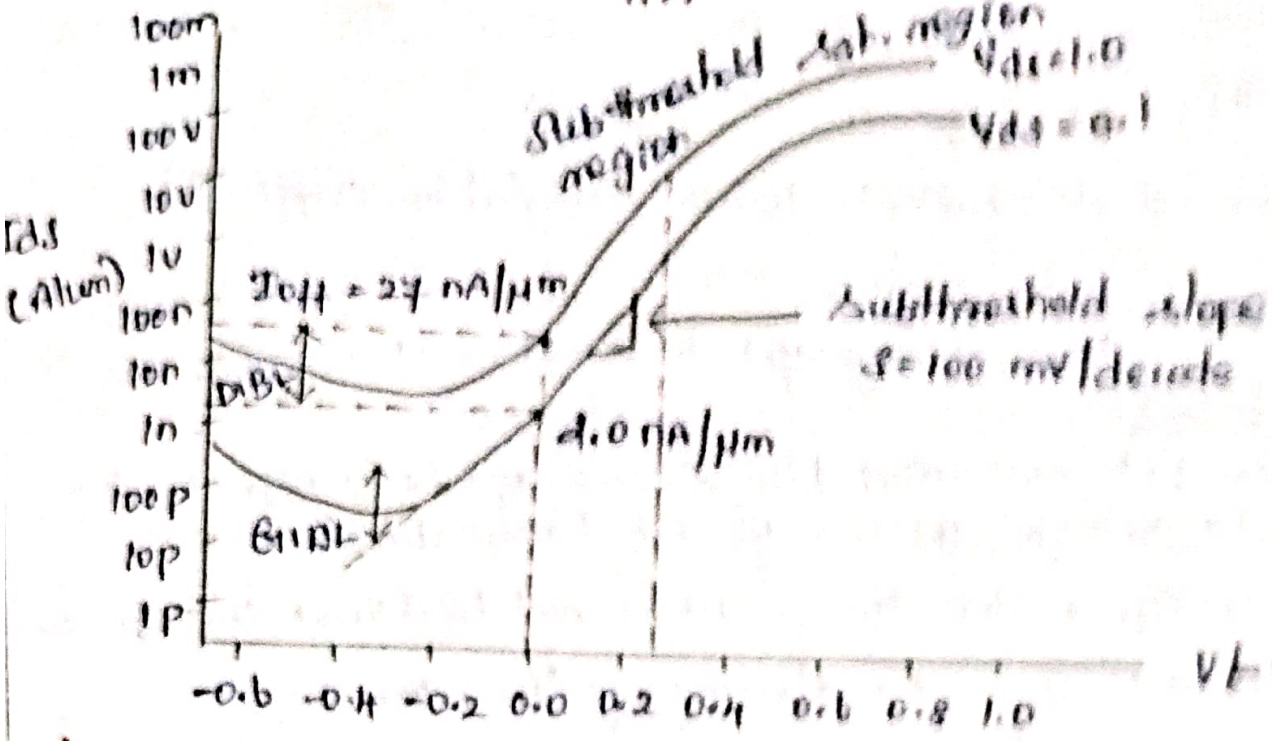
Ideal Vs Simulated nMos I-V plot:-



Sub-threshold leakage:-

→ Sub-threshold leakage exponential with V_{gs}.

$$I_{ds} = I_{dso} \frac{V_{gs} - V_{to} + \eta V_{ds} - \gamma V_{sb}}{nV_T} \left(1 - e^{-\frac{V_{ds}}{V_T}}\right)$$



→ This condition is also known as leakage and often this result is undesired current when a transistor is normally OFF. I_{dso} is the current at threshold

Application:-

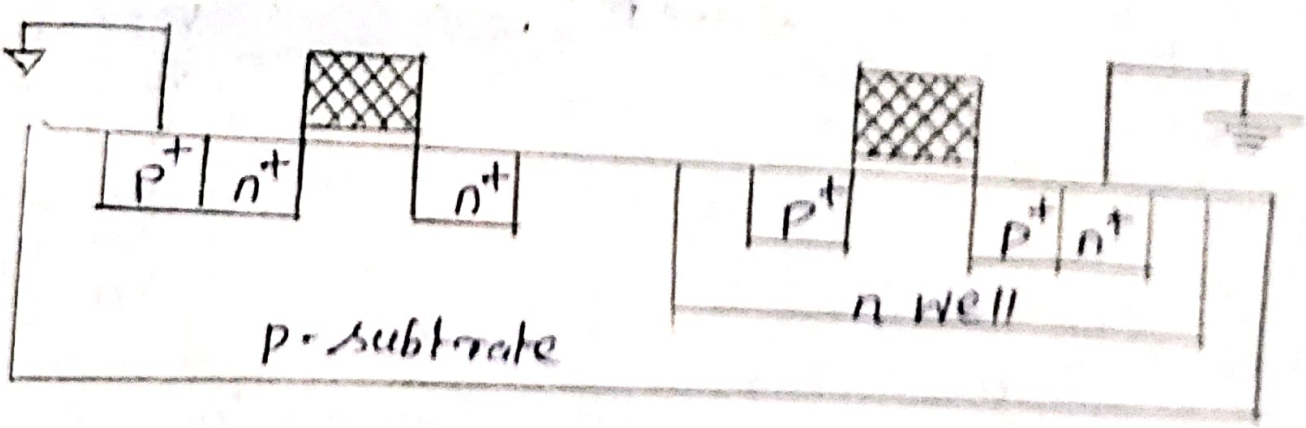
→ This is used in very low power analog circuit

→ This is used in dynamic circuit and DRAM.

Advantage:-

→ Leakage increases exponentially as V_t decreases or as temperature rises.

Junction Leakage:-



Reverse biased p-n junction have some leakage

- Ordinary diode leakage
- Band-to-Band Tunneling (BTBT)
- Gate induced drain leakage (GIDL)

Reverse-biased p-n junction have some leakage

$$I_D = I_S\left(e^{\frac{V_D}{V_T}} - 1\right)$$

Where, I_D = diode current
V_d = diode voltage

Tunneling:-

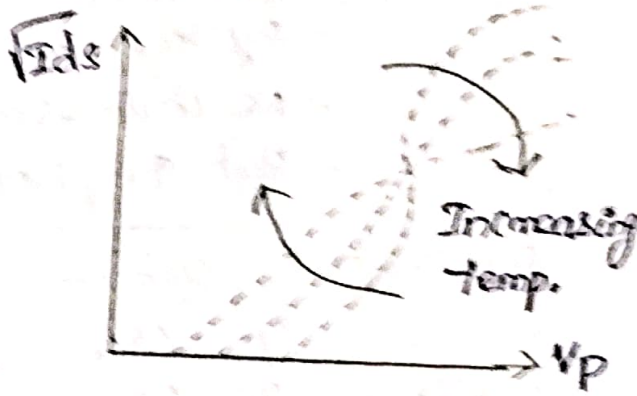
According to the Quantum mechanics, there will a finite probability that carrier will tunnel through the gate oxide this result in gate leakage current flowing into the gate and it exponentially decreased with oxide thickness.

Temperature Sensitivity:-

→ Increasing temperature
- Reduce Mobility
- Reduce V_t.

→ I_{ON} decreases with temperature

→ I_{OFF} increases with temperature



Geometric Dependenses:-

→ Geometric dependensis layout design. VLSI chip for particular dimensions but the actual dimensions get varied with respect to some parameter. Hence MOSFET is also depends on dimensions of design

$$L_{eff} = L_{drawn} + X_L - 2L_D$$

$$W_{eff} = W_{drawn} + X_W - 2W_D$$

Where,

W_{drawn} & L_{drawn} = Drawn transistor with width and length.

(iv) DC Transfer Characteristics:-

DC transfer characteristics include the following.

i. Complementary CMOS Inverter DC Char.
ii. Beta Ratio Effect
iii. Noise Margin
iv. Ratioed Inverter Transfer function
v. Pass Transistor DC Char.
vi. Tristate Inverter.

Complementary CMOS Inverter DC characteristics:-

→ DC response: V_{out} vs V_{in} for a gate

Ex: Inverter

- When $V_{in}=0$ $V_{out}=V_{DD}$
- When $V_{in}=V_{DD}$ $V_{out}=0$
- In between V_{out} depends on transistor size and current.
- By KCL, must settle such that $I_{dsn} = |I_{dsp}|$
- We could solve equation.
- But graphical solution gives more insight.

operation	cut-off	Linear	Saturation
nMOS	$V_{gsn} < V_{tn}$ $V_{in} < V_{tn}$	$V_{gsn} > V_{tn}$ $V_{in} > V_{tn}$ $V_{dsn} < V_{gsn} - V_{tn}$ $V_{out} < V_{in} - V_{tn}$	$V_{gsn} > V_{in}$ $V_{in} > V_{tn}$ $V_{dsn} > V_{gsn} - V_{tn}$ $V_{out} > V_{in} - V_{tn}$
PMOS	$V_{gsp} > V_{tp}$ $V_{in} > V_{DD} + V_{tp}$	$V_{gsp} < V_{tp}$ $V_{in} < V_{DD} + V_{tp}$ $V_{dsp} > V_{gsp} - V_{tp}$ $V_{out} > V_{in} - V_{tp}$	$V_{gsp} < V_{tp}$ $V_{in} < V_{DD} + V_{tp}$ $V_{dsp} < V_{gsp} - V_{tp}$ $V_{out} < V_{in} - V_{tp}$

$V_{tp} = -V_{tn}$, $\beta_n = \beta_p$. I_{dsn} and I_{dsp} in terms of V_{dsn} and V_{dsp} for various values of V_{gsn} and V_{gsp}.

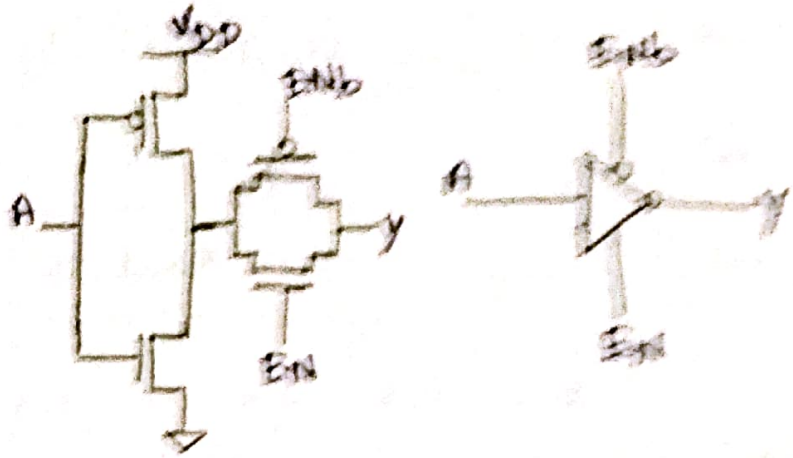
Application:-

- It is used as a multiplexing element, a logic structure, a latch element and an analog switch.

Tristate Inverter:-

→ Tristate inverter is constructed by cascading a transmission gate with an inverter.

- When EN=0 and ENb=1, output Y of the inverter is in tristate
- When EN=1 and ENb=0, output Y of the inverter is the complement of input A.



Application:-

It is used in various types of clocked logic, latches, bus driver, I/O structures & multiplexers.

(M) Device Models:-

SPICE provides different varieties of MOS transistor models as given below

i. Level 1 Models.
ii. Level 2 and 3 Models
iii. BSIM Models.
iv. Diffusion capacitance model

Level 1 Models:-

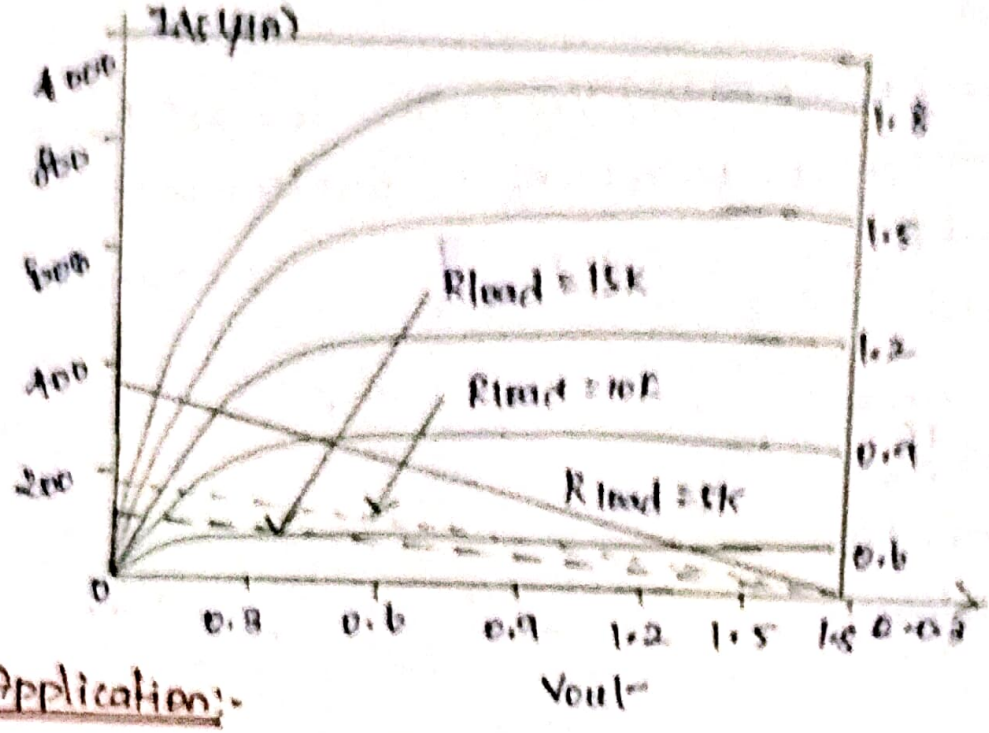
SPICE level 1 model is also called Shichman-Hodges model. This is closely related to Shockley model. The basic current model is given by

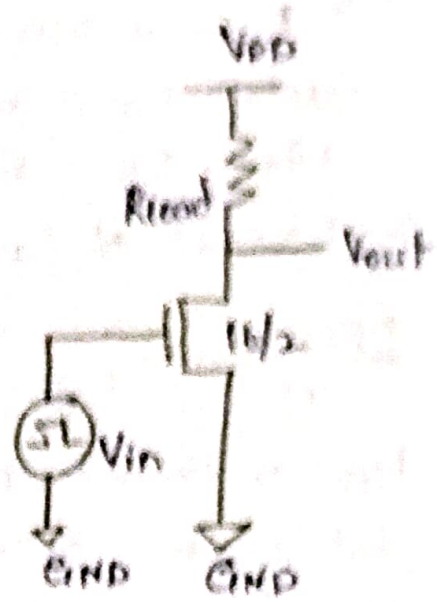
$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \text{ - cut off} \\ k_p \dfrac{W_{eff}}{L_{eff}} (1+LAMDA \cdot V_{ds})(V_{gs} - V_t - V_{ds}/2) V_{ds} & V_{ds} < V_{gs} - V_t \text{ - Linear} \\ k_p \dfrac{W_{eff}}{L_{eff}} (1+LAMDA \cdot V_{ds})(V_{gs} - V_t)^2 & V_{ds} \geq V_{gs} - V_t \text{ - saturation} \end{cases}$$

$$V_t = VTO + GAMMA\left(\sqrt{PHI + V_{sb}} - \sqrt{PHI}\right)$$

Ratioed Inverter Transfer function:-

- → $V_{in} = V_{DD}$, the output becomes small. When $V_{in} = 0$, V_{out} rises to V_{DD}
- → In the transfer characteristics for a given V_{in}, V_{out} is obtained with $I_{dsn} = |V_{dsp}|$





Application:-

This is used to limited circumstances where they offer critical benefits such as smaller area.

Pass Transistor DC Characteristics:-

When an nMOS or a pMOS is used alone as an imperfect switch a pass transistor is obtained. Pass gate or transmission gate consists of an nMOS transistor and a pMOS transistor in parallel with gate controlled by complementary signals. When transmission gate is ON, the transmission gate passes both '0' and '1' as well.

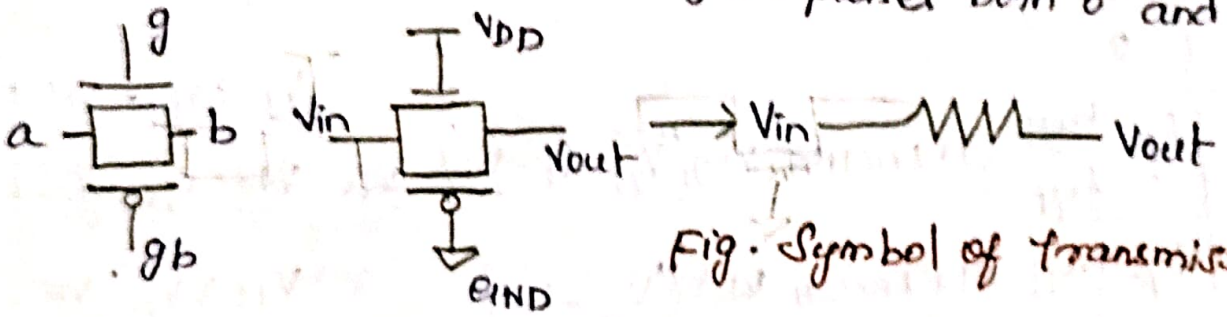


Fig. Symbol of transmission gate

When $V_s = 0$, $V_{gs} > V_{tn}$, the transistor is ON and I flows.

When $V_s = V_{DD} - V_{tn}$, V_{gs} falls to V_{in} and transistor is OFF

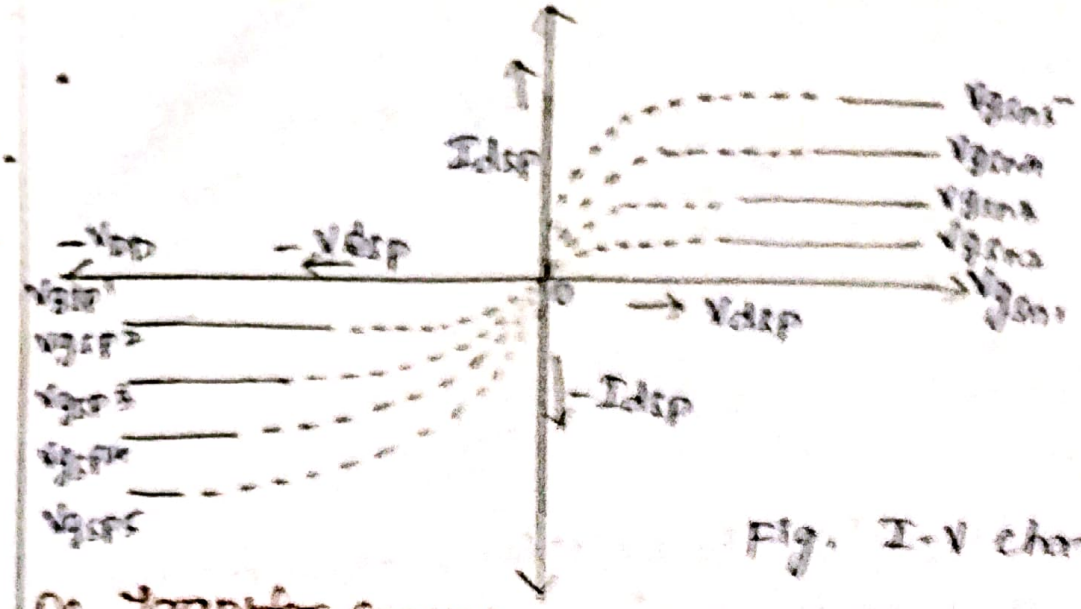
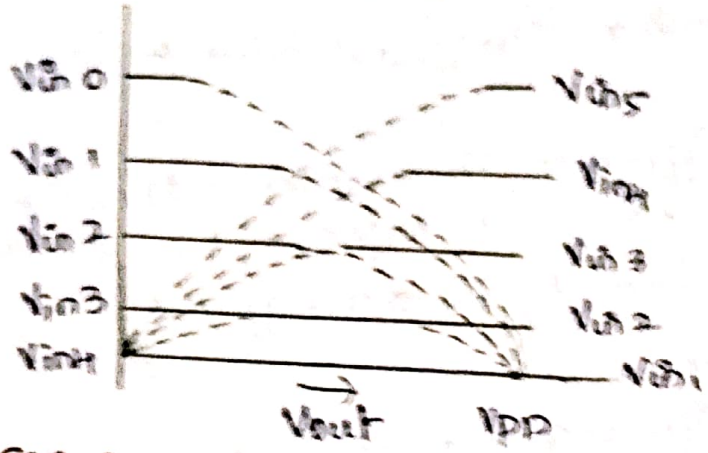
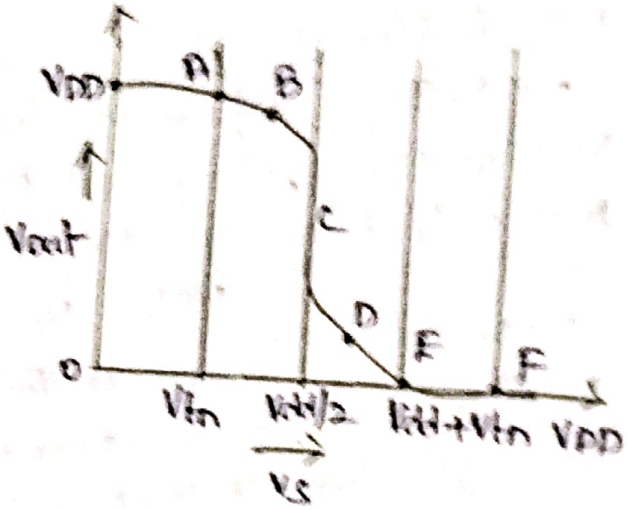


Fig. I-V characteristics

DC transfer curve:-





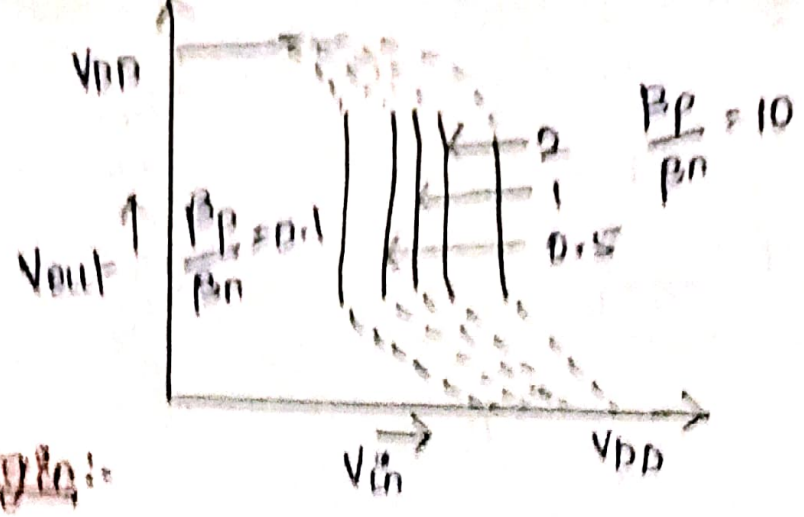
CMOS Inverter operation:-

Region	p-device	n-device	output
A	Linear	cut-off	$V_{out} = V_{DD}/2$
B	Linear	Saturated	$V_{out} > V_{DD}/2$
C	Saturated	Saturated	Vout drops
D	Saturated	Linear	$V_{out} < V_{DD}/2$
E	cut-off	Linear	$V_{out} = 0$

Beta Ratio output:-

→ If $\beta_n = \beta_p$, the inverter threshold voltage V_{inv} is $V_{DD}/2$. Inverters with different values of beta ratio, β_n/β_p are called skewed inverters. If $\beta_n/\beta_p > 1$, the inverter is HI-skewed. the inverter is Lo-skewed. If $\beta_n/\beta_p = 1$, the inverter has normal skew or unskewed. If the input is $V_{DD}/2$, the output will be greater than $V_{DD}/2$.

→ Other gates: collapse into equivalent inverter.



Noise Margin:-

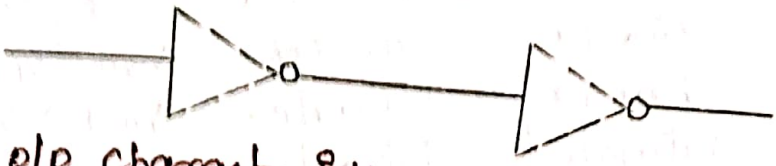
→ Noise margin is related to DC Voltage characteristics. Two parameters are used to specify noise margin or noise immunity, namely low noise margin (NM_L) and HIGH Noise margin (NM_H).

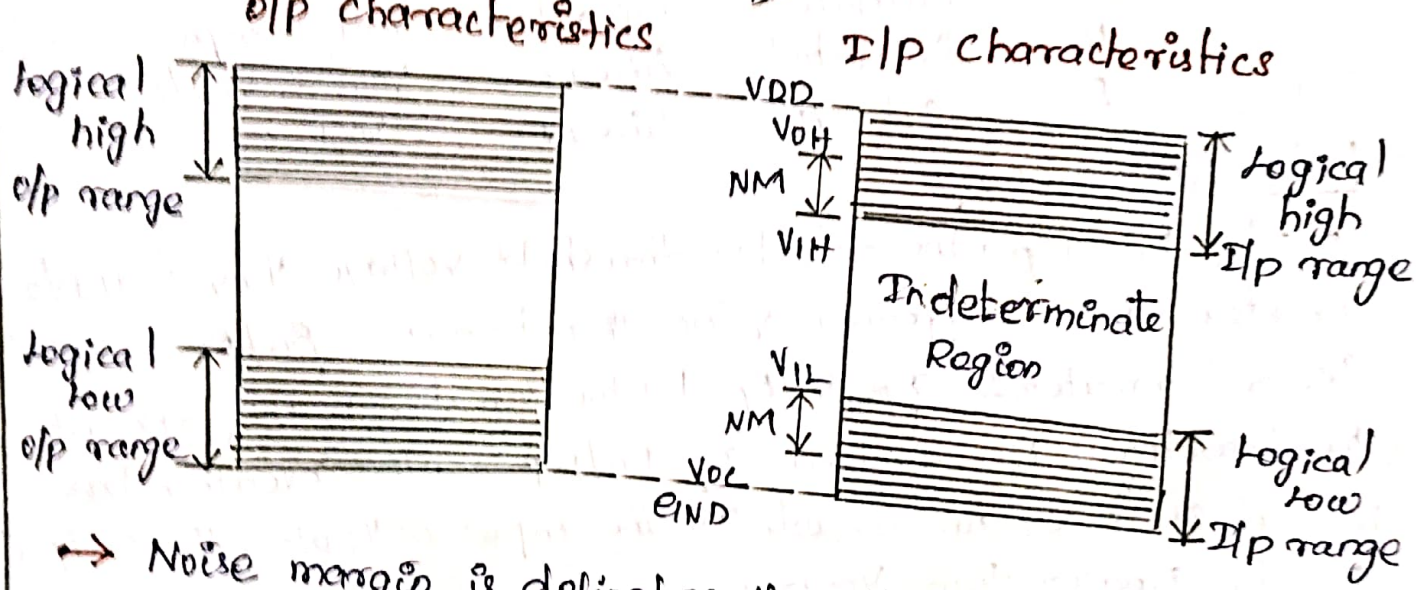
→ NM_L is defined as the difference in maximum LOW input voltage (V_{IL}) recognized by the receiving gate and the maximum low output voltage (V_{OL}) produced by the driving gate.

$$NM_L = V_{IL} - V_{OL}$$

→ NM_H is defined as the difference in minimum HIGH output voltage (V_{OH}) of the driving gate and minimum HIGH input voltage (V_{IH}) recognized by the receiving gate.

$$NM_H = V_{OH} - V_{IH}$$





→ Noise margin is defined as the allowable noise at the input of the gain so that the output will not be corrupted.

Where, Kp → Model parameter

Weff → effective Width Leff → effective length

LAMDA = models channel length Modulation

V_{To} = zero-bias threshold voltage V_{To}

GAMMA = body effective coefficient γ

PHI = surface potential.

Level 2 and 3 Models:-

These models added the effect of velocity saturation, mobility degradation, sub-threshold condition and drain induced barrier lowering. Level 2 model is based on Grove-Frohman equations Level 3 model is based on empirical equation that provides similar accuracy, faster, simulation time & presently not used.

BSIM:-

Berkeley short channel IGFET MOS this widely used in circuit simulation and it has the following features

1. Continous & differentiable IV characteristics
2. Sensitivity to parameter
3. Detailed Modulae.
4. Short channel effect consideration
5. Multiple gate capacitance Model.
6. Diffusion gate capacitance Model.

Diffusion Capacitance Model:-

→ Diffusion capacitance determines the parasitic delay of a gate; In SPICE models, the diffusion capacitance b/w source and body is given by

$$C_{sb} = AS \cdot CJ \cdot \left(1+\frac{V_{sb}}{PB}\right)^{-MJ} + PS \cdot CJSW \cdot \left(1+\frac{V_{sb}}{PHP}\right)^{-MJSW}$$

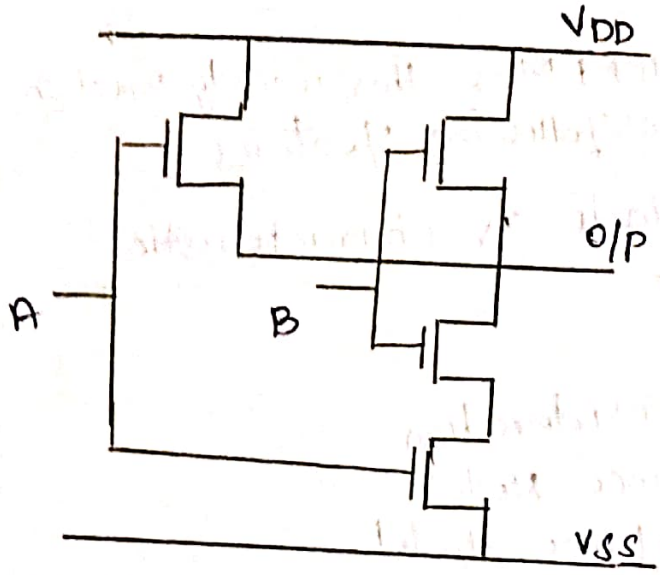
Where, CJ, CJSW, PB, PHP, MJ and MJSW are SPICE model parameter.

