BHARAT INSTITUTE OF ENGINEERING & TECHNOLOGY

BHARAT INSTITUTE OF ENGINEERING AND TECHNOLOGY

POLYTECHNIC

MOHADA,BERHAMPUR,GANJAM

LECTURE NOTES ON

VLSI & EMBEDDED SYSTEM

(TH-2)

3rd Year, 5th Semester

PREPARED BY: -

ER. PINAKI BISOYI

LECTURER IN E&TC DEPT.

VLSI ENGINEERING

MODULE 1:

Issues and Challenges in VLSI Design:

Integration density and performance of integrated circuits have gone through an astounding revolution in the last couple of decades. In the 1960s, Gordon Moore, then with Fairchild Corporation and later cofounder of Intel, predicted that the number of transistors that can be integrated on a single die would grow exponentially with time. This prediction, later called *Moore's law*, has proven to be amazingly visionary. Its validity is best illustrated with the aid of a set of graphs. Figure 1 plots the integration density of both logic IC's and memory as a function of time. As can be observed, integration complexity doubles approximately every 1 to 2 years. As a result, memory density has increased by more than a thousand fold since 1970.

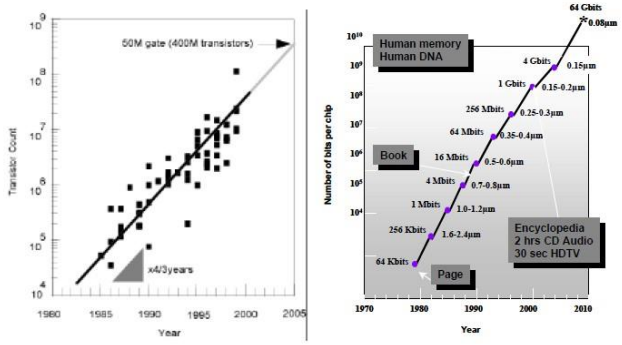


Figure 1: Evolution of integration complexity of logic ICs and memories as a function of time.

Typically used abstraction levels in digital circuit design are, in order of increasing abstraction, the device, circuit, gate, functional module (e.g., adder) and system levels (e.g., processor), as illustrated in Figure 2. A semiconductor device is an entity with a very complex behavior.

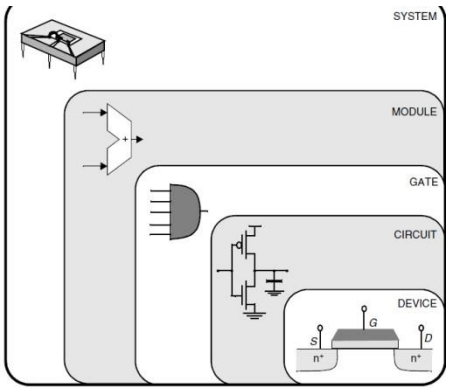
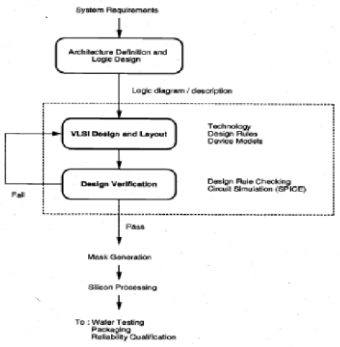


Figure 2 Design abstraction levels in digital circuits.

VLSI Design Flow:



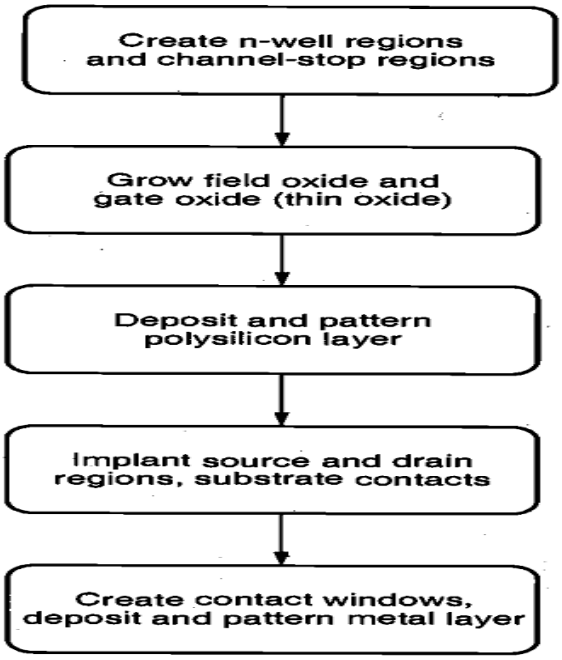
UNIT -2

FABRICATION OF MOSFET

SIMPLIFIED PROCESS SEQUENCE FOR FABRICATION-

- CMOS fabrication technology requires both NMOS and PMOS transistor to be built on the same chip substrate.
- To accommodate both NMOS & PMOS devices, special regions must be created in which the semiconductor type is opposite to the substrate type. These special regions are called wells or tubs.
- So, a n well is formed in a P substrate and a P well is formed in a n substrate.

  The simplified process sequence for the fabrication of CMOS-
- The process starts with the creation of the n well regions for PMOS and p well regions for NMOS by ion implantation into the substrate.
- Ion implantation is the process of adding impurities to a silicon wafer.
- Then a thick oxide is grown in the regions surrounding the NMOS and PMOS active regions. The thin gate oxide is subsequently grown on the surface through thermal oxidation.
- Again a polysilicon layer is deposited on the surface of the oxide layers and selectively removed to form the gate.
- These steps are followed by the creation of n+ and P+ regions.
- At last metallization is done means creation of metal interconnects.
- Metallization is the process by which the components of IC's are interconnected by aluminum conductor.
- Channel stop implant is used to prevent the formation of any unwanted channels between two neighboring regions. Hence channel stop implants act to electrically isolate neighboring devices built on the same substrate.

