

# MODULE :- 1

## INTRODUCTION

Advaith PR  
Asst. Prof  
Dept of ECE  
SVIT

### A Brief History :-

The incredible growth in transistors have come from steady miniaturization of transistors and improvements in manufacturing processes. As transistors become smaller, they also become faster, dissipate less power and are cheaper to manufacture.

The level of integration of chips has been classified as small scale, medium scale, large scale and very large scale.

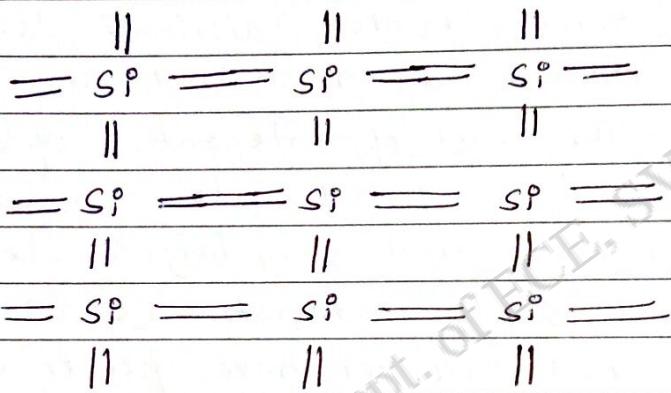
- Small scale integration (SSI) Circuits such as 7404 inverter have fewer than 10 gates with a conversion of roughly half a dozen transistors per gate.
- Medium scale integration (MSI) Circuits such as 74161 counter have upto 10000 gates.
- Large scale integration (LSI) Circuits such as a simple 8 bit microprocessors have upto 10,000 gates.
- The term very large scale integration (VLSI) is widely used to describe most integrated circuits.

"A law proposed by Moore states that the transistors become faster, consume less power and are cheaper to manufacture each year."

## MOS Transistors :-

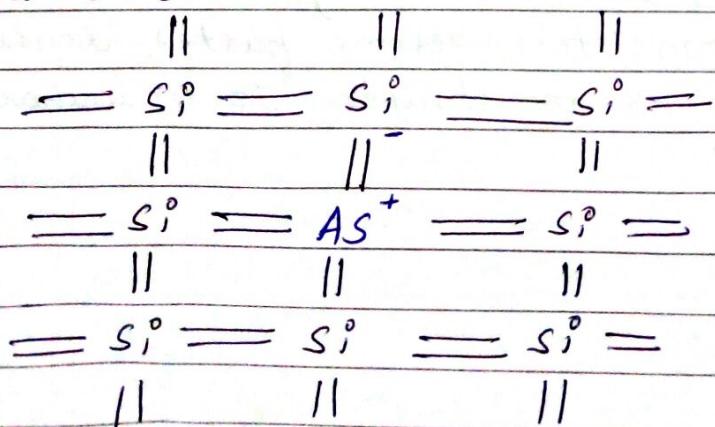
Silicon a semiconductor, forms the basic starting material for a large class of integrated circuits. Pure silicon consists of a three dimensional lattice of atoms.

Silicon is a group IV Element, so it forms a covalent bonds with four adjacent atoms as shown below.



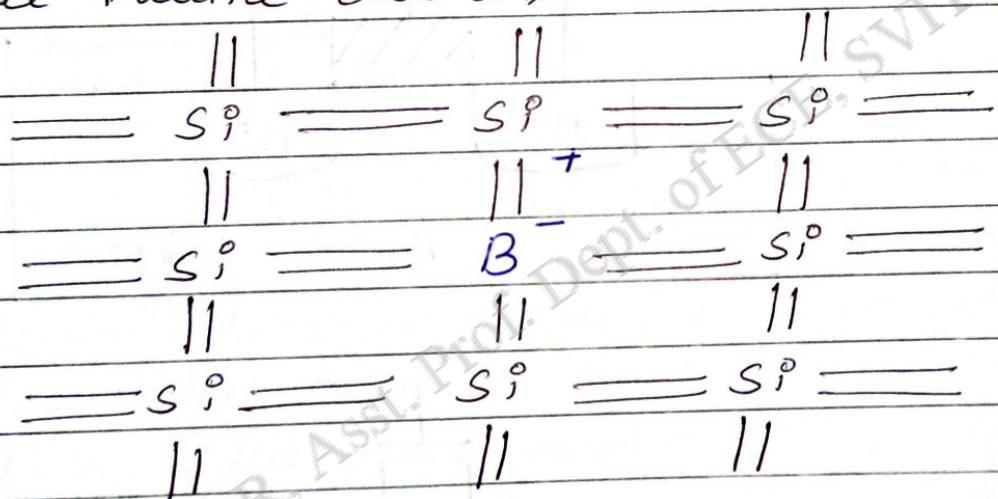
The conductivity can be raised by introducing small amounts of impurities into the silicon lattice. These impurities are called as dopants.

A Group V dopant such as arsenic has five valence electrons. It replaces a silicon atom in the lattice and bonds with four neighbours, hence the fifth valence electron is loosely bound to the arsenic atom as shown below.



Thermal vibration of the lattice at 9100m temperature is enough to set the Electron free to move, leaving a positively charged  $\text{As}^+$  ion and a free Electron. The free Electron can carry current so the conductivity is higher. We call this as n type semiconductor because the free carriers are negatively charged electrons.

A Group III dopant such as boron has three valence electrons as shown below.



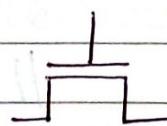
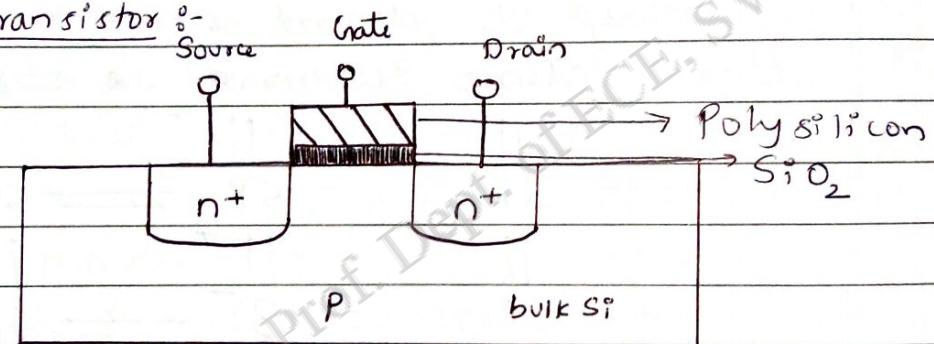
The dopant atom can borrow an Electron from a neighbourhood silicon atom, which in turn becomes short by one Electron, and so forth. Hole can propagate throughout the silicon lattice. Since the hole acts as a positive charge carrier, we call this kind of semiconductor as p type semiconductor.

MOS (Metal Oxide Semiconductor) structure is created by superimposing several layers of conducting and insulating materials to form a sandwich like structure.

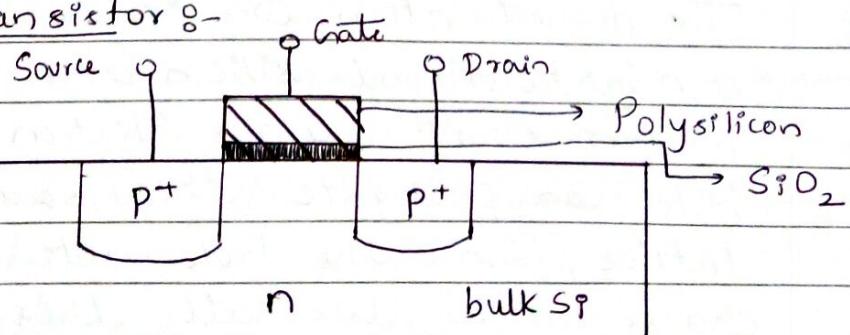
These structures are manufactured using a series of chemical processing steps involving oxidation of silicon, diffusion of impurities into silicon to give silicon certain conducting characteristics, deposition and etching of aluminium or other metals to provide interconnections in the same way that a printed wiring board is constructed.

CMOS technology provides two types of transistors, n type transistor (nMOS) and a p type transistor (pMOS).

n mos transistor :-



p mos transistor :-



Each transistor consists of a stack of a conducting gate, an insulating layer of silicon dioxide ( $\text{SiO}_2$ ), and silicon wafer also called as substrate, body or bulk.

An nMOS transistor is built with a p type body and has regions of n type semiconductor adjacent to the gate called as source and drain. The body (substrate or bulk) is typically grounded.

A pMOS transistor is built with a n type body (substrate or bulk) and has regions of p type semiconductor adjacent to the gate called as source and drain.

In a CMOS technology with both flavors (pmos & nmos) of transistors, the substrate is either ntype or p type. The other flavor of transistor must be built in a special well in which dopant atoms will be locally added to form the body of the opposite type.

X. Gate is the control input, it affects the flow of electrical current between the source and drain.

Let us consider an nmos transistor, the body is generally grounded so the pn junctions of the source and drain to body are reverse biased. If the gate is also grounded, no current flows through the reverse biased junctions. Hence we say that the transistor is OFF.

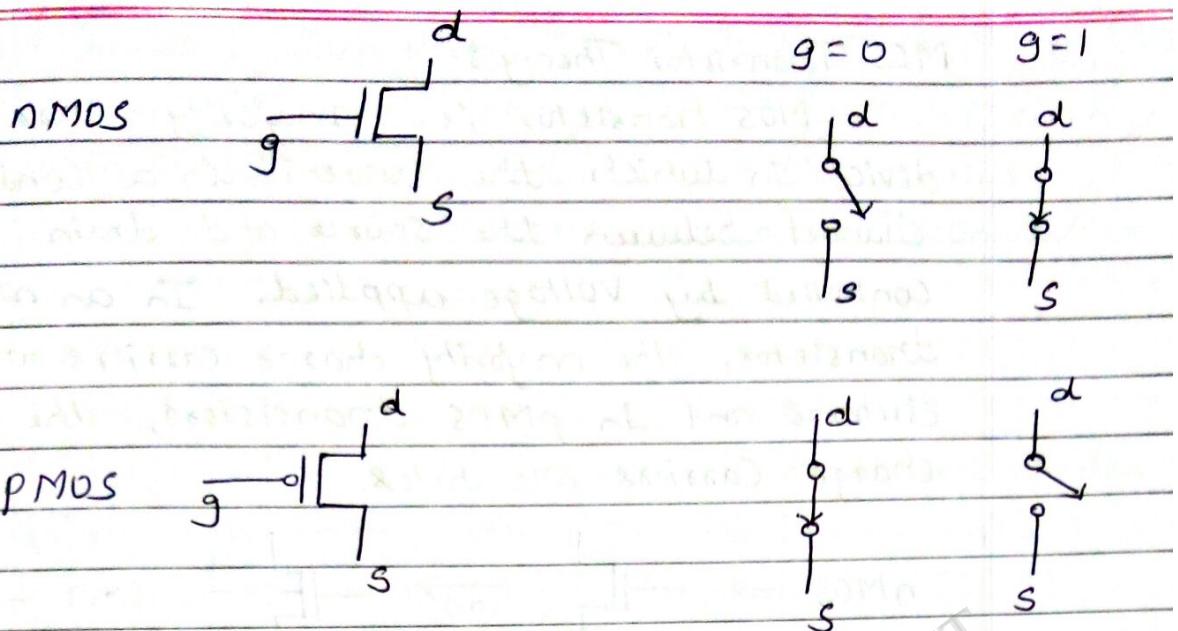
A sufficient gate voltage creates an electric field that starts to attract free electrons to the underside of Si-SiO<sub>2</sub> interface. If the gate voltage is sufficiently raised, the electrons outnumber the holes and a thin region under the gate called as channel is formed and acts as n-type semiconductor. Hence a conducting path of electron carriers is formed from source to drain and current can flow. Now the transistor is ON.

Let us consider a pMOS transistor, the situation is reversed. The body is held at a high potential. When the gate is also at high potential, the source and drain junctions are reverse biased and no current flows and the transistor is turned OFF.

When the gate voltage is lowered, positive charges are attracted to the underside of the Si-SiO<sub>2</sub> interface. A sufficiently low gate voltage inverts the channel and a conducting path for positive charge carriers is formed from source to drain, so the transistor is ON.

In summary, the gate of an MOS transistor controls the flow of current between the source and drain.

- When the gate of nMOS transistor is 1, the transistor is ON and there is conducting path between source and drain.
- When the gate of nMOS transistor is low(0), the transistor is OFF and negligible amount of current flows from source to drain.



Where  $g = \text{gate}$

$d = \text{drain}$

$s = \text{source}$

- When the gate of pMOS transistor is 0, the transistor is ON and there is conducting path between drain and source.
- When the gate of pMOS transistor is 1, the transistor is OFF and negligible current flows from source to drain.

In simple terms, nMOS transistor turns ON for high gate input and turns OFF for low gate input.

pMOS transistor turns ON for low gate input and turns OFF for high gate input.

Advaith PR  
Asst. Prof  
Dept. of ECE  
SVIT

## CMOS logic :-

(a) The Inverter :-

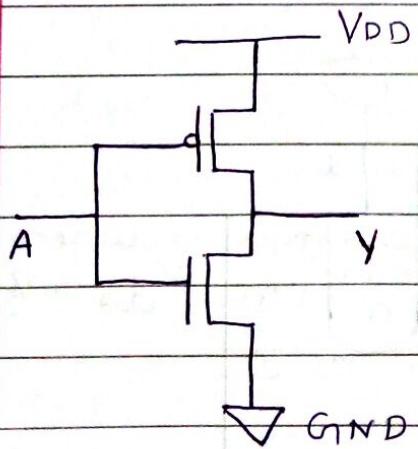


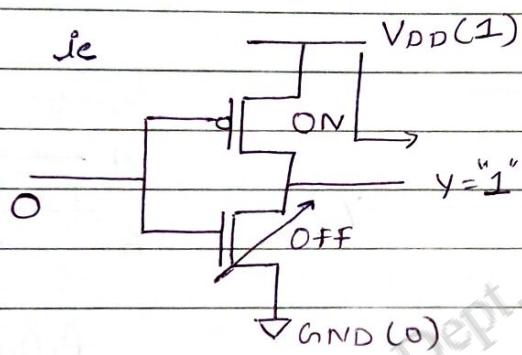
fig (ii)

fig (i)

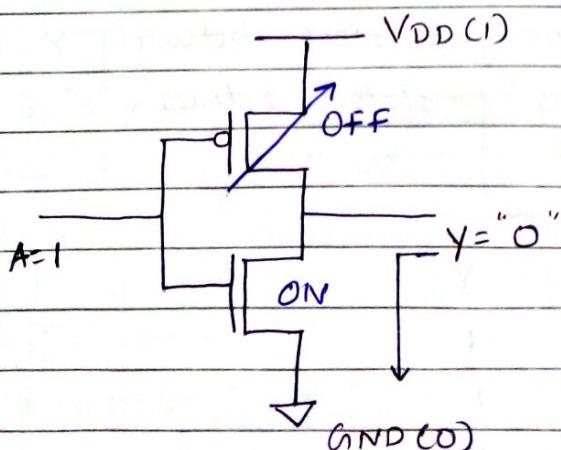
→ fig (i) shows a CMOS inverter or NOT gate using one nMOS transistor and one pMOS transistor.