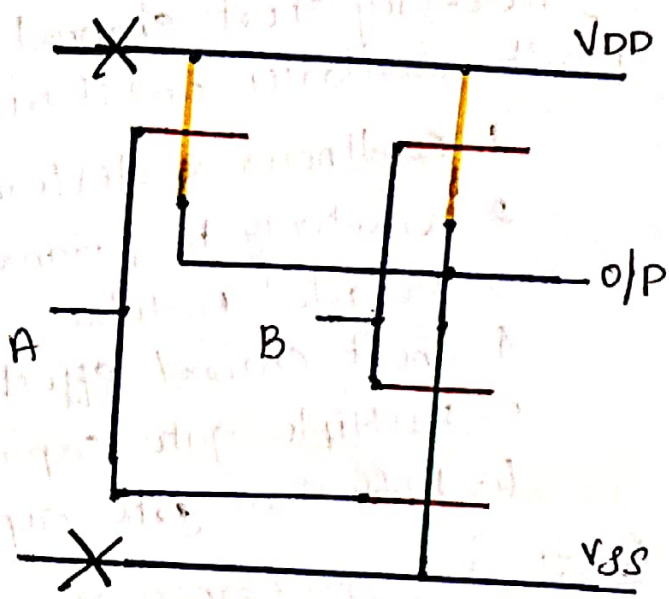
→ The diffusion area and perimeter are used to calculate the junction leakage current.

③ STICK DIAGRAM:-

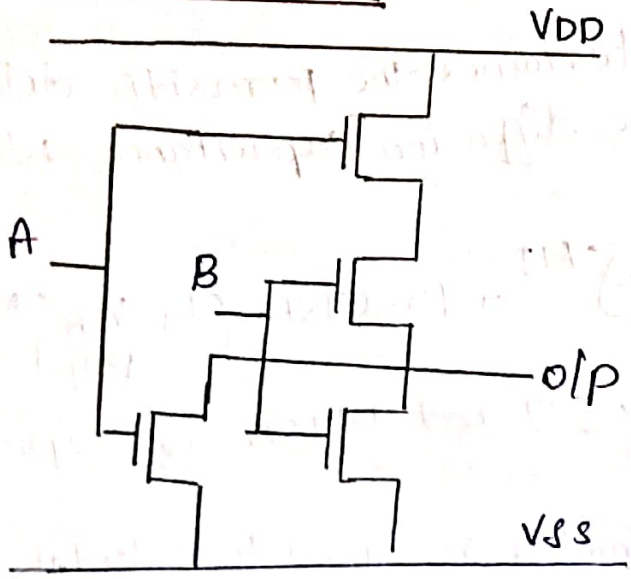
Layer	Colour Stick Notation	Colour codes
n-diffusion		Green
p-diffusion		Yellow
Poly-silicon		Red
metal-1		Blue
Metal-2		Dark blue or purple
contact cut	●	black
Via	●	black
demarcation line	- - - - - - - -	brown
Vdd or Vss contact	✕	black

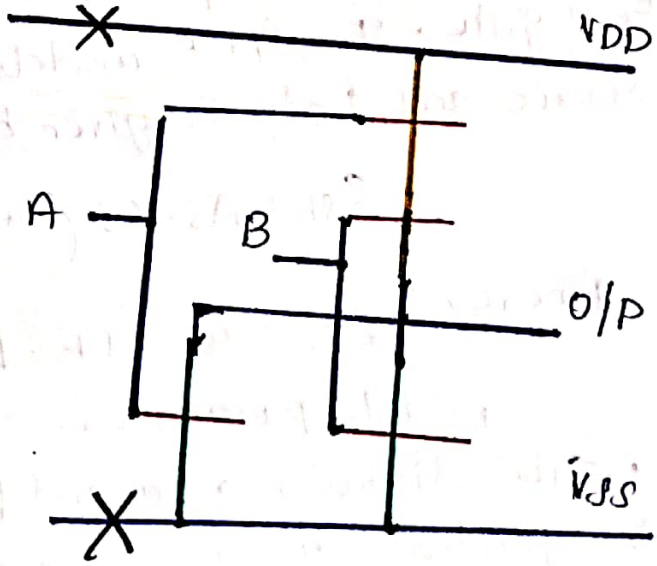
CMOS NAND Gate:-





CMOS NOR Gate:-





UNIT - 2

COMBINATIONAL LOGIC DESIGN

PART - B

Explain about static CMOS design in detail:

Static CMOS design are an extension of static CMOS inverter to multiple inputs. Each gate output is connected to either V_{DD} or V_{SS} via a low resistance path.

Advantages:

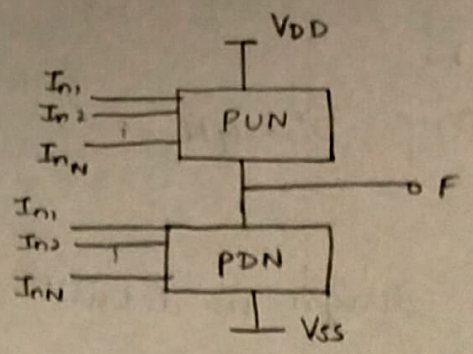
i) Low Sensitivity to noise
ii) Good Performance
iii) Low Power Consumption
iv) Large fan-in
v) Similar Circuit topology

Examples:

* Complementary CMOS
* Ratioed Logic
* Pass transistor Logic

Complementary CMOS:

* It is a combination of Pull-Up Network (PUN) and Pull down Network (PDN)
* The function of PUN is to connect output to V_{DD} when the output of logic gate is 1
* The function of PDN is to connect output to V_{SS} when the output of logic gate is 0
* In Steady state, either PUN or PDN will conduct.



A transistor can be thought of as a switch controlled by its gate signal.

* Strong Zeros - NMOS - PDN
* Strong Ones - PMOS - PUN

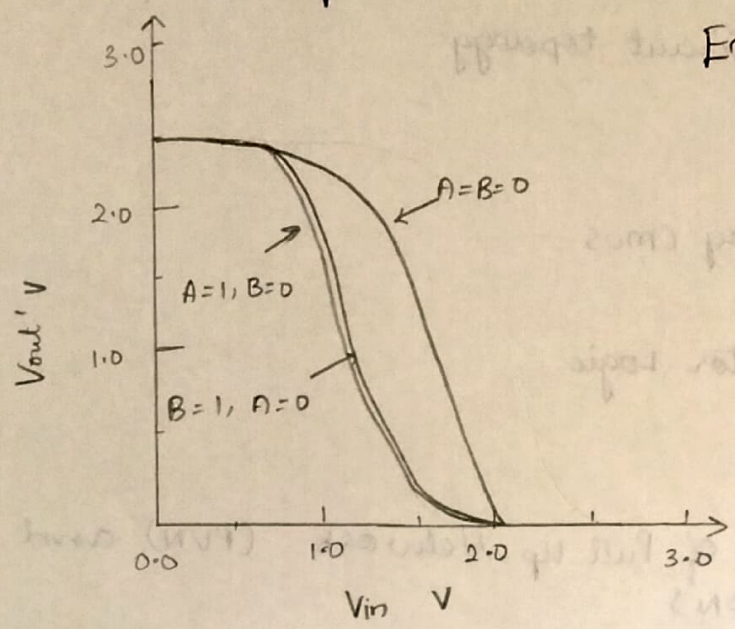
Static Properties:

It has rail to rail Swing with $V_{OH} = V_{DD}$ and $V_{OL} = GND$

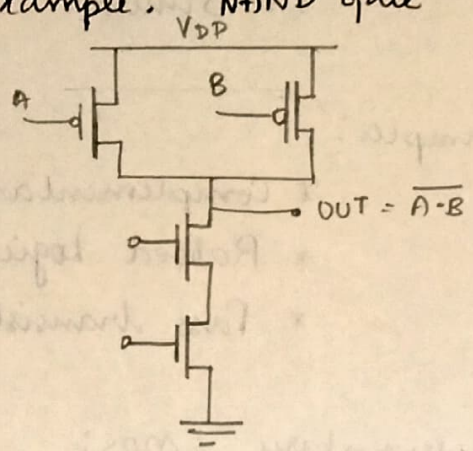
It has no static Power dissipation

The analysis of DC voltage transfer characteristics and noise margins are more complicated than inverter

It depends on data input patterns applied to gate



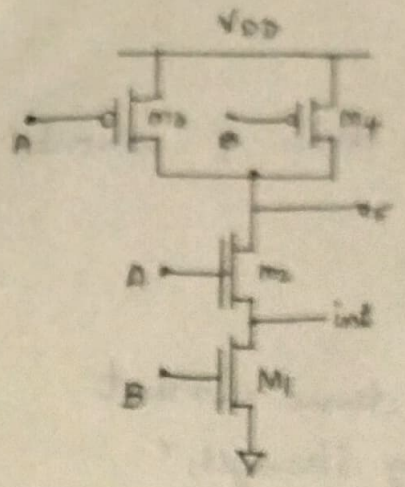
Example: NAND Gate

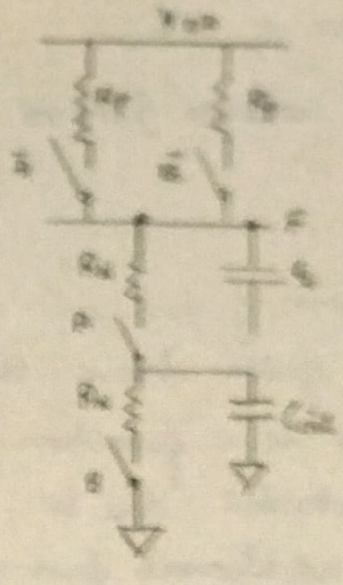


Delay Dependancies:

Each transistor is modeled as a resistor in series with an ideal switch.

The resistance value depends on power Supply voltage and an equivalent large signal resistance, Scaled by the ratio of device width over length.





Delay is dependent on the pattern of inputs

Low to high transition

- both inputs go low
  - delay is 0.69 $R_p/2\ C_L$
- One input goes low
  - delay is 0.69 $R_p C_L$

High to low transition

- both inputs go high
  - delay is 0.69 $2 R_n C_L$

Design Techniques:

i) Transistor Sizing: This lowers the resistance of device in series and lowers the time constants. It is effective only when the load is dominated by fan-out

ii) Progressive Transistor Sizing: Each transistor is scaled up uniformly. $M_1 > M_2 > M_3 > M_4$. It reduces the dominant resistance.

iii) Input Recording: Some signals is more critical than others. An input signal to a gate is called critical, if it is the last signal of all inputs to assume stable value

iv) Logic Restructuring: Manipulating the logic equations can reduce the fan-in requirements and thus reduce the gate delay.

t_p as a function of fan-in :

Gates with a fan-in greater than 4 should be avoided

t_p as a function of fan-out:

All gates have the same drive current

Slope is a function of "driving Strength"

Fan-in : Quadratic due to increasing resistance and Capacitance

Fan-out : Each additional fan-out gate adds two gate Capacitance to C_L

$$t_p = a_1 FI + a_2 FI^2 + a_3 FO$$

Power Consumption :

i) Static Component: This is only a function of topology of the logic network.

ii) Dynamic Component: This is due to the timing behaviour of circuit. This is also called glitching.

Design Techniques to reduce Switching Activity :

Logical restructuring

Input ordering

Time multiplexing resources

Glitch reduction

The following techniques are used to optimize the Static CMOS circuits.

Bubble Pushing

Compound gates

Input ordering Delay Effect

Asymmetric Gates

Skewed Gates

P/N Ratios

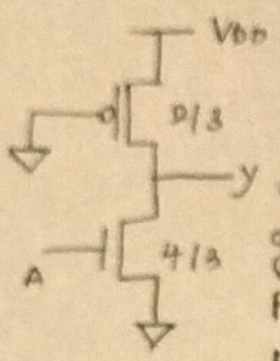
Multiple Threshold Voltages

Ratioed logic:

This uses weak pull-up and strong Pull down
They are designed by following logics
i) Pseudo - nMOS logic
ii) Ganged CMOS logic
iii) Source Follower Pull-Up logic
iv) Cascode Voltage Switch logic

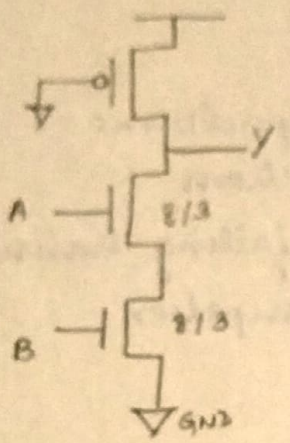
Pseudo - nMOS Logic:

Pull-up network is same, Pull down network is replaced by pMOS.
PMOS transistor is always ON.



Inverter

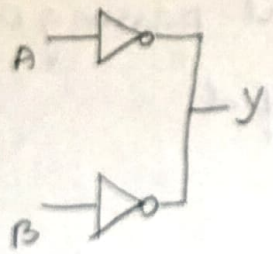
$g_u = 4/3$
$g_d = 4/9$
$g_{avg} = 8/9$
$P_u = 18/9$
$P_d = 6/9$
$P_{avg} = 12/9$

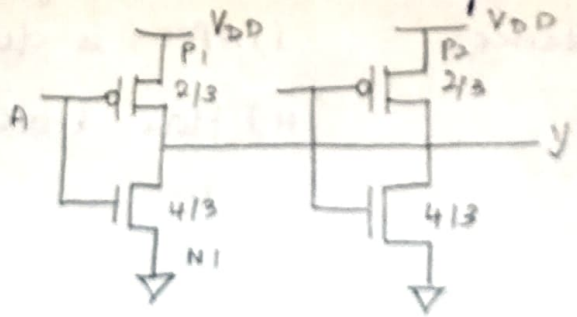


$g_u = 8/3$
$g_d = 8/9$
$g_{avg} = 16/9$
$P_u = 30/9$
$P_d = 10/9$
$P_{avg} = 20/9$

Ganged CMOS Logic:

Also called Symmetric NOR.
When one I/p is 0 and other 1, it will act as pseudo nMOS.
When both I/p's are 0, both pMOS turns ON.
When both I/p's are 1, both pMOS turns OFF.





$g_u = 1$
$g_d = 2/3$
$g_{avg} = 5/6$

Source Follower:

Similar to pseudo nMOS gate except pull-up is controlled by inputs

When one i/p is ON, Source follower pulls node x to $\frac{V_{DD}}{2}$

This will partially turn OFF P1 and o/p will be OFF

Cascode Voltage Switch Logic:

Also called differential Cascode Voltage Switch Logic

Transistor are connected in Series

It has performance of ratioed circuits without static power consumption.

This uses both true and complementary i/p S/c's

Disadvantages:

Reduced input capacitance

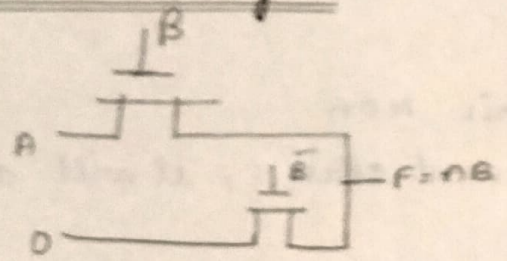
Slow rising transitions

Contention on the falling transitions

Static Power dissipation

Non-zero V_{OL}

Pass Transistor Logic:



If B is high, then output F will be input A.

If B is low, the bottom transistor will be turned on & o/p will be 0

Advantages:

i) Low Capacitance

Disadvantages:

i) nMOS is strong at passing 0

ii) Has body effect

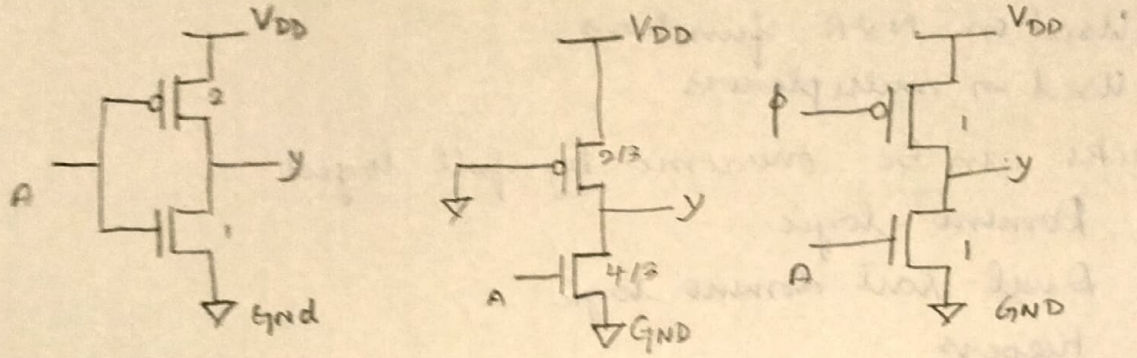
Differential Pass Transistor Logic
Efficient Pass Transistor Logic
CMOS with Transmission Gates
are Some of the pass Transistor Logic circuits.

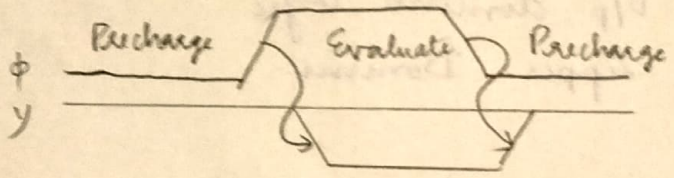
DYNAMIC CIRCUITS :

Dynamic circuits overcome the drawbacks of ratioed circuits by the use of clocked pull-up transistors.

It operates in 2 modes :

* Precharge - clock ϕ is 0
* Evaluation - clock ϕ is 1





In dynamic inverter,

During Precharge, input - 1 , PMOS & nMOS will be ON
contention take place.
input - 0, avoids contention

The additional transistor is called a foot.

Advantages:

Lower input capacitance
No Contention during switching
Zero static power dissipation

Disadvantages:

Require careful clocking
Consume significant dynamic power
Sensitive to noise

Applications:

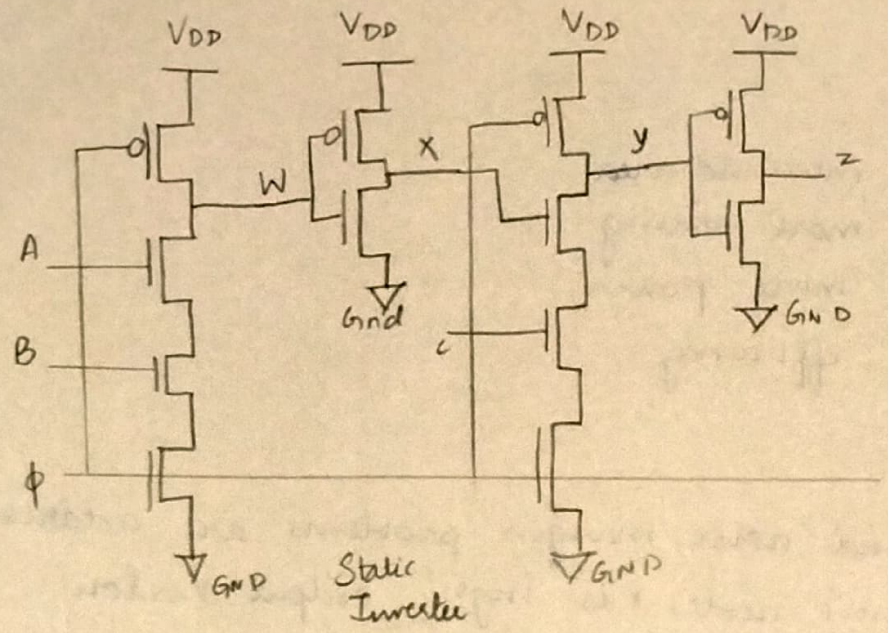
Used in NOR functions
Used in multiplexers

Drawbacks can be overcome by foll. logics:

Domino logic
Dual-rail domino logic
Keepers
Multiple-O/p domino logic
NP and Zipper Domino

DOMINO Logic:

The dynamic - static pair together is called a domino gate.
Single clock can be used to precharge
The static inverter is usually a HI-Skew gate
Evaluation occurs sequentially.



Dual-Rail Domino Logic:

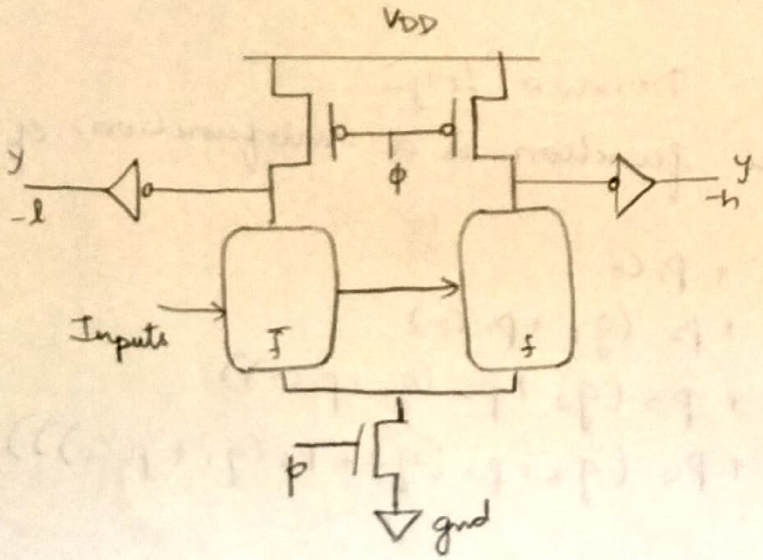
It can compute all inverting and non-inverting logic functions

It can be reviewed as dynamic form of CVSL.

It can encode each signal with a pair of wires.

Iput = 1, _h is asserted

output = 0, _l is asserted

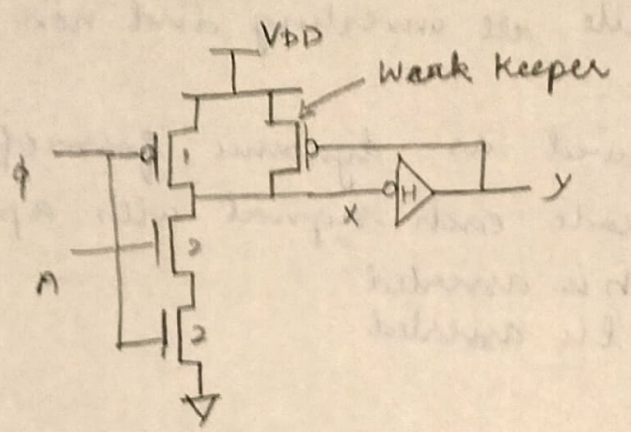


Disadvantages:

Requires more ~~ad~~ area
Requires more wiring
Requires more power
Has less efficiency

Keepers:

Leakage and noise margin problems are addressed
If dynamic node, x is high, output y is low
If x falls, y rises
Noise Margin can be improved by using strong keepers.

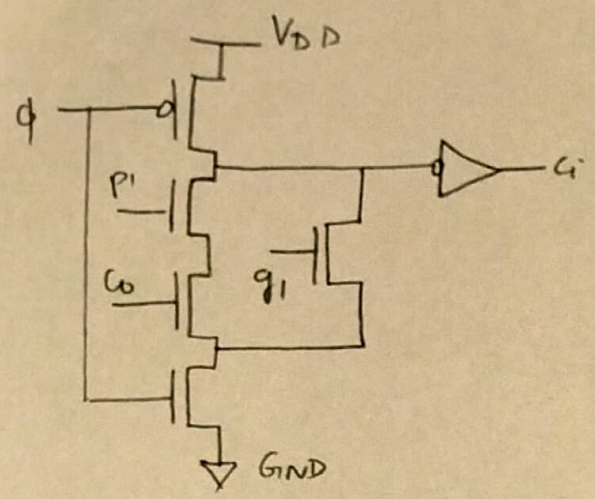


Multiple-Output Domino Logic:

It is one function is a subfunction of another sub function

$$C_1 = g_1 + p_1 C_0$$
$$C_2 = g_2 + p_2 (g_1 + p_1 C_0)$$
$$C_3 = g_3 + p_3 (g_2 + p_2 (g_1 + p_1 C_0))$$
$$C_4 = g_4 + p_4 (g_3 + p_3 (g_2 + p_2 (g_1 + p_1 C_0)))$$

Example: Look ahead carry adder



NP and Zipper Domino:

The HI-Skew inverting static gates are replaced
If ϕ is 0, I & III stages precharge high & II - low
ϕ is 1, all stages evaluate

Disadvantages:

Logical effort is worst
Susceptible to noise

LOW POWER DESIGN PRINCIPLES:

Low power dissipation is greater than static power dissipation when systems are active.

For high performance systems, dynamic power consumption per chip is limited to 150W.

Commonly used metrics in low power design are power, power-delay product and energy-delay product.

There are two power reduction techniques.

Dynamic Power reduction
Static Power reduction

* Dynamic Power Reduction:

It can be reduced by decreasing the following factors.

i) Activity Factor
ii) Switching Capacitance
iii) Power Supply
iv) Operating frequency.

i) Activity Factor:

Static logic has a low activity factor
Clocked nodes have an activity factor of 1
Turnoff the clock network wherever possible.
Sense chip temperature and cut back activity, if temperature becomes too high.
Spike may occur, leads to inductive noise.

ii) Switching Capacitance:

Reduced by choosing small transistors

Small gates can be used on non-critical paths.

Interconnect switching capacitance can be reduced by proper floor planning

iii) Power Supply:

Choosing low power supply reduces power consumption

Voltage can be adjusted on operating mode.

iv) Operating Frequency:

In DSP, two multipliers running at half speed can be replaced by a single multiplier at full speed to reduce power consumption.

* Static Power Reduction:

It involves minimizing I_{static}.

Analog current source and pseudo-nMOS gates are turned off.

Reverse Body Bias (RBB) can be used during idle mode to reduce leakage

Forward Body Bias (FBB) can be used during active mode to increase performance.

Adaptive Body bias (ABB) can compensate and achieve more uniform transistor performance.

Sub-Threshold leakage can be controlled through the body voltage using body-effect.

POWER DISSIPATION:

Static CMOS gates are very power-efficient
They dissipate nearly zero power while idle

Instantaneous Power: (P_t):

It is propotional to ~~and~~ current I_{DD} and voltage V_{DD}

$$p(t) = i_{DD}(t)\, V_{DD}$$

Energy (E):

The energy consumed over time interval T is the integral of instantaneous power.

$$E = \int_0^T i_{DD}(t)\, V_{DD}\, dt$$

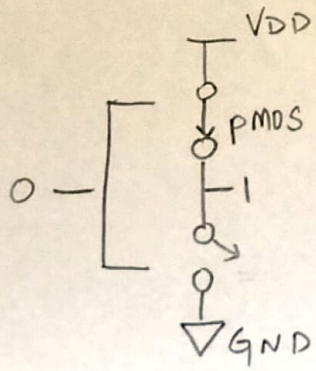
Average Power (P_{avg}):

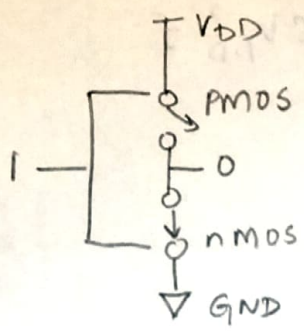
$$P_{avg} = \frac{E}{T} = \frac{1}{T}\int_0^T i_{DD}(t)\, V_{DD}\, dt$$

The 2 Components:
Static Power dissipation
Dynamic Power dissipation

Static Dissipation:

Occurs due to following factors
i) Sub-threshold condition through OFF transistors
ii) Tunneling current through gate oxide
iii) Leakage through reverse biased diodes
iv) Contention current in ~~rat~~ ratioed circuits.





I/p	nMOS	pMOS	O/P
0	off	ON	1
1	ON	off	0

In both cases, one of the transistor is off.
Ideally, no current flows through the OFF transistor
It is the product of the total leakage current (I_{static}) and the supply voltage (V_{DD})

$$P_{static} = I_{static} \times V_{DD}$$

Dynamic Dissipation:

Occurs due to following factors:

i) charging and discharging of load capacitance.

ii) Short circuit current while both pMOS and nMOS networks are partially ON.

During charging, current flow is from V_{DD} to load
During discharging, current flow is from load to GND.

Avg. Power dissipation,

$$P_{dynamic} = \frac{1}{T}\int_0^T i_{DD}(t) V_{DD}\, dt = V_{DD}/T \int_0^T i_{DD}(t)\, dt$$

integrating,

$$P_{dynamic} = \frac{V_{DD}}{T}\left[T f_{sw} C V_{DD}\right] = C V_{DD}^2 f_{sw}$$

Dynamic Power dissipation,

$$P_{dynamic} = \alpha C V_{DD}^2 f$$

STATIC LATCHES AND REGISTERS:

1. Bistability principle
2. Multiplexer - based Latches
3. Master - slave edge triggered Registers
4. Low - voltage static Latches
5. Static SR Flip Flops

BISTABILITY PRINCIPLE:

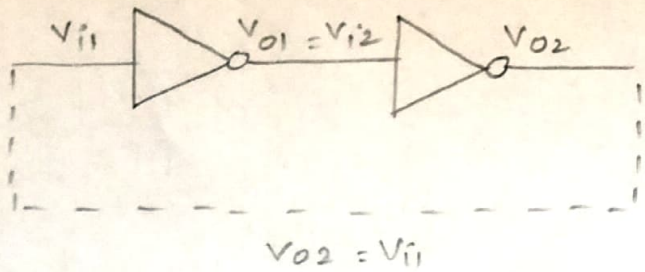
Bistable circuits have two stable states (0 and 1). If the output of second inverter is connected to input of first inverter as shown by dotted lines, then three operation points (A, B and C) exists. If cross coupled inverter is coupled at point C, small deviation δ occurs which is amplified and bias C point moves to operation point A or B. So point C is unstable or metastable.

It can be used as a memory. A trigger is used to change the logic state.