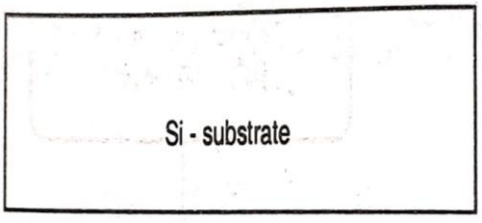


BASIC STEPS OF FABRICATION-

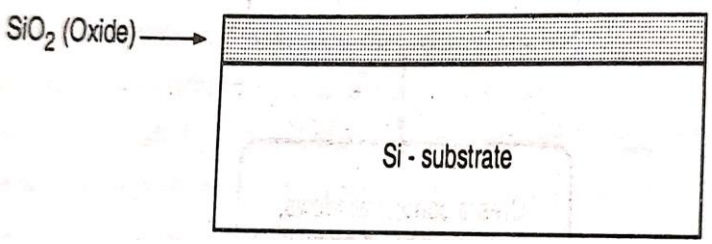
- The fabrication cycle of VLSI chips consists of a sequential set of basic steps which are wafer preparation, oxidation, lithography and etching.
- During fabrication process, the devices are created on the chip. So, IC may be viewed as a set of patterned layers.
- A layer must be patterned before the next layer of material is applied on the chip.
- Pattering uses the process of lithography. The process used to transfer a pattern to a layer on the chip is called lithography.
- The lithography sequence must be repeated for every layer.

Steps:

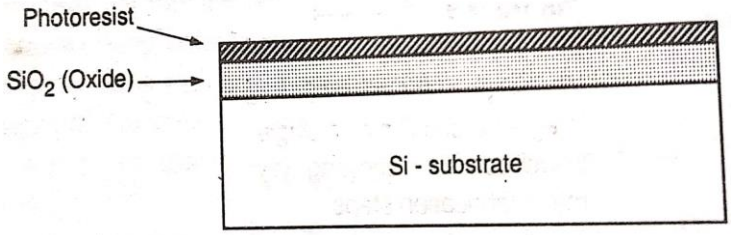
- First we take a Si substrate.



- The sequence starts with the thermal oxidation of the silicon surface. Due to which oxide layer formed of 1mm thickness.



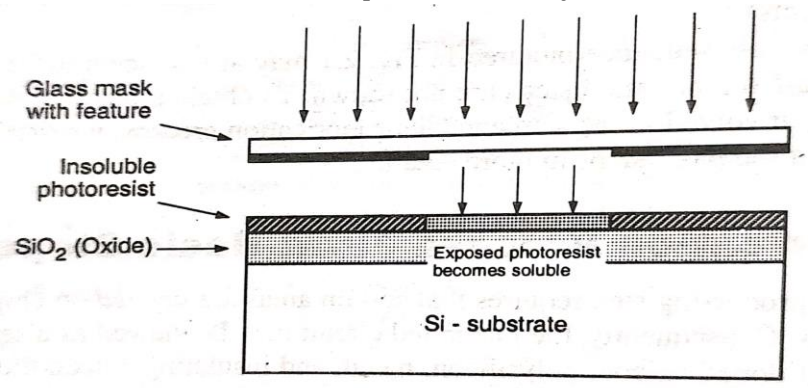
- The entire oxide surface is then covered with a layer of photoresist.



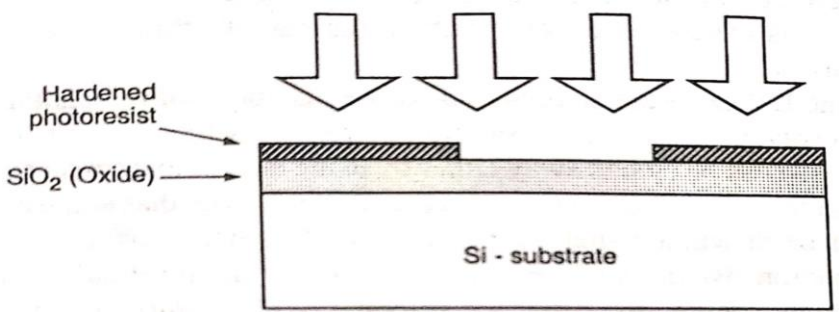
- Photoresist is a light sensitive material. It is of 2 types.

1) Positive photoresist
2) Negative photoresist

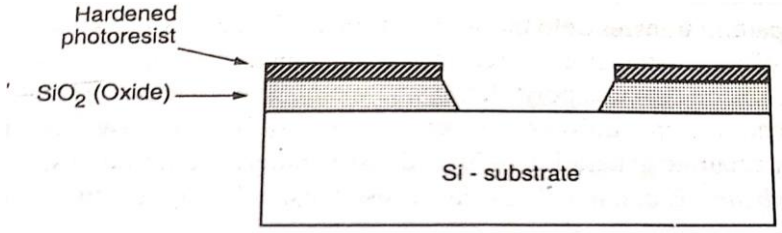
- Positive photoresist is initially insoluble and becomes soluble after exposure to UV light.
- Negative photoresist is initially soluble and becomes insoluble after exposure to UV light.
- Here we use positive photoresist. So, we have to cover some of the areas on the surface and selectively expose the photoresist.
- The areas becomes soluble, which are exposed to UV rays.