→ The horizontal bar at the top indicates  $V_{DD}$  and the triangle at the bottom indicates GND.

→ When the input A is "0", the nMOS transistor is "OFF" and pMOS transistor is "ON", thus the output "Y" is pulled up to "1", because it is connected to  $V_{DD}$ .



→ When the input A is "1", the nMOS transistor is "ON" and pMOS transistor is "OFF", thus the output "Y" is pulled down to "0", because it is connected to ground.



Truth Table:-

| A | Y |
|---|---|
| 0 | 1 |
| 1 | 0 |

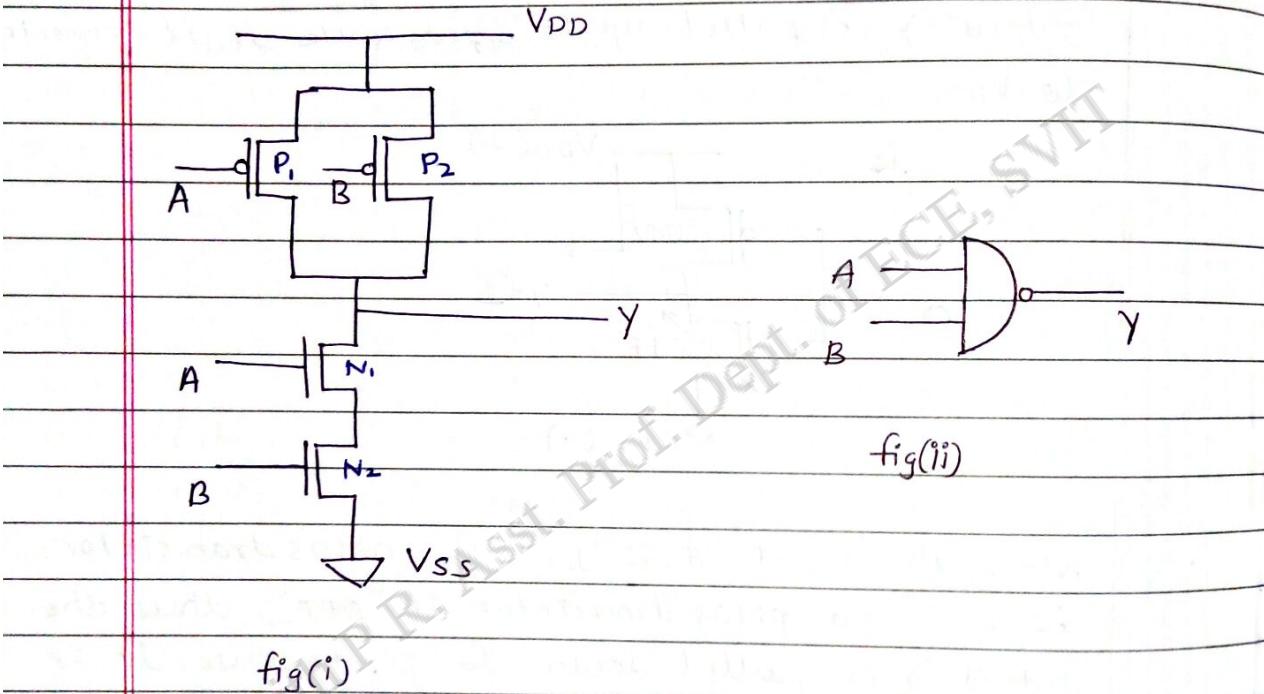
(b) The NAND gate :-

Boolean expression for NAND gate is

$$Y = \overline{A \cdot B}$$

In CMOS, "•" operation, PMOS are connected in parallel. NMOS are connected in series

"+" operation, PMOS are connected in series. NMOS are connected in parallel



fig(ii)

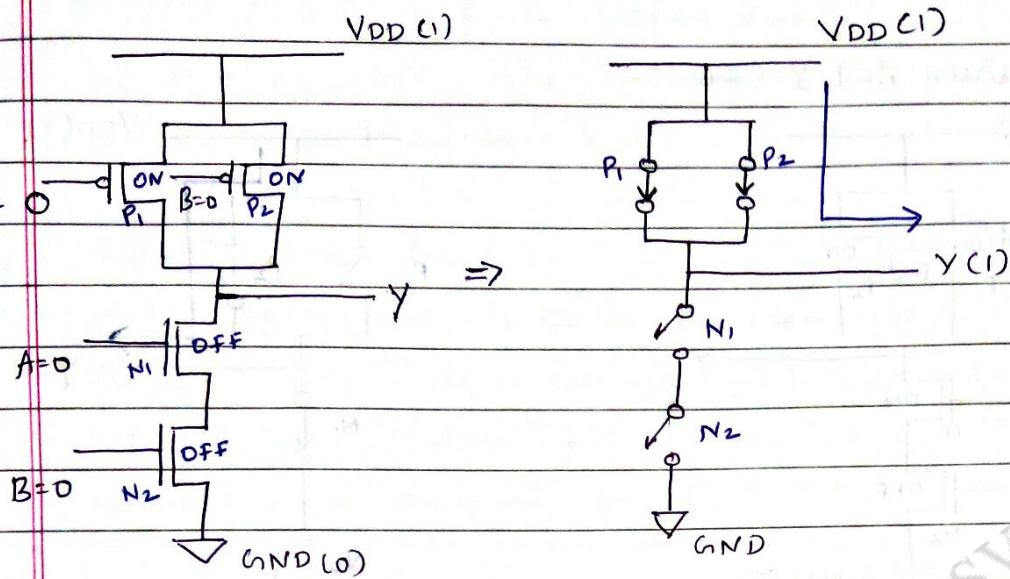
fig(i)

→ fig (i) shows a 2 input NAND gate. It consists of two series NMOS transistors between "Y" & "Gnd" and two parallel PMOS transistors between "Y" & "VDD".

Truth table :-

| A | B | Y |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

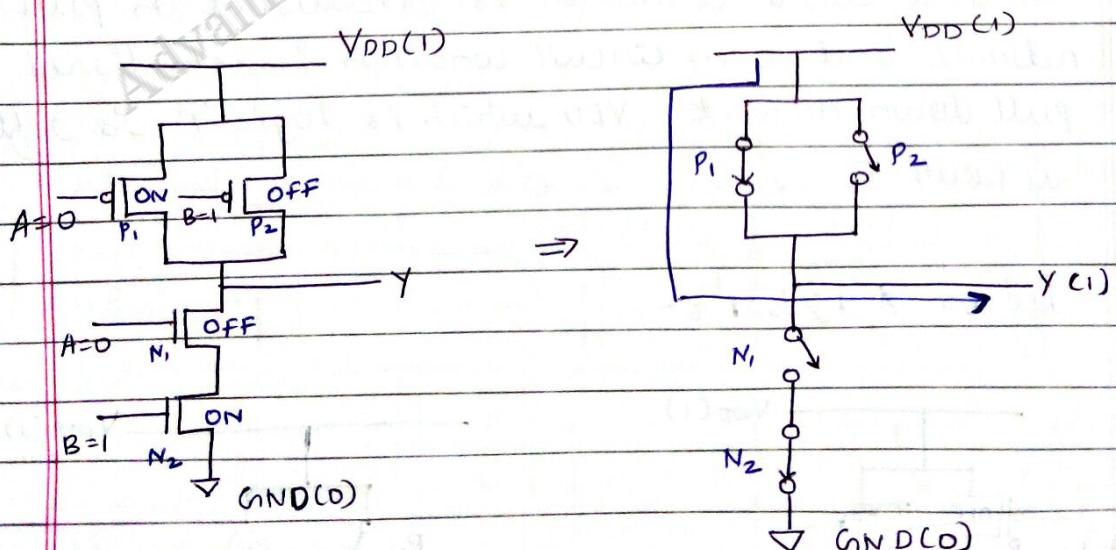
(i) When  $A=0; B=0$  :-



→ When  $A=0; B=0$ , P<sub>1</sub> & P<sub>2</sub> are ON & N<sub>1</sub> & N<sub>2</sub> are OFF  
hence P<sub>1</sub> & P<sub>2</sub> acts as closed switch & N<sub>1</sub>, N<sub>2</sub> acts as open switch.

→ Hence V<sub>DD</sub> which is logic "1" is reflected as output.

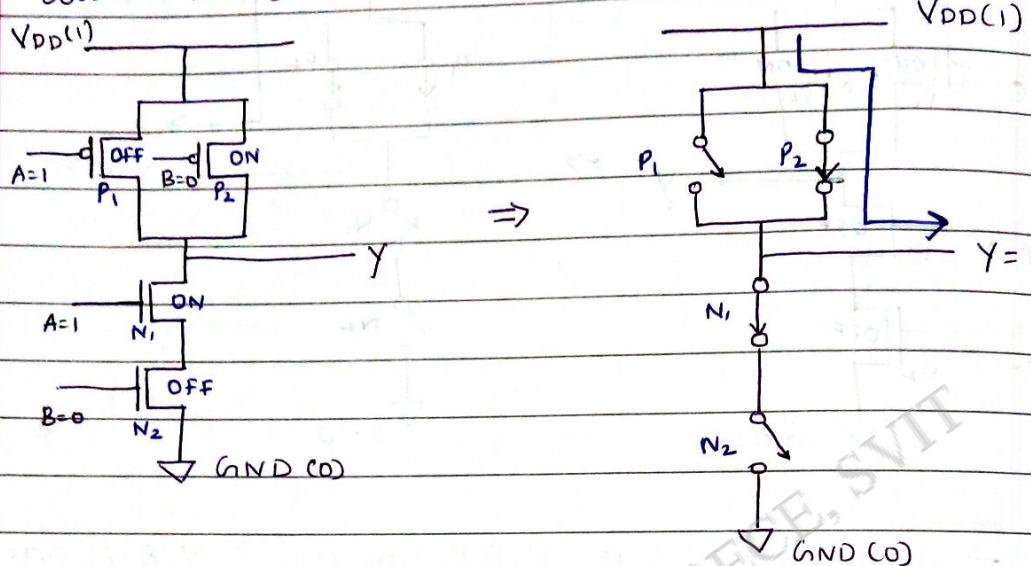
(ii) When  $A=0; B=1$  :-



→ When  $A=0; B=1$ , P<sub>1</sub> is "ON" & P<sub>2</sub> is "OFF". N<sub>1</sub> is "OFF" and N<sub>2</sub> is "ON". Therefore P<sub>1</sub> acts as closed switch, P<sub>2</sub> acts as open switch. N<sub>1</sub> acts as open switch, N<sub>2</sub> acts as closed switch.

→ Hence V<sub>DD</sub> which is at logic "1" is reflected as output "Y". ∴ "Y=1".

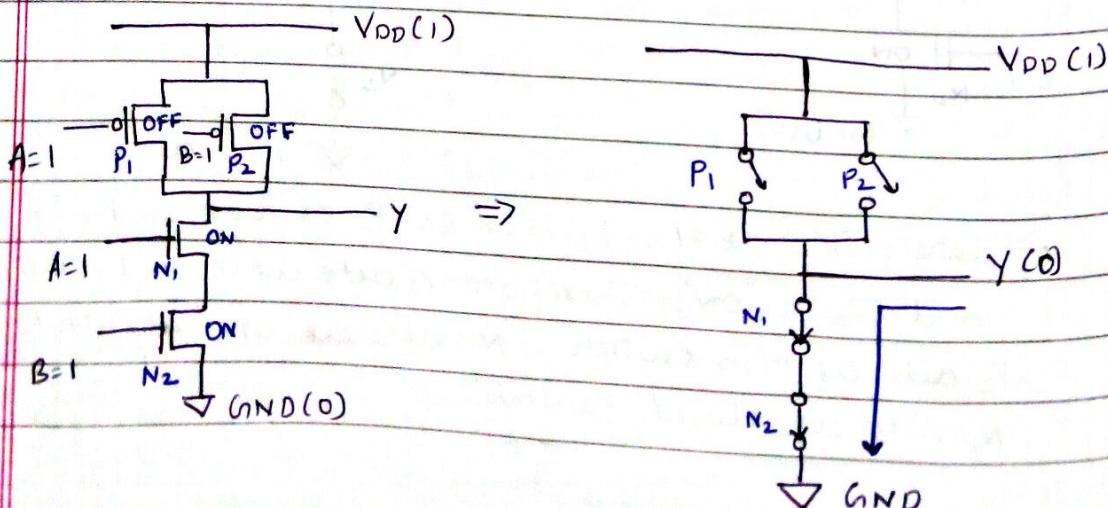
(iii) When A=1; B=0 :-



→ When A=0 & B=1, P<sub>1</sub> transistor is "OFF" and P<sub>2</sub> transistor is "ON". N<sub>1</sub> transistor is "ON" and N<sub>2</sub> transistor is "OFF". Therefore P<sub>1</sub> acts as open switch, P<sub>2</sub> acts as closed switch. N<sub>1</sub> acts as closed switch & N<sub>2</sub> acts as open switch.

→ Since a closed connection is established in pull up network and open circuit condition is established in pull down network, V<sub>DD</sub> which is logic "1" is reflected as output.