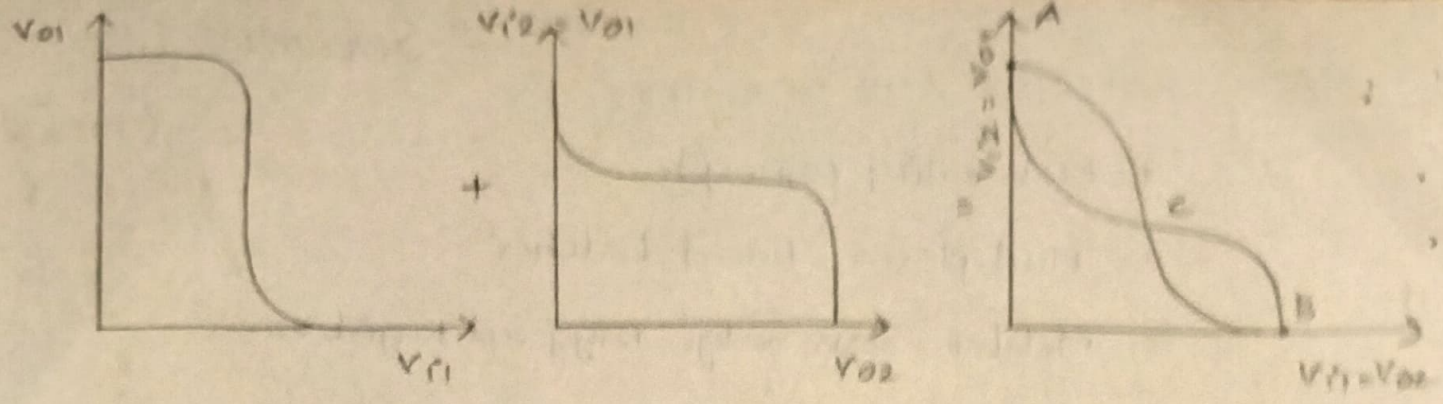
The changing of logic state is made by following ways

i) cutting the feedback loop (multiplexer based latches

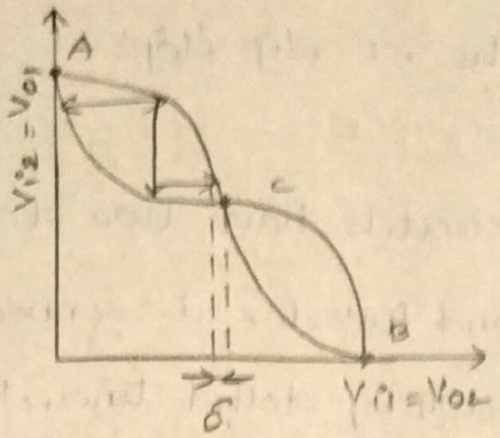
ii) overpowering the feedback loop.



Two cascaded inverters



VOLTAGE TRANSFER CHARACTERISTICS

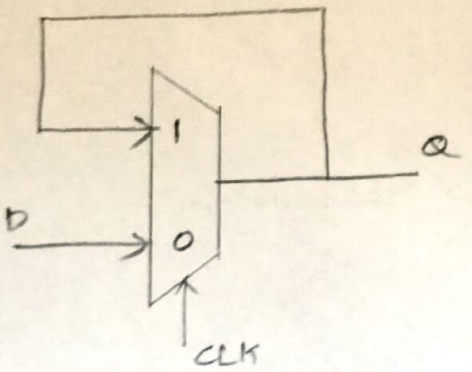


METASTABLE OPERATION POINT

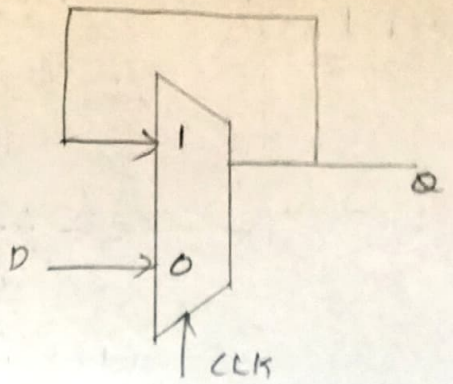
MULTIPLEXER - BASED LATCHES:

In negative latch, the D input to the output 0 input of the multiplexer is selected when the clock is low. D input is present at the output.

When CLK is high the 1 input of the multiplexer is selected. Since the output of 1 input of multiplexer as feedback, the output remains same when the clock is high. At this time, the output is stable.



NEGATIVE LATCH

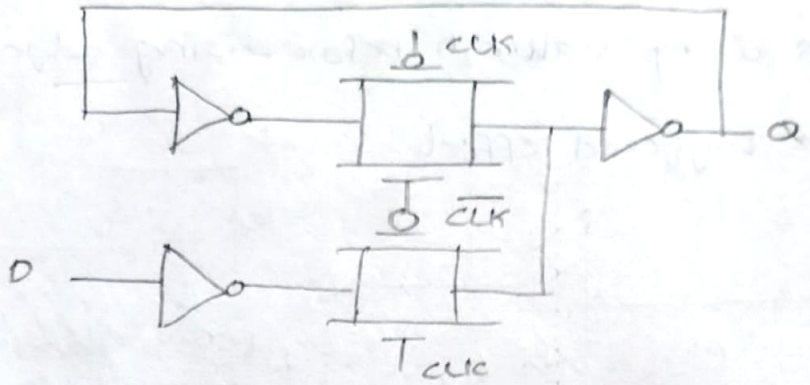


POSITIVE LATCH

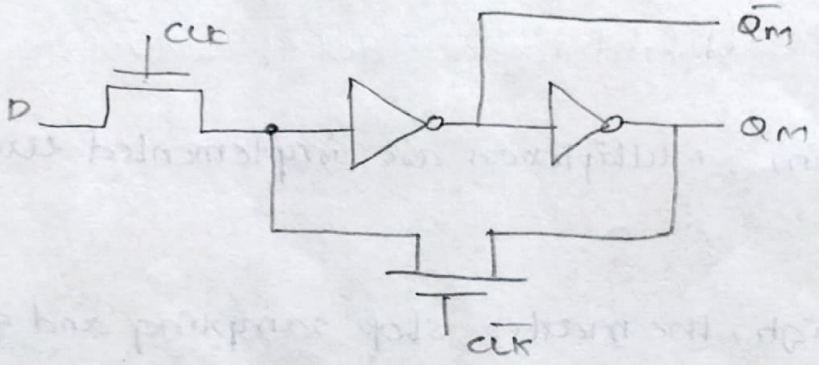
1. MULTIPLEXER BASED NEGATIVE AND POSITIVE LATCHES

In below diagram, when clock is high, the top transmission gate is off and the bottom transmission gate is ON. Now feedback loop is open. D input is copied to output Q.

It loads 4 transistors to clock signal. It can be reduced by 2 using nMOS only as pass transistor.



2. POSITIVE LATCH USING TRANSMISSION GATE



3. MULTIPLEXER BASED POSITIVE LATCH USING NMOS PASS TRANSISTOR

In above diagram, when CLK is high, the D input is copied to output.

when CLK is low, the feedback is enabled and in hold mode.

Disadvantages:

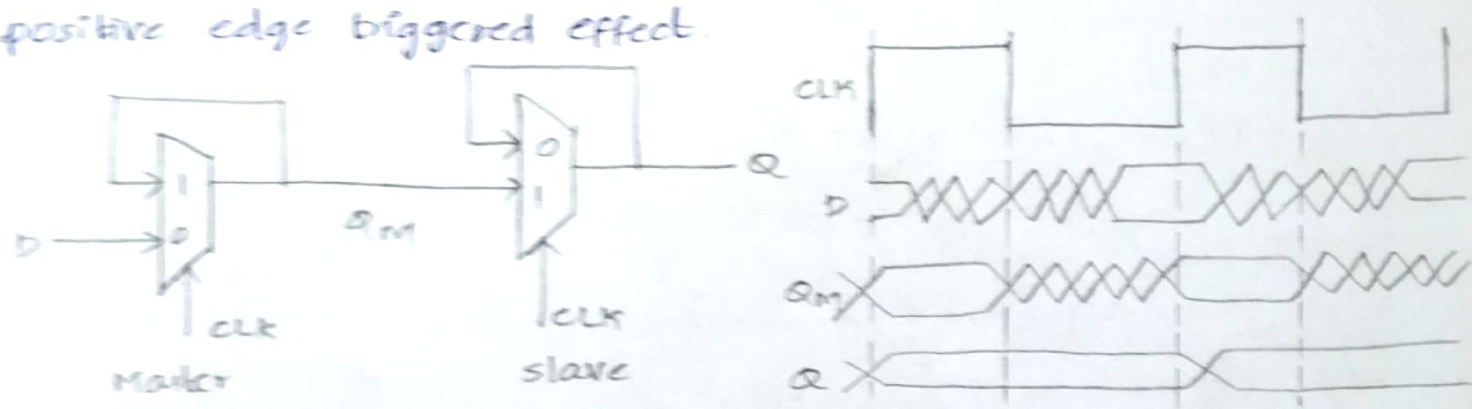
High static power dissipation in first inverter.

MASTER SLAVE EDGE TRIGGERED REGISTER:

Negative latch act as the master and the positive latch act as slave.

During low phase of clock, the master is transparent and the D input is passed to the master stage output. The slave is hold mode keep its previous value using feedback.

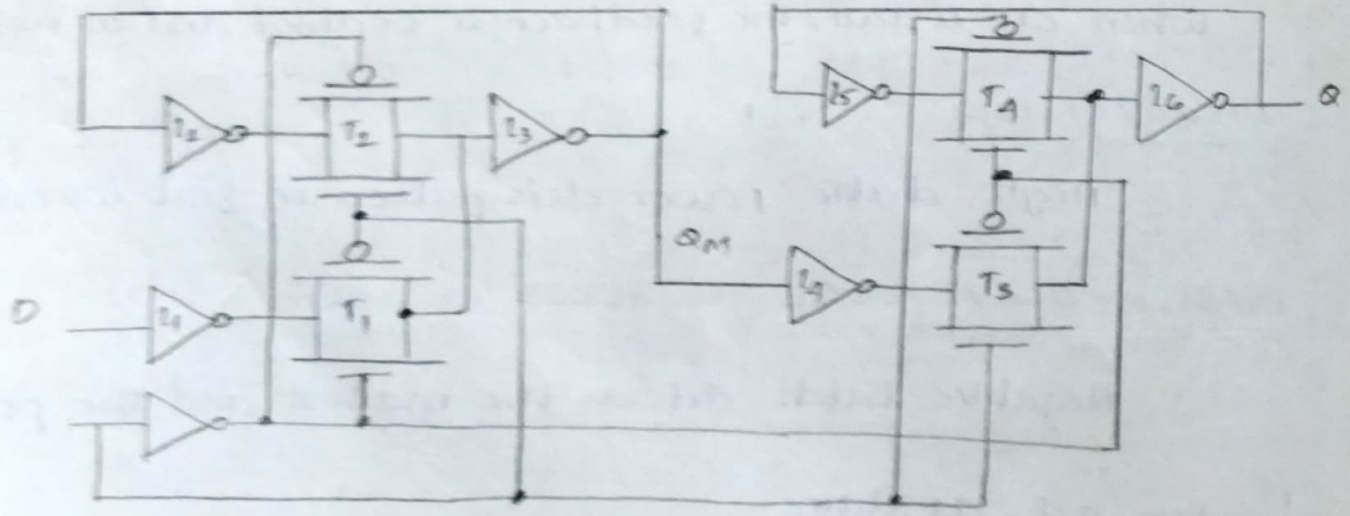
During high phase of clock, the slave samples the output from the master stage and master remains in hold mode. The output Q is i/p value D before rising edge of clock to get positive edge triggered effect.



Master slave Edge Triggered Register and its waveform

In below diagram, Multiplexers are implemented using transmission gates.

When clk is high, the master stop sampling and goes to hold mode. Now T_1 is OFF and T_2 is ON and cross coupled inverter I_2 and I_3 hold the state Q_M. At this time T_3 is ON and T_4 is OFF and Q_M is copied to the output Q.



Master slave positive Edge triggered Register

TIMING PROPERTIES OF MASTER SLAVE EDGE TRIGGERED REGISTERS:

SETUP TIME: The time before the rising edge of the clock that the input data D must be valid.

PROPAGATION DELAY: The time taken for the value of Qm to propagate to the output Q.

HOLD TIME: The time that the input must be held stable after the rising edge of the clock.

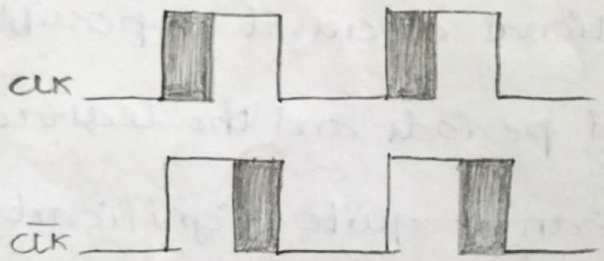
DRAWBACKS OF TRANSMISSION GATE REGISTER:

i. High capacitive load in clock signal

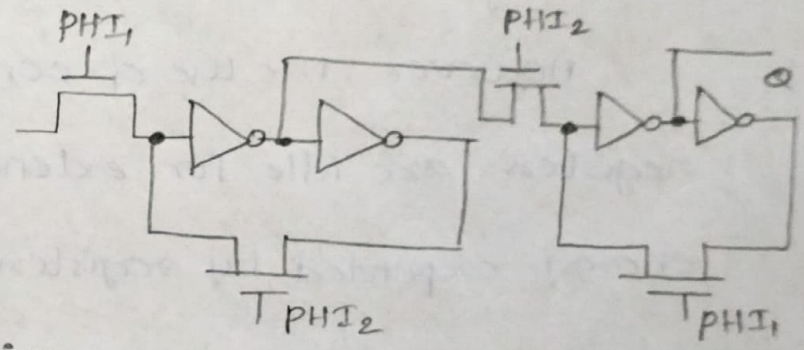
ii. More power dissipation.

NON IDEAL CLOCK SIGNALS:

Normally delay in inverter is zero. variations exist in wires that are used to route the two clock signals. This effect is called clock skew. This causes overlap.



clock signals overlapping

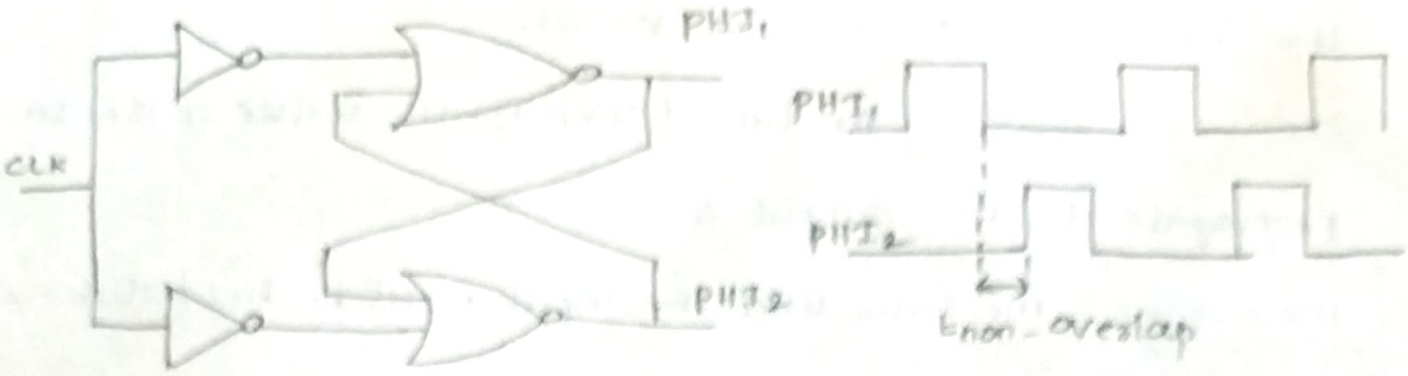


Pseudostatic two phase D register

The clock overlap produces two failures such as the output is undefined state. It can be avoided by using two non-overlapping clocks PHI1 and PHI2.

During non overlap time, high impedance state and loop

gain is zero. If the condition holds for long time it is called pseudostatic.



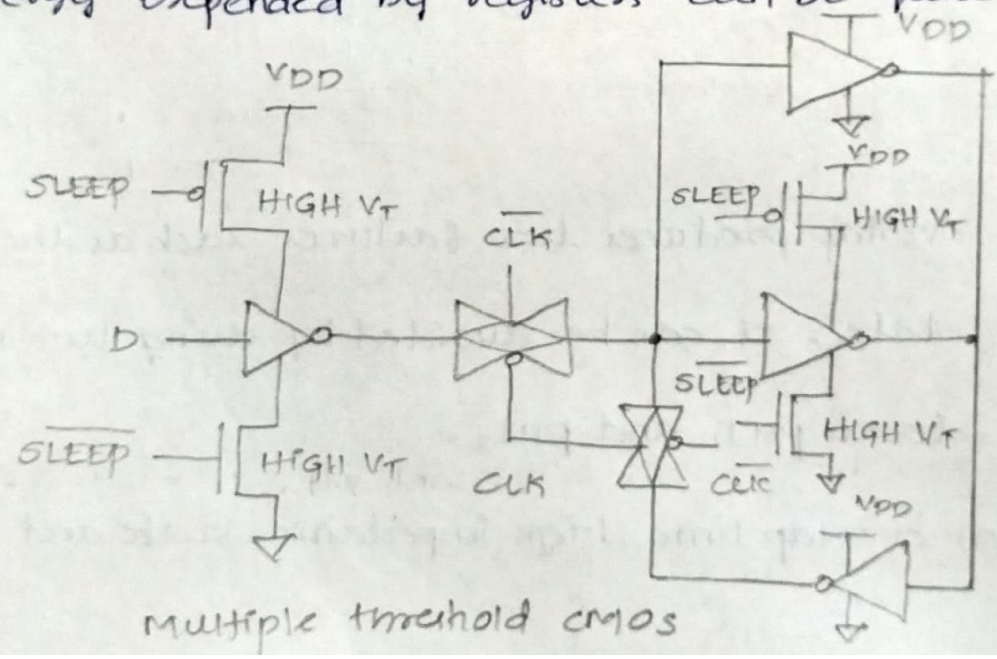
Two phases Nonoverlapping clocks

LOW VOLTAGE STATIC LATCHES:

The scaling of supply voltages is critical for low power operation. unfortunately certain latches don't function at reduce the supply voltage.

At very low power supply voltages, the input to the inverter cannot be raised above the switching threshold resulting in incorrect evaluation.

However, the use of conditional clocks, it is possible that registers are idle for extended periods and the leakage energy expended by registers can be quite significant.

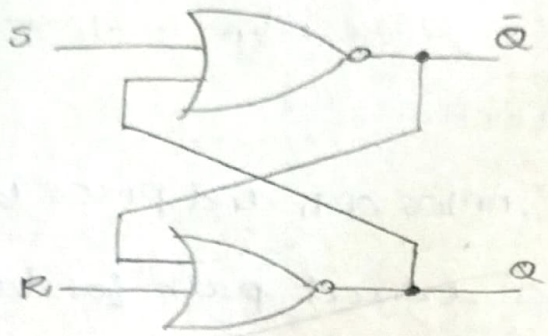


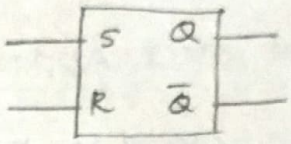
Multiple threshold CMOS

STATIC SR FLIP-FLOPS:

A bistable state can change state by overpowering the feedback loop using SR flipflop.

The second input of NOR gate is connected to trigger inputs S and R, that makes it possible to force the output Q and $\bar{Q}$. These outputs are complementary.





S	R	Q	$\bar{Q}$
0	0	Q	$\bar{Q}$
1	0	1	0
0	1	0	1
1	1	0	0

Forbidden state

SR FLIP FLOP USING NOR GATES

DISADVANTAGES OF STATIC LATCHES AND REGISTERS:

i. stored value remains valid as long as supply voltage is available.

ii. complexity

DYNAMIC LATCHES AND REGISTERS:

Dynamic circuits are based on temporary storage of charge on parasitic capacitors.

The presence of charge denotes a 1 and absence denotes 0. capacitor has some leakage. The stored value can be kept in for some milliseconds. To preserve signal integrity, periodic refresh is necessary.

DYNAMIC TRANSMISSION GATE EDGE TRIGGERED REGISTERS:

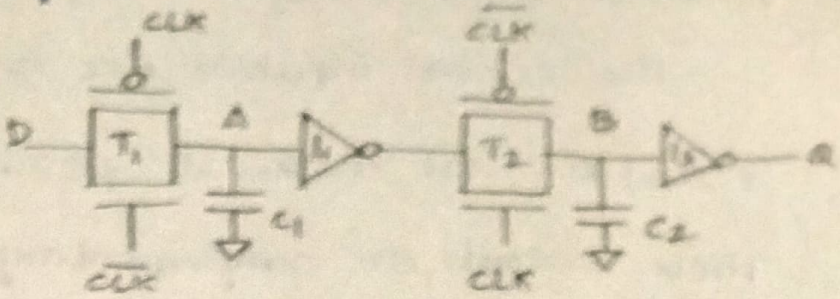
when clock is 0 input data is sampled on storage

node 1, the slave is in hold mode and node 2 is in high impedance state. on rising edge of clock, the transmission gate T_2 turns on and sampled value at node 1 propagate to o/p Q.

Advantages:

very efficient

zero hold time



Dynamic edge triggered register

IMPACT OF NON OVERLAPPING CLOCK:

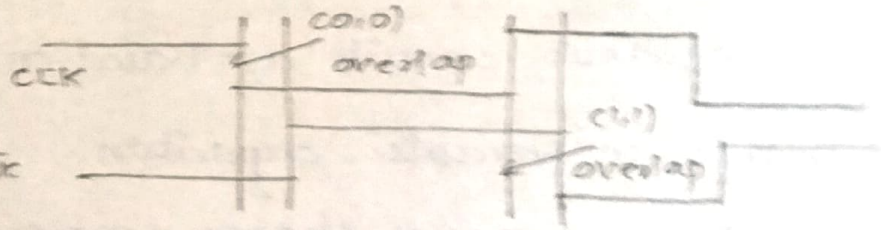
During 0-0 overlap period, nMOS of T_1 and pMOS T_2 are simultaneously ON. This creates a direct path for data to flow from D input to output Q. This is called race condition

During 1-1 overlap period, an input output path exist pMOS of T_1 and nMOS of T_2 take care of hold time constraint. The overlap period is given by

$$t_{overlap\ 0-0} < t_{T1} + t_{I1} + t_{T2}$$

The constraint for 1-1 overlap is given by

$$t_{hold} > t_{overlap\ 1-1}$$



Impact of Nonoverlapping clock

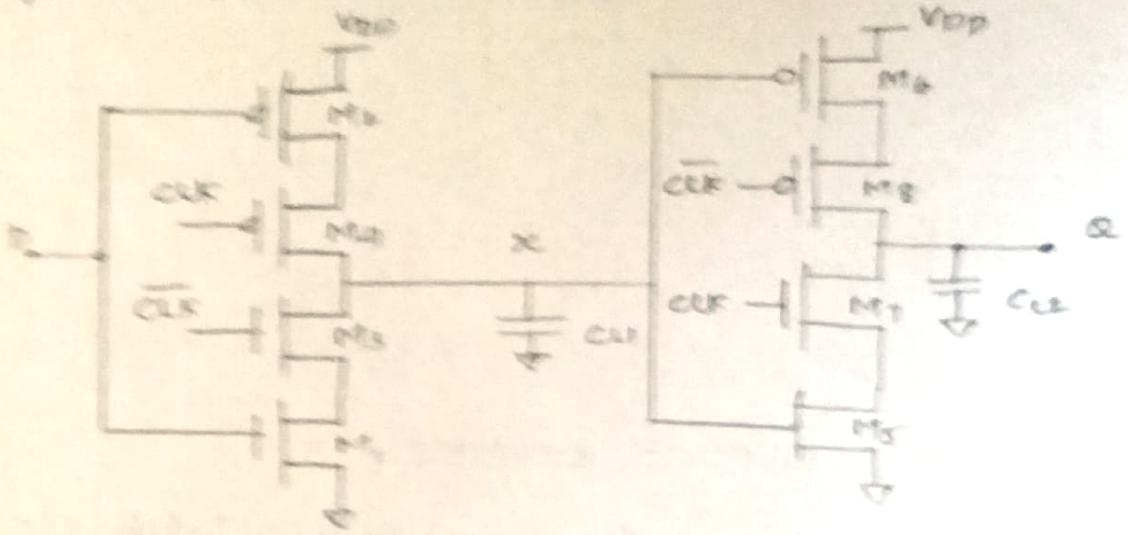
C^2MOS WITH CLOCK SKEW INSENSITIVE APPROACH:

C^2MOS REGISTER:

A positive edge triggered register based on master slave concept insensitive to clock overlap. This is called C^2MOS. It has two phases.

If clock is 0, the master act as inverter. This inverts the D input on internal node x. The master stage is evaluation mod and slave stage is in high impedance mode. M_7 and M_8 are off condition and output Q retains its previous value stored C_{L2}.

If clock is 1, the master stage is in hold mode, M_3 and M_4 are off condition. The slave stage is evaluation mode making M_7 and M_8 in ON condition. The value stored C_{L1} propagate to o/p



CMOS master slave positive edge triggered Register

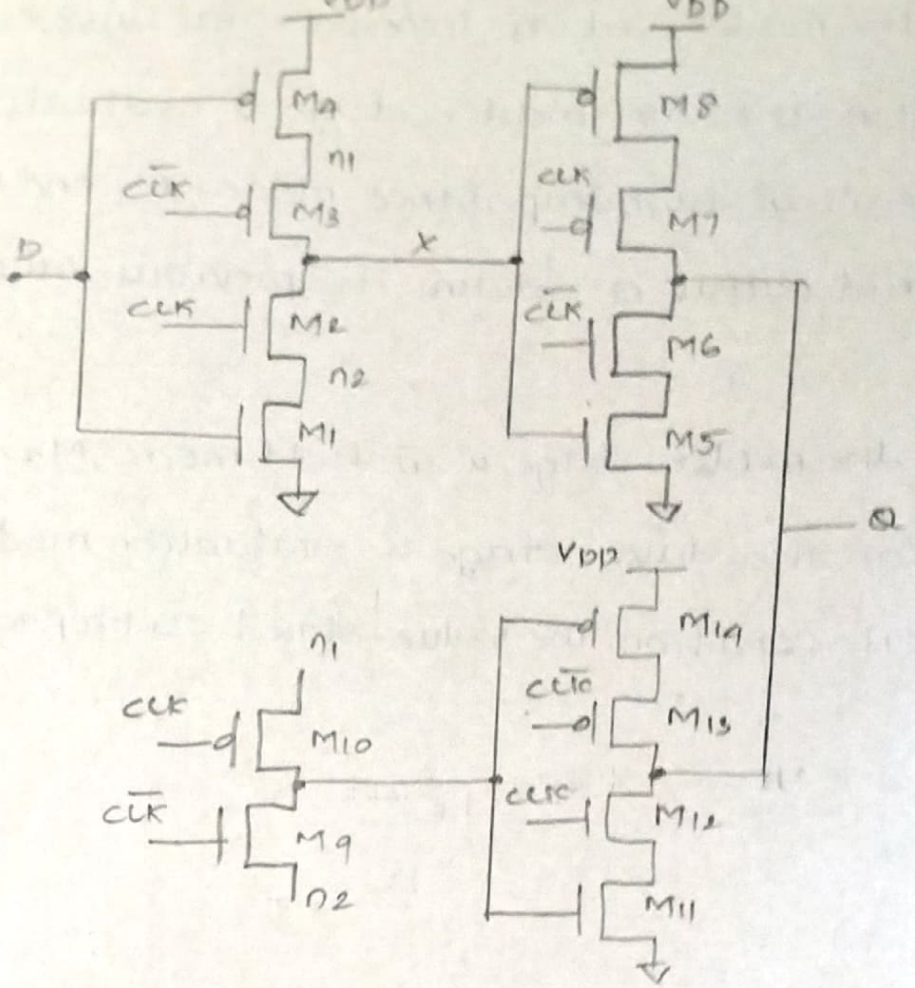
DUAL EDGE REGISTER:

When in sequential circuits samples the input on both clock edges.

When clock is 1, M_1 and to M_4 is sampling inverted D i/p on node x. M_9 and M_{10} are off condition and node y is held stab

On falling edge of clock, M5 to M8 turns on and drives the inverted value of x to the output Q.

When clock is 0, M_1, M_4, M_9, M_{10} turned on. On rising edge of clock, the bottom slave conducts and drives the inverted value to on y to node Q. The data changes on both edg

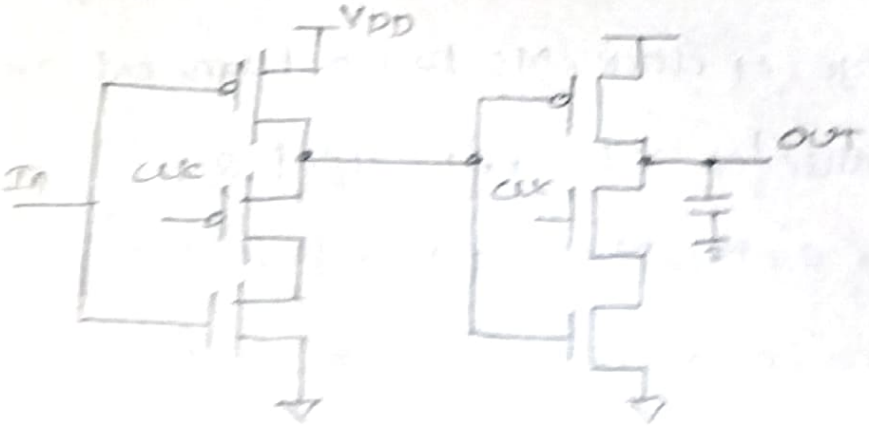


C²MOS based dual edge triggered register

TRUE SINGLE PHASE CLOCKED REGISTER:

It uses single phase clock. In positive latch, clock is ↑ (high), the latch is transparent mode and corresponds to two cascaded inverters. The input propagates to output.