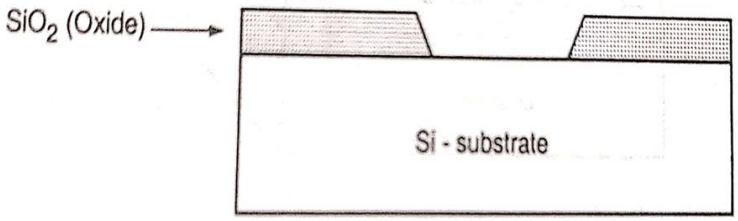
- Then the soluble areas can be etched away. Etching is the process of material being removed from the surface.



- The two major types of etching are wet etching and dry etching.
- The etching process that involves using liquid chemicals to take off the substrate material is called wet etching. Ex- Hydrofluoric Acid, Nitric acid, Acetic acid
- The dry etching is known as plasma etching. Etchant gases are used to remove the substrate material. Ex. Tetra fluoromethane, sulfur hexafluoride, Nitrogen trifluoride, Chlorine gas, Fluorine gas
- **Negative photoresists are more sensitive to light, but their photolithographic resolution is not as high as that of the positive photoresists. Therefore, negative photoresists are used less commonly.**
- The silicon dioxide regions which are not covered by hardened photoresist can be etched away either by using a chemical solvent or by using a dry etching process.

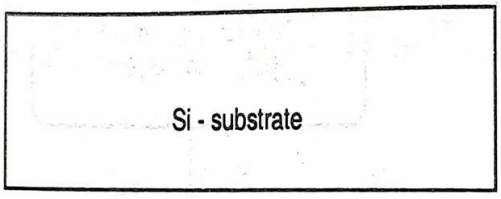


- After that the unexposed portions of the photoresist can be removed by a chemical leaving the patterned SiO_2.

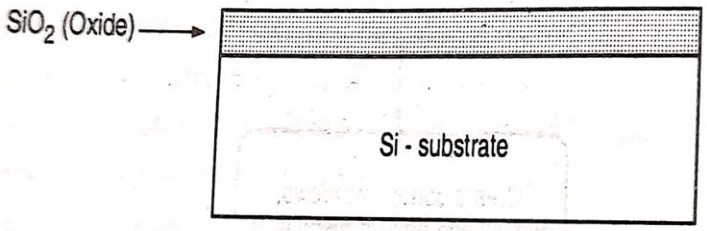


FABRICATION PROCESS OF NMOS TRANSISTOR-

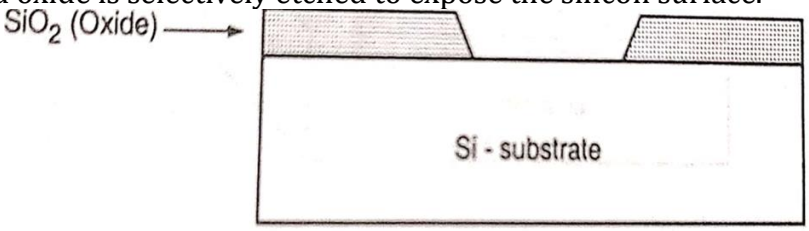
- First we take a p type silicon substrate.



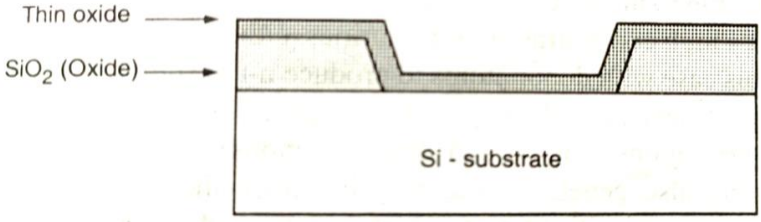
- The process starts with the oxidation of the silicon substrate



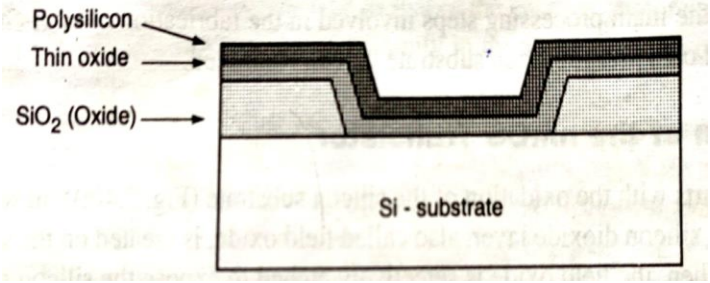
- Then the field oxide is selectively etched to expose the silicon surface.



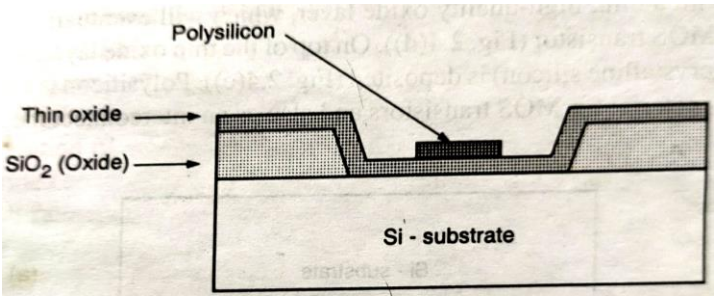
- Again the surface is covered with a thin oxide layer.