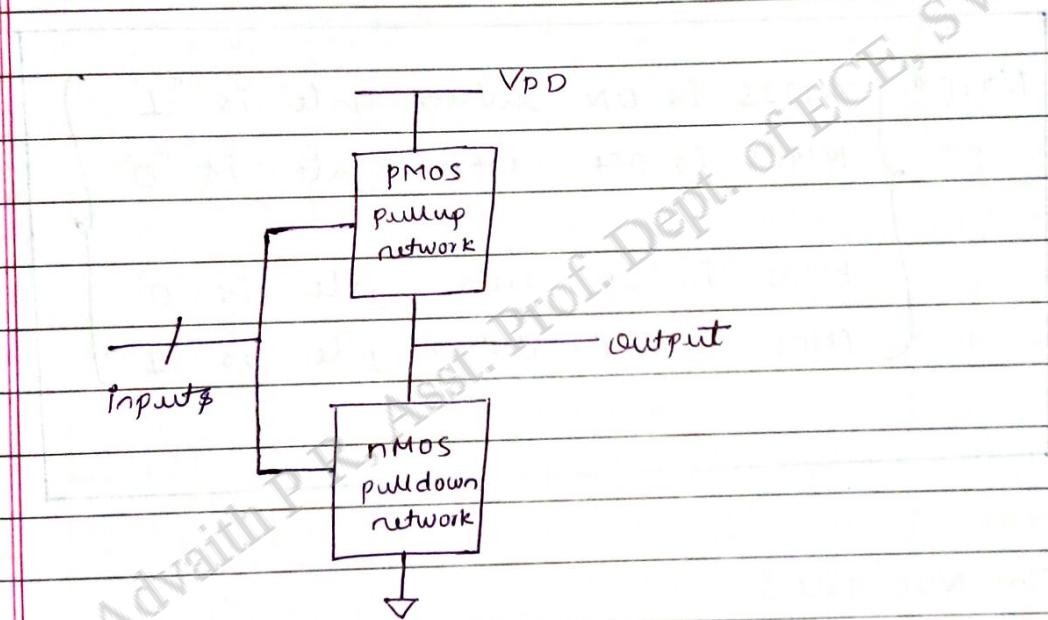
(iv) When A=1; B=1 :-



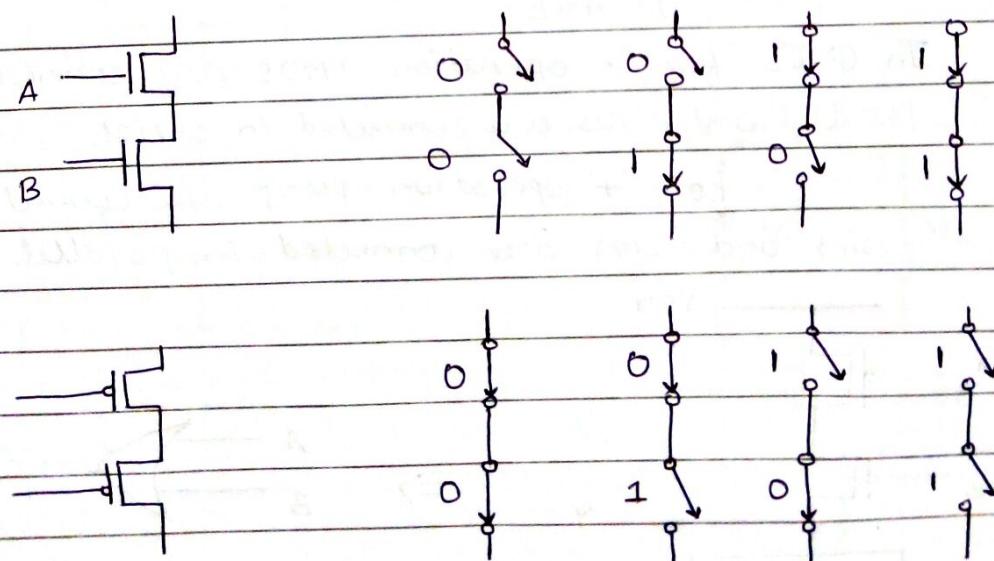
- When "A=1" & "B=1",  $P_1$  &  $P_2$  transistors are "off".  $N_1$  &  $N_2$  transistors are "ON". Therefore  $P_1$  &  $P_2$  acts as open switch,  $N_1$  &  $N_2$  acts as closed switch.
- Because of this, the output  $Y$  is pulled down to "zero" and hence " $Y=0$ ".

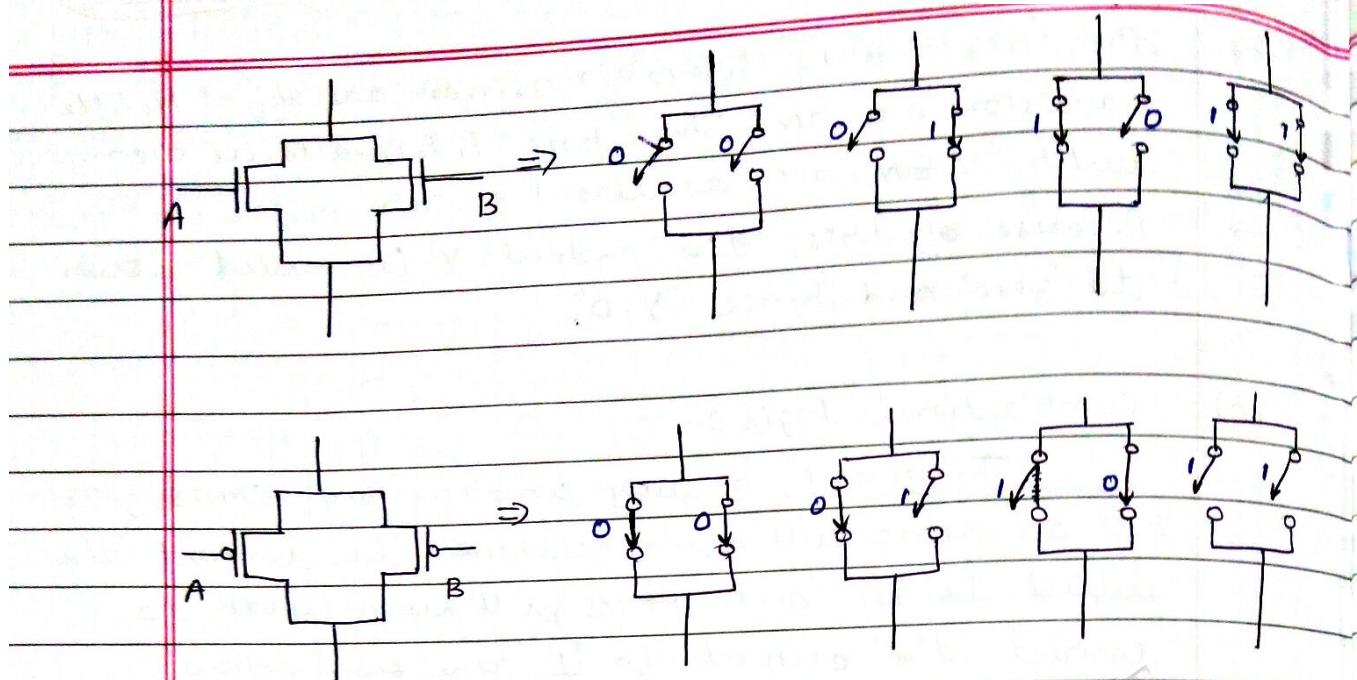
### (c) Combinational logic :-

In general, a fully complementary CMOS gate has an nMOS pull down network to connect the output to "0", and pMOS pull up network to connect the output to "1", as shown below



Example :-





**NOTE :-**

- { NMOS is ON when gate is "1"
- NMOS is OFF when gate is "0"
- { PMOS is ON when gate is "0"
- PMOS is OFF when gate is "1"

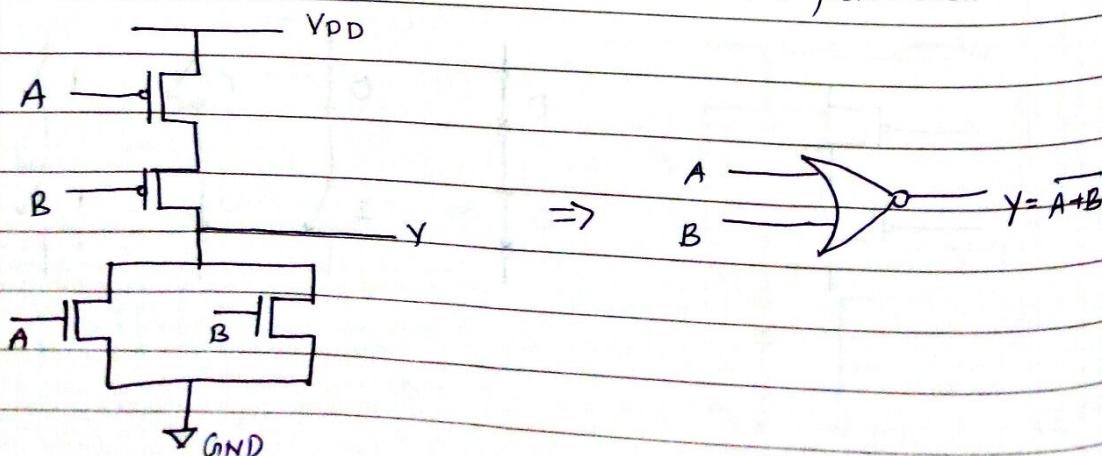
#### (d) The NOR gate :-

Boolean expression of NOR gate

$$Y = \overline{A+B}$$

In CMOS for "•" operation PMOS are connected in parallel and nMOS are connected in series.

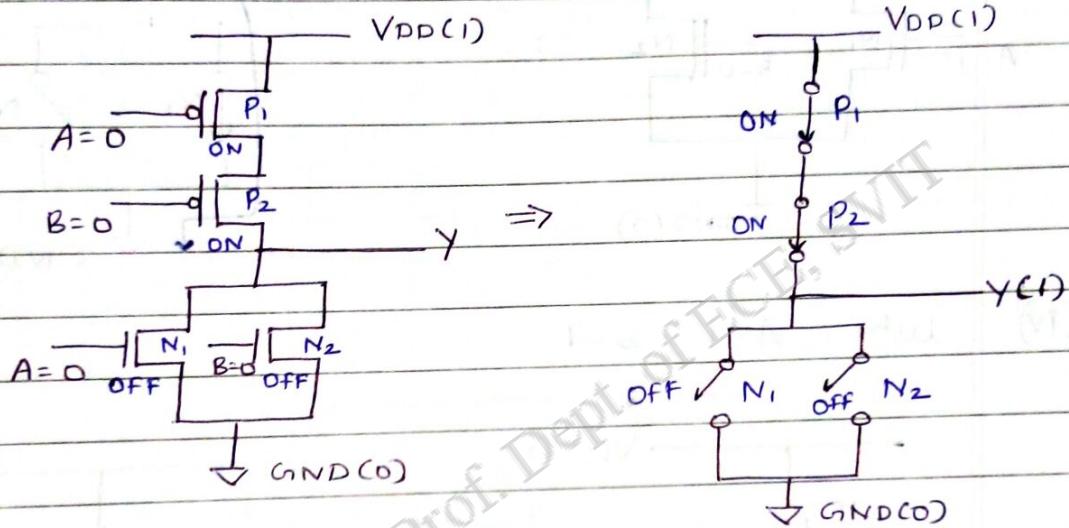
for "+" operation PMOS are connected in series and nMOS are connected in parallel.