When clock is low, the latch is hold mode. Now pull up networks are active and pull down circuits are disabled. so no signal never propagate.



True single phase positive latch

2. PIPELINES:

pipelining is a designing technique used to increase the operation of datapaths in digital processor.

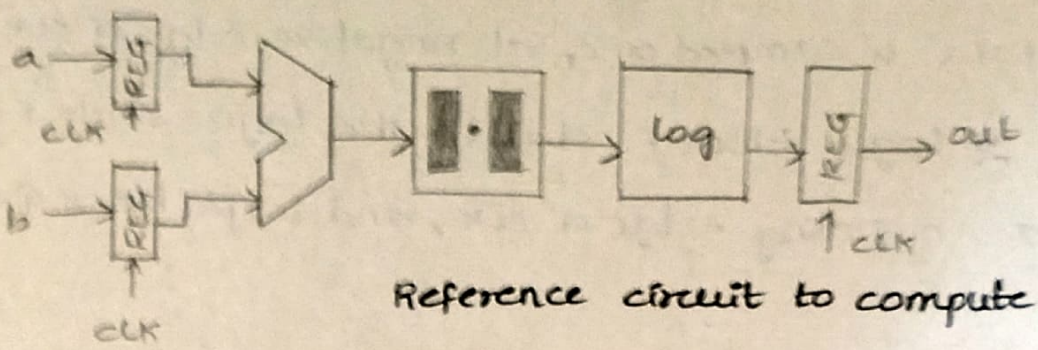
example:

$$T_{min} = t_{c-q} + t_{pd\text{-}logic} + t_{su}$$

where, t_{c-q} - propagation delay

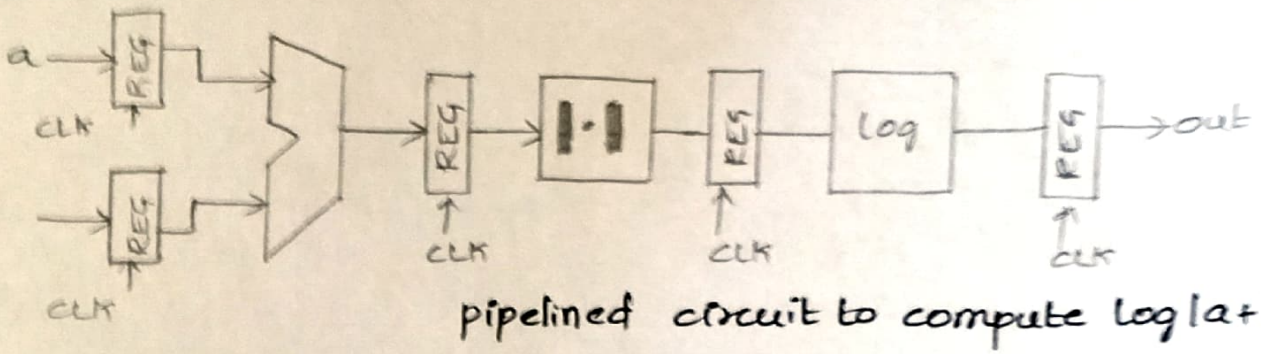
t_{su} - setup time of register

$t_{pd,logic}$ - worst case delay path through combinational nets



Reference circuit to compute $\log|a+b|$

The delay in logarithm function is much larger than delay in registers.



pipelined circuit to compute $\log|a+b|$

clock period	Adder	Absolute value	Logarithm
1	a_1+b_1		
2	a_2+b_2	$\lvert a_1+b_1\rvert$	
3	a_3+b_3	$\lvert a_2+b_2\rvert$	$\log(\lvert a_1+b_1\rvert)$
4	a_4+b_4	$\lvert a_3+b_3\rvert$	$\log(\lvert a_2+b_2\rvert)$
5	a_5+b_5	$\lvert a_4+b_4\rvert$	$\log(\lvert a_3+b_3\rvert)$

$$T_{min,pipe} = t_{c-q} + \max(t_{pd}\text{'add'}\ t_{pd}\text{'abs'}\ t_{pd}\text{'log}) + t_{su}$$

$$T_{min\text{-}pipe} = \frac{T_{min}}{3}$$

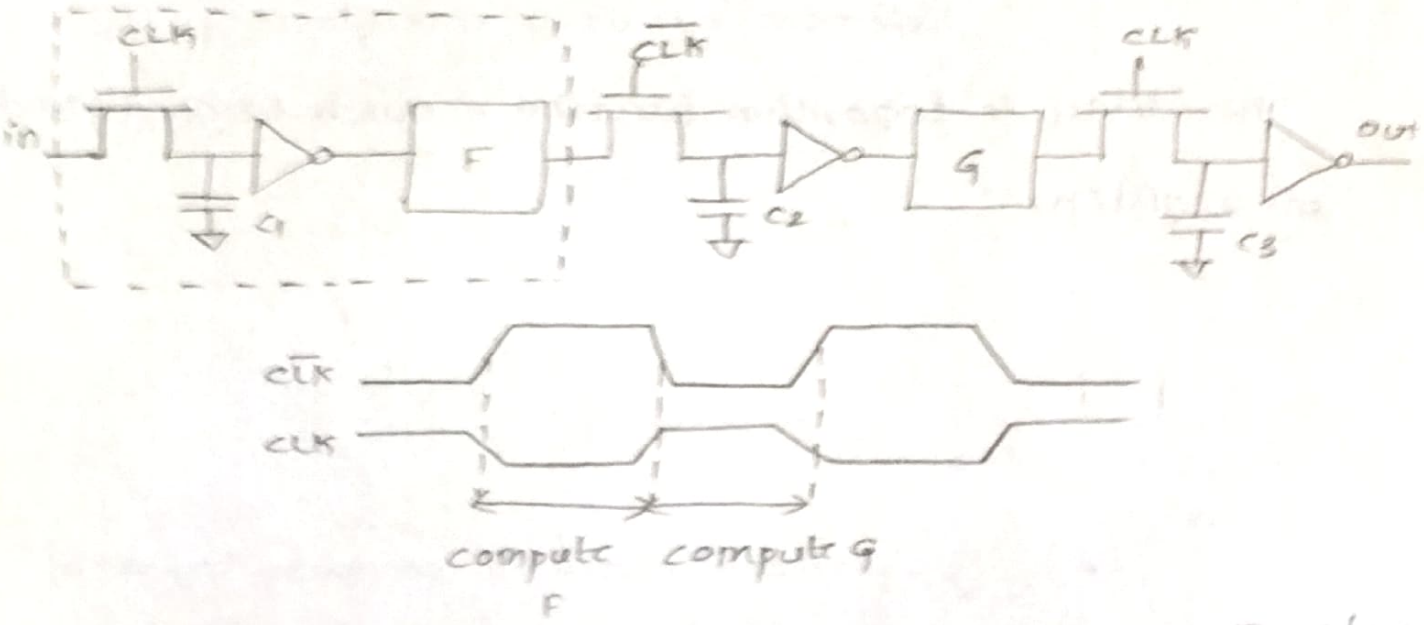
LATCH - VERSUS REGISTER BASED PIPELINES:

Pipelined circuits are constructed by using level-sensitive latches instead of edge triggered registers.

It is implemented by using pass transistors based positive and negative latches. The logic is introduced between master and slave.

When CLK and $\overline{CLK}$ are non-overlapping, the correct pipeline operation is obtained.

Input data is sampled on C_1 at negative edge of $\overline{CLK}$ and computation of logic block F starts. The logic state F is stored on C_2 on falling edge of $\overline{CLK}$, and logic block G starts.

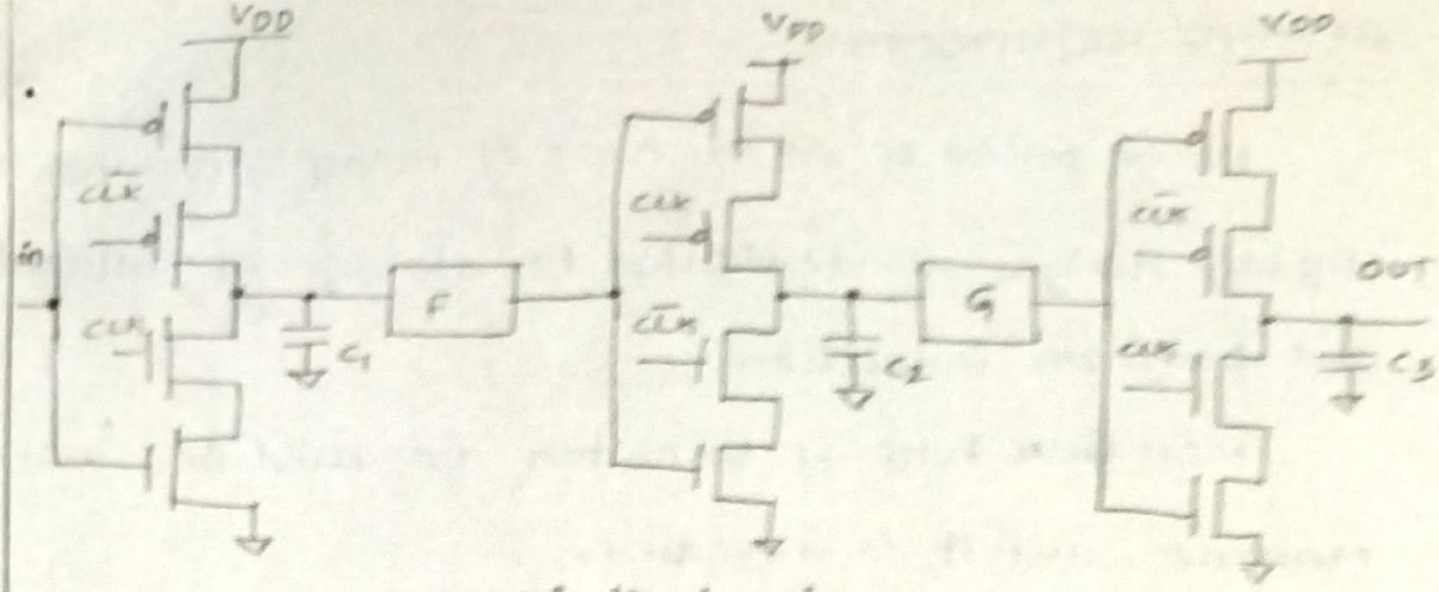


Two phase pipelined circuit using Dynamic Registers

NORA - CMOS BASED PIPELINES:

The C²MOS based pipelined circuit is race free as long as all the logic functions F between the latches are noninverting.

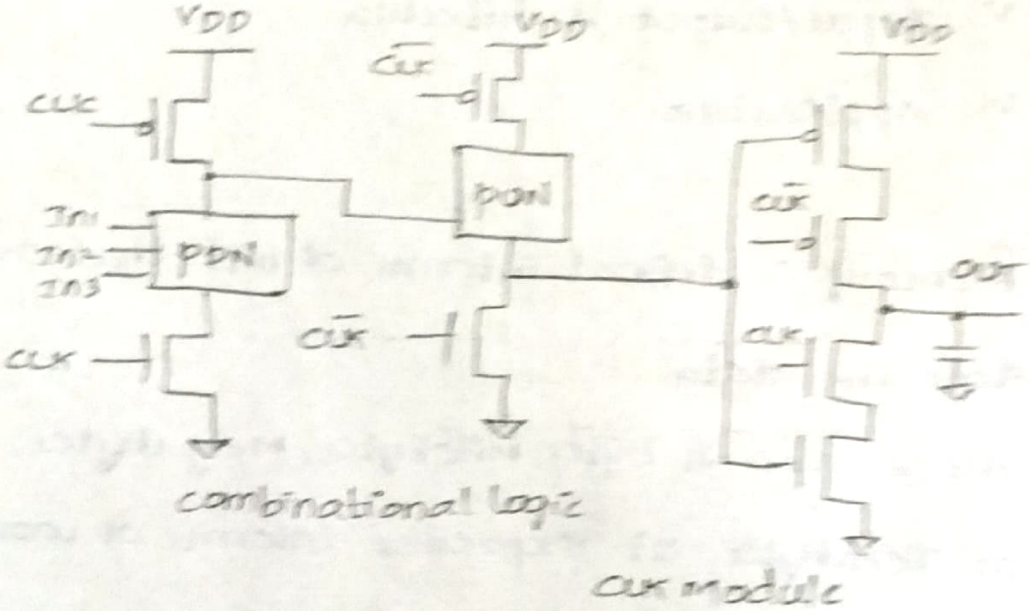
During (0-0) overlap between CLK and $\overline{CLK}$ all C²MOS latches simplify to pull up networks.



pipelined datapath using C²MOS latches

The only way a signal can race from stage to stage when logic function F is inverting. It combines C²MOS pipeline register and NORA dynamic logic function blocks.

A block that is evaluation during clk=1 called clk module while the inverse is called $\overline{clk}$ module.



	clk block		$\overline{clk}$ block	
	Logic	latch	Logic	Latch
clk = 0	precharge	Hold	evaluate	Evaluate
clk = 1	Evaluate	Evaluate	precharge	Hold

Operation mode

3. MEMORY ARCHITECTURE:

Large portion of silicon area of many contemporary digital designs is dedicated to storage of data values and program instructions.

More than half of transistors are used for cache memories and it is increased.

Memory is classified based on

i) size

ii) Timing parameters

iii) Function

iv) Access pattern

v) Input/output Architecture

vi) Application

SIZE:

size of memory is defined interms of bits that are needed to store the data.

The memory sizes in bytes, kilobytes, megabytes, Gigabytes or Terabytes. It expresses interms of word.

TIMING PARAMETERS:

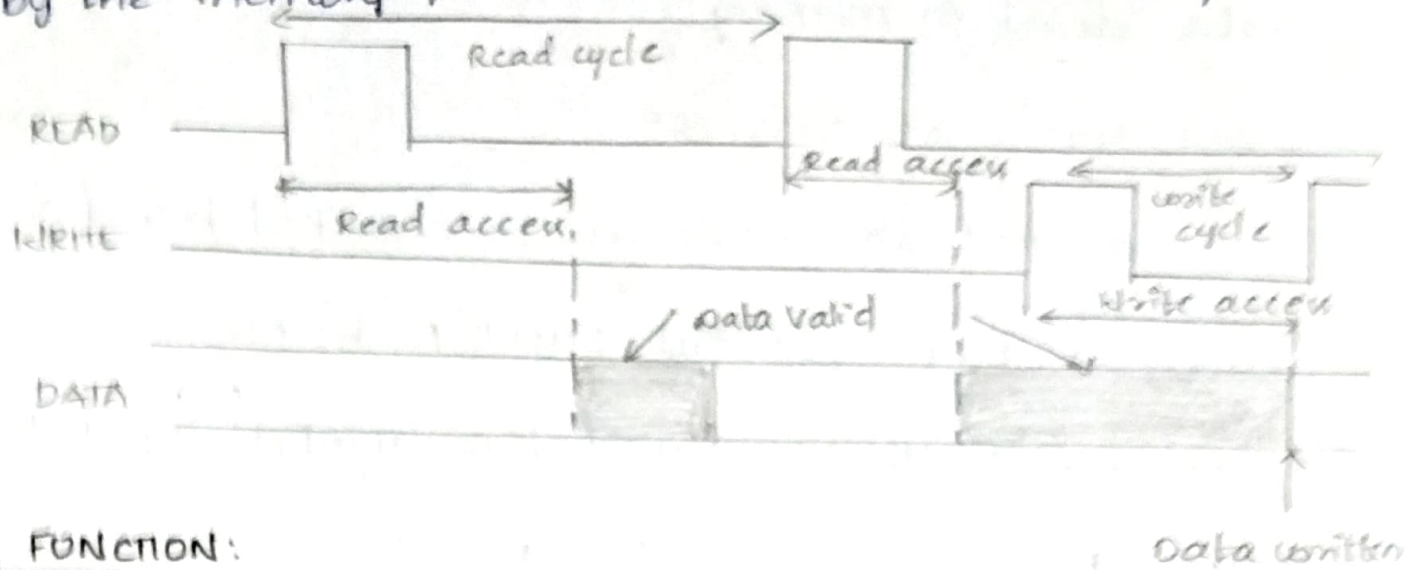
Time takes to retrieve or read data from memory is called read-access time.

Read access time is defined as delay between the read request and the moment the data is available

at output.

Write access time is the time elapsed between a write request and final writing of the input data.

Read or write cycle time is minimum time required by the memory for successive read or write operation.



FUNCTION:

ROM and (Read and write Memory) RWM are important types of semiconductor memory based on functions. RWM have advantage of both read and write with access time and are more flexible memories.

Data are stored either in flipflops are called static and dynamic memories respectively. RWM are volatile memories in which data lost when supply is turned off. ROM are non-volatile memories in which data cannot lost.

ACCESS PATTERN:

Memory locations can be read or written in a random order called Random Access memory. Memory types with faster access time, smaller area, or a memory with a

special functionality belongs to First-In First-out (FIFO), Last-In First-out (LIFO) used as a stack and shift register.

content Access memory (CAM) also called an associative memory. The MATCH signal remains low if no data stored in memory.

INPUT/OUTPUT ARCHITECTURE:

semiconductor memories are classified based on the number of data input and output ports. Memories with higher bandwidth requirement have multiple input and output ports are called multiport memories.

APPLICATIONS:

Most large size memories were packaged as standalone IC's. Integration of multiple functions on a single die, an ever larger fraction of memory is now integrated on same die as logic functionality called embedded.

TYPES:

i. N-word Memory Architecture

ii. Array structured memory Architecture

iii. Hierarchical Memory Architecture

iv. content Addressable Memory Architecture

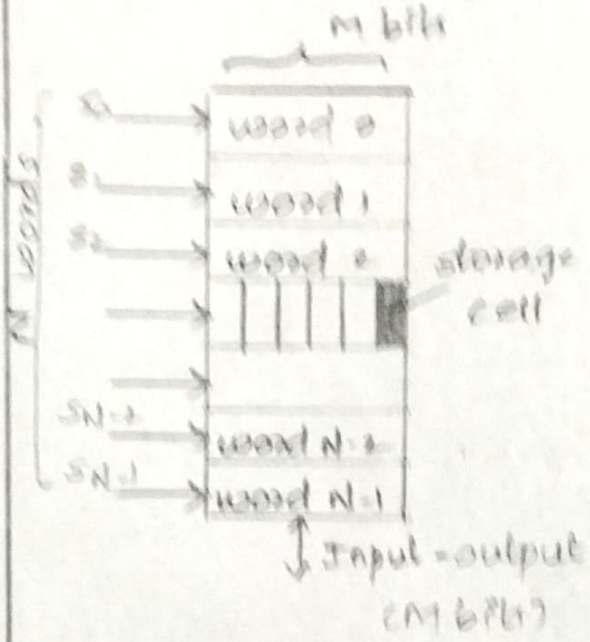
N-WORD MEMORY ARCHITECTURE:

Memory words are stacked in a linear fashion one word at a time is selected for reading or writing with the aid of a select bit (S_0 to S_{n-1}).

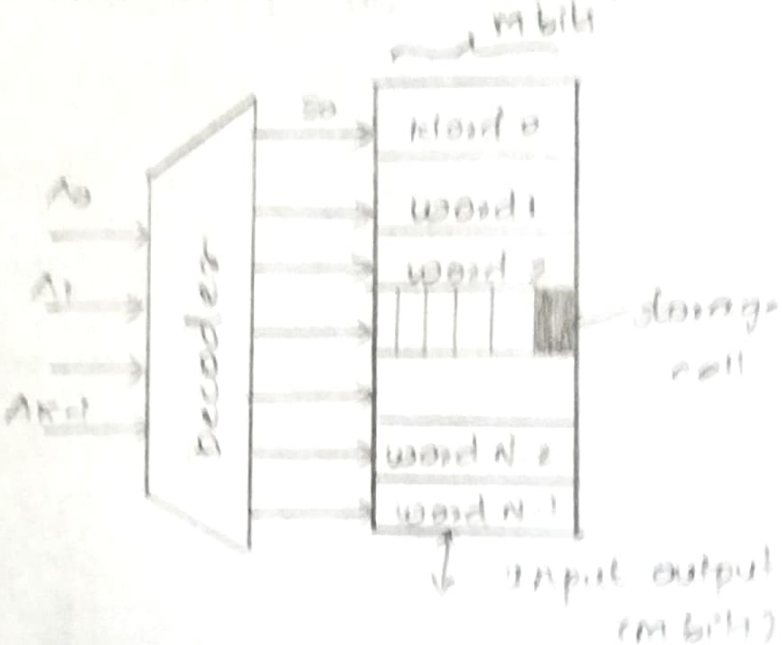
If this module is a single-port memory, only one signal S_i can be high at any time.

If storage cell is a D flip-flop, then select signal is used to activate or clock the cell. It is simple and works well for very small memories.

A decoder is selected to reduce the no of select signals



N-word Memory architecture



N-word Memory architecture using Decoder

Disadvantages:

i. poor memory aspect ratio

ii. Not easy to implement

iii. odd shape factor

iv. slow operation

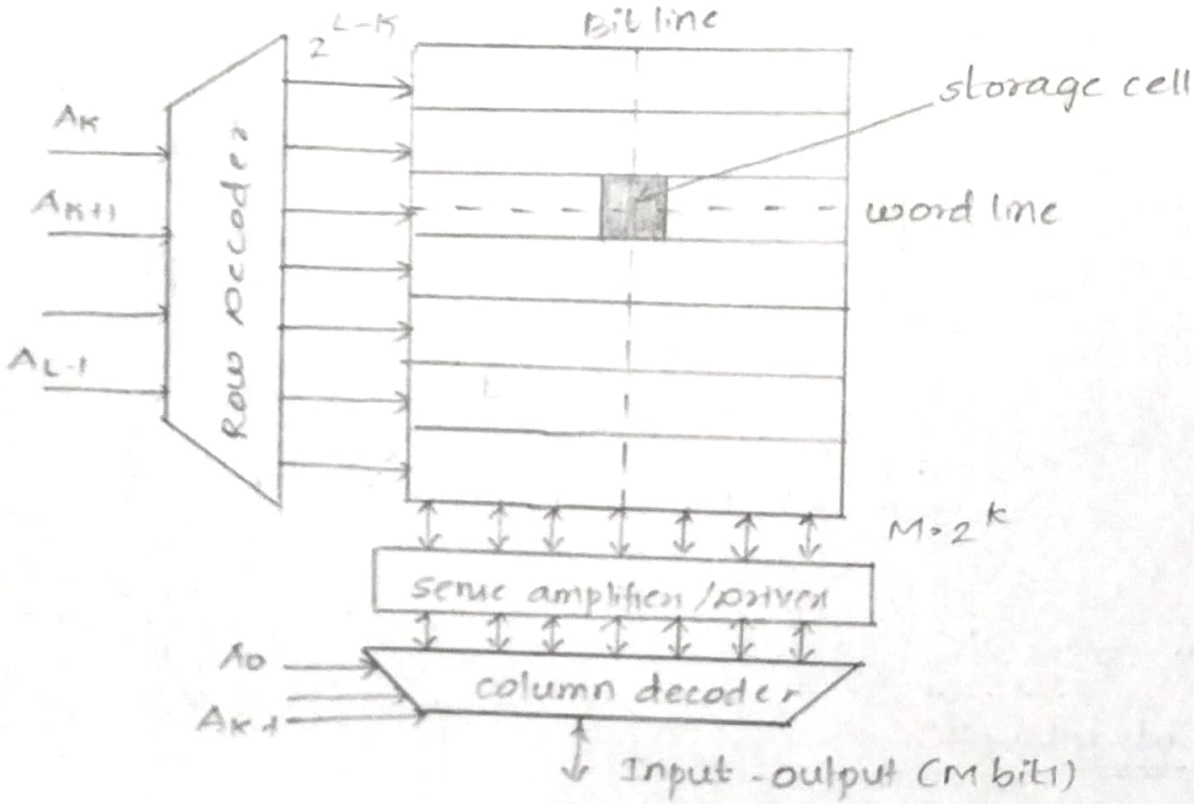
v. High delay

ARRAY STRUCTURED MEMORY ARCHITECTURE:

To overcome the problem in N-word memory, memory arrays are organized in vertical and horizontal dimensions with same order of magnitude. The aspect ratio near to unity.

Multiple words are stored in a single row and are selected simultaneously. The address word is paritioned into a column address (A_0 to A_{K-1}) and row address (A_K to A_{L-1})

The row address enables one row of memory, while column picks one particular word.



Array structured Memory Architecture

HIERARCHICAL MEMORY ARCHITECTURE:

Array structured memory architecture works well for memories upto range of 64 Kbits to 256 Kbits.

Larger memories start to suffer from serious speed degradation as the length, capacitance and resistance of word and bit lines become larger.

For larger memories add one extra dimension to address space. The memory is partitioned into p smaller blocks.

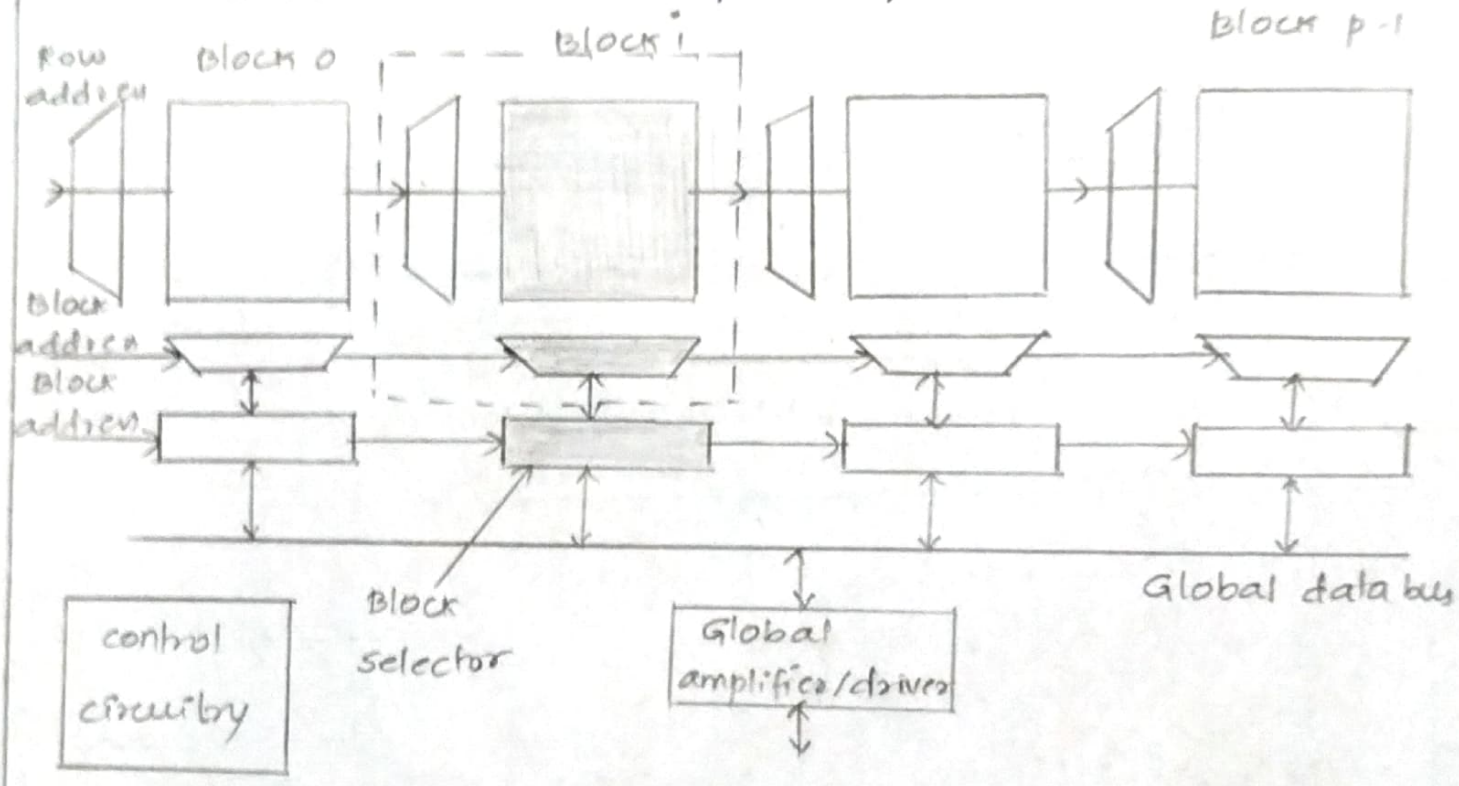
Each individual block is identical. An extra address word called block address, selects one of p blocks to be read or written.

Advantages:

Faster access time

Large power saving

Non-active blocks are put in power saving mode.



CONTENT ADDRESSABLE MEMORY ARCHITECTURE:

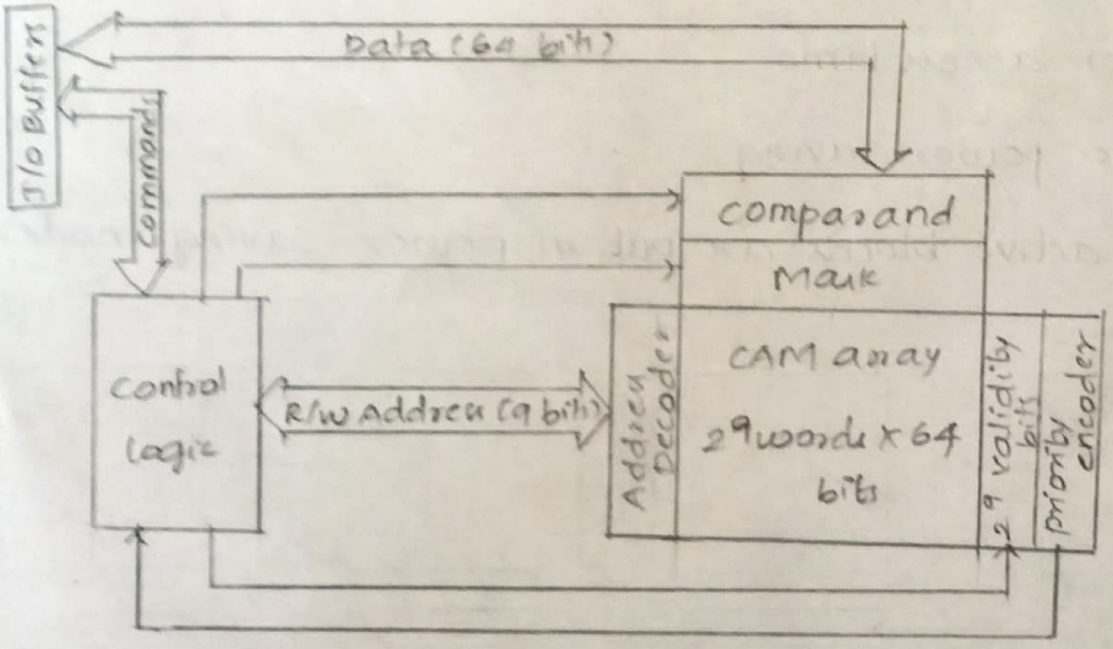
It supports three modes of operation namely read, write and match.