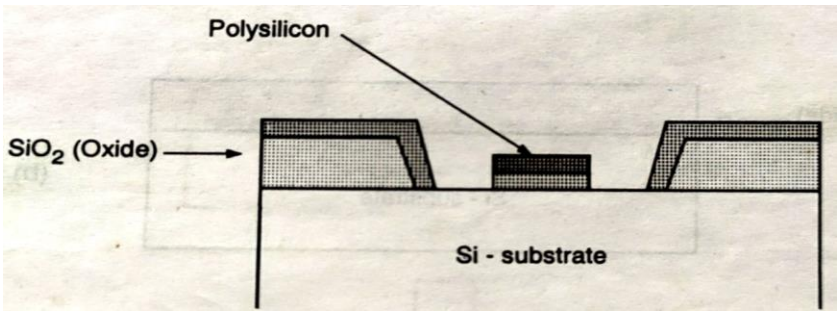
- On top of the thin oxide layer, a layer of polysilicon is deposited.



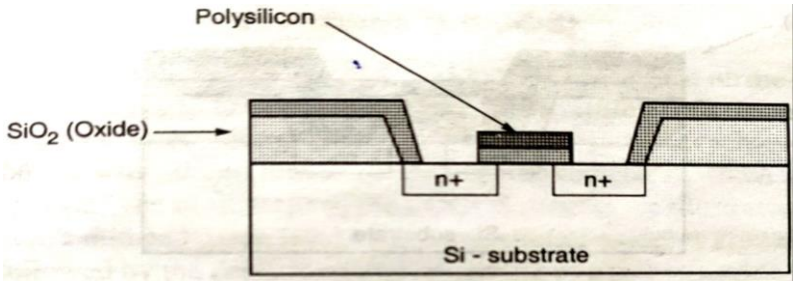
- After deposition, the polysilicon layer is patterned and etched to form gate of the MOSFET.



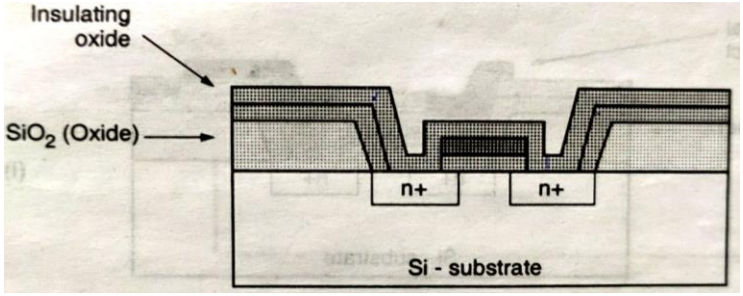
- The thin gate oxide not covered by polysilicon is also etched away, which exposes the silicon surface on which the source and drain junctions are to be formed.



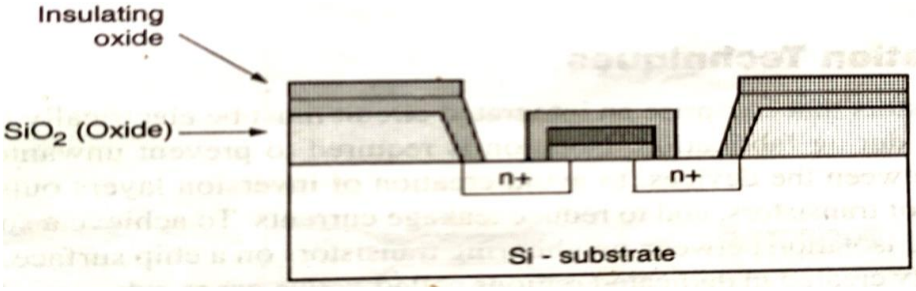
- The entire silicon surface is then doped with a high concentration of impurities, ultimately creating two n type regions.



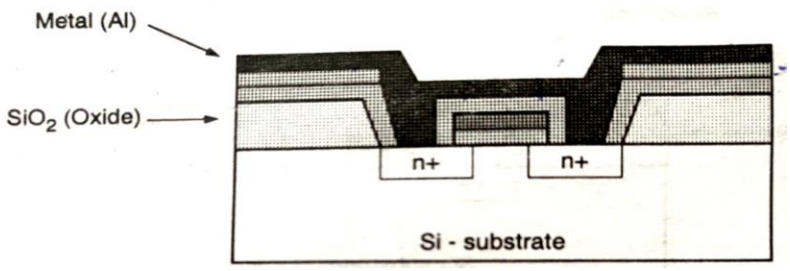
- Once the source and drain regions are completed, the entire surface is again covered with an insulating layer of silicon dioxide.



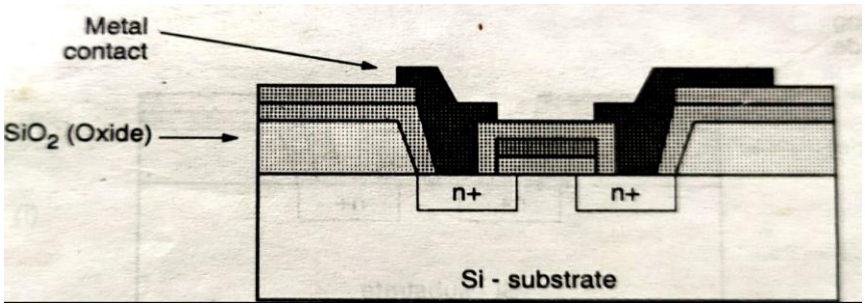
- The insulating oxide layer is then patterned in order to provide contact windows for the drain and source.



- Then the surface is covered with evaporated aluminum which will form the interconnects.