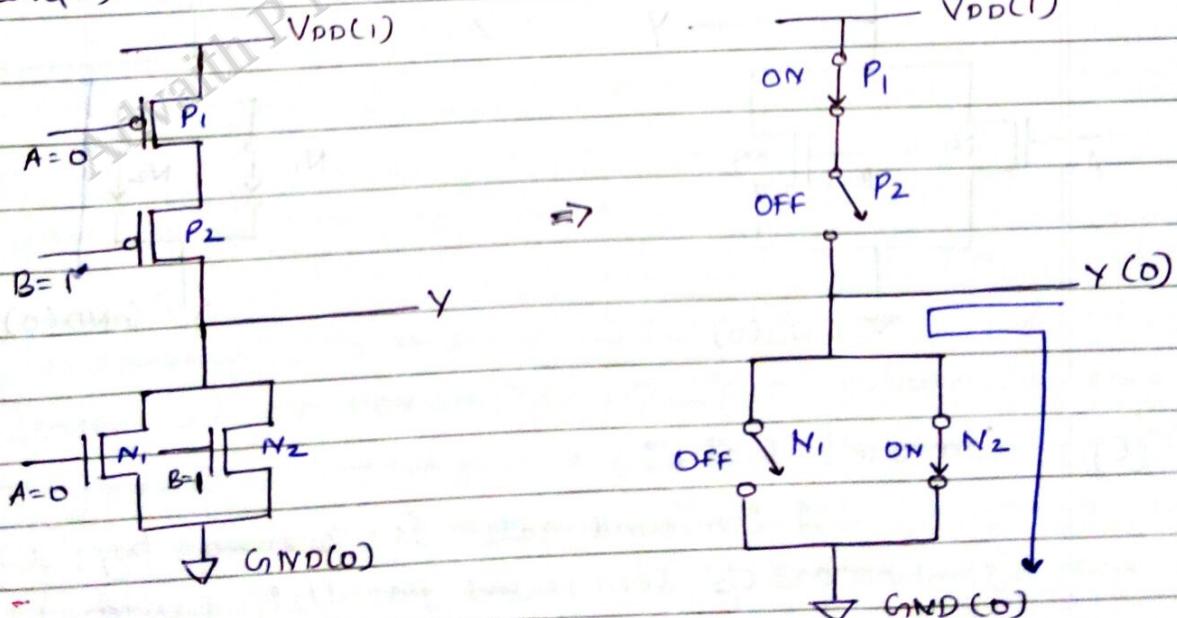
Truth table :-

| A | B | Y |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 0 |

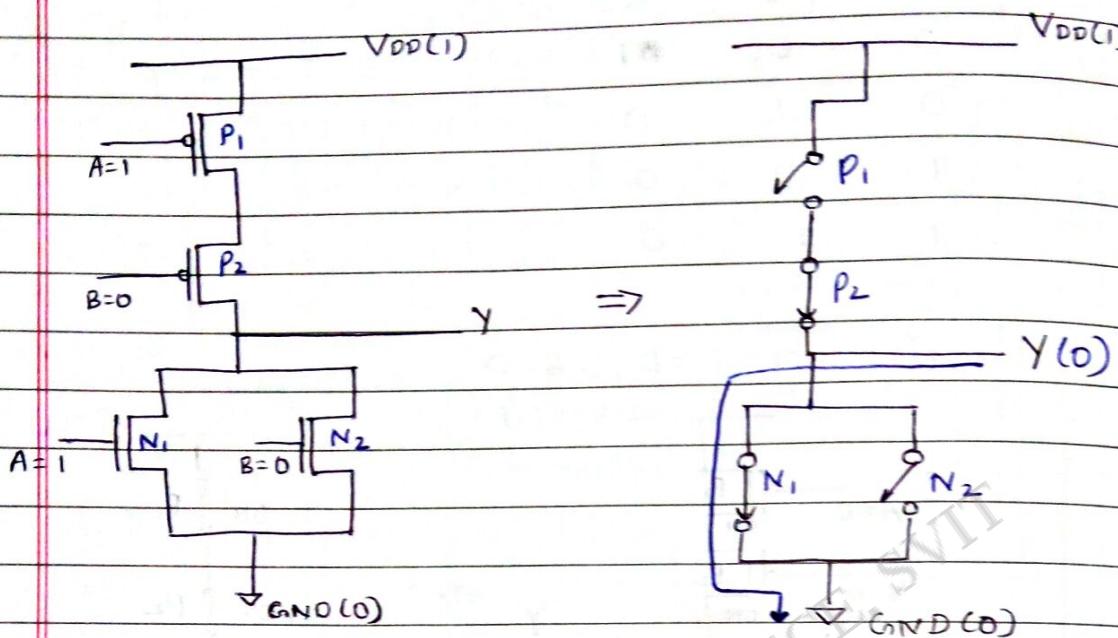
(i) Case (i) when  $A=0 ; B=0$



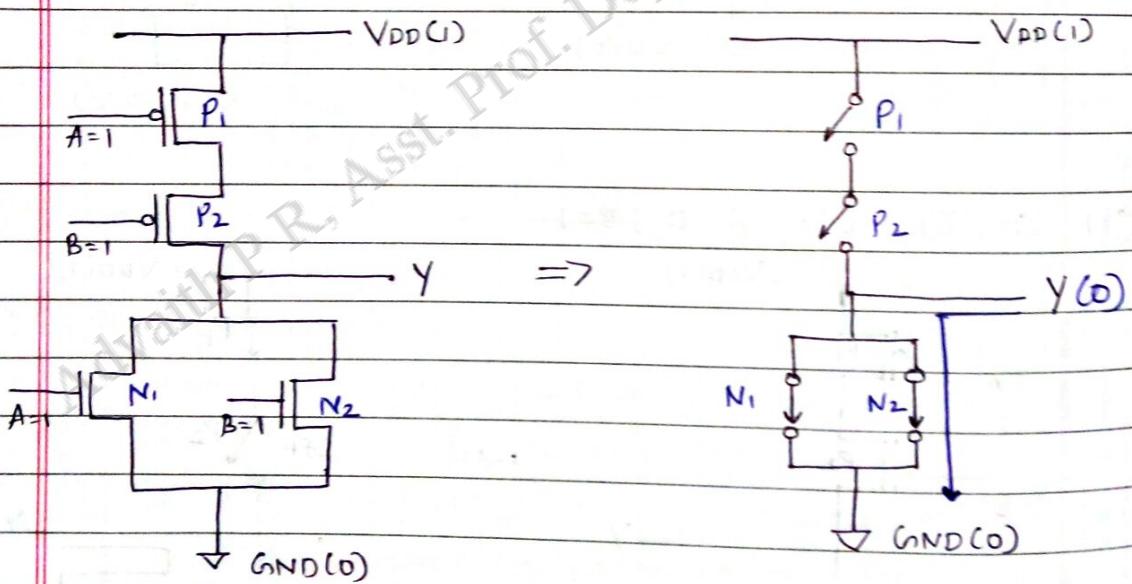
(ii) Case (ii) when  $A=0 ; B=1$



(iii) Case(iii) when  $A=1 ; B=0$



(iv) when  $A=1 ; B=1$



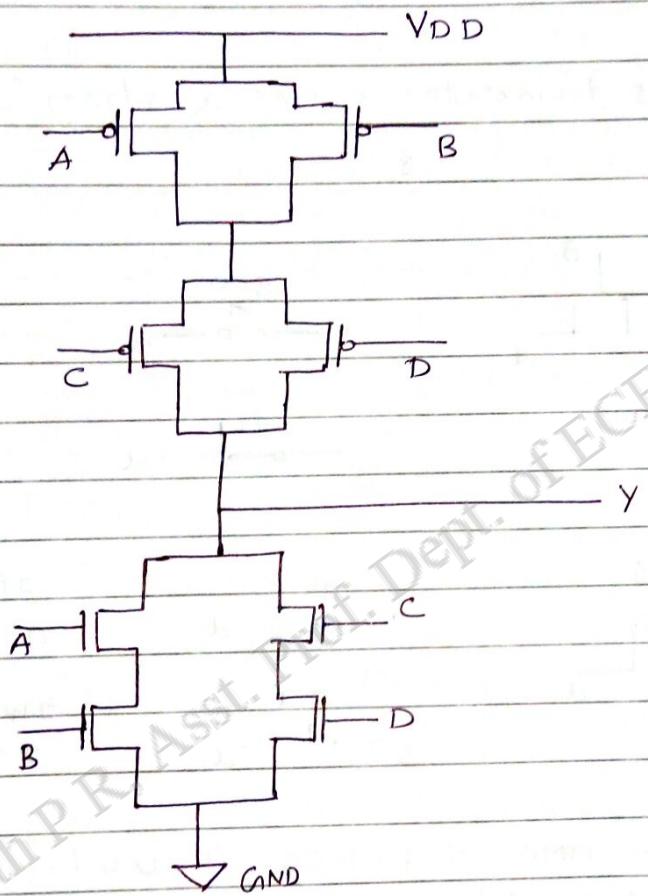
(e) Compound gates :-

A Compound gate is formed by using a combination of series and parallel switch structures.

$$\text{Ex:- } Y = \overline{(A \cdot B)} + (C \cdot D)$$

Always  $\cdot$  operation pmos are connected in parallel  
nmos are connected in series

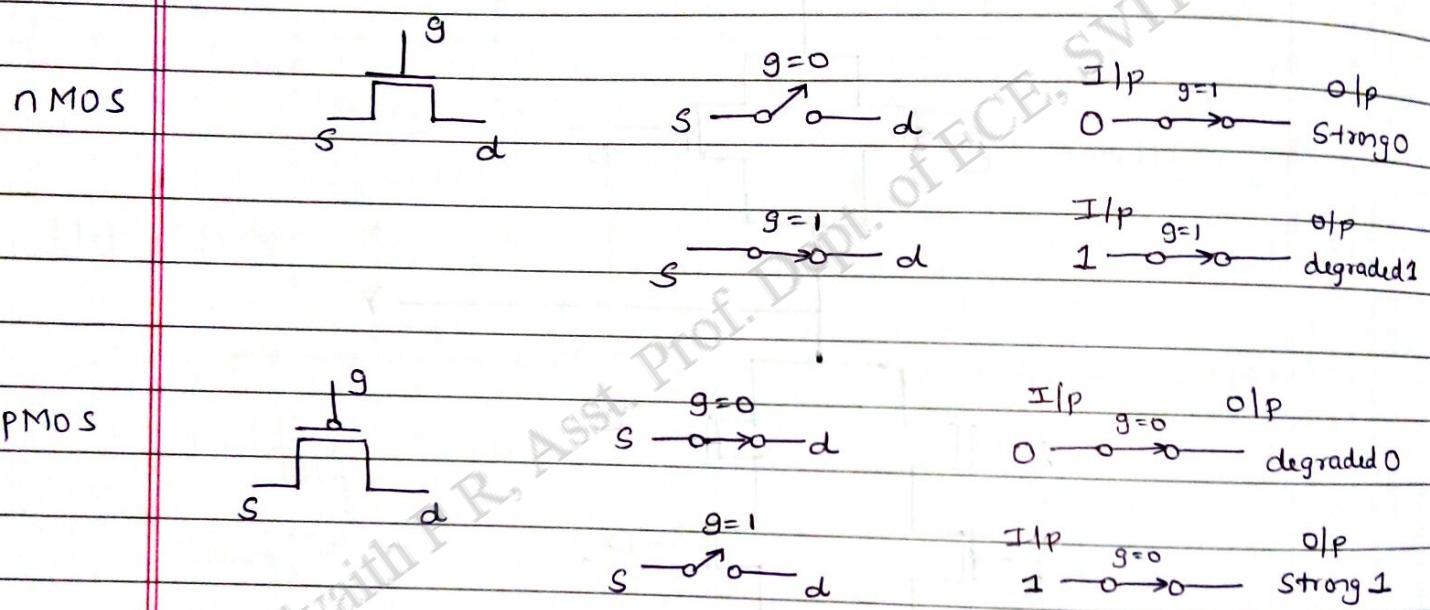
for  $(+)$  operation pmos are connected in series  
nmos are connected in parallel



### (b) Pass Transistors and Transmission Gates :-

- The strength of a signal is measured by how closely it approximates an ideal voltage source.
- In general, the stronger a signal, the more current it can source or sink.
- The power supplies, or rails ( $V_{DD}$  &  $GND$ ) are the source of the strongest "1's" and "0's".

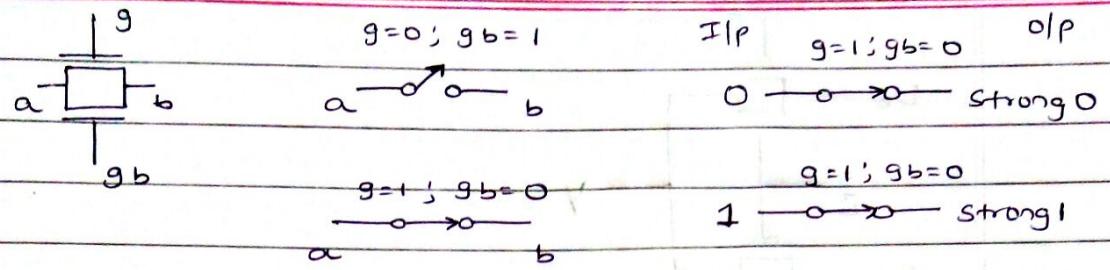
- An nMOS transistor is almost perfect switch when passing a "0" and thus we say it passes a strong "0".
- However, the nMOS transistor is imperfect at passing a "1". The high voltage level is somewhat less than  $V_{DD}$ . We say it passes a "degraded" or "weak" 1.
- A pMOS transistor passes a strong "1" but degraded "0".



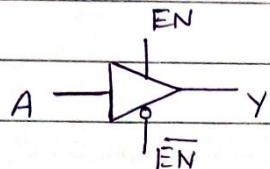
- When an nMOS and pMOS is used alone as an imperfect switch, we call it as "pass transistor".

### (9) Transmission gate :-

By combining an nMOS and pMOS transistor in parallel, we obtain a switch which turns on when "1" is applied to "g" in which 0's and 1's are both passed in an acceptable fashion. This combination of parallel nMOS and pMOS is called as transmission gate or pass gate.



Tristates :-

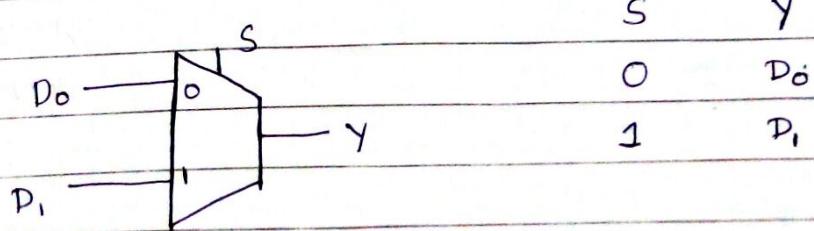


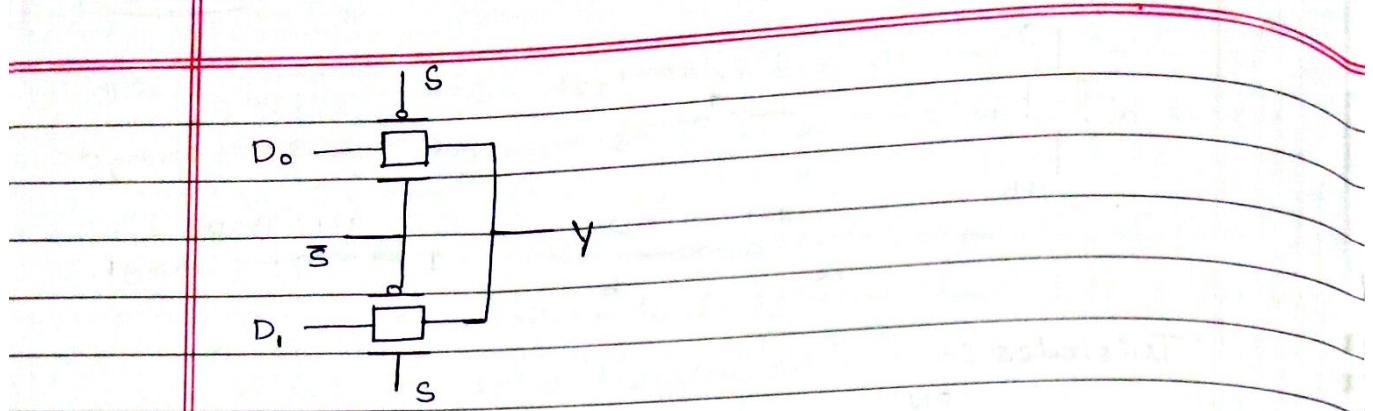
- When the Enable input "EN" is "1", the output  $y$  equals the input  $A$ .
- When enable is zero,  $y$  is left floating ("Z" value).

| EN | A | $y$ |
|----|---|-----|
| 0  | 0 | Z   |
| 0  | 1 | Z   |
| 1  | 0 | 0   |
| 1  | 1 | 1   |

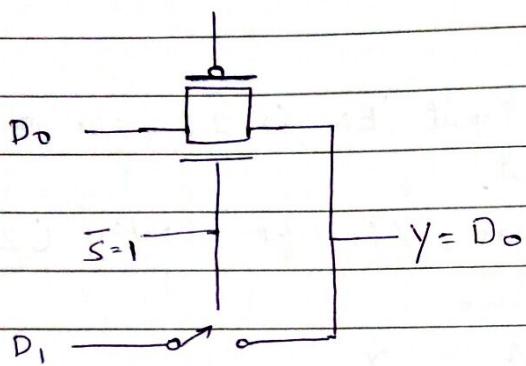
(h) Multiplexers :-

- Multiplexers are key components in CMOS memory elements and data manipulation structures.
- A multiplexer chooses the output to be one of several inputs based on select signal.
- A two input or 2:1 multiplexer, chooses input  $D_0$  when the select is "0" and input  $D_1$  when the select is "1".





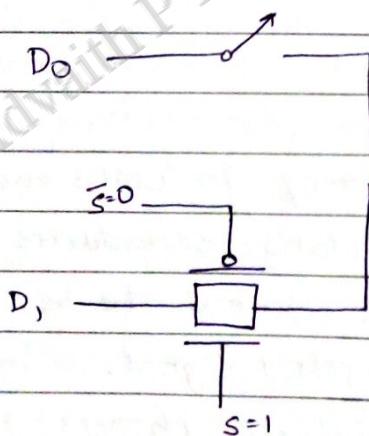
When  $S = 0$ ,



$\rightarrow$  When  $S = 0$ ,  $\bar{S} = 1$  hence only the transistor linked with  $D_o$  turns ON and the input  $D_o$  will be passed on to output.

$\rightarrow$  The transistors connected to  $D_i$  are OFF.