The read and write mode access and manipulate data in CAM array as in ordinary Memory.

The match mode is unique to associate memory.

The comparand block is filled with data pattern to match and the mask word indicates which bits are significant.

Every row that matches the pattern is passed to validity block. In CAM array, 9 bits are required to indicate the highest row that matched.



Content Addressable Memory Architecture

VLSI DESIGN

UNIT-IV Designing Arithmetic Building Blocks

1. Ripple Carry Adders:

* The most commonly used arithmetic operation is addition. Adder is the speed-limiting element so careful optimization is needed for adder design

* The optimization is made either at logic or circuit level

* Logic level optimization uses rearrange the Boolean equations. circuit optimization manipulate transistor sizes and circuit topology to optimize the speed

* A full adder has three inputs and two outputs A, B are adder input, C_i is carry input, S is Sum output, C_o is carry output. The output S and C_o are defined by some intermediate Signals: G (generate), D (delete), P (propagate)

G=1 ensures that carry bit will be generated (1)

D=1 ensures that carry bit will be deleted (0)

P=1 guarantees that an incoming carry C_i will propagate C_o

Truth Table Full Adder

A	B	C_i	S	C_o	carry Status
0	0	0	0	0	delete
0	0	1	1	0	delete
0	1	0	1	0	propagate
0	1	1	0	1	propagate
1	0	0	1	0	propagate
1	0	1	0	1	propagate
1	1	0	0	1	generate/propagate
1	1	1	1	1	generate/propagate

$S = A \oplus B \oplus C_i$

$= ABC_i + \bar{A}\bar{B}C_i + A\bar{B}\bar{C}_i + \bar{A}B\bar{C}_i$

$C_o = AB + BC_i + AC_i$

$G = AB$

$D = \bar{A}\bar{B}$

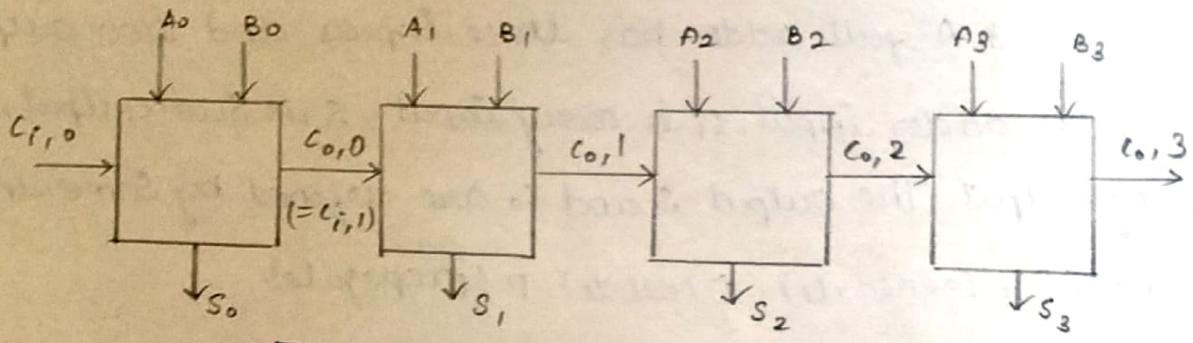
$P = A \oplus B$

Now S and C_o is written as a function of P and G as,

$C_o(G, P) = G + PC_i$

$S(G, P) = P \oplus C_i$

* G and P are functions of A and B only are not dependent upon C_i. Similarly expressions for $S(D, P)$ and $C_o(D, P)$ can be derived.



Four bit Ripple carry Adder

* An N-bit adder can be constructed by cascading N Full adder (FA) circuits in series, connecting $C_{o,k-1}$ to $C_{i,k}$ for $k = 1$ to $N-1$, and first carry-in $C_{i,0}$ to 0.

* In ripple carry adder the carry bit ripples from one stage to the other.

* For some I/P signal, no ripple carry occurs at all. But for some carry has to ripple all the way from LSB to MSB.

* The propagation delay in this path is called the critical path is defined as worst case delay over all possible input patterns.

* Delay proportional to the no. of bits in input words N is given by,

$$t_{adder} \propto (N-1)\, t_{carry} + t_{sum}$$

where

t_{carry} - propagation delay from C_i to C_o

t_{sum} - propagation delay from C_i to S

* Inverting all inputs to a full adder results in inverted values for all outputs. This is called the inverter property. It is expressed as,

$$\bar{S}(A, B, C_i) = S(\bar{A}, \bar{B}, \bar{C_i})$$

$$\bar{C_o}(A, B, C_i) = C_o(\bar{A}, \bar{B}, \bar{C_i})$$

2. CARRY LOOKAHEAD ADDERS:

The following are the types of carry lookahead adder

* Monolithic lookahead adder
* Logarithmic lookahead adder

Monolithic Lookahead adder:

* For Fast adders it is essential to get around the rippling effect of the carry. This effect is present in both carry bypass and carry select adder. The following relation holds for each bit position in an N-bit adder.

$$C_{o,k} = f(A_k, B_k, C_{o,k-1}) = G_k + P_k C_{o,k-1}$$

* The dependency between $C_{o,k}$ to $C_{o,k-1}$ is eliminated by expanding $C_{o,k-1}$

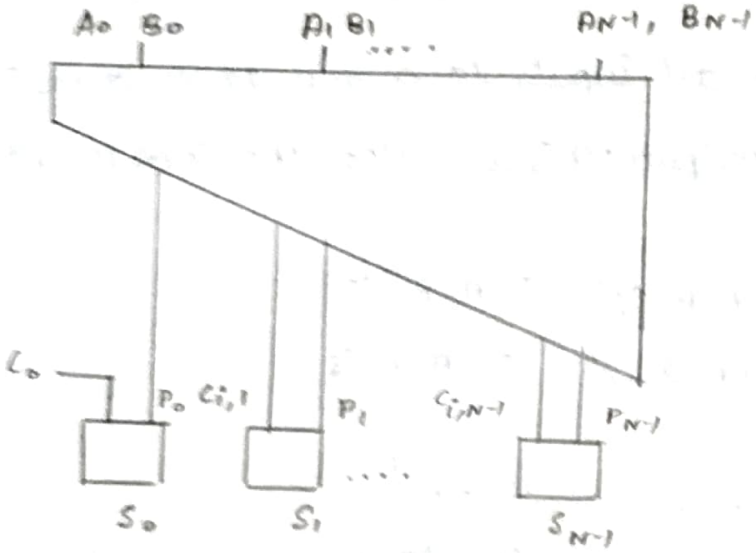
$$C_{ok} = G_k + P_k (G_{k-1} + P_{k-1} C_{o,k-2})$$

The above eqn is suggested for small values for $N (\leq 4)$

For greater values of N, the eqn is expressed as

$$C_{0,K} = G_K + P_K(G_{K-1} + P_{K-1}(\dots + P_1(G_0 + P_0 C_{i,0})))$$

* The value of $C_{i,0}$ equal to 0. The figure shows the block diagram of carry lookahead adder.



Carry look ahead adder

* This high level model has some hidden dependencies. Fig. shows diagram of mirror implementation of four bit lookahead adder.

* The circuit exploits self duality and recursivity of the carry lookahead adder to build a mirror structure.

* The large fan-in of the circuit makes it prohibitively slow for larger values of N.

* The fan-out of some of the signals tends to grow excessively, slowing down the adder even more.

Disadvantages:

* For a group of N bits, the transistor implementation has N+1 parallel branches and N+1 transistors in stack

* wide gates and large stacks display poor performance.

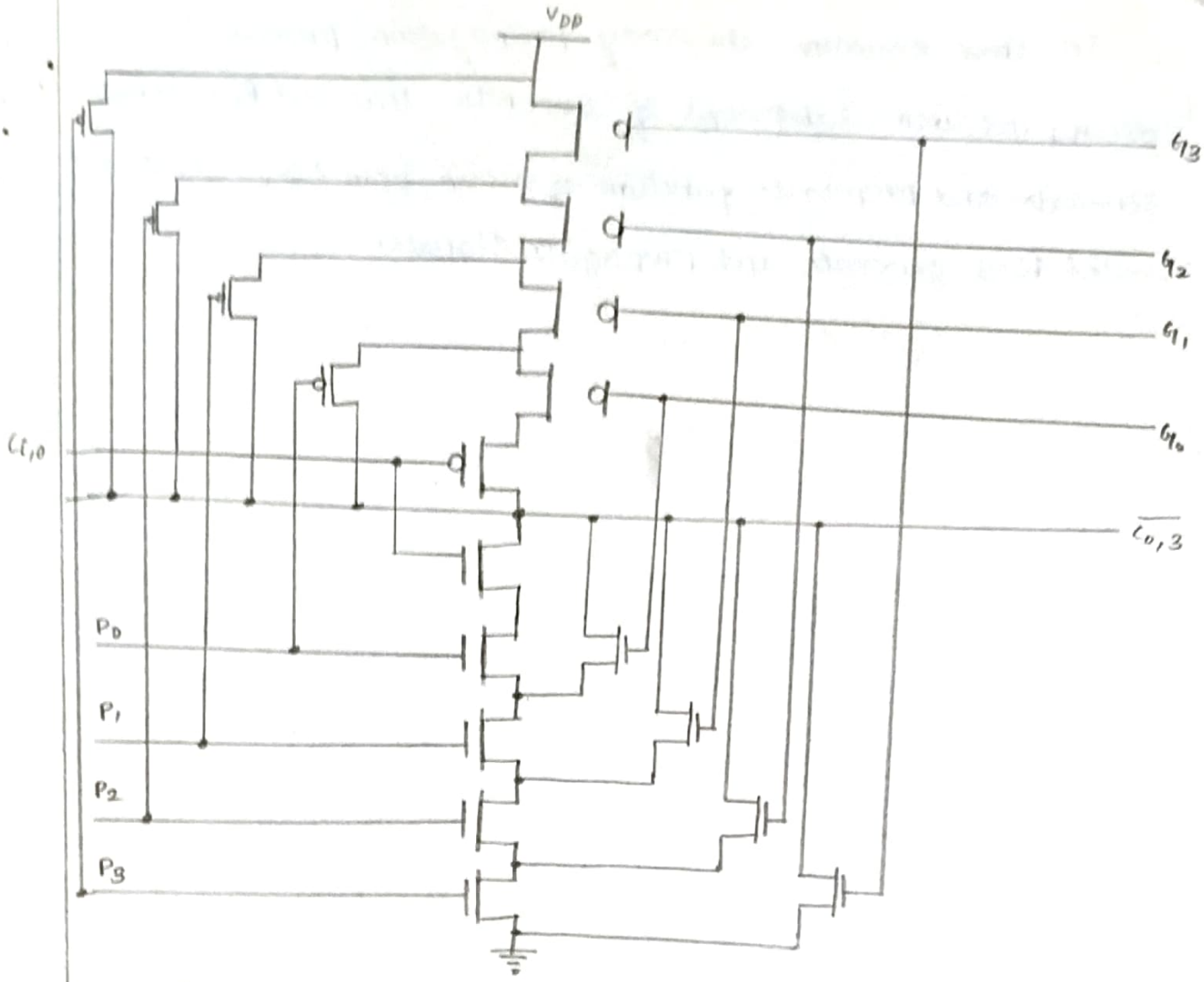


Fig: Mirror implementation of four bit lookahead adder

Logarithmic Lookahead Adder:

* For high speed adders, the carry propagation and generation are organized into recursive trees.

* A more effective implementation is obtained by hierarchically decoming the carry propagation into subgroups of N bits

$$C_{0,0} = G_0 + P_0 C_{i,0}$$

$$C_{0,1} = G_1 + P_1 G_0 + P_1 P_0 C_{i,0} = (G_1 + P_1 G_0) + (P_1 P_0) C_{i,0}$$

$$= G_{1:0} + P_{1:0} C_{i,0}$$

$$C_{0,2} = G_2 + P_2 C_{0,1}$$

$$C_{0,3} = G_{3:2} + P_{3:2} C_{0,1}$$

In above equation the carry-propagation process is decomposed into subgroups of two bits. $G_{i:j}$ and $P_{i:j}$ denote the generate and propagate function of group from i to j. These are called block generate and propagate signals.

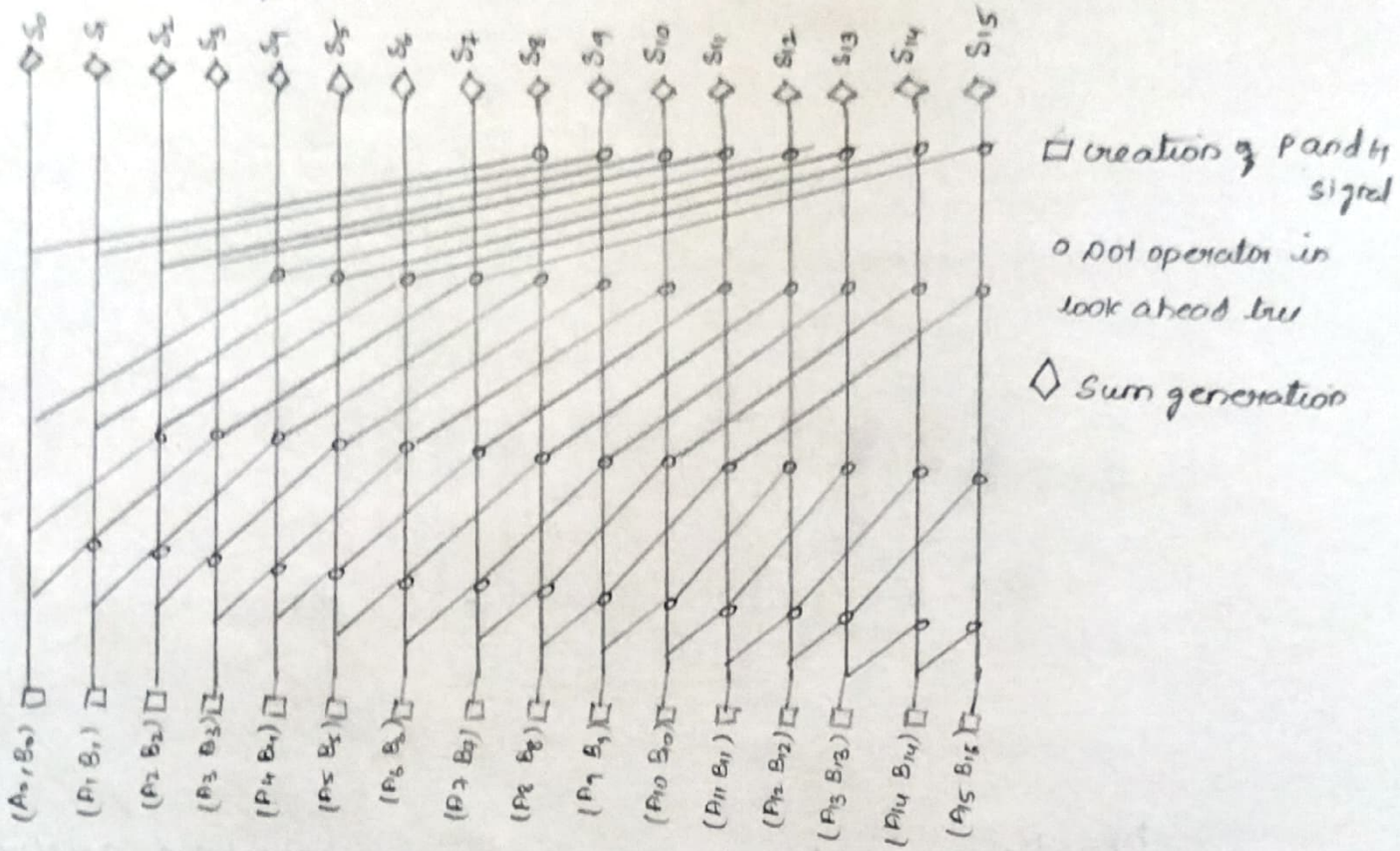


Fig: Kogge-Stone 16-bit lookahead logarithmic adder

Advantages of Kogge-Stone Structure:

* Regular interconnect structure
* Easy implementation
* constant fan-out throughout the tree

Disadvantage of Kogge-Stone Structure:

* Replication of carry trees to generate the intermediate carries comes at a large cost in terms of both area and power.

3. MULTIPLIERS:

* Multiplications are expensive and slow operation. performance of many computational circuits is decided by the speed at which a multiplication operation can be executed.

* Multipliers have complex adder arrays.

* Consider two unsigned binary numbers X and Y that are M and N bits wide respectively. Express X and Y in binary representation as,

$$X = \sum_{i=0}^{M-1} x_i 2^i \qquad Y = \sum_{j=0}^{N-1} y_j 2^j$$

where

$x_i, y_j \in (0,1)$

* The multiplication operation is given by, $Z = X \times Y = \sum_{k=0}^{M+N-1} z_k 2^k$

$$= \left(\sum_{i=0}^{M-1} x_i 2^i\right)\left(\sum_{j=0}^{N-1} y_j 2^j\right) = \left(\sum_{j=0}^{N-1} x_i y_j 2^{i+j}\right) \sum_{i=0}^{M-1}$$

* Multiplication is performed using a single two-input adder with M and N bits wide. The Multiplication takes M cycles using an N-bit adder.

* Fast multiplication is done similar to manually computing a multiplication operation. All the partial products are generated at the same time organized in an array. This structure is called array multiplier. It has three functions.

* Partial-product generation
* Partial-product accumulation
* Final Addition.

```
        1 0 1 0 1 0      Multiplicand
X             1 0 1 1    Multiplier
-----------------------
        1 0 1 0 1 0
      1 0 1 0 1 0        Partial products
    0 0 0 0 0 0
+ 1 0 1 0 1 0
-----------------------
  1 1 1 0 0 1 1 1 0      Result
```

Fig: Binary Multiplication

(i) Partial - Product generation:

* Fig shows partial- product generation logic

Partial product results from the logical AND of multiplicand X with a multiplier bit y_i. Each row in the partial-product array is either a copy of the multiplicand or a row of zeroes.

* If multiplier consists of all ones, all the partial product exists, while in the case of all zeros, there is none.

This reduces the no. of generated partial product by half.

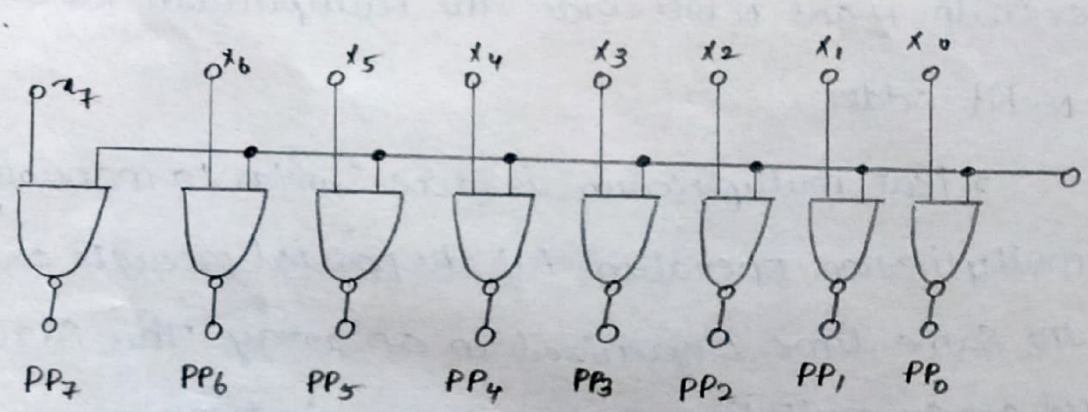


Fig: Partial product Generation logic

*Consider an 8-bit multiplier of the form 01111110, which produces six non zero partial- product rows.

Reduce the number of nonzero rows by recording this number $(2^7+2^6+2^5+2^4+2^3+2^2)$ into different format. This type of transformation is called Booth's reading. The format is

$$y = \sum_{j=0}^{(N-1)/2} y_j 4^j \text{ with } (y_j \in \{-2,-1,0,1,2\})$$

* The 1010...10 represents the worst case multiplier input because it generates the most partial product. Multiplication with {0,1} is equivalent to AND operation multiplying with {-2,-1,0,1,2} this requires combination of inversion and shift logic.

Fig: Modified Booth's Recoding

Multiplier bits	Recoded bits
000	0
001	+ Multiplicand
010	+ Multiplicand
011	+2 x Multiplicand
100	-2 x Multiplicand
101	- Multiplicand
110	- Multiplicand
111	0

(ii) Partial Product Accumulation.

* Once the partial products are generated, it has to be summed. This accumulation is a multioperand addition. This is done by using a no. of adders that will form an array called array multiplier. The following 3 methods are used for this accumulation

* Array multiplier
* Carry save multiplier
* Tree multiplier

* Array multiplier :

* Fig shows 4X4 bit array multiplier. Generation of N partial product requires N X N 2-bit AND gates. Multipliers most area is utilized to add the N partial products.

* This requires N-1 N bit adders. The overall structure is compacted into a rectangle to get a very efficient layout.

FA represent Full Adder, HA stands for Half Adders with two inputs.

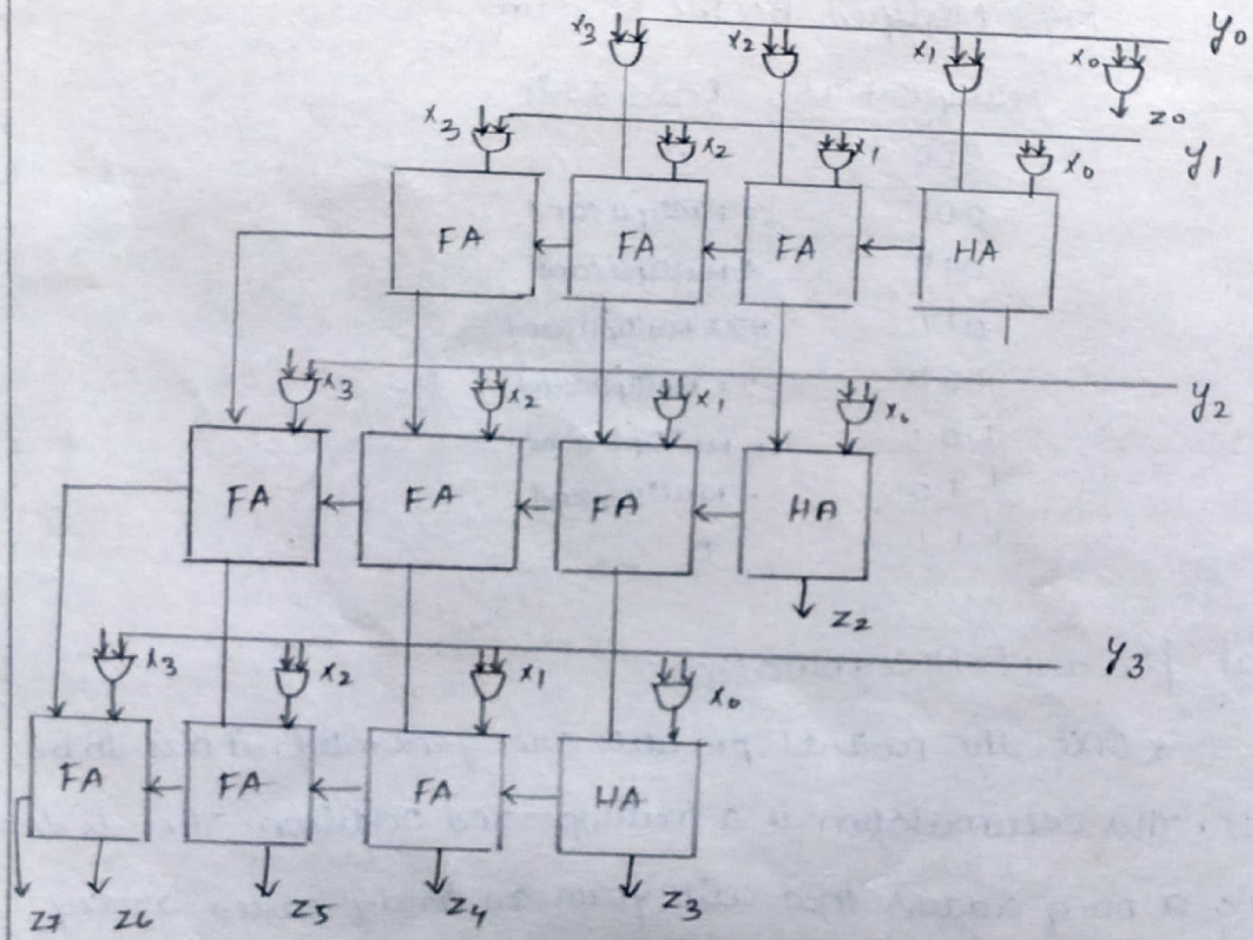


Fig: 4X4 Bit array multiplier

* Propagation delay in an array organized circuit is difficult to find. The critical path yields an approximate expression for propagation delay as,

$$t_{mult} \approx [(N-1)+(N-2)]t_{carry} + (N-1)t_{sum} + t_{and}$$

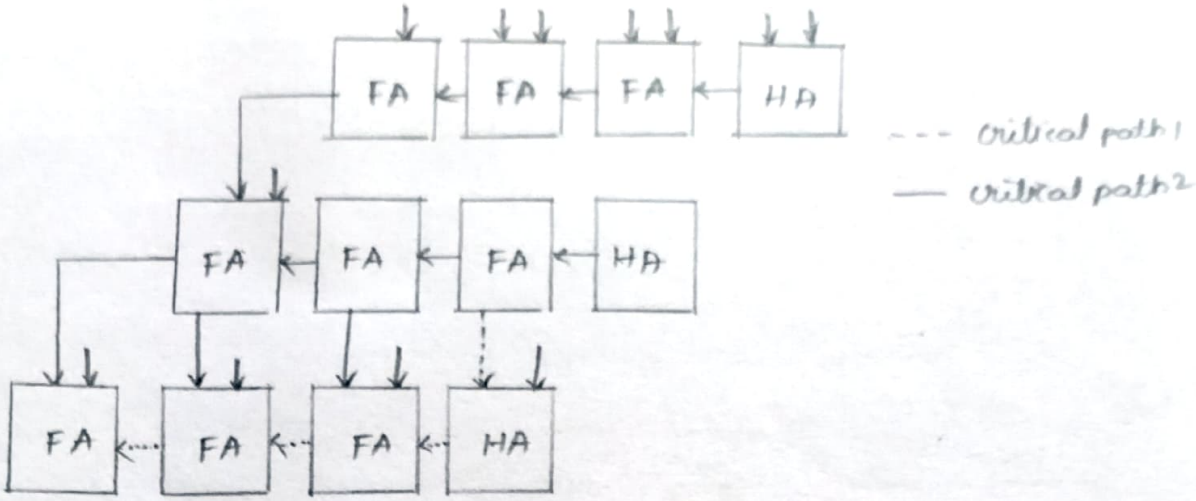


Fig: Ripple carry based 4X4 Multiplier

* All critical paths have to be attacked at same time. Minimization of t_{mult} requires the minimization of both t_{carry} and t_{sum}.

* Carry save multiplier:

* A more efficient realization is obtained by noticing that the multiplication does not change when the output carry bits are passed diagonally downwards instead of only to right

* An extra adder called vector-merging adder is added to generate final result. The resulting multiplier is called a carry save multiplier.

The worst case critical path is given as below by assuming

$t_{add} = t_{carry}$

$t_{mult} = t_{and} + [N-1]t_{carry} + t_{merge}$

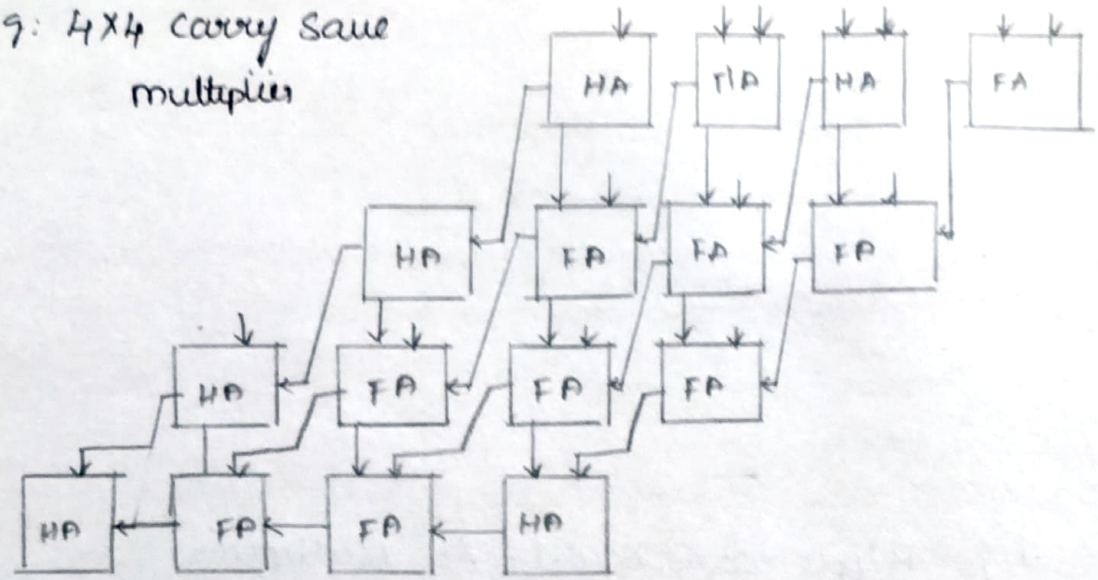
Advantages:

Shorter worst case critical path

Disadvantage:

Increased area cost

Fig: 4X4 carry save multiplier



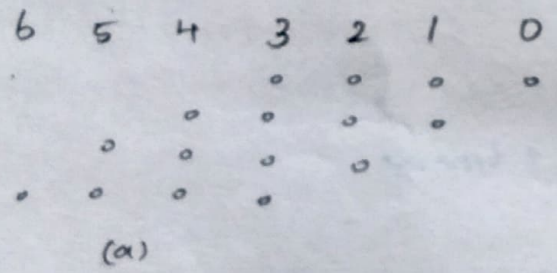
* Tree Multiplier:

* Partial Sum adders are arranged in a tree like fashion to reduce both the critical path and number of adder cells.