- Finally the metal layer is patterned and etched, completing the interconnection of the MOS transistors on the surface.

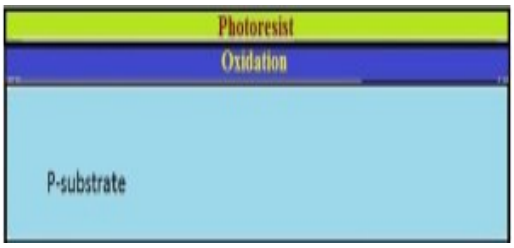


CMOS N-WELL FABRICATION PROCESS FLOW-

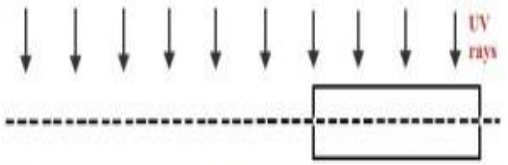
- For less power dissipation requirement CMOS technology is used for implementing transistor.
- The ne well technology and p-well technologies are used for fabrication of CMOS.
  Now let's discuss the steps of CMOS n-well fabrication.
- First we select a substrate as a base for fabrication. So, here we select a p-type substrate.

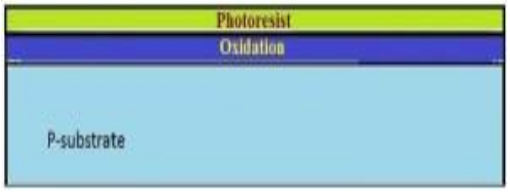
- Silicon dioxide layer formed by oxidation process on the Si substrate.

- For selective etching the SiO_2 layer is subjected to photolithography process. In this process, the wafer is coated with a uniform film of a photosensitive material known as photoresist.

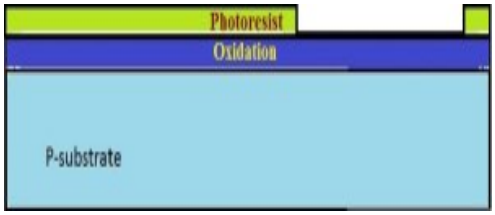


- The photoresist layer selectively exposed to UV rays.

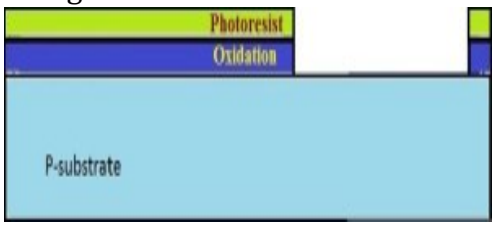




- The soluble photoresist is removed.



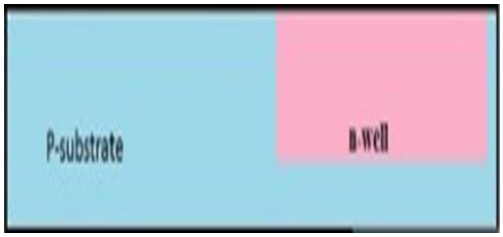
- The exposed silicon dioxide region is removed.



- The remaining photoresist layer is removed.

- N well is formed using ion implantation or diffusion.

- The remaining silicon dioxide is removed.

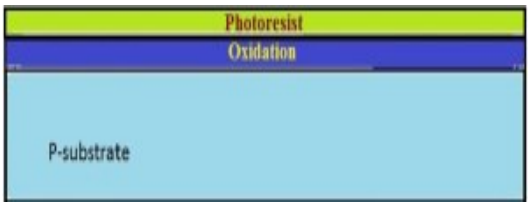


CMOS FABRICATION PROCESS BY N-WELL ON P SUBSTRATE-

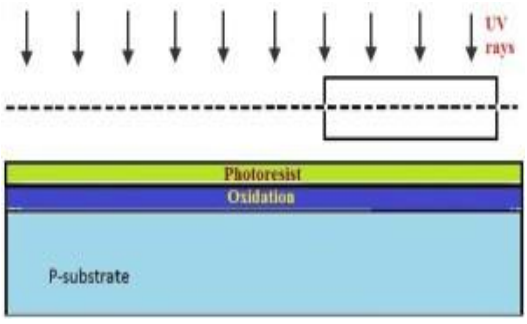
- For N well process first we take a P type substrate.

- Substrate is oxidized in high temperature.

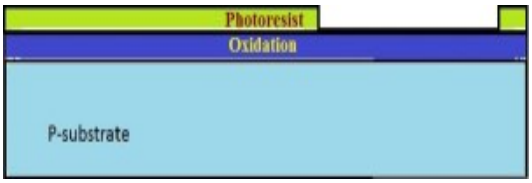
- Apply photoresist on the surface of the silicon dioxide.



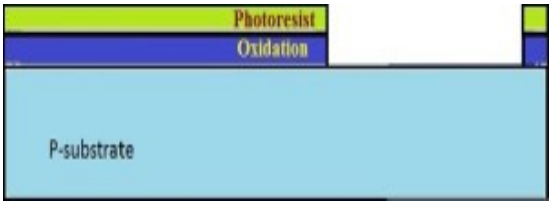
- Selectively expose the photoresist to the UV rays.



- The soluble photoresist is removed.

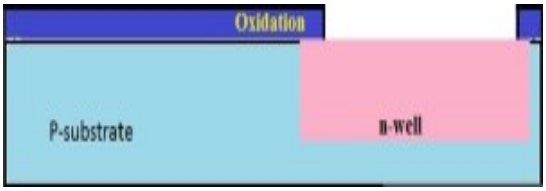


- The exposed Silicon dioxide region is removed.