When  $S = 1$ ,



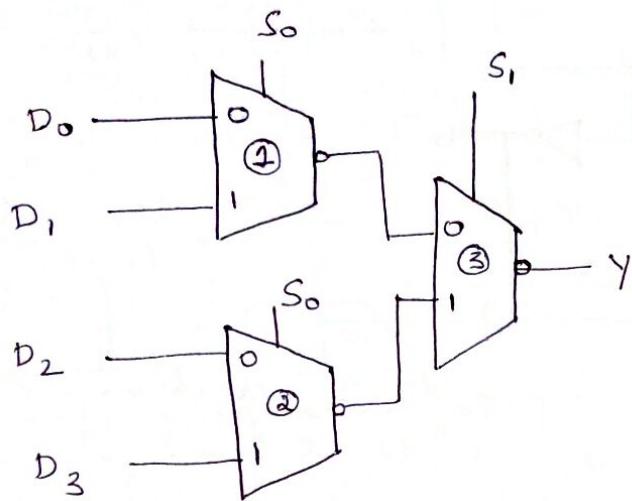
$\rightarrow$  When  $S = 1$ ,  $\bar{S} = 0$  hence only the transistors linked with  $D_i$  turns ON and the input  $D_i$  will be passed on to output.

$\rightarrow$  The transistors connected to  $D_o$  are OFF.

Advaith PR  
Assistant Professor  
Dept. of ECE  
SVIT

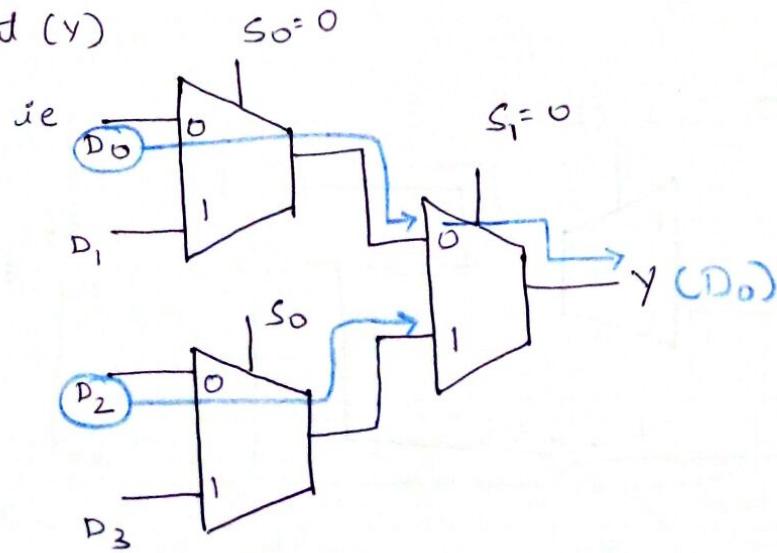
Larger multiplexers can be built from multiple 2 input multiplexers.

Eg:- 4:1 multiplexer can be built by using three 2to1 multiplexers.



| $S_1$ | $S_0$ | $Y$   |
|-------|-------|-------|
| 0     | 0     | $D_0$ |
| 0     | 1     | $D_1$ |
| 1     | 0     | $D_2$ |
| 1     | 1     | $D_3$ |

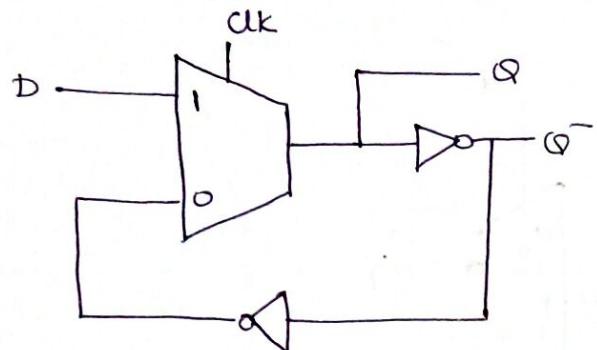
(i) when  $S_0=0$  &  $S_1=0$ ,  $D_0$  will be selected from mux "1" and  $D_2$  will be selected from mux "2". Since "S" is 0, the top input will be selected ie  $D_0$  will be given as output (Y)



Similarly, based on select lines, the output is given with  $D_1$ ,  $D_2$  &  $D_3$  as given in truth table.

### Hatches and flip flops :-

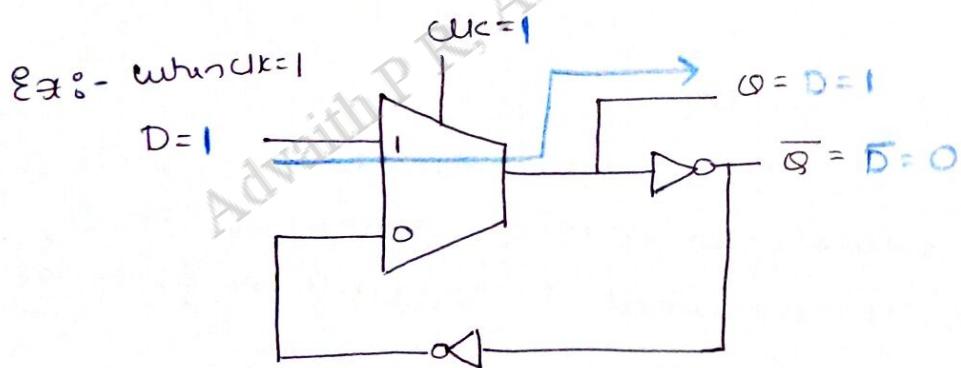
A D latch can be built by using one 2 input multiplexer and two inverters as shown below.



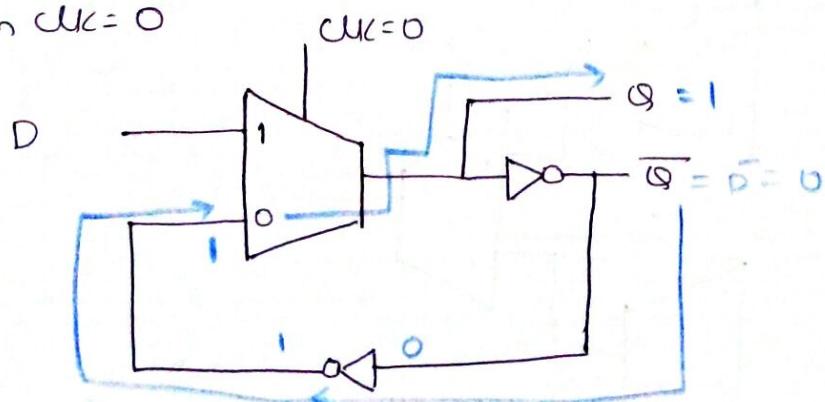
→ It consists of a data input,  $D$ , a clock input  $clk$  and true & complementary outputs  $Q$  &  $\bar{Q}$ .

→ When  $clk=1$ ,  $Q=D$  &  $\bar{Q}=\bar{D}$ .

→ When  $clk=0$ , a feedback path around the inverter pair is established to hold the current state of  $Q$ .

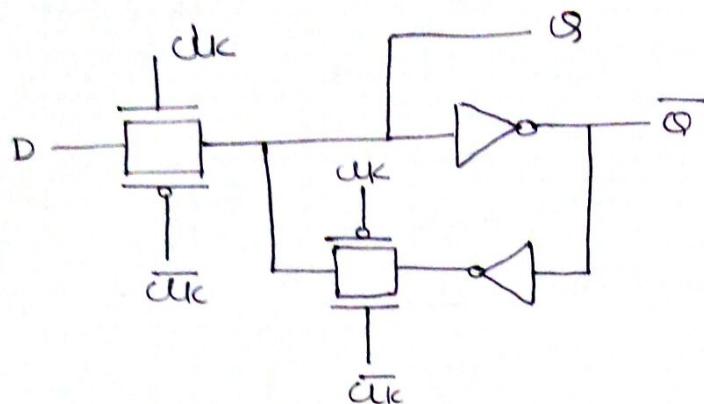


when  $clk=0$

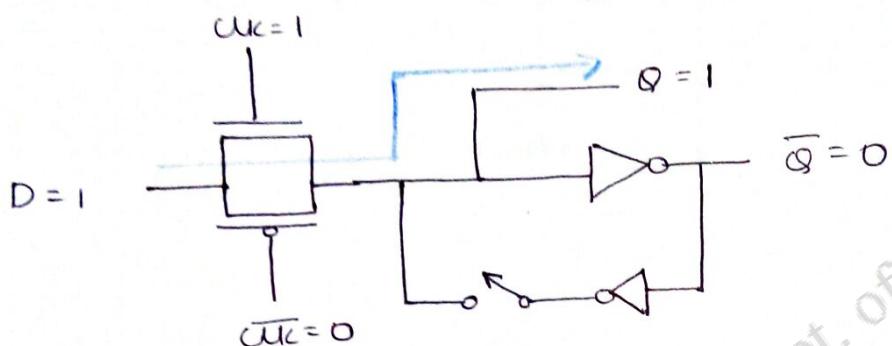


i.e when  
 $clk=0$ ,  
previous state  
is maintained

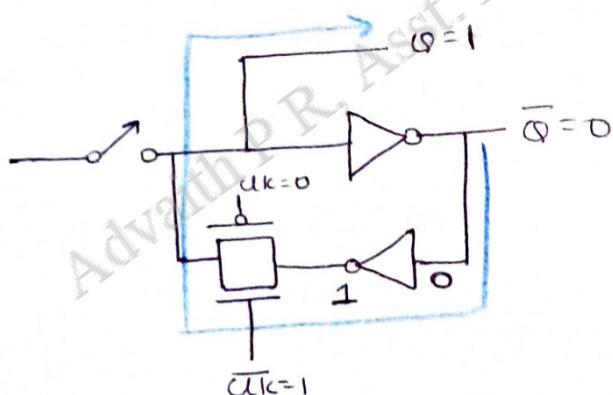
We can construct the multiplexers using transmission gates as shown below.



Ex:- When  $\text{clk}=1$  &  $D=1$ , then

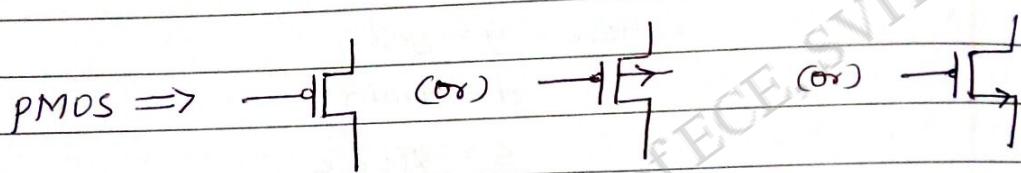
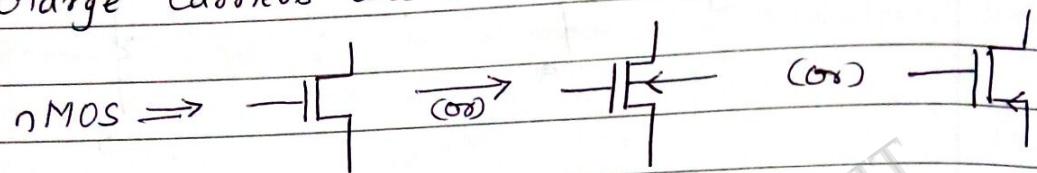


When  $\text{clk}=0$ , then



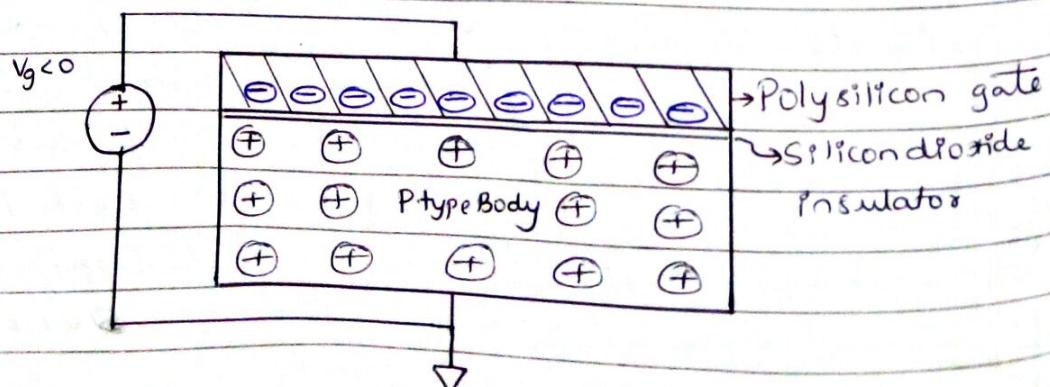
## MOS Transistor Theory :-

MOS transistor is a majority - carrier device in which the current in a conducting channel between the source and drain is controlled by voltage applied. In an nMOS transistor, the majority charge carriers are electrons and in pMOS transistors, the majority charge carriers are holes.



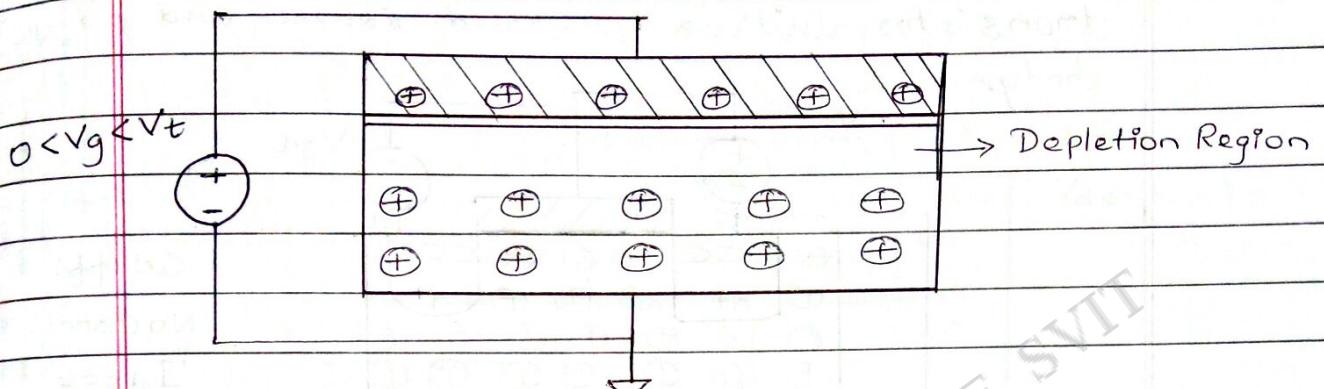
The behaviour of the MOS transistor can be understood by first examining an isolated MOS structure with a gate and body but no source and drain.

The top layer of the MOS structure is a good conductor called the gate. Normally gate of MOS transistors is made of metal or polysilicon. The middle layer is a very thin insulating film of  $\text{SiO}_2$  called the gate oxide. The bottom layer is the doped silicon body (substrate or bulk).