* Consider four partial products each of which is four bits wide as shown in Fig (a). All other columns are less complex as shown in fig b.

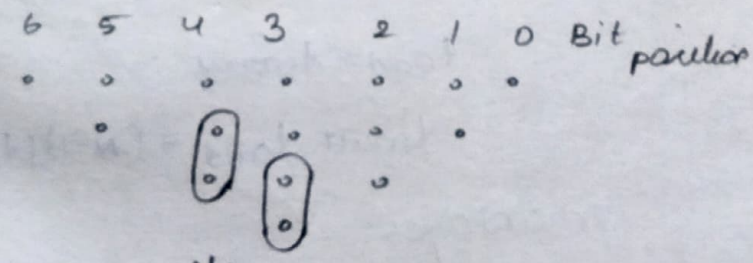
* The first type of operator used to cover the array is full adder, which takes three inputs and produces 2 outputs: the Sum, located in the same column and carry, located in next one. Full adder called 3-2 compressor

Partial products:



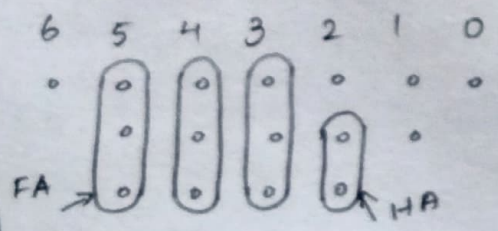
(a)

First Stage



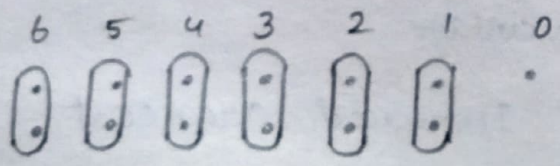
(b)

Second Stage:



(c)

Final adder



(d)

Disadvantage:

* very irregular
* complicated layout

* Final Addition:

The final step for computing the multiplication is to combine the result in the final adder. The choice of the adder style depends on the structure of the accumulation array.

4. DIVIDERS:

The following methods are used to divide integers

* combinational divider
* sequential divider
* Dividing by a constant

Combinational divider:

*Consider two unsigned binary numbers, D mbit dividend, d n-bit divisor. The division consists of finding two unsigned binary numbers, c the quotient and r the remainder, with r<d and d≠0, such that

$$D = c \times d + r$$

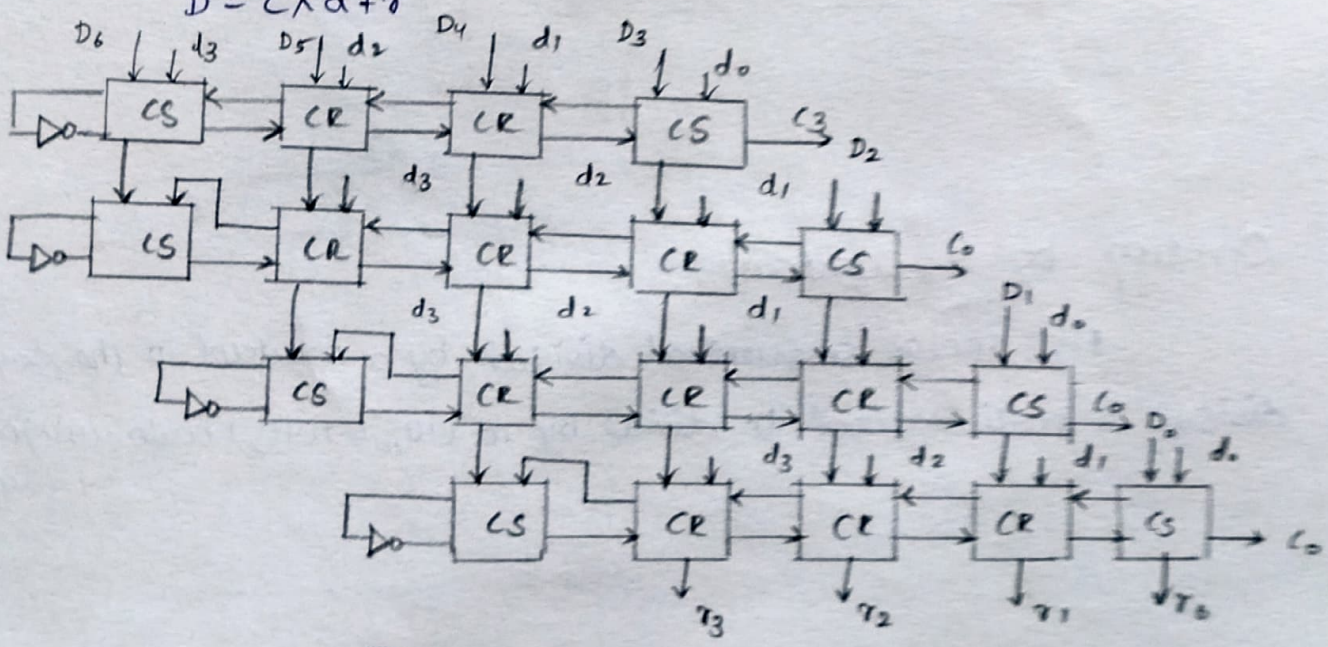


Fig combinational division

* Fig. has 7 bit dividend by a 4-bit unsigned divisor. Shows CR and CS cell respectively. The length of the dividend is $m = 2n-1$.

* If result is +ve, quotient bit is 1 and difference passes to next stage as a MSB of the modified dividend.

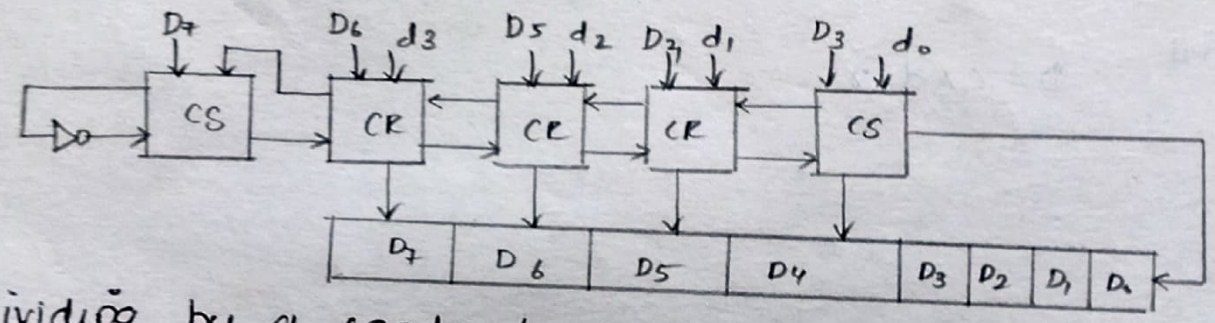
* If result is -ve, the quotient is 0 and dividend unchanged passes to next stage. The remainder $r_3 r_2 r_1 r_0$ is obtained

Sequential divider:

* Sequential division that devides D_1 of $2n-1$ bits by d_1 of n bits using n clock pulses. The circuit has three CR cells, two cells, one 4 bit latch to store the divisor d and 8 bit register for dividend D.

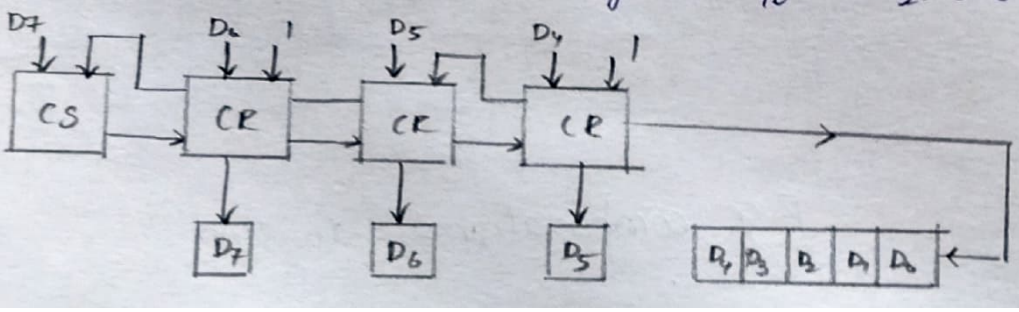
* The first register ($D_7 D_6 D_5 D_4$) reading and writing. The second ($D_3 D_2 D_1 D_0$) must be shift register serial I/p and o/p.

* D_7 is 0 before starting to divide and divisor is shifted so MSB bit of d is 1.



Dividing by a constant:

Fig: shows sequential division by a constant 10. The sequential division is particularized to divide by 10 ($10_{10} = 1010_2$) with unsigned 7 bit integer d



5. SHIFTERS:

* Shift operation is an arithmetic operation. This requires adequate hardware support. Shifters are used in floating units, scalars and multiplication by constant numbers.

* Programmable shifter is more complex and requires active circuitary.

* Fig. shows one bit left-right shifter. Depending on the control signals, the input word is either shifted left, or right or unchanged.

This approach is complex and ~~too~~ slow for large shift values. They are two types:

* Barrel Shifter
* logarithmic Shifter

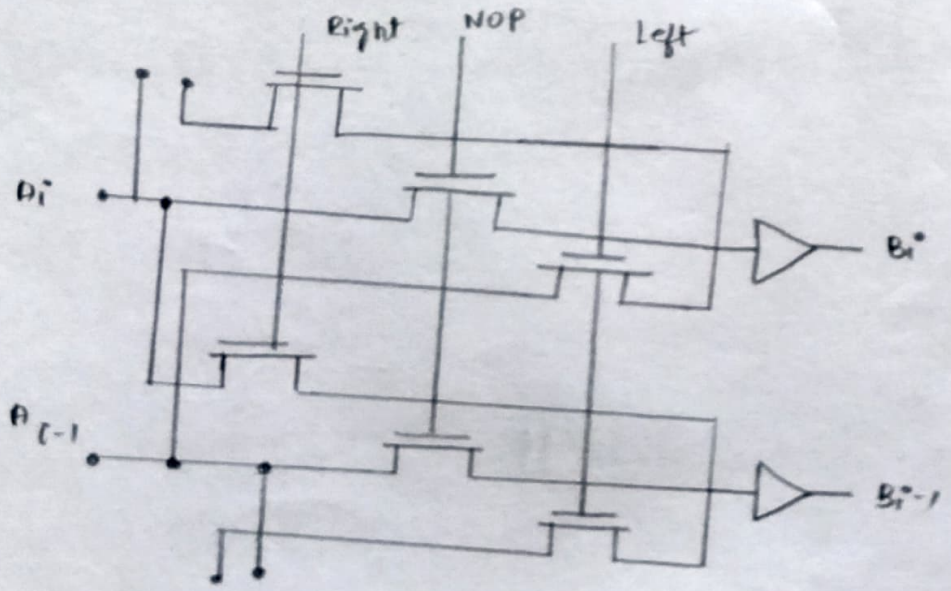


Fig: One bit left right shifter

Barrel Shifter:

* It consists of an array of transistors, in which the no. of rows equals to word length of data, no. of columns equals to maximum shift width. The control caries are routed diagonally through the array.

Advantage:

* Signal has to pass through at most one transmission gate
* propagation delay is constant
* used for small shift values.

Disadvantage:

* capacitance at the input of buffers rises linearly with maximum shift width

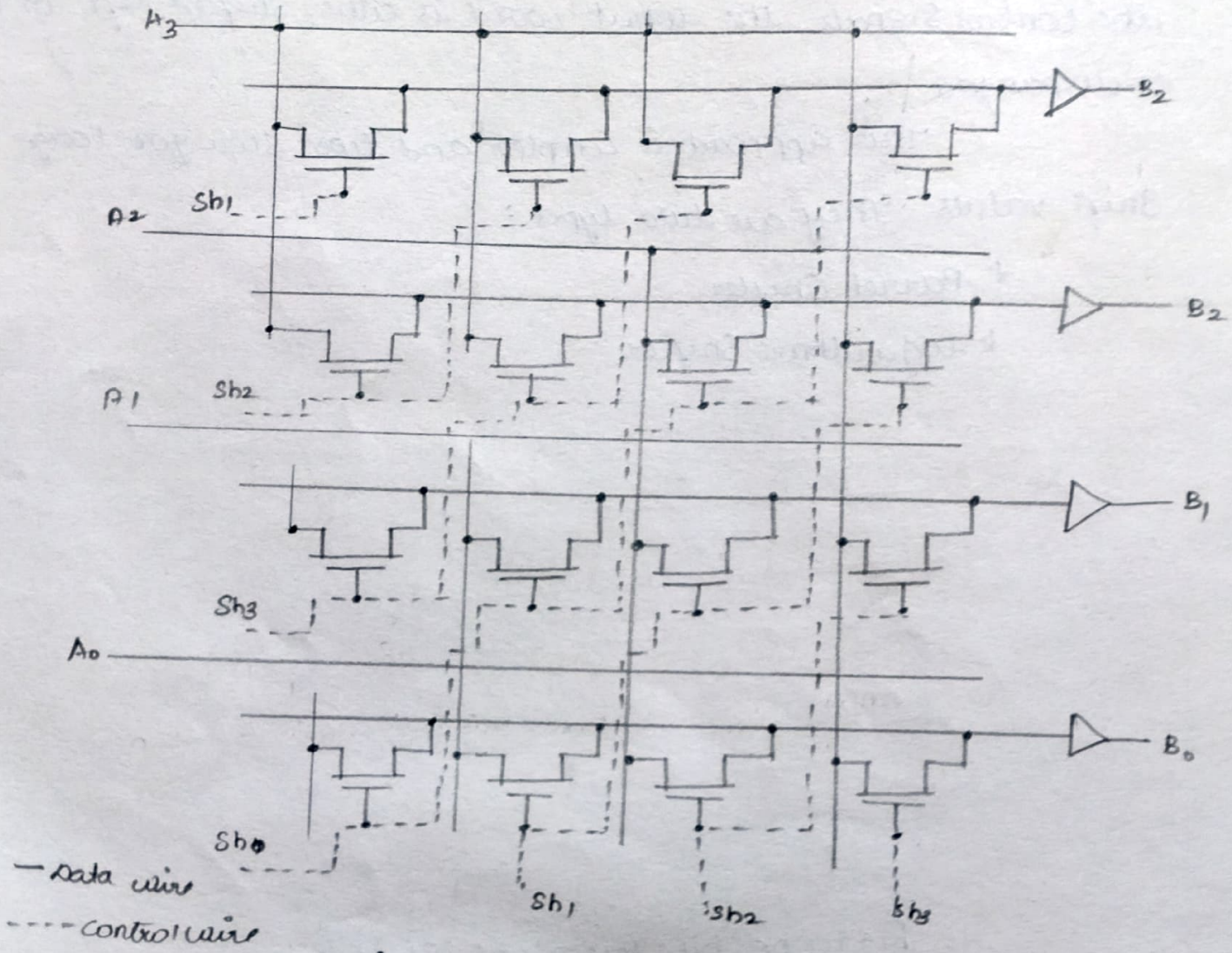


Fig: Barrel Shifter

* Layout size in barrel shifter is not dominated by the active transistors, but the no. of wires running through the cell.

* The size of the cell is bounded by the pitch of the metal wires.

* When selecting a shifter, the format in which to present the shift value is provided. Barrel shifter needs a control wire for every shift bit.

* The shift values required comes in an encoded binary format represented as 11 for shift over three bits

Logarithmic Shifter:

* It uses staged approach for shifting. The total shift value is decomposed into shift over powers of two.

* A shifter with a maximum shift width of M consists of $\log_2 M$ stages, where i^{th} stage either shifts overs 2^i or passes the data unchanged.

The control word for this shifter is already encoded and no separate decoder is required.

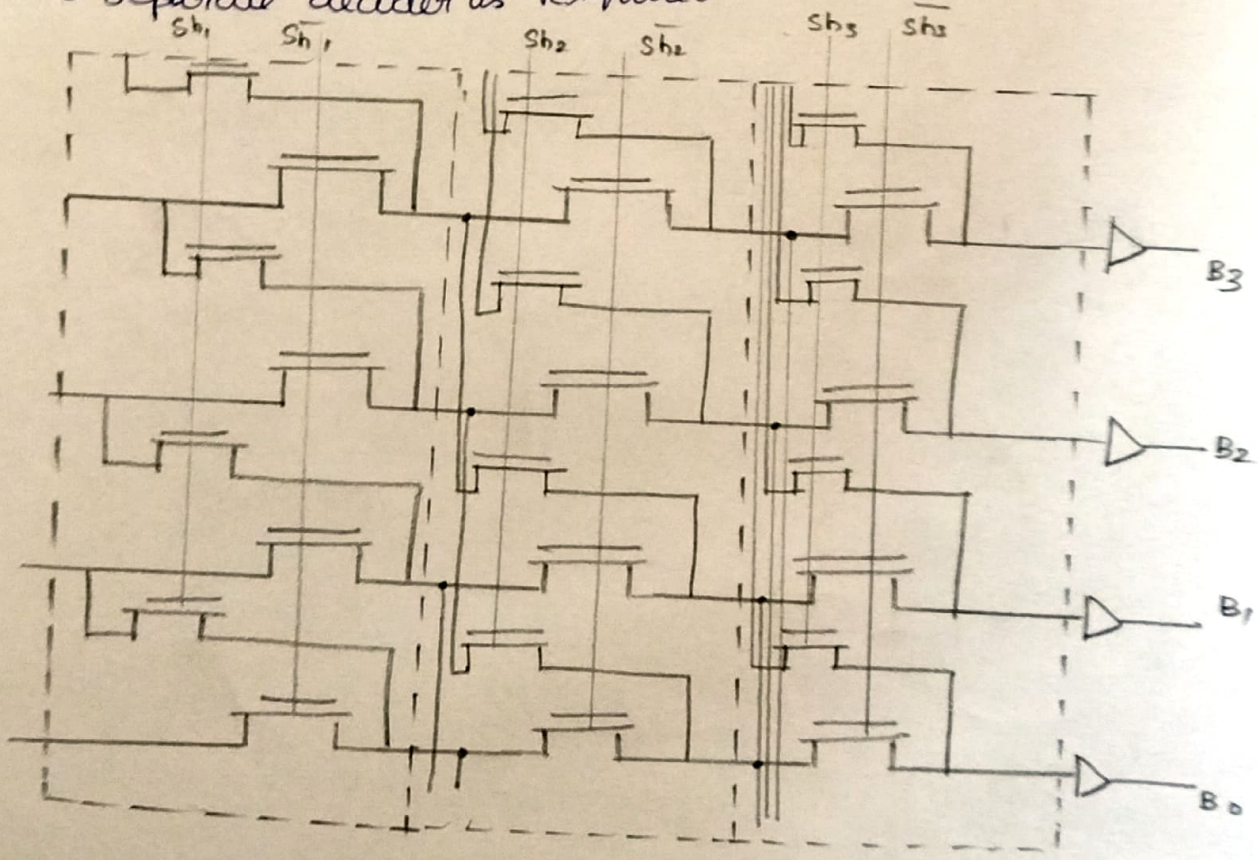


Fig: 4.24 (book) Logarithmic shifter with maximal shift width of seven bits to the right

* Speed of the logarithmic shifter depends on the shift width in logarithmic way.

* An M bit shift register $\log_2 M$ stages. Also intermediate buffers are careful introduced.

* For larger shift values, the logarithmic shifter becomes more effective, in terms of both area and speed.

* The shifter is easily parameterized, allowing for automatic generation.

UNIT - V

1. FPGA BUILDING BLOCK ARCHITECTURES.

All programmable ASICs or FPGAs Contain a basic logic cell. The basic logic cell is replicated in a regular array across the chip.

Three different type of basic logic cells:

i). Multiplexer based logic cells.

ii). Look-up table based logic cells.

iii) Programmable array logic based logic cells.

Four programming techniques for the Selection of these logic cells.

i). Actel ACT

ii) Xilinx LCA

iii) Altera FLEX

iv) Altera MAX.

1.1. Actel ACT:

Actel ACT has three logic family.

a). Act 1 b). Act 2 c). Act 3.

Act 1 → Uses one type of logic module

Act 2 & Act 3 → Uses two different type of logic modul

1.1.1. Act 1 logic Module:

This logic function is build using an Actel logic module by Connecting logic Signals to Some or all the logic module inputs and by Connecting any remaining logic module inputs to V_{DD} or GND.

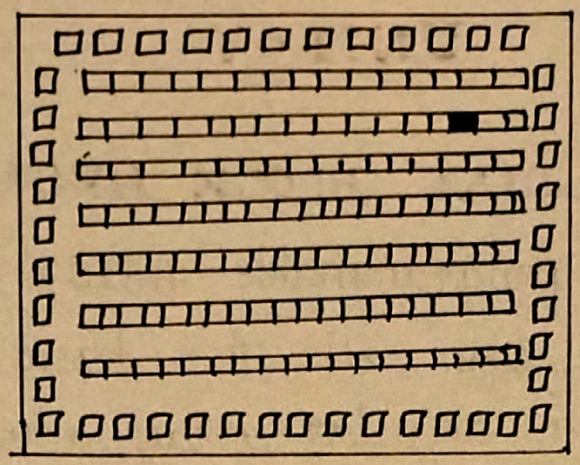


fig: functional behaviour of Actel ACT 1.

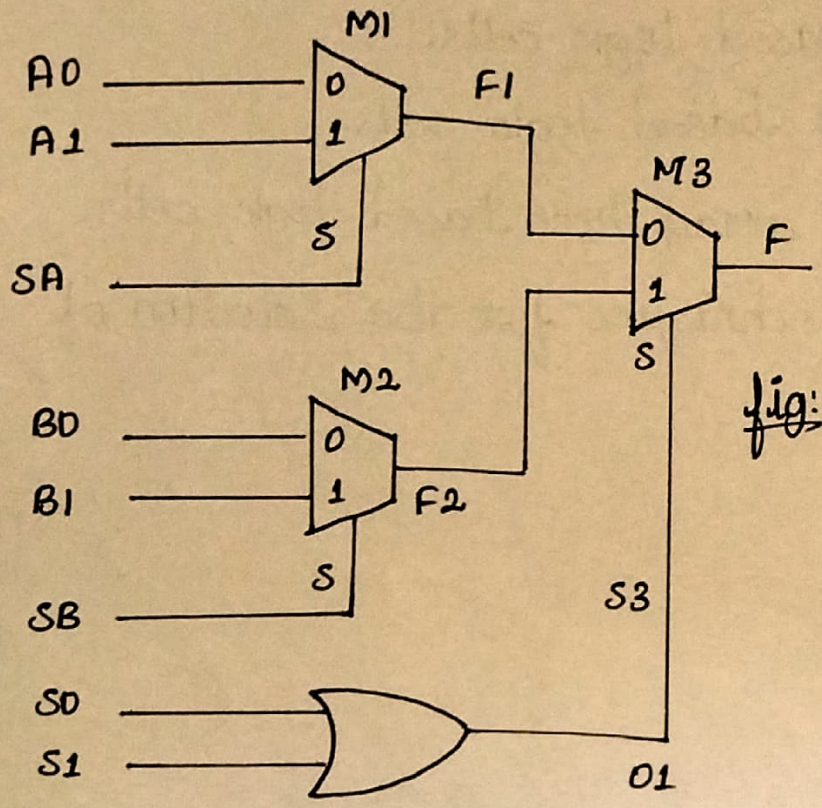


fig: Circuit level implementation of Actel ACT 1.

1.1.2. Act 2 and ACT 3 logic Modules.

The Act 2 and Act 3 architectures uses two different types of logic modules.

i). Sequential element implementation that can be Configured as a flip flop. in fig. ⓐ.

ii). Sequential element Configured as positive edge triggered D flipflop. in fig ⓑ.

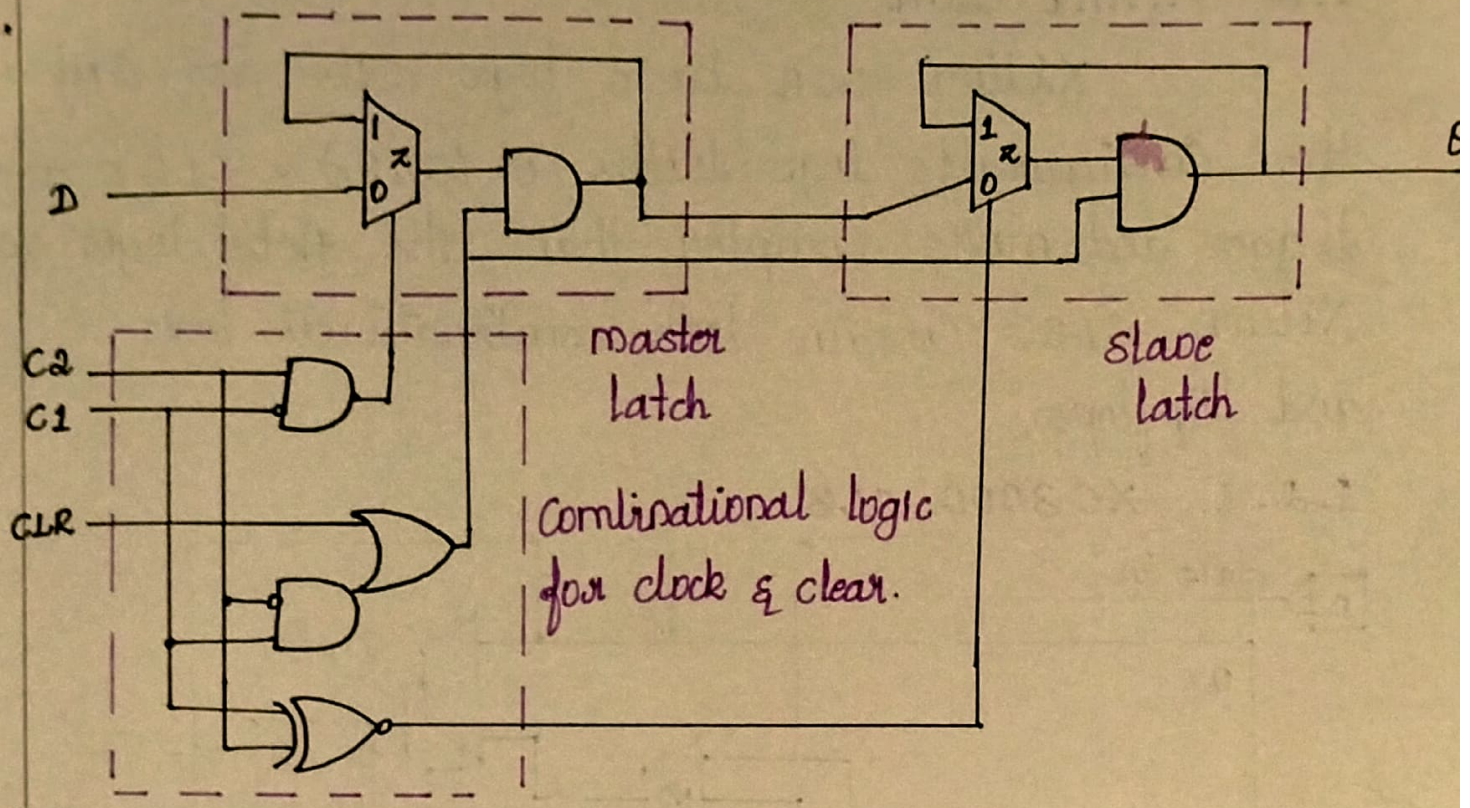


fig ⓐ. Sequential element implementation that can be Configured as a flipflop.

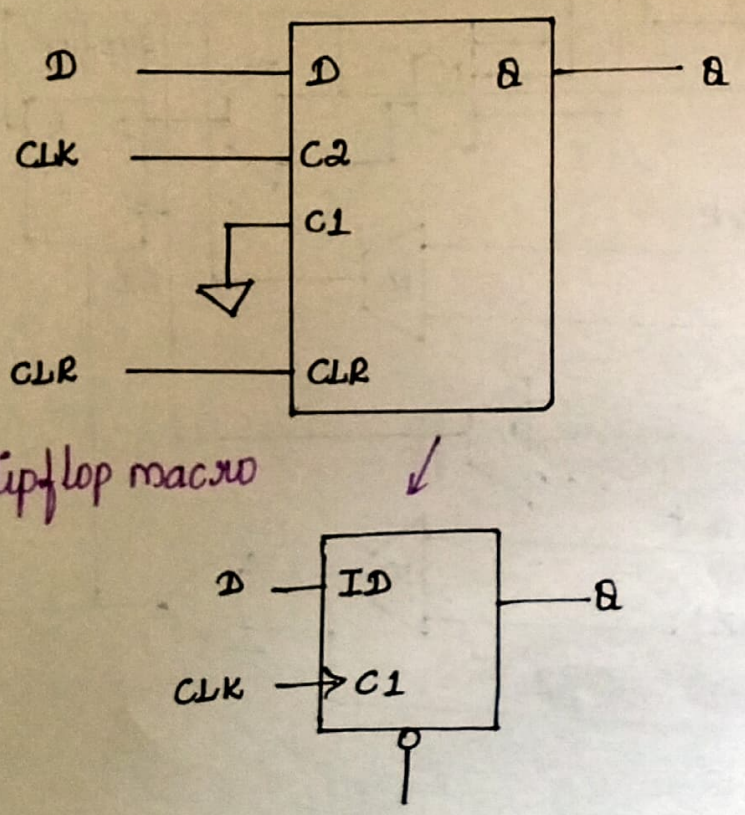


fig ⓑ. Sequential element Configured as positive edge triggered D flip flop.