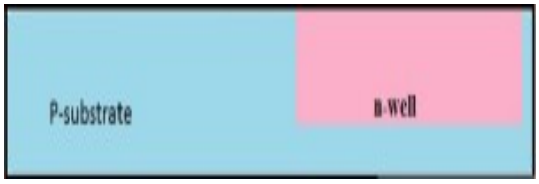


- The entire photoresist layer is stripped off.

- By using ion implantation or diffusion process N-well is formed.

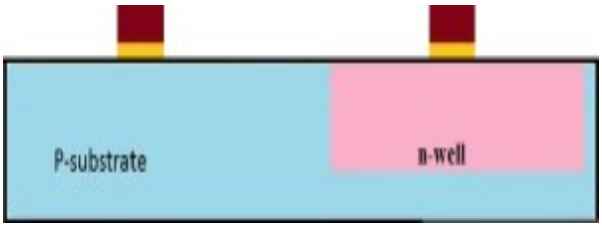


- The remaining silicon dioxide is removed.

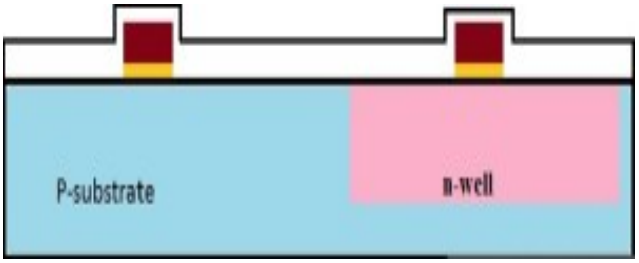


- A thin layer of gate oxide is deposited on the surface of the substrate. Then apply the polysilicon on the surface.

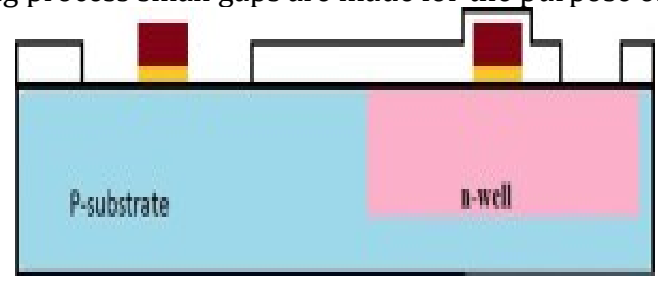
- Gate oxide and polysilicon layers are selectively removed.



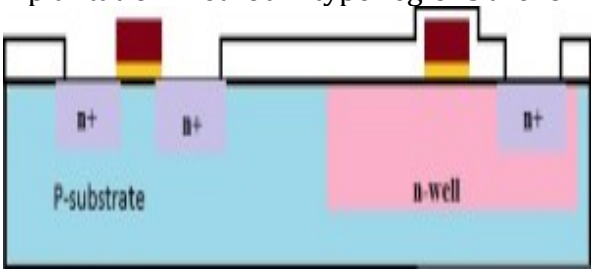
- An oxide layer is formed on the surface.



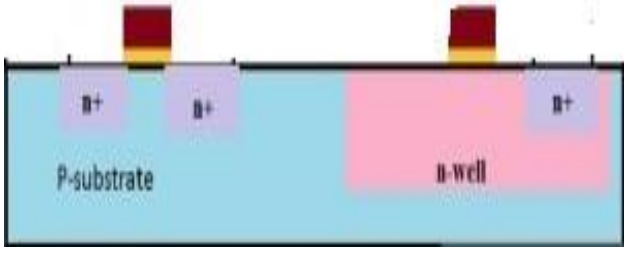
- By using the masking process small gaps are made for the purpose of N-diffusion.



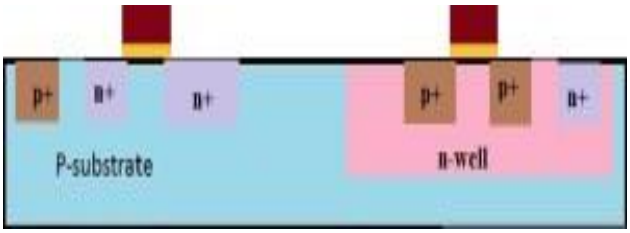
- Using diffusion or ion implantation method n type regions are formed.



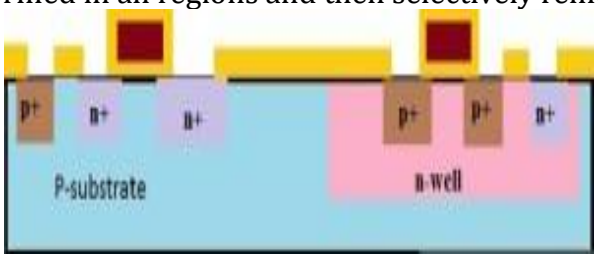
- The remaining oxide layer is stripped off.



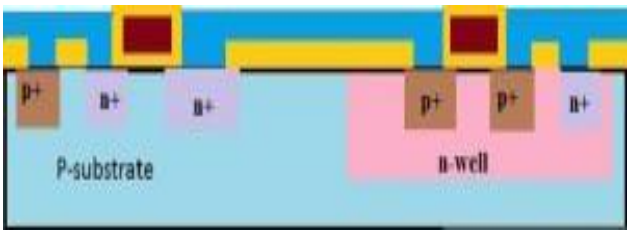
- Similar to the above process, the p type regions are formed.