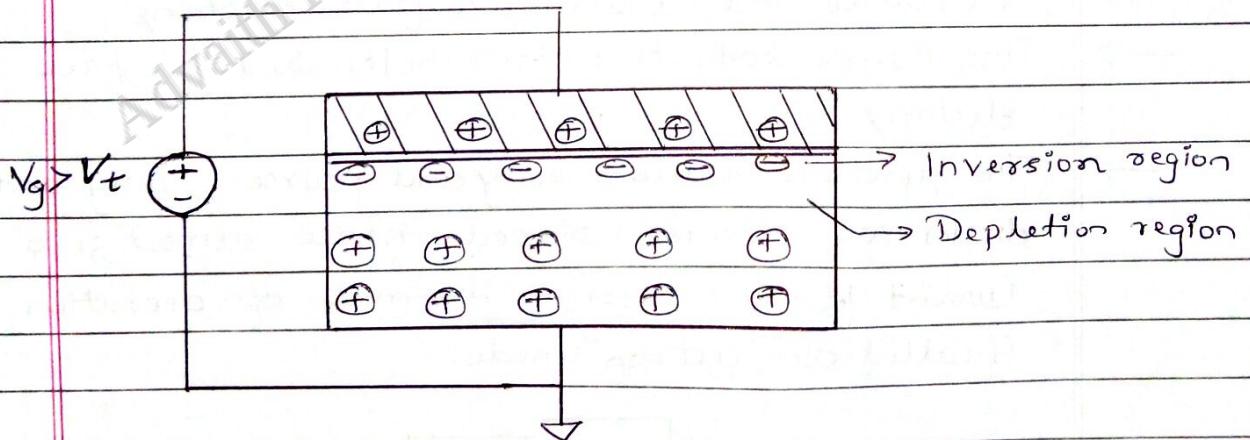
(a) MOS structure demonstrating Accumulation

A negative Voltage is applied to the gate, so there is negative charge on the gate. Positively charged holes are attracted to the region beneath the gate. This is called as "Accumulation".



(b) MOS structure demonstrating Depletion.

When a low positive voltage is applied to the gate, resulting in some positive charge on the gate. The holes in the body are repelled from the region directly beneath the gate, resulting in a Depletion region forming below the gate.

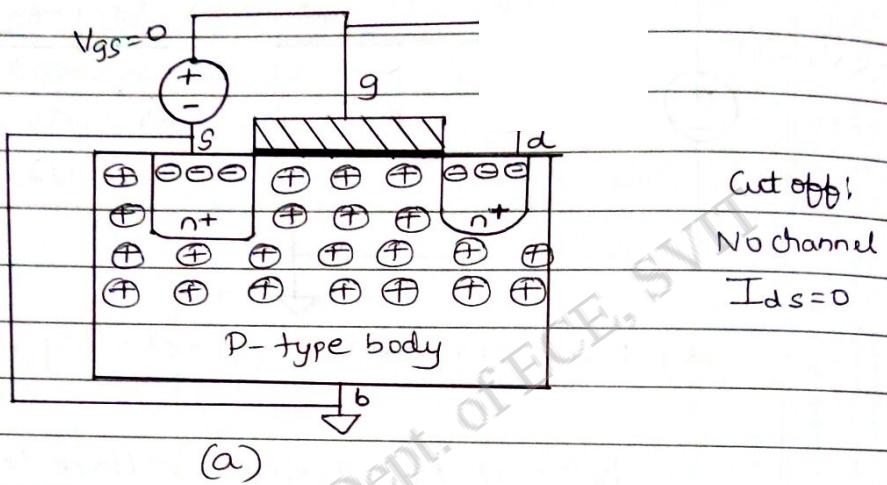


(c) MOS structure demonstrating inversion

A higher positive potential exceeding a critical threshold voltage  $V_t$  is applied, attracting more positive charge to the gate. The holes are repelled further and a small

Number of free electrons in the body are attracted to the region beneath the gate. This conductive layer of electrons in the p type body is called inversion layer.

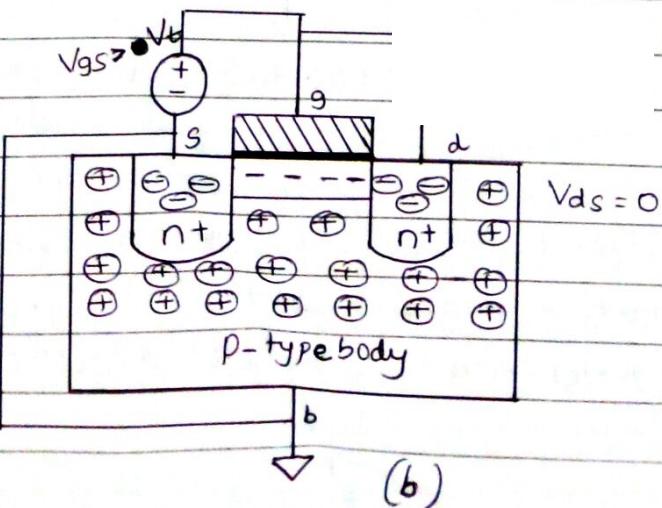
The below figure shows an NMOS transistor with a grounded source and p type body.



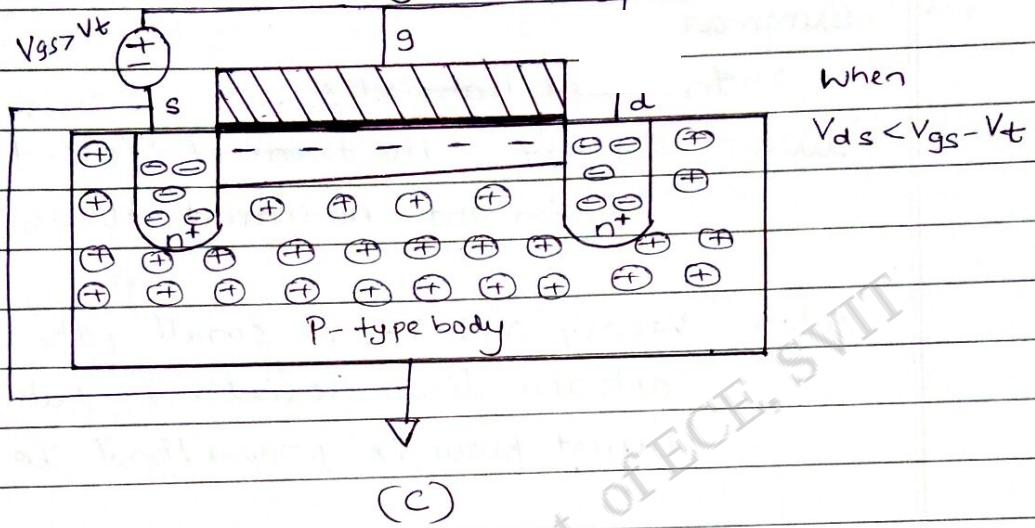
The transistor consists of MOS stack between two n type regions called source and drain.

In the above figure, the gate to source Voltage  $V_{GS}$  is less than the threshold voltage.

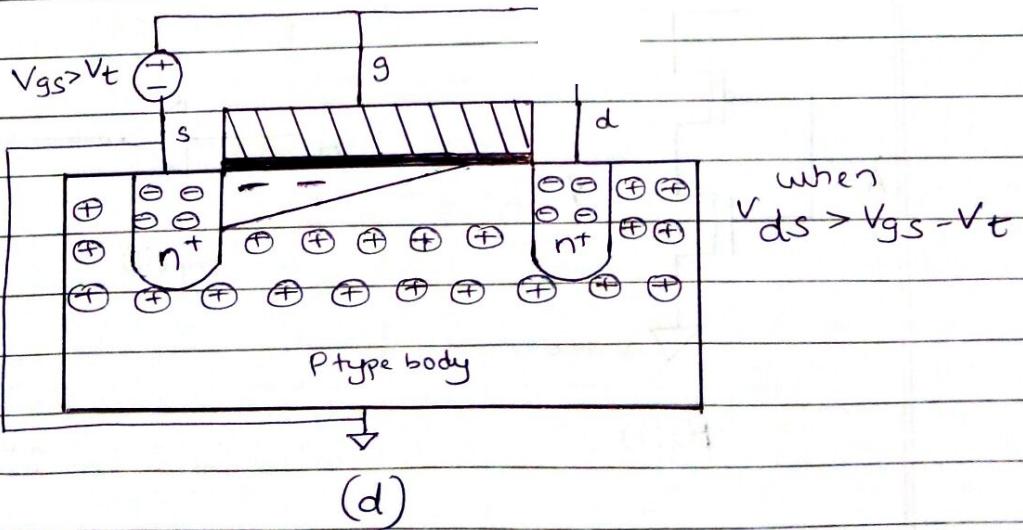
- The source and drain have free electrons.
- The P type body has free holes but no free electrons.
- The junctions between body and source, body and drain are reverse biased, hence almost zero current flows ( $I_{DS} = 0$ ). This mode of operation is called as "cut off" mode.



- In figure b, the gate voltage is greater than threshold voltage ( $V_t$ ). Now an inversion region of electrons called the channel connects the source and drain, creating a conductive path.
- The number of carriers and the conductivity increases with the gate voltage.



- From figure (c), when a small positive positive voltage  $V_{ds}$  is applied to the drain, current  $I_{ds}$  flows through channel from drain to source. This mode of operation is termed as "linear region" or "resistive region" or "non saturated region" or "unsaturated region".



- When  $V_{ds}$  becomes sufficiently large, the channel is no longer inverted near drain and becomes "pinched off".
- This state of mosfet is termed as "saturation" because " $I_{ds}$ " becomes saturated.

Summary:-

In nmos transistor,  
when  $V_{gs} < V_t$ , the transistor is in the "cut off" region and no current flows

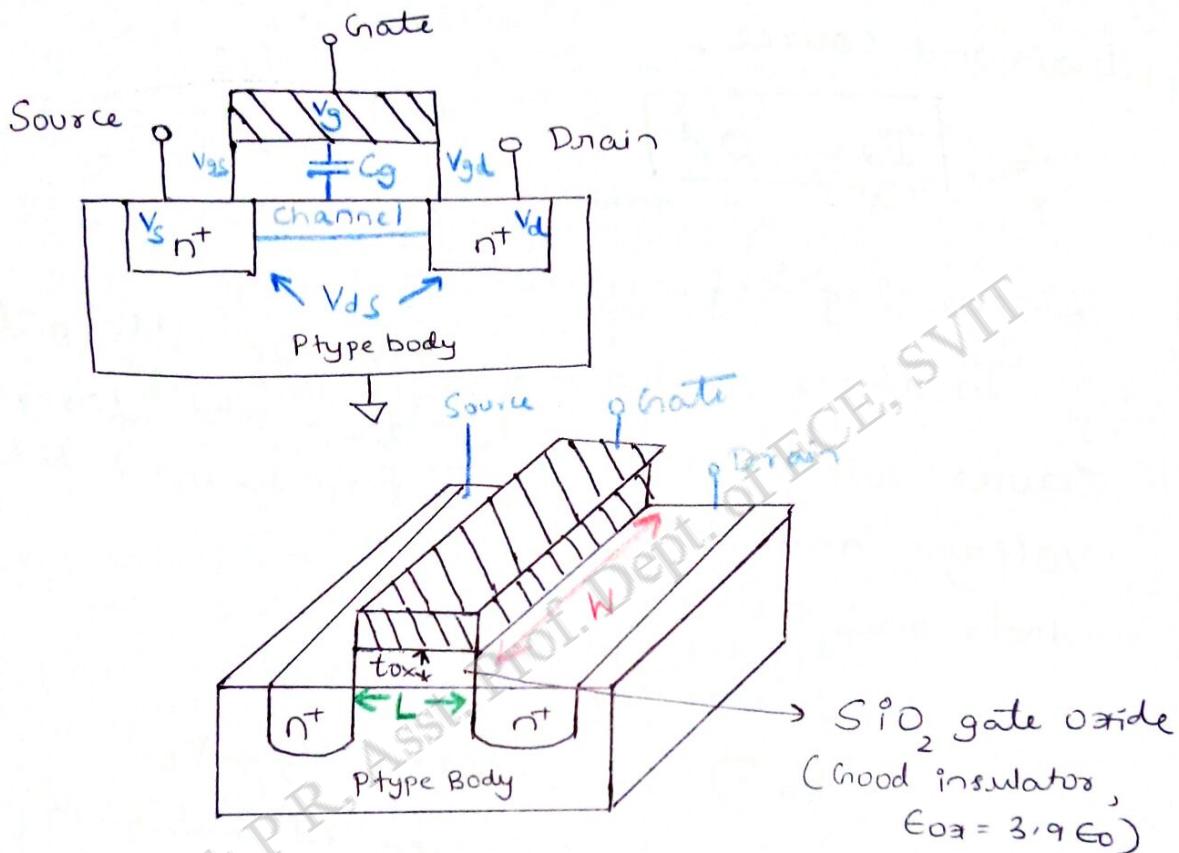
when  $V_{gs} > V_t$  and  $V_{ds}$  is small, the transistor acts as a linear resistor in which the current flow is proportional to  $V_{ds}$ .

when  $V_{gs} > V_t$  and  $V_{ds}$  is large, the transistor goes into "saturation" and current becomes independent of  $V_{ds}$ .

## Ideal I-V characteristics :-

MOS transistors have three regions of operation

- Cut off region
- Linear region
- Saturation region



- The distance between source and drain is called as channel length ( $L$ ).
- Width of gate is ( $W$ ).
- $t_{ox}$  = Oxide thickness
- $\epsilon_{ox}$  = Oxide permittivity ( $3.9 \epsilon_0$ )

Let us derive  $I_{DS}$  for three regions, namely cut off region, linear region and saturation region.

### (a) Cut off region :-

In cut off region,  $V_{gs} < V_t$  ie the gate to source voltage is lesser than the threshold voltage.

Hence there is no formation of channel between drain and source.

$$\therefore \boxed{I_{ds} = 0A}$$

### (b) Linear Region :-

In linear region,  $V_{gs} > V_t$  ie the gate to source voltage is greater than the threshold voltage and hence channel is formed between drain and source.

In linear region

$$V_{gs} > V_t \text{ & } V_{ds} < V_{gs} - V_t$$

As soon as voltage is applied between drain & source, the electrons starts to flow and sets up current  $I_{ds}$ .

$$\text{WKT } I = \frac{Q}{t}$$

$$\text{ie } I_{ds} = \frac{Q}{t_f} \longrightarrow ①$$

where  $Q$  = charge

$t_f$  = transit time which is the time taken by the electron to reach drain from source.