1.2 Xilinx LCA

Xilinx LCA basic logic cells are called the Configurable logic blocks or (CLBs). CLBs are bigger and more Complex than the Actel logic cells. Xilinx CLBs Contain both combinational logic and flipflops.

1.2.1. XC 3000 CLB.

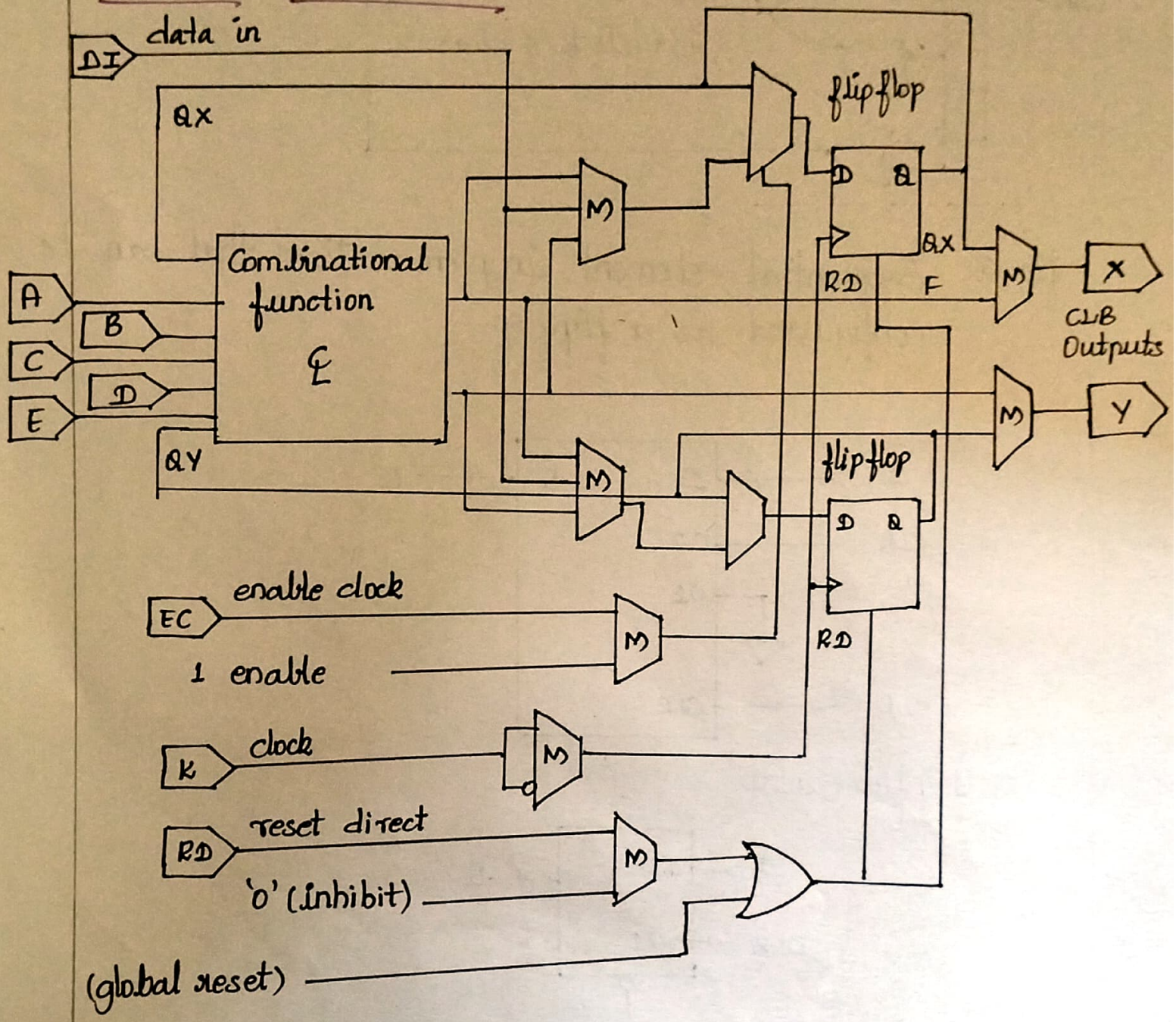


fig: XC3000 CLB.

1.2.2. The above figure has,

five logic inputs (A – E)

A Common clock input

An asynchronous direct-reset input (RD)

An enable (EC)

Two CLB outputs X and Y are connected independently to the flip flop output QX and QY

A 32-bit look up table (LUT), stored in 32 bits of SRAM, provides the ability to implement combinational logic.

1.2.2 XC4000 Logic block.

This is a complicated basic logic cell containing 2 four-inputs LUTs that feed a three input LUT. This has special fast carry logic hard-wired between CLBs.

MUX Control logic maps four control inputs (C1 – C4) into the following four inputs.

i) HI – LUT input

ii) DIN – Direct IN

iii) EC – Enable clock

iv) S/R – Set/Reset Control.

The control inputs is (C1 – C4) is used to control the use of the F' and G' LUTs as 32 bits of SRAM.

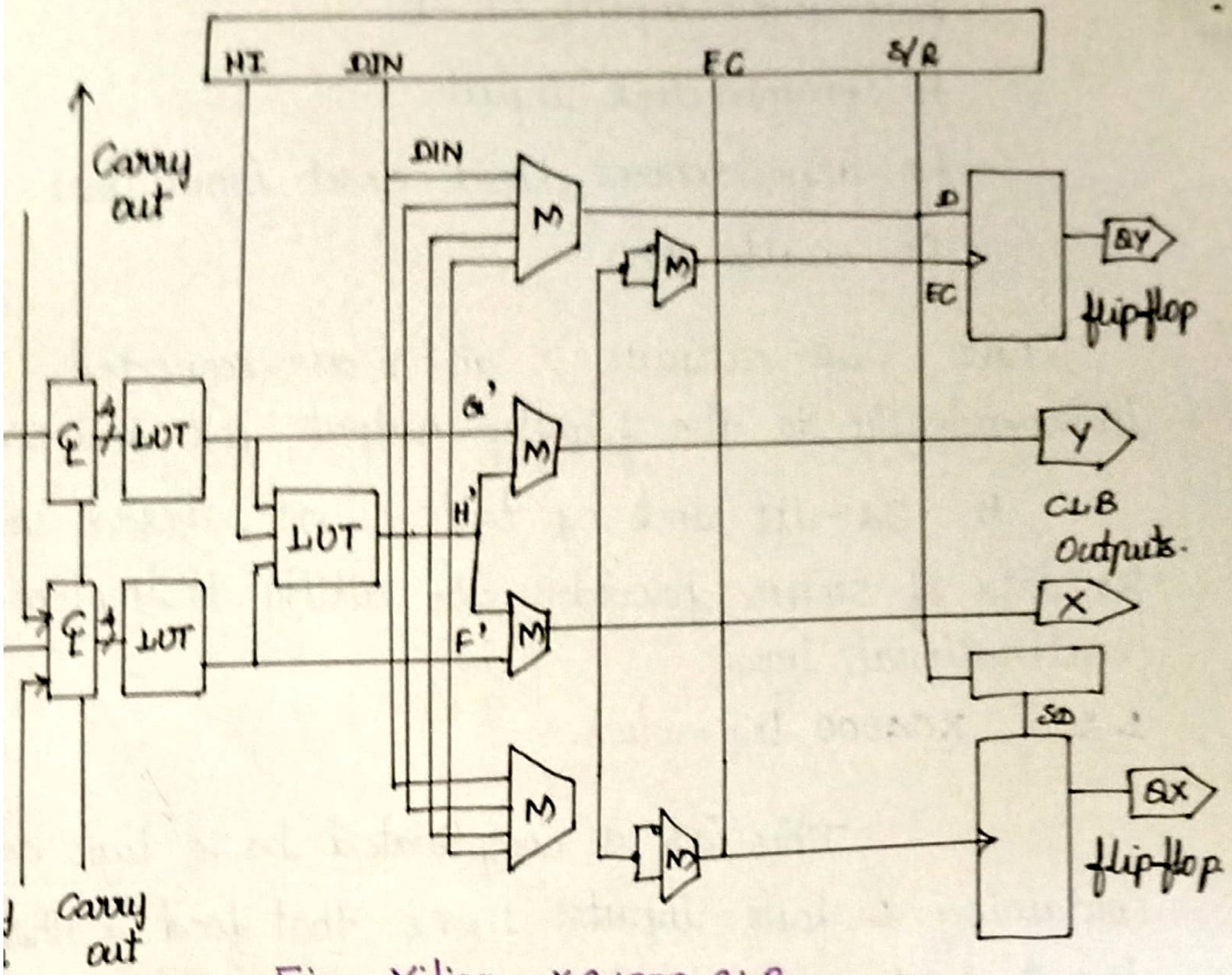


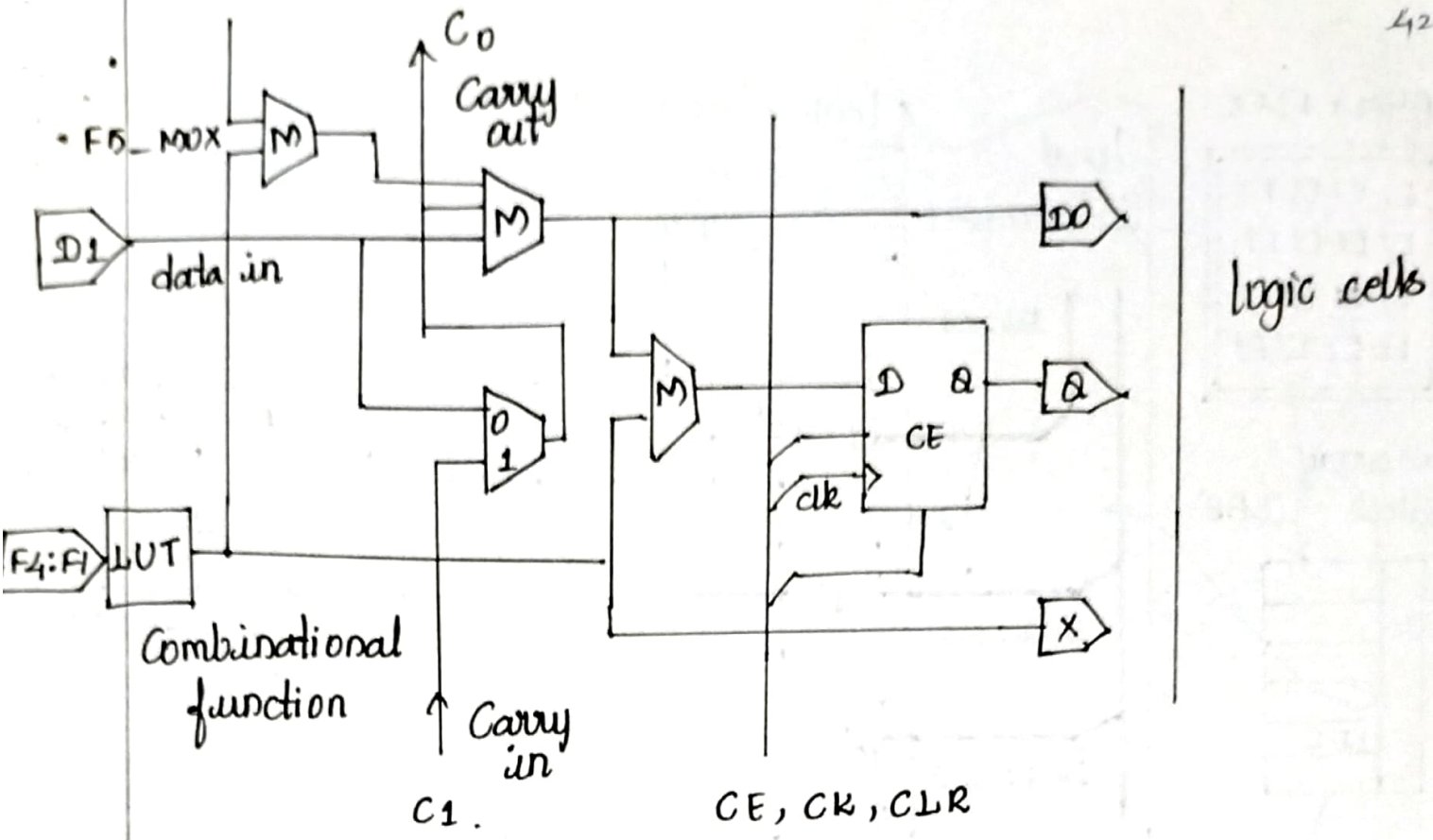
Fig: Xilinx XC4000 CLB.

1.2.3. XC 5200 logic block L.

This is a Simpler logic cell.

The arithematic carry logic is Separate from the LUTs.

A limited Capality to cascade function is provided to gang two LC's in parallel to provide the equivalent of the five input LUT.



Xilinx CLB Analysis:

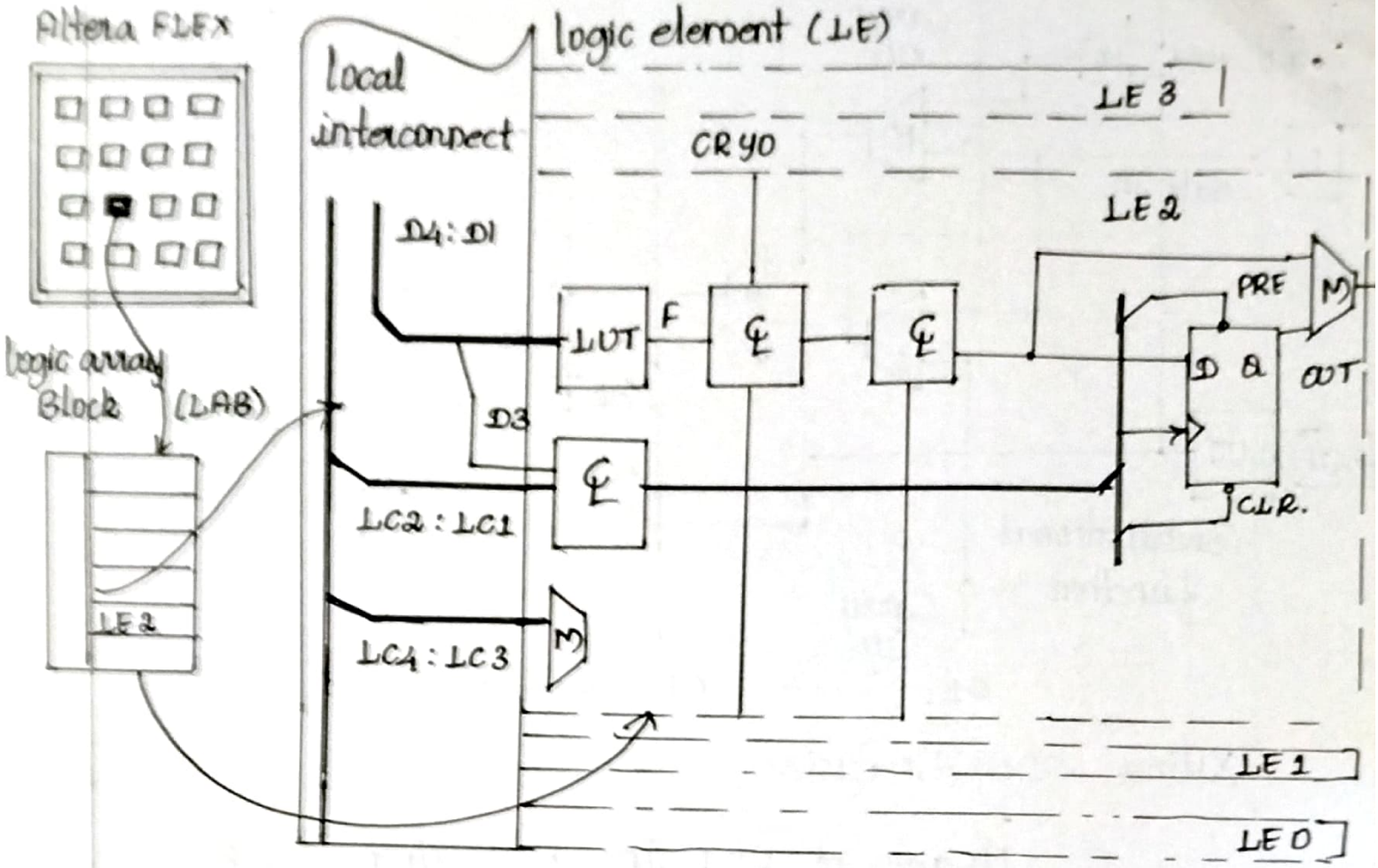
Usage of LUT in a Xilinx CLB to implement Combinational logic is both an advantage and disadvantage.

1.3.1. Altera FLEX.

LE (logic element) is used in Altera FLEX 8000 Series of FPGAs.

FLEX Cell resembles XC5200LC architecture and based on SRAM programming technology

The FLEX LE uses four-input LUT, a flipflop, cascade logic and carry logic.



Altera flex Architecture.

1.4.1. Altera MAX.

The figure a) shows a Simple two level logic circuits that implements Sum of products.

This representation is Simplified by drawing the input lines to a Multiple-input AND Gate as if they were one horizontal wire.

This structure is Called programmable array logic.

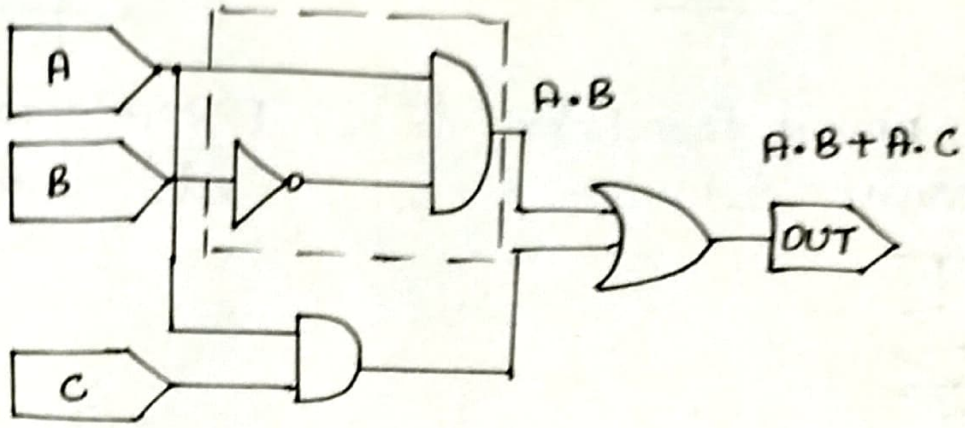


fig. a). Two level logic Array.

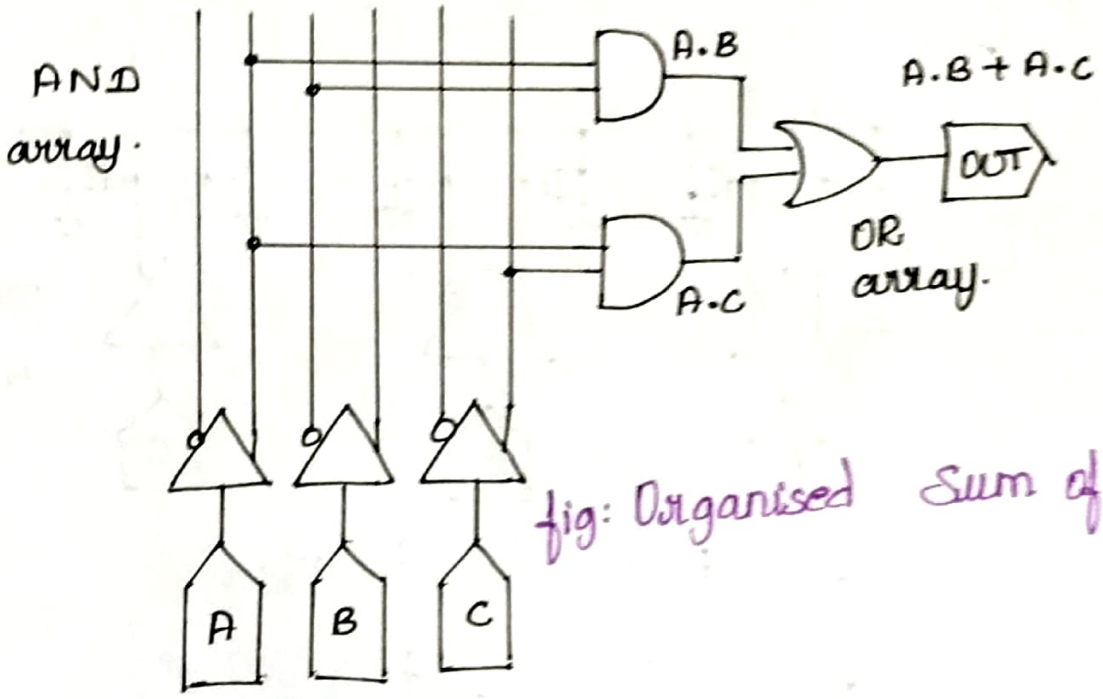


fig: Organised Sum of products

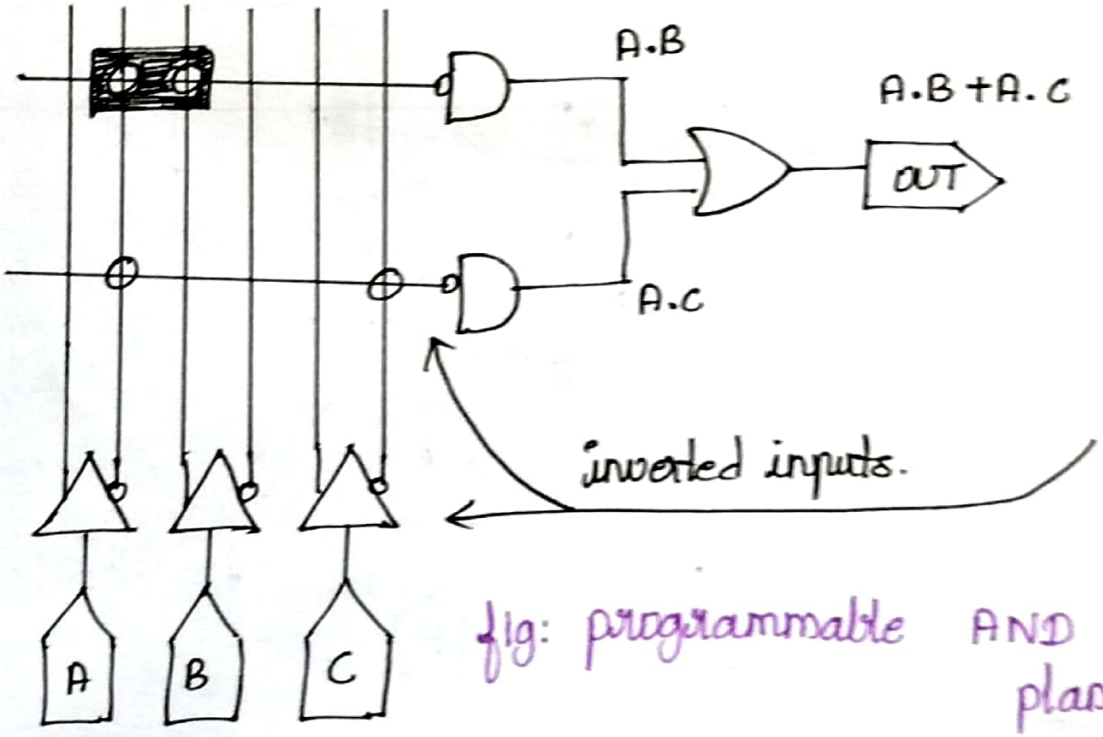


fig: programmable AND plane.

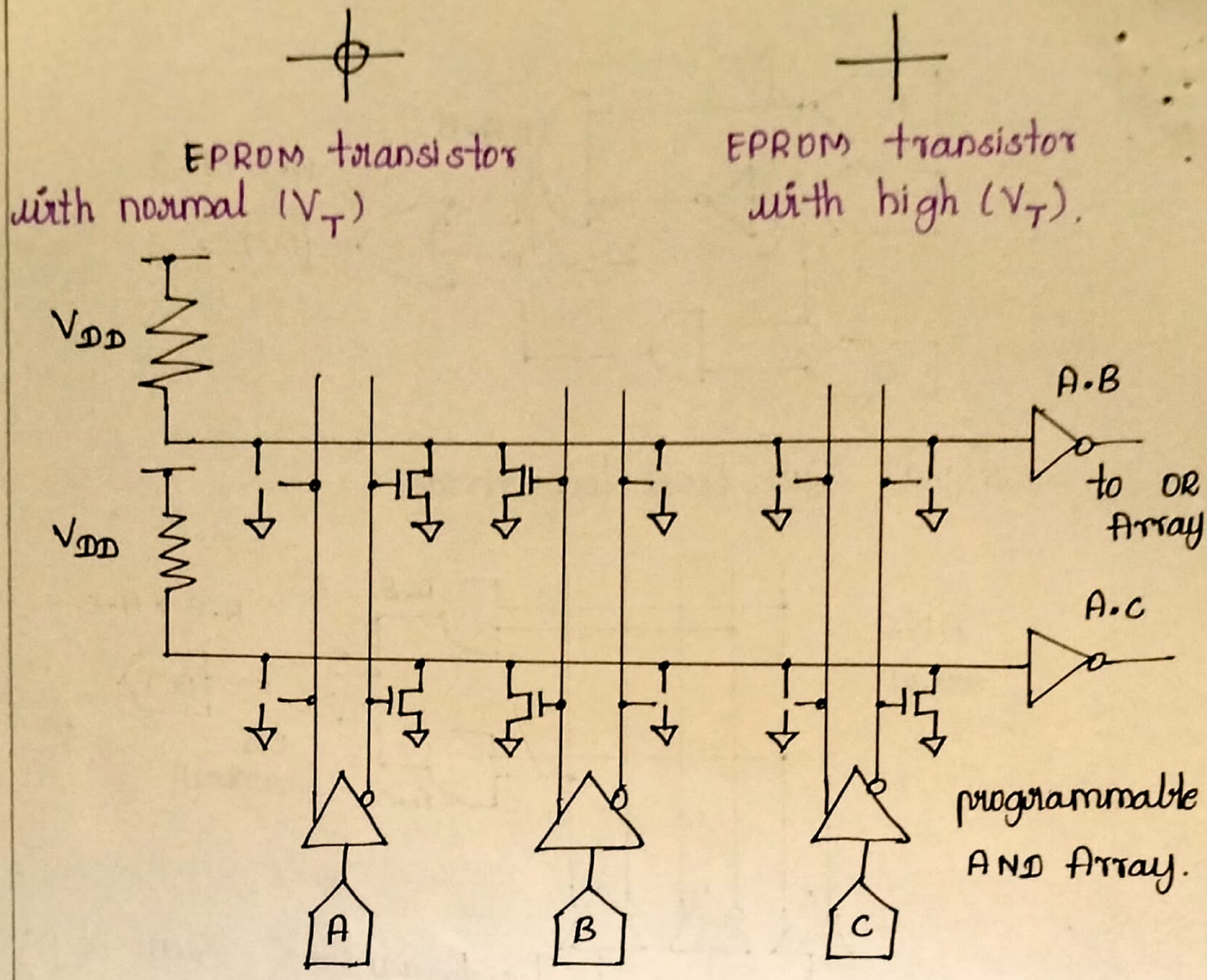


fig: EPROM logic Array.

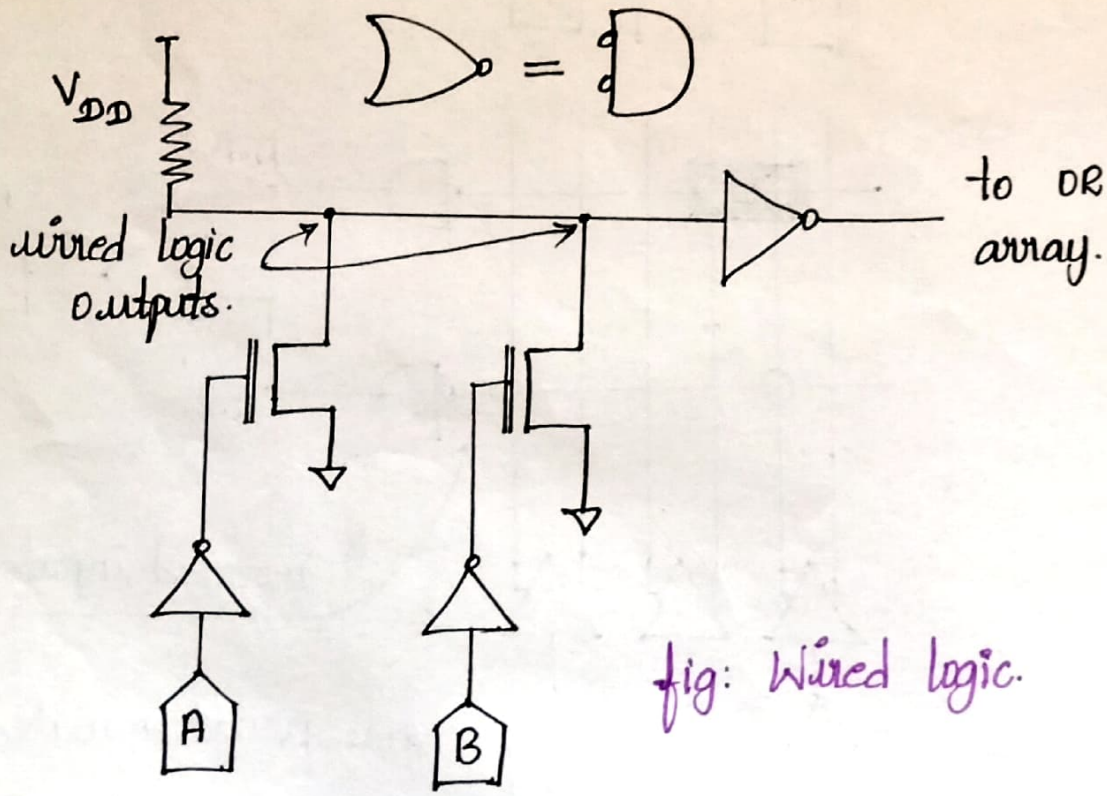


fig: Wired logic.