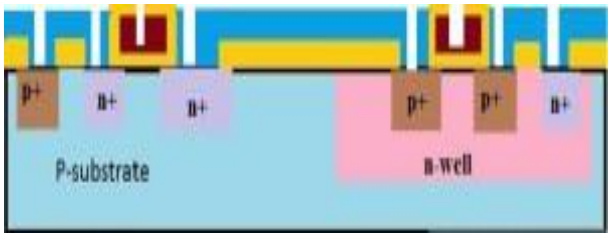
- A thick-field oxide is formed in all regions and then selectively removed.



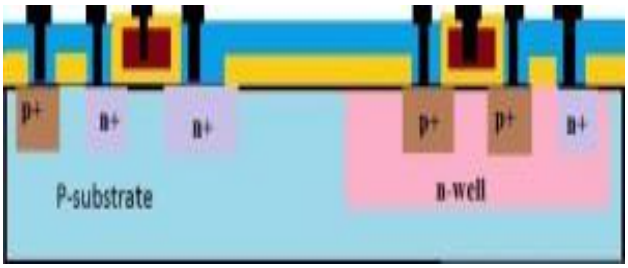
- Then the surface is covered with evaporated aluminum.



- The excess metal is removed from the surface.

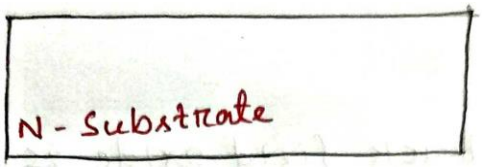


- The terminals of the PMOS and NMOS are made from respective gaps.

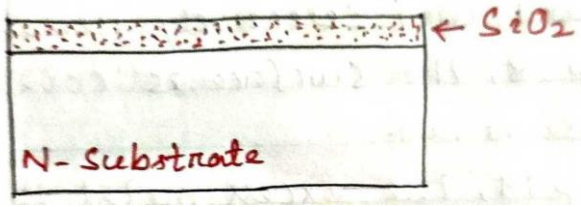


CMOS FABRICATION PROCESS BY P-WELL ON N SUBSTRATE-

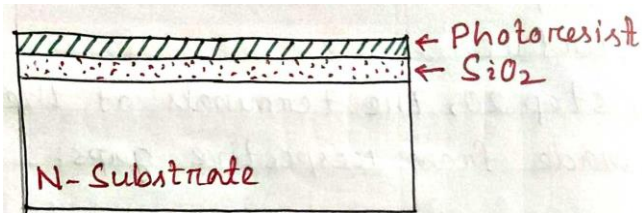
- For P well process first we take an N type substrate.



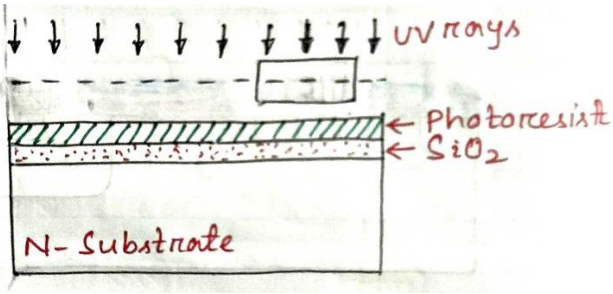
- Substrate is oxidized in high temperature.



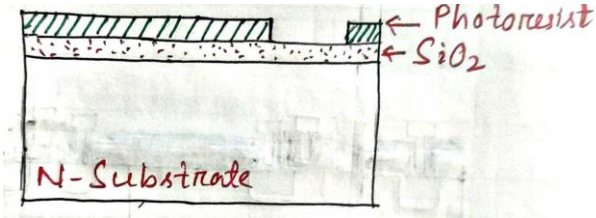
- Apply photoresist on the surface of the silicon dioxide.



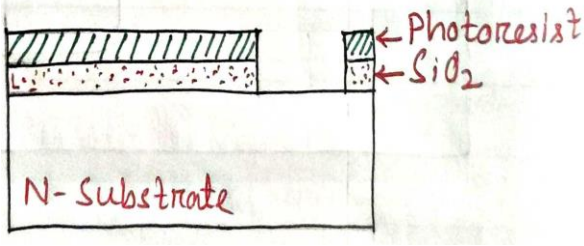
- Selectively expose the photoresist to the UV rays.



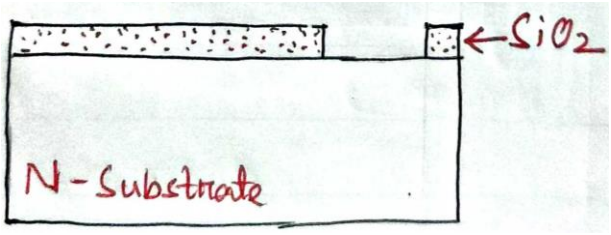
- The soluble photoresist is removed.



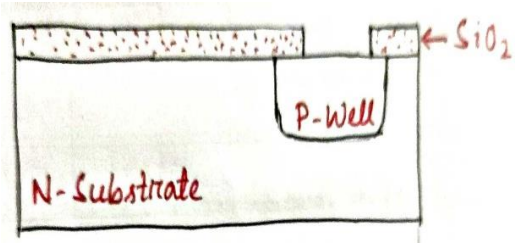
- The exposed Silicon dioxide region is removed.



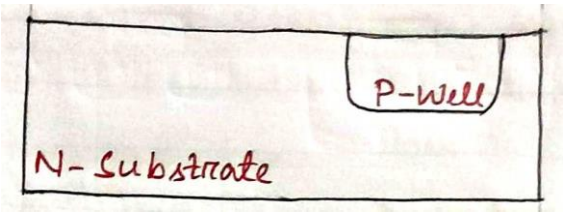
- The entire photoresist layer is stripped off.