We all know that

$$\text{time} = \frac{\text{Distance}}{\text{Velocity}}$$

$$\text{i.e } t_f = \frac{L}{V} \longrightarrow \textcircled{2}$$

where  $L$  = channel length

$$V = \text{Velocity}$$

$$\text{where } V = \mu E \longrightarrow \textcircled{3}$$

where  $\mu$  = mobility of electrons

$E$  = Electric field

By  $E = \frac{\text{Voltage}}{\text{distance}}$

$$E = \frac{V_{ds}}{L} \longrightarrow \textcircled{4}$$

Substituting Equation  $\textcircled{4}$  in  $\textcircled{3}$ , we get

$$V = \mu \left[ \frac{V_{ds}}{L} \right] \longrightarrow \textcircled{5}$$

Substituting Equation  $\textcircled{5}$  in Equation  $\textcircled{2}$ , we get

$$t_f = \frac{L}{\mu \left[ \frac{V_{ds}}{L} \right]}$$

$$t_f = \frac{L^2}{\mu V_{ds}} \longrightarrow \textcircled{6}$$

To find  $I_{ds}$ , we need to know the value of  $Q$  as well since  $I_{ds} = \frac{Q}{t_f}$

To find  $Q$ ,

$$\text{we know that } Q = C_g V \longrightarrow ⑦$$

Voltage here is the difference between gate channel voltage ( $V_{gc}$ ) and threshold voltage ( $V_t$ ).

$$\text{where } V_{gc} = \frac{V_{gs} + V_{gd}}{2}$$

$$0^\circ \quad V = V_{gc} - V_t$$

$$V = \left( \frac{V_{gs} + V_{gd}}{2} \right) - V_t$$

$$\text{since } V_{ds} = V_{gs} - V_{gd}, \quad || \text{ ie } V_{gd} = V_{gs} - V_{ds} ||$$

We can write

$$V = \frac{V_{gs} + V_{gs} - V_{ds} - V_t}{2}$$

$$V = \frac{2V_{gs}}{2} - \frac{V_{ds}}{2} - V_t$$

$$V = (V_{gs} - V_t) - \frac{V_{ds}}{2} \longrightarrow ⑧$$

$$\text{we all know that } C_g = \frac{\epsilon A}{d}$$

where  $A$  = Area of gate

$\epsilon$  = permittivity

$d$  = distance between two plates ie {distance between gate & body}

$$0^\circ \quad C_g = \frac{\epsilon_{ox}(L \cdot W)}{t_{ox}}$$

$$C_g = \frac{\epsilon_{ox}}{t_{ox}} \cdot W \cdot L \quad || \quad C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} ||$$

$$C_g = C_{ox} \cdot W \cdot L \longrightarrow ⑨$$

Substituting Equn ⑧ & ⑨ in ⑦, we get

$$Q = C_{ox} WL \cdot (V_{gs} - V_t) - \frac{V_{ds}}{2}$$

$$Q = C_{ox} WL \left\{ (V_{gs} - V_t) - \frac{V_{ds}}{2} \right\} \longrightarrow ⑩$$

Substituting Equation ⑩ and Equn ⑥ in Equn ①, we get

$$I_{ds} = \frac{Q}{t_b}$$

i.e

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

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

since  $C_{ox}$ ,  $W$ ,  $L$ ,  $\mu$  are constants, we can take it as  $\beta$

$$\therefore I_{ds} = \beta V_{ds} \left\{ (V_{gs} - V_t) - \frac{V_{ds}}{2} \right\} \longrightarrow ⑪$$

(\*)

$$I_{ds} = \beta \left[ (V_{gs} - V_t) V_{ds} - \frac{V_{ds}^2}{2} \right]$$

(c) Saturation region:-

In saturation region

$$V_{GS} > V_T \quad \& \quad V_{DS} \geq V_{GS} - V_T$$

Substituting  $V_{DS} = V_{GS} - V_T$  in Eqn ⑪,

$$I_{DS} = \beta \left[ (V_{GS} - V_T)(V_{GS} - V_T) - \frac{(V_{GS} - V_T)^2}{2} \right]$$

$$I_{DS} = \beta \left[ (V_{GS} - V_T)^2 - \frac{(V_{GS} - V_T)^2}{2} \right]$$

$$I_{DS} = \beta \left[ \frac{2(V_{GS} - V_T)^2 - (V_{GS} - V_T)^2}{2} \right]$$

$I_{DS} = \frac{\beta}{2} [(V_{GS} - V_T)^2]$

→ ⑫

Observing at Equation ⑫,  $I_{DS}$  is independent on  $V_{DS}$

i.e for nmos,

$$I_{DS} = \begin{cases} 0 & ; V_{GS} < V_T ; \text{ cut off} \\ \beta \left[ (V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2} \right] & ; V_{GS} > V_T \\ & \quad V_{DS} < V_{GS} - V_T ; \text{ linear} \\ \frac{\beta}{2} [(V_{GS} - V_T)^2] & ; V_{GS} > V_T \\ & \quad V_{DS} > V_{GS} - V_T ; \text{ Saturation} \end{cases}$$

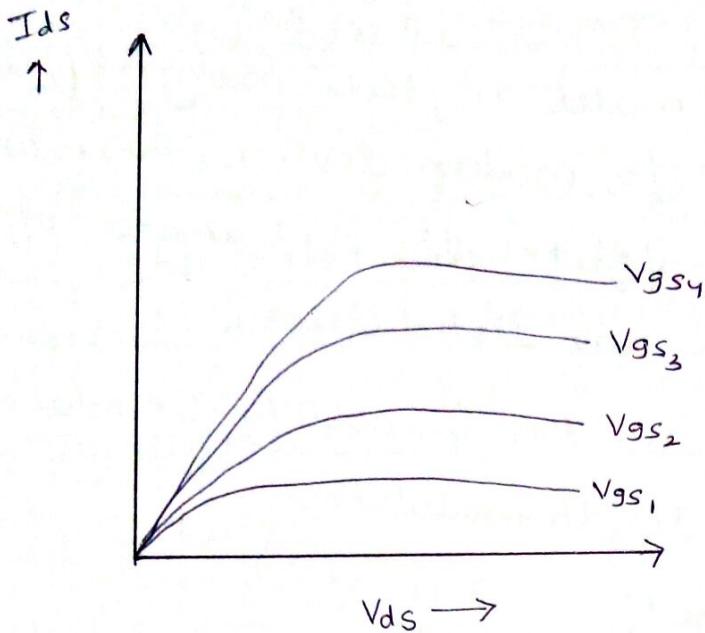
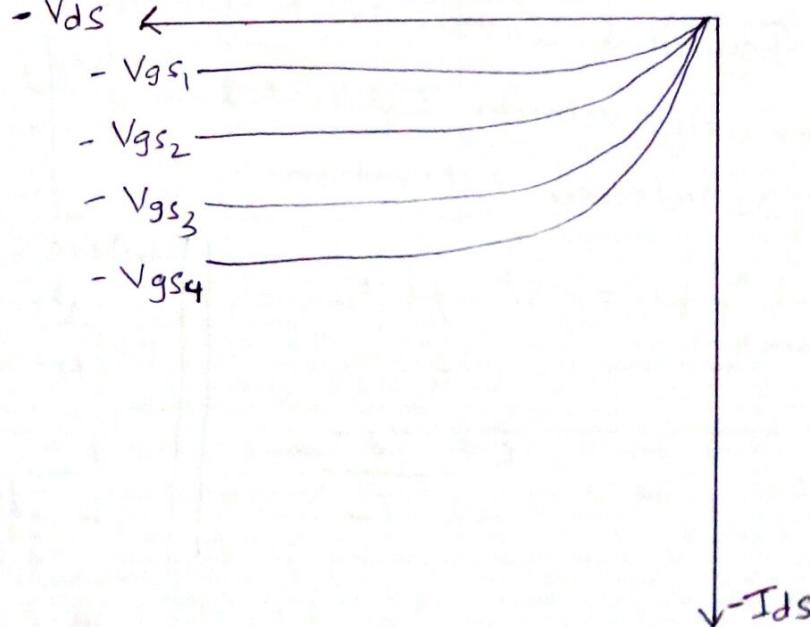


fig:- I-V characteristics of ideal nmos transistor

According to the figure, current is zero for gate voltage below  $V_t$ . For higher gate voltages, current increases linearly with  $V_{ds}$ .

As  $V_{ds}$  reaches the saturation point ( $V_{ds} = V_{gs} - V_t$ ), current rolls off and eventually becomes independent of  $V_{ds}$  when the transistor is saturated.

PMOS transistors behave in much the same way, but the signs are reversed and I-V characteristics is in the third quadrant.



## Non Ideal I-V characteristics :- { non ideal I-V Effects }

The ideal I-V model neglects many effects that are important to modern devices. There are certain effects that affects the performance of the device, some of them are listed below.

- (i) Velocity saturation & mobility degradation
- (ii) Channel length modulation
- (iii) Body effect
- (iv) Subthreshold conduction
- (v) Junction leakage
- (vi) Tunneling
- (vii) Temperature dependence
- (viii) Geometry dependence

### (i) Velocity saturation & mobility degradation :-

Under strong lateral electric fields produced by  $V_{DS}$ , the charge carriers are accelerated and the collisions between the charge carriers and the silicon lattice increases.

Beyond a certain level of electric field, the carrier drift velocity saturates at  $V_{sat}$ . This is called as velocity saturation.

$$\text{WKT } V = \mu E$$

&

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

where

$V$  = Velocity

$\mu$  = Mobility

$E$  = Electric field

$L$  = length of the channel.

With the decrease in the length of the channel, the effective electric field for the same  $V_{ds}$  increases.

With increase in the strength of the lateral electric field, the charge carriers of the lateral electric field short channel device becomes velocity saturated.

Mobility degradation :-

WKT

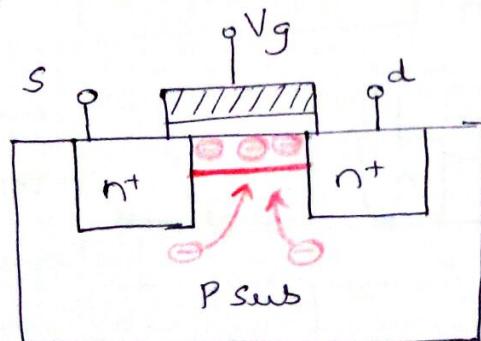
$$V = \mu E$$

where  $V$  = velocity

$\mu$  = Mobility

$E$  = Electric field

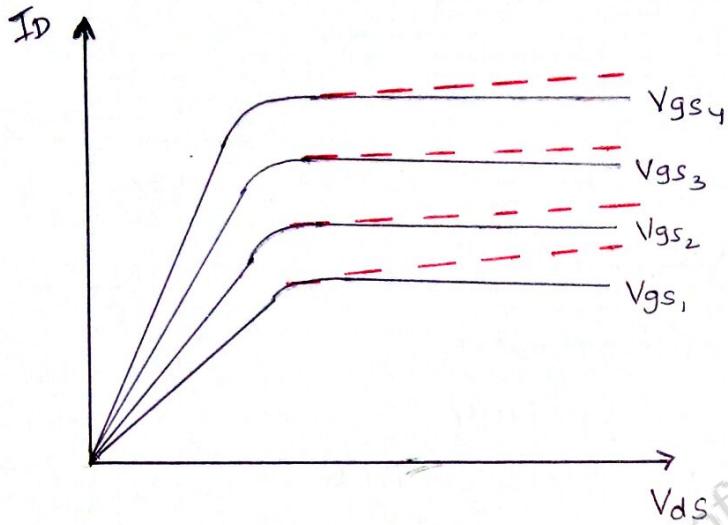
→ As we increase  $V_{gs}$ , there will be more number of electrons attracted towards the channel. As the channel length is less, & more electrons are getting accumulated in the channel, the mobility of electrons gets degraded (reduced) because of collisions. This effect is termed as mobility degradation.



Advaith PR  
Asst. Prof.  
Dept. of ECE  
SVIT

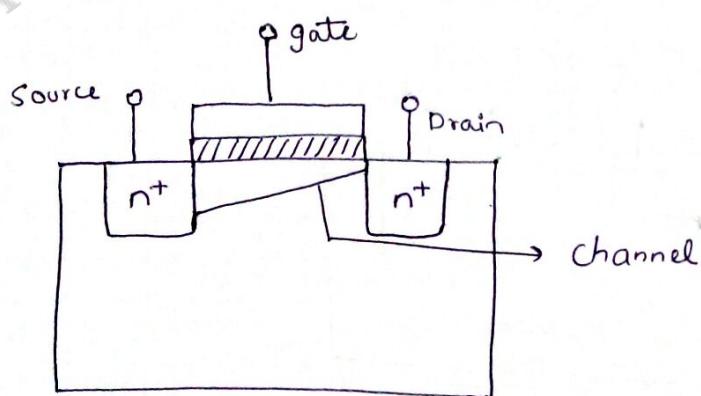
## (ii) Channel length modulation :-

From the previous discussions, when  $V_{ds}$  is greater than or equal to  $V_{gs} - V_t$ , the mosfet becomes saturated, as shown below.



i.e for the given value of  $V_{gs}$ , even if  $V_{ds}$  is increased beyond  $V_{gs} - V_t$ , the drain current ( $I_D$ ) remains constant

But in reality, if  $V_{ds}$  is increased beyond  $V_{gs} - V_t$ , the drain current ( $I_D$ ) increases slightly, this phenomenon is called as "channel length modulation". { represented by red dotted line in the above graph }.



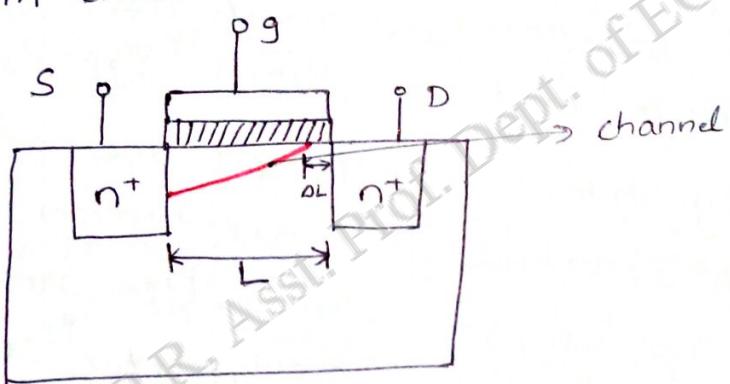
As we increase the drain to source voltage ( $V_{ds}$ ), the drain terminal becomes more positive than the gate terminal.

ie the gate terminal attracts lesser electrons, when compared to source terminal.

ie the electron charge density is lesser towards drain terminal when compared to source terminal. This is clearly shown in the figure, the channel thickness is more towards source region when compared to drain region.

As we increase  $V_{DS}$  even more, the channel towards the drain side becomes even more negligible. This point is called as "pinch off" point.

If we increase  $V_{DS}$  even more, the "pinch off" point starts to move from drain end to source end, as shown below.



now the effective length of the channel is

$$L_{eff} = L - \Delta L$$

Even though the channel becomes pinched off, a small amount of current makes its way from drain to source. Hence in the graph it is clear that the  $I_{DS}$  increases slightly as we increase  $V_{DS}$  than remaining constant.