2). STANDARD CELL DESIGN

Each standard cell in a library is rectangle with Same height but different widths.

The bounding box (BB) of a logical cell is the Smallest rectangle that encloses all of the geometr of the cell.

BB is determined by the well layers.

Cell Connectors or terminals are placed on the cell abutment box (AB).

physical connector is a piece of metal to which wires are connected. standard cells are constructed in such a manner that they are placed next to each other horizontally with the cell touching AB

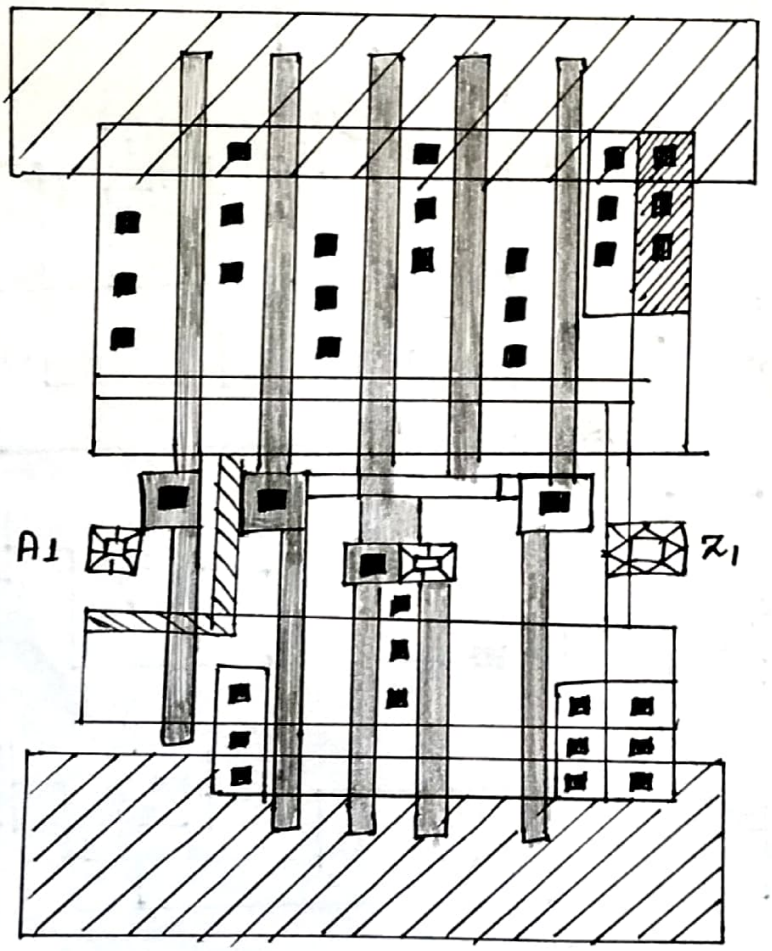


figure: Standard cell.

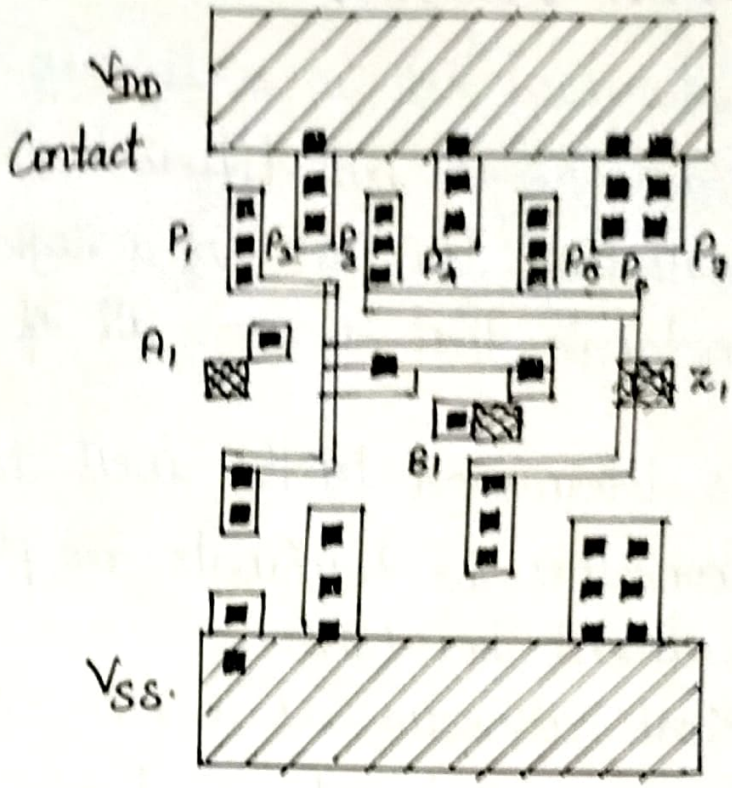


figure: m1 and Contact layers.

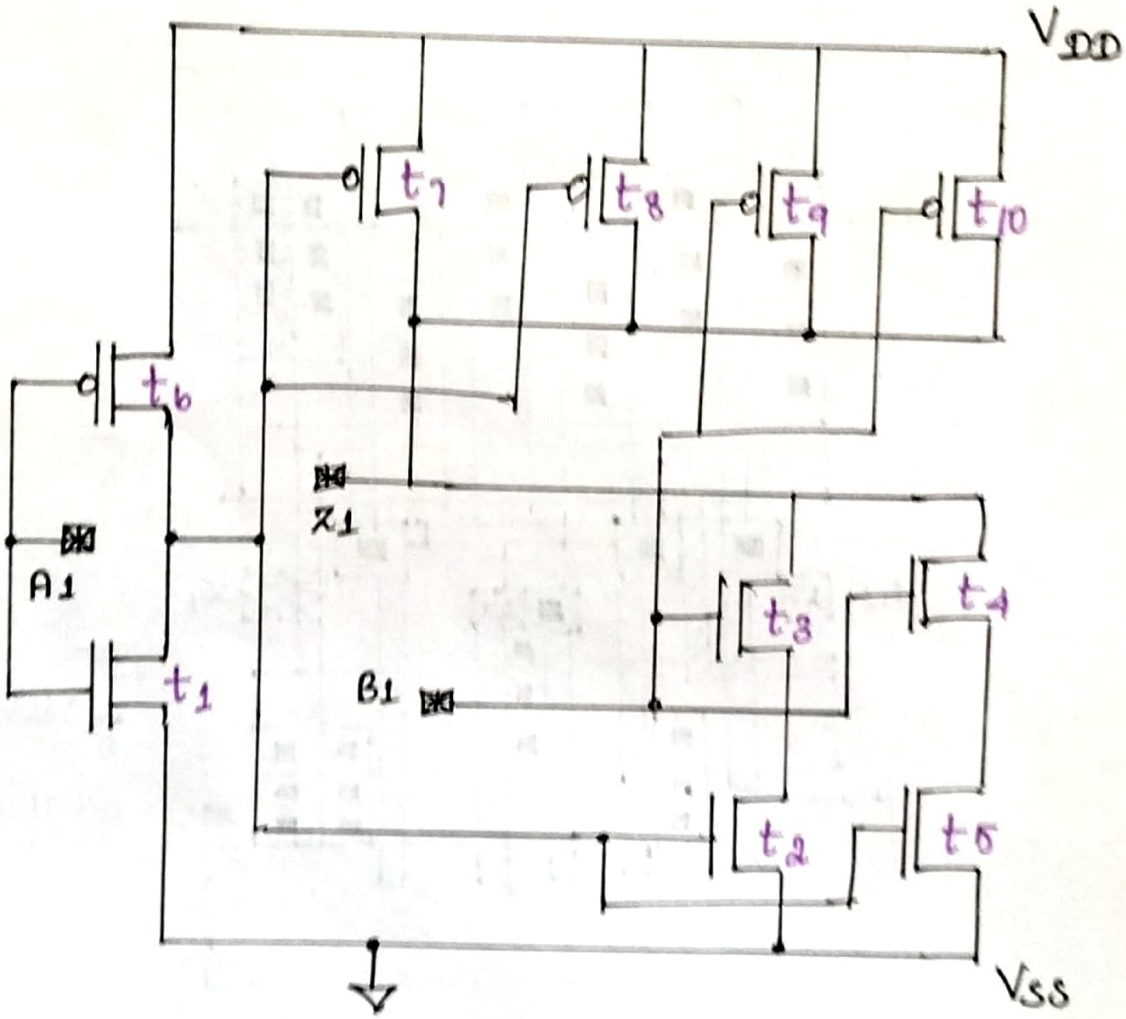


fig: Equivalent Schematic of standard cells.

ASIC CELL LIBRARIES.

Cell library is very important in ASIC design. for MGAs and CBICs there are three choices to have the cell library.

i) ASIC manufacturer will supply a cell library.
ii) Cell library is bought from a third-party library vendor.
iii) Build your own cell library.

In the first choice, an ASIC manufacturer provides the library that makes to use a set of design tools approved by the ASIC manufacturer to enter and stimulate our design. This library is normally a phantom library.

The second and third choices need to make a buy or build decision. This is called Customer owned tooling (COT).

The third choice is to develop a cell library in house.

Each cell in an ASIC cell library contains the following.

i). Physical layout: In programmable ASIC cell layout is part of programmable ASIC design.

ii). Behavioural Model: Needed because Simulation at the detailed timing level takes too long for a Complete ASIC design.

iii) Verilog / VHDL model: This models are required in addition to the models for a particular logic Simulations.

iv) Timing Model: Needed to determine the performance of the Critical pieces of an ASIC. library engineers Simulate the delay of each cell a process known as characterization.

v). Test strategy:

vi). Circuit Schematic: cell designer can perform Simulation for Compler cells. Called layout verses Schematic (LVS) check.

vii) Cell icon: Each cell needs a cell icon together with Connector and naming information.

viii). Wire load Model: For the statistical estimate of the Capacitance for a net in a given Size circuit block. This needs a look up table known as wire load model.

ix). Routing Model: For this Simpler representation called phantom is needed. Phantom of a physical layout Contains all neccessary informatn.

3. ARRAY BASED DESIGN

3.1. Gate array based ASICs.

Gate Array (GA) based ASIC has predefined transistors on the Silicon wafer.

The predefined pattern of transistors on a Gate array is the base array.

The base array is made up of Smallest element Called a primitive Cell.

The logic cells in a gate array library are Called 'macros'.

3.2. Channeled Gate array

The channeled Gate array has space between the rows of transistors for wiring.

The channeled Gate array is Similiar to a CBIC. Both use rows of cells Separated by channels used for interconnect.

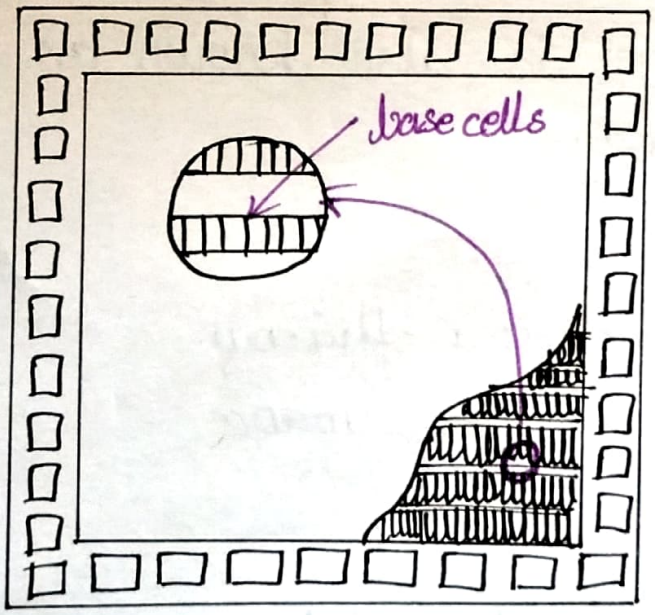


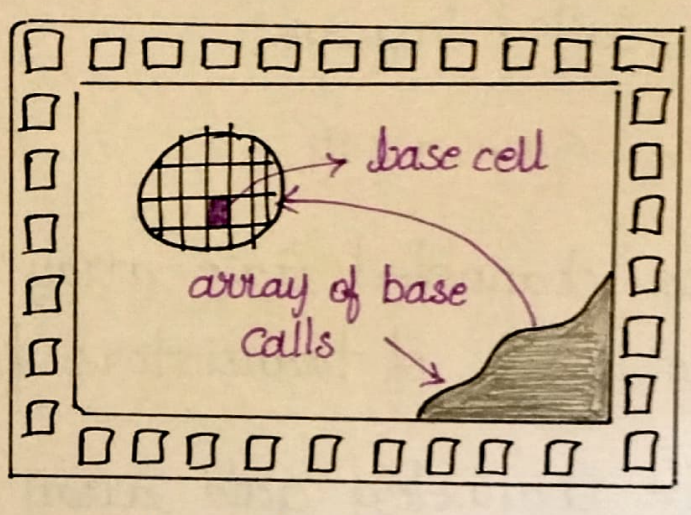
fig: channeled Gate array.

3.3. Channelless Gate Array:

channelless gate array is widely used and is also known as channel free gate array, Sea of gate array, or SOG array.

The features are:

- ⇒ Top few mask layers are customized interConnec
- ⇒ Manufacturing lead time is between two days and two weeks.



3.4. Structured Gate Array:

Structured Gate Array can either be channelled or channelless.

It is also known as masterslice or masterimage.

Advantages:

Improved area efficiency.

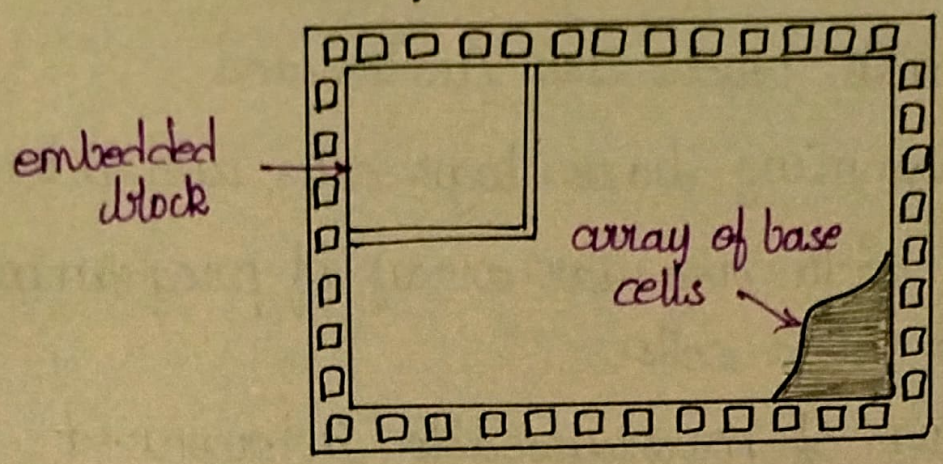
Increased performance

Lower Cost

Faster turnaround.

Disadvantages:

Embedded function is fixed.



3.5. Programmable logic Devices.

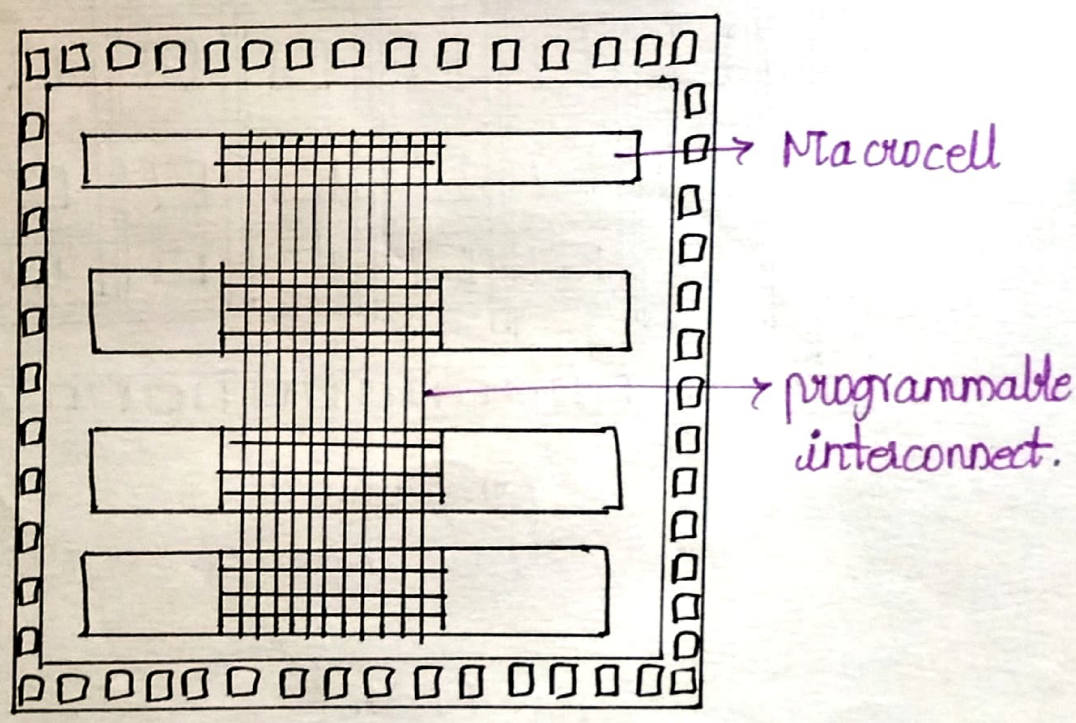
features:

i). No Customized mask layers or logic cells

ii) fast design turnaround.

iii) Single large block of programmable interconnect.

iv). Matrix of large macro cells.



3.6 Field programmable Gate Arrays.

Characteristics:

1) No mask layers are Customized
2). Programming basic logic cells and interconnect.
3). Core with regular array of programmable basic logic cells.
4). Matrix of programmable interconnect
5). Programmable I/O Cells Surround the core
6). Design turnaround is a few hours.

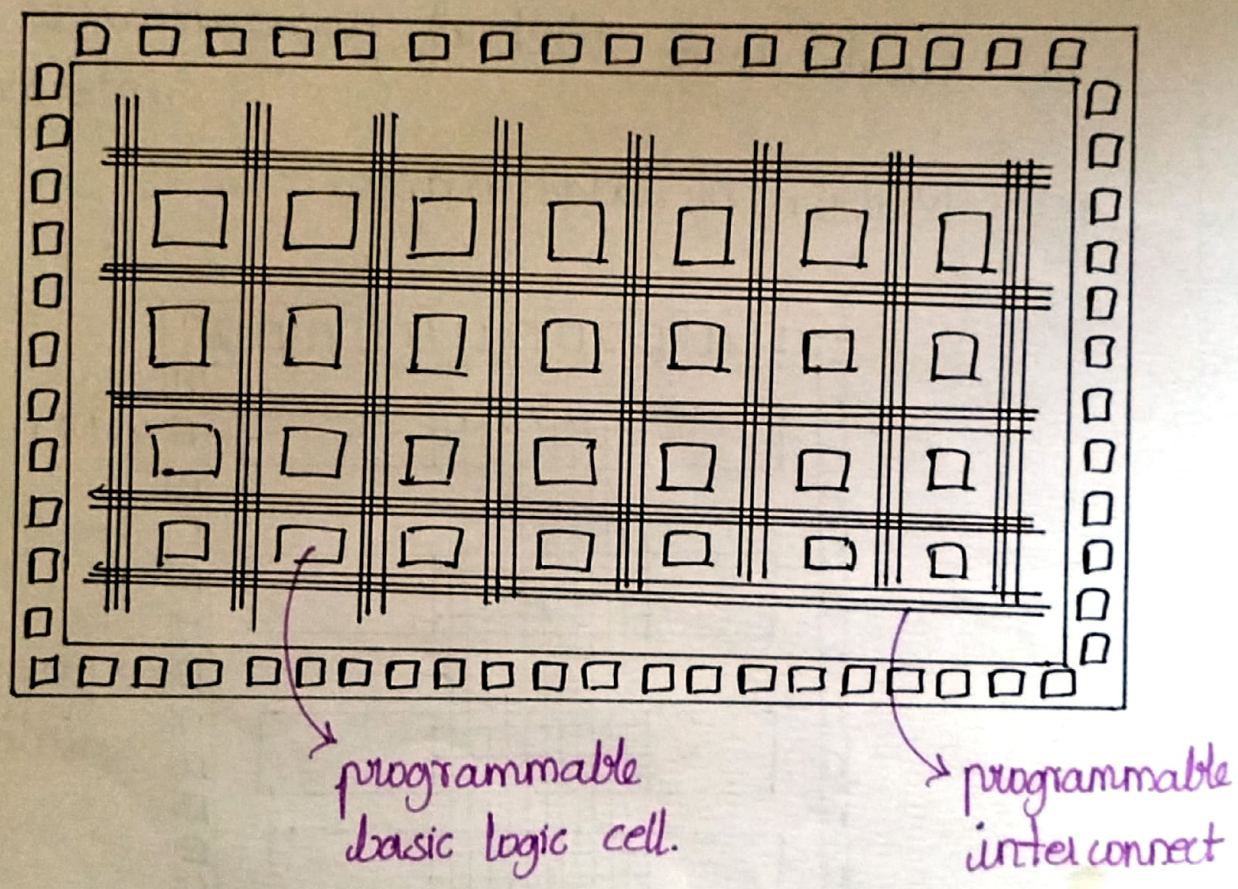


fig: FPGA

• ASIC DESIGN FLOW.

1). DESIGN ENTRY: Enter the design into an ASIC design systems, either using HDL or Schematic entry

2). LOGIC SYNTHESIS: Use VHDL or Verilog and a logic Synthesis tool to produce a netlist

3). SYSTEM PARTITIONING: Divide a large System into ASIC sized pieces.

4). PRELAYOUT STIMULATION: check whether the design functions are correct.

5). FLOORPLANNING: Arrange the blocks of the netlist on the chip.

6). PLACEMENT: Decide the locations of cell in a block.

7). Routing: Make the connections between cells and blocks.

8). Extraction: Determine the resistance and Capacitance of the interconnect.

9). Post layout stimulation: check to see the design still works with the added loads.

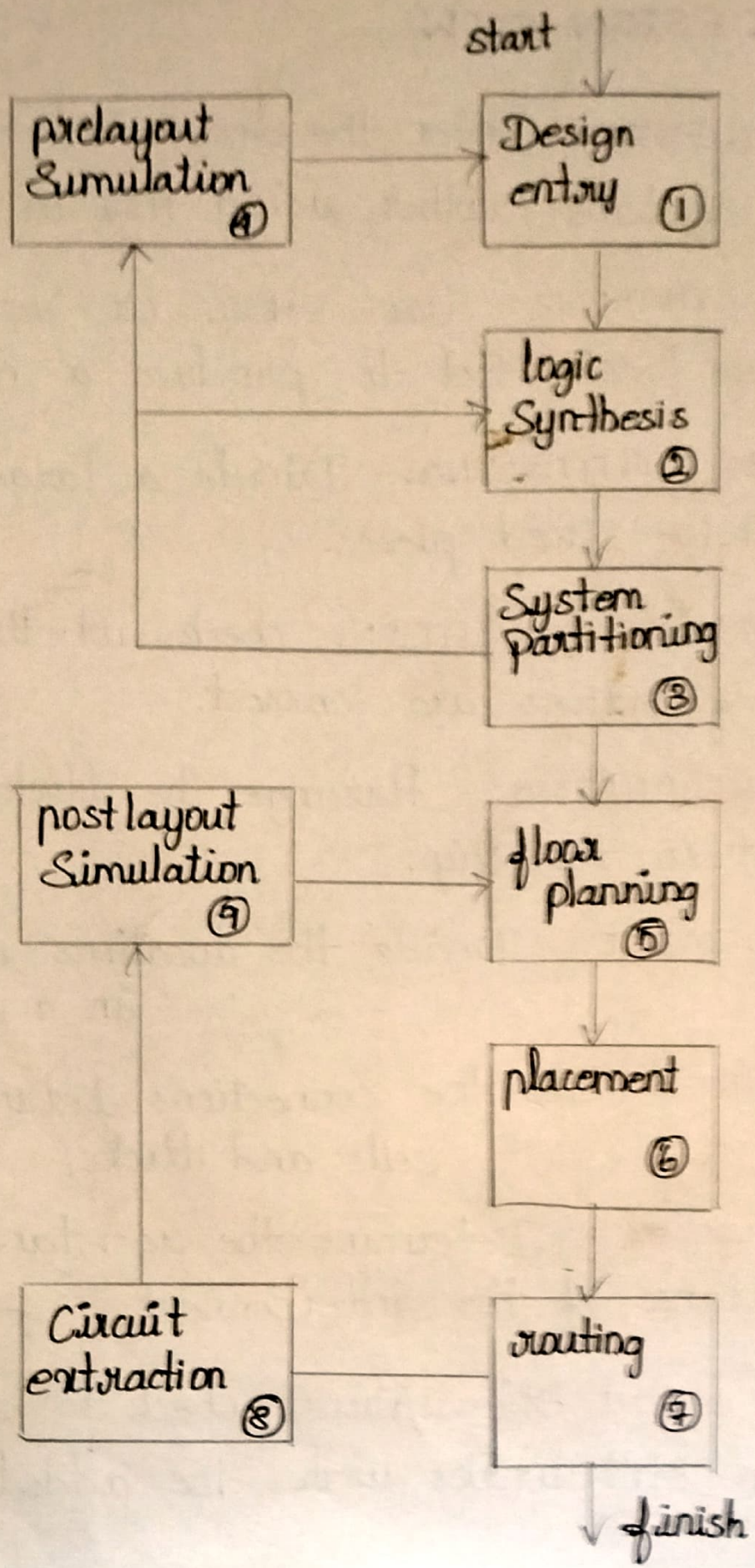
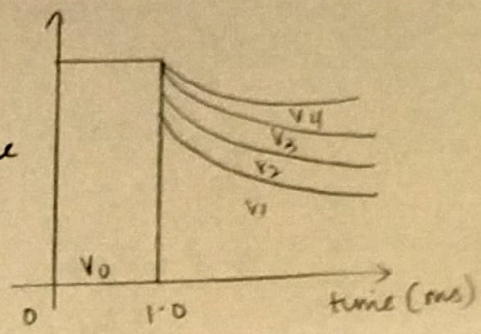


fig: ASIC Design flow.

Let the incident initial voltage at each node V_{DD} assuming to be 1 volt at time $t=0$

At time, the node is connected to ground. So $V_0 = 0$ Volt. we need to find the voltage from V_1 to V_4 as a function of time



Elmore obtain some delay in voltage reduction from V_1 to V_4 by using the RC ladder ckt. The current in any branch K,

$$I_K = -C_K \, dV_K/dt$$

The voltage at that node I can be represented as,

$$V_i = -\sum_{K=1}^{n} R_{Ki} \cdot C_K \frac{dV_K}{dt}$$

The node voltages have different values at each point in the time but the waveforms are similar.

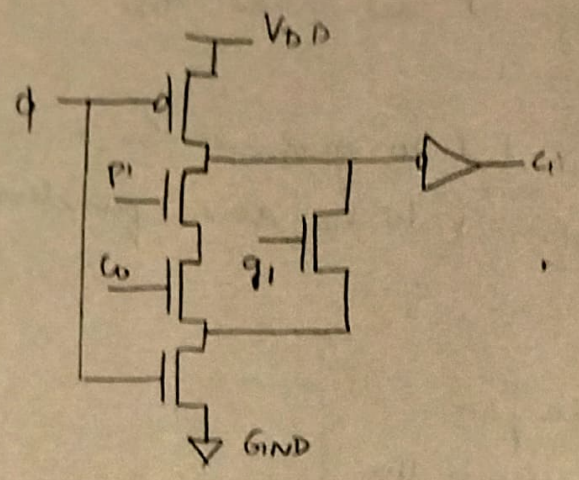
The linear first order differential equation with the following equation.

$$V_i(t) = e^{-t/\tau_{Di}}$$

$$\tau_{Di} = \sum_{K=1}^{n} R_{Ki} \cdot C_K$$

The time constant τ_{Di} is called the elmore's delay or elmore's time constant.

The elmore's delay is different for each node.



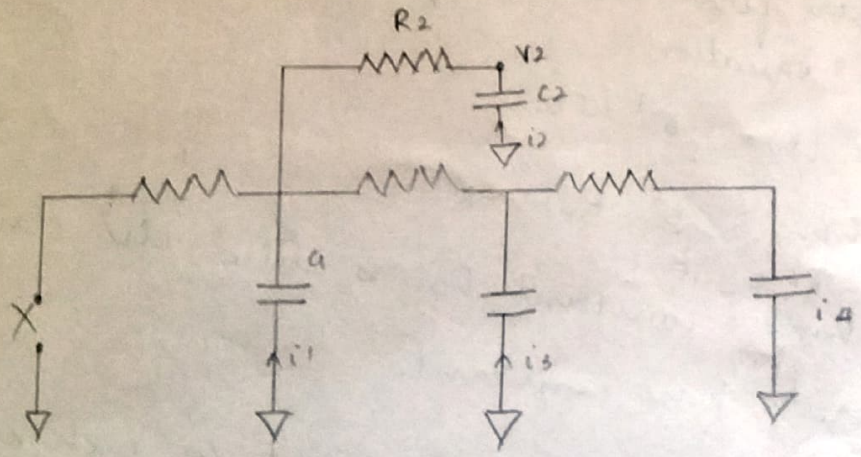
NP and Zipper Domino:

The HI-Skew inverting static gates are replaced
If φ is 0, I & III stages precharge high & II - low
φ is 1, all stages evaluate

Disadvantages:

Logical effort is worst
Susceptible to noise

ELMORE IS CONSTANT:



RC CIRCUIT