### (iii) Body Effect:-

We know that the MOSFET is a four terminal device consisting of gate, drain, source and body terminals.

usually body terminal of the mosfet is connected to source and in turn it is connected to ground.

But in some practical applications, there is a possibility that body & source is not connected to ground.

In such scenarios, the difference in potential between the body and the source terminal causes a change in the threshold voltage of the MOSFET.

This effect of change in threshold voltage is called the "Body Effect" or "Back gate Effect".

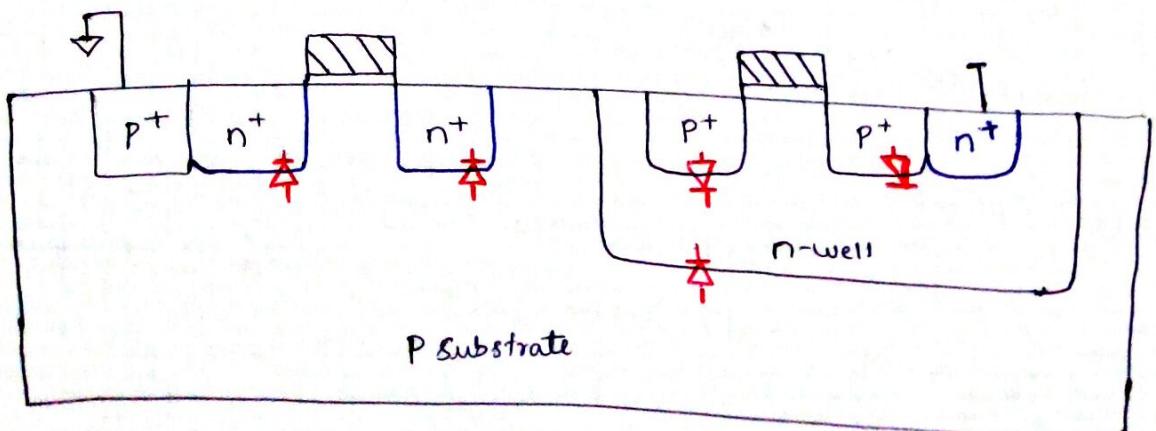
#### (iv) Subthreshold conduction :-

The ideal transistor I-V model assumes current flow from source to drain when  $V_{GS} > V_t$ .

In reality transistors current does not abruptly cut off below threshold, but conduction happens because of leakage current which often results in undesired current when a transistor is nominally OFF.

#### (V) Junction leakage :-

The p-n junctions between diffusion and the substrate or well form diodes as shown below.



There is a well to substrate diode, diffusion to substrate diode. The substrate and well are tied to GND or VDD to ensure these diodes remain reverse biased.

However, reverse biased diodes still conduct a small amount of current  $I_D$ , which is given by the formula

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

where  $I_D$  = diode current

$I_S$  = diode reverse bias saturation current

$V_D$  = diode voltage

$V_T$  = Threshold voltage

This junction leakage current (reverse diode current) may pose effect on the performance of the device. Generally the leakage current ranges between  $0.1$  to  $0.01 \text{ FA}/\mu\text{m}^2$

#### (vi) Tunneling :-

There is a finite probability that carriers will tunnel through the gate oxide. This results in gate leakage currents flowing into the gate.

For oxides thinner than about  $15-20 \text{ Å}$ , tunneling current becomes a factor and may become comparable to subthreshold leakage in advanced processes.

#### (vii) Temperature dependence :-

Transistor characteristics are influenced by temperature.

Carrier mobility: decreases with temperature

Threshold voltage: decreases with temperature

Junction leakage: Increases with temperature

Most wear out mechanisms are temperature dependent so transistors are more reliable at lower temperatures

### (viii) Geometry Dependence :-

The layout designer draws transistors with width and length  $W_{\text{drawn}}$  and  $L_{\text{drawn}}$ . The actual gate dimension may differ by some factors  $X_W$  &  $X_L$ .

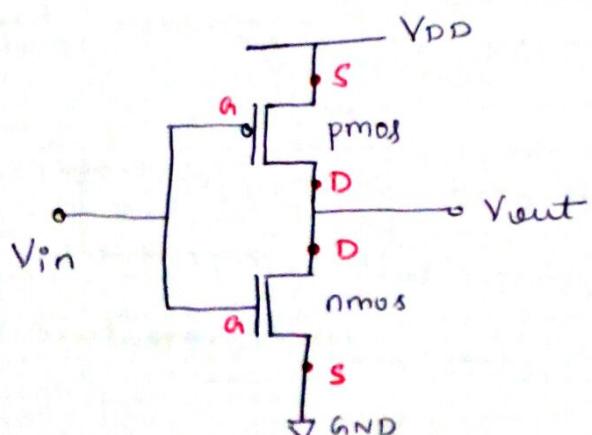
For ex:- The manufacturer may create masks with narrower polysilicon or may over etch the polysilicon to provide shorter channels without changing overall design rules or metal pitch.

If we scale down the length ( $L$ ) & width ( $W$ ) of transistor, care should be taken that other characteristics of MOSFET is not affected.

### DC Transfer Characteristics :-

The DC transfer characteristics of a circuit relate the output voltage to the input voltage.

### DC characteristics of CMOS Inverter :-



- Voltage b/w gate & source of nmos is  $V_{GSn}$
- Voltage b/w drain & source of nmos is  $V_{DSn}$
- Voltage b/w gate & source of pmos is  $V_{GSp}$
- Voltage b/w drain & source of pmos is  $V_{DSP}$

- Threshold voltage of nmos ( $V_{Tn}$ ) is always positive
- Threshold voltage of pmos ( $V_{Tp}$ ) is always negative

For nmos :-

$$V_{gsn} = (V_g - V_s)_n = V_{in} - 0 = V_{in}$$

$$V_{dsn} = (V_d - V_o)_n = V_{out} - 0 = V_{out}$$

For pmos :-

$$V_{gsp} = (V_g - V_s)_p = V_{in} - V_{DD}$$

$$V_{dsp} = (V_d - V_o)_p = V_{out} - V_{DD}$$

The reason for writing the above equation is, the DC transfer characteristics is a plot of output voltage to the input voltage.

Regions of operation of nmos and pmos :-

### nmos

$V_{gsn} < V_{Tn}$  → Cut off region

$V_{gsn} > V_{Tn}$  → Either linear  
or  
saturation

→  $V_{dsn} < V_{gsn} - V_{Tn}$  → linear

→  $V_{dsn} \geq V_{gsn} - V_{Tn}$  → saturation

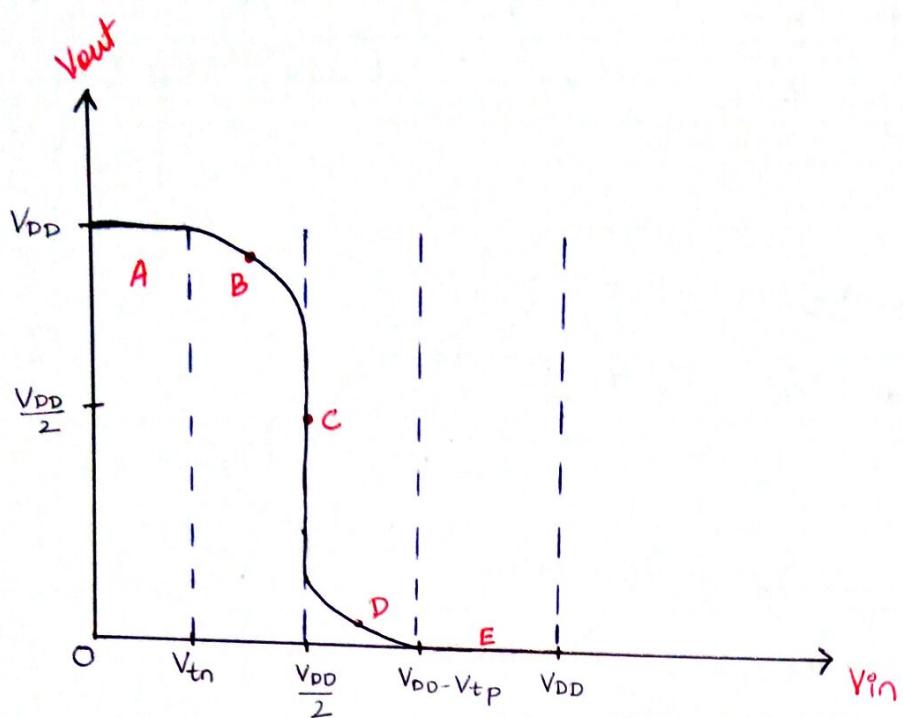
### pmos

$V_{gsp} > V_{Tp}$  → Cut off region

$V_{gsp} < V_{Tp}$  → Either linear  
or  
saturation

→  $V_{dsp} > V_{gsp} - V_{Tp}$  → linear

→  $V_{dsp} \leq V_{gsp} - V_{Tp}$  → saturation



In region A, the nmos transistor is OFF so the pmos transistor pulls the output to  $V_{DD}$ .

In region B, the nmos transistor starts to turn on, pulling the output down.

In region C, both transistors are in saturation.

In region D, the pmos transistor is partially on.

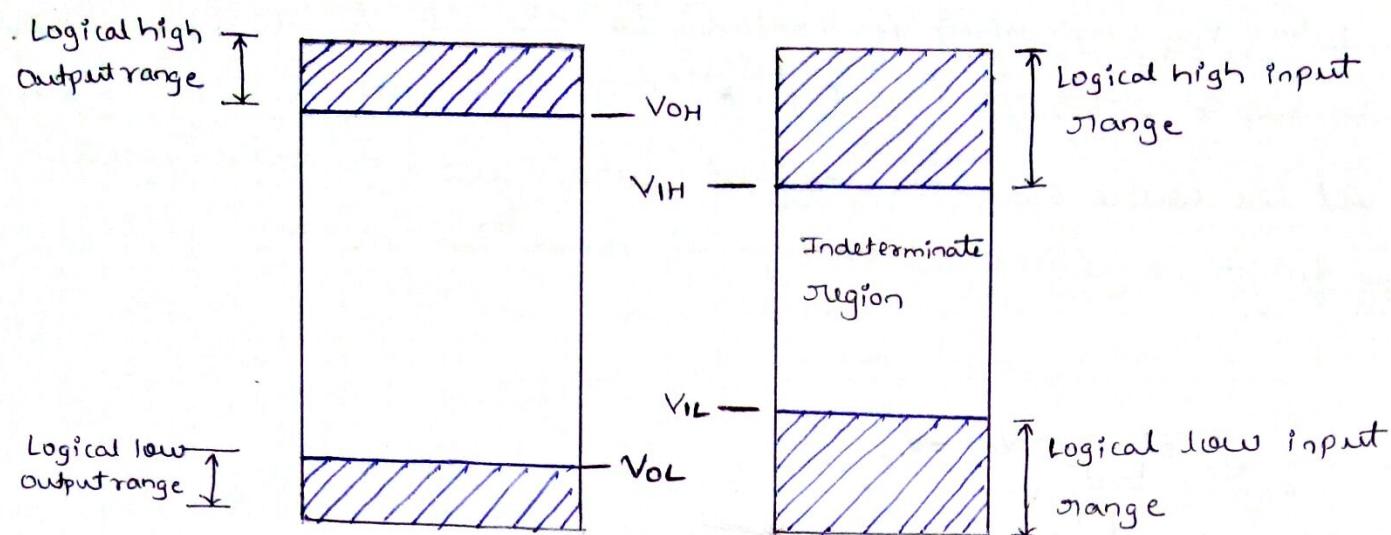
In region E, the pmos transistor is completely OFF and nmos transistor is completely ON. Hence the output is completely pulled down to zero.

### Noise Margin :-

Noise margin parameter allows us to determine the allowable noise voltage on the input of a gate so that the output will not be corrupted.

The specification most commonly used to describe noise margin uses two parameters,

the low noise margin  $NM_L$  and the high noise margin  $NM_H$ .



where,  $V_{IH}$  = minimum High input voltage

$V_{IL}$  = maximum Low input voltage

$V_{OH}$  = minimum High output voltage

$V_{OL}$  = maximum Low output voltage

$$NM_L = V_{IL} - V_{OL} \quad NM_H = V_{OH} - V_{IH}$$

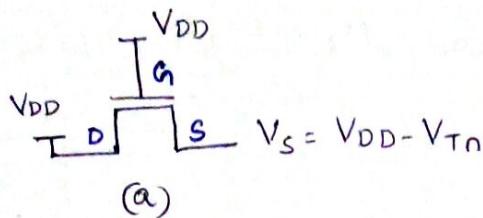
The difference between maximum low input voltage to maximum low output voltage is called as noise margin low.

The difference between minimum high output voltage to minimum high input voltage is called as noise margin high.

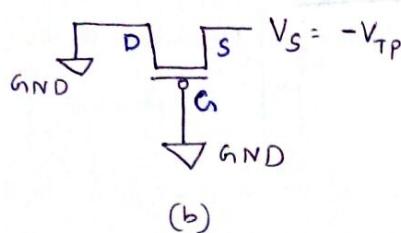
Inputs between  $V_{IL}$  &  $V_{IH}$  are said to be in the Indeterminate region or forbidden zone and do not represent legal digital logic levels.

Pass Transistors :-

Transistor used to pass logic "0" or logic "1" is termed as pass transistor.



When  $V_{DD}$  (logic high) is applied to the gate of nmos transistor, connection is established between drain and source. At the source side, the output which we get has a term  $V_{Tn}$  which is threshold voltage of nmos transistor.



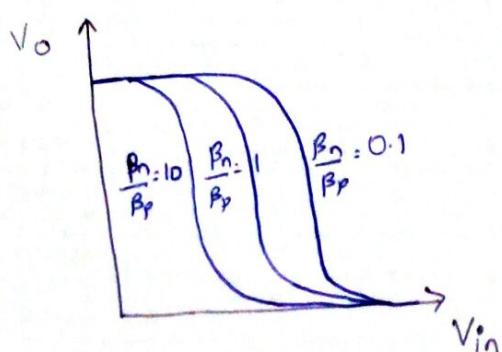
PMOS transistor turns on for low logic gate, but the output which we get from source has a term  $V_{TP}$  which is threshold voltage of PMOS.

$\beta_n / \beta_p$  Ratio :-

$$\beta_n = \frac{M_n C_{ox} W_n}{L_n} \quad \& \quad \beta_p = \frac{M_p C_{ox} W_p}{L_p}$$

For the CMOS inverter  $\frac{\beta_n}{\beta_p} = 1$  is desirable, since it allows the sourcing and sinking currents to be equal. In other words, the charging and discharging of the load capacitor will be equal.

But if  $\frac{\beta_n}{\beta_p} \neq 1$ , then the dimensions of the p-transistor must be larger than that of N-transistor.



From the graph, it is observed that as  $(\frac{\beta_n}{\beta_p})$

decreases, the transition region shifts from left to right