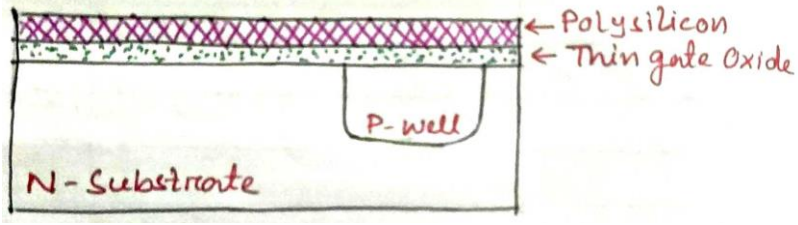
- By using ion implantation or diffusion process P-well is formed.



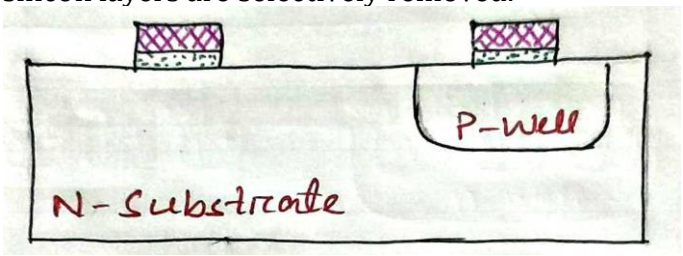
- The remaining silicon dioxide is removed.



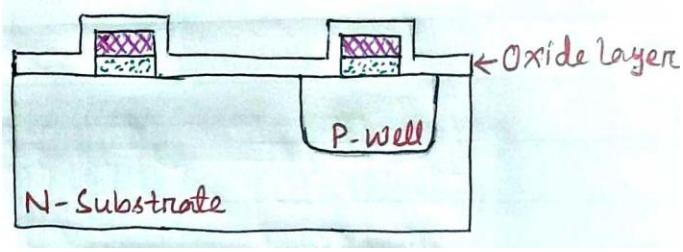
- A thin layer of gate oxide is deposited on the surface of the substrate. Then apply the polysilicon on the surface.



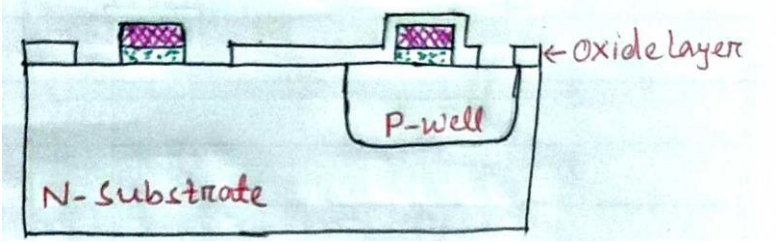
- Gate oxide and polysilicon layers are selectively removed.



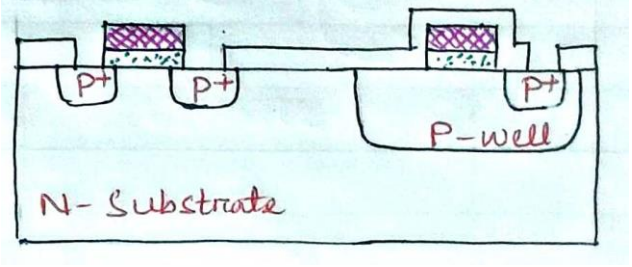
- An oxide layer is formed on the surface.



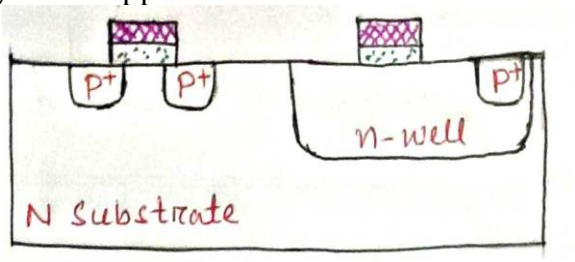
- By using the masking process small gaps are made for the purpose of P-diffusion.



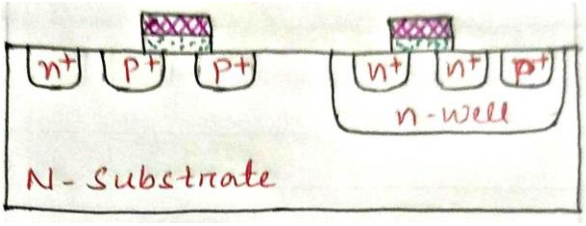
- Using diffusion or ion implantation method P type regions are formed.



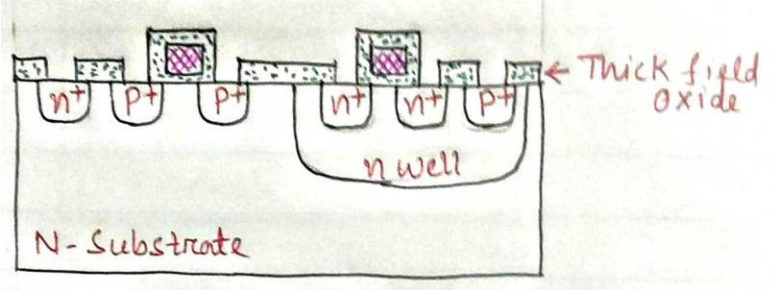
- The remaining oxide layer is stripped off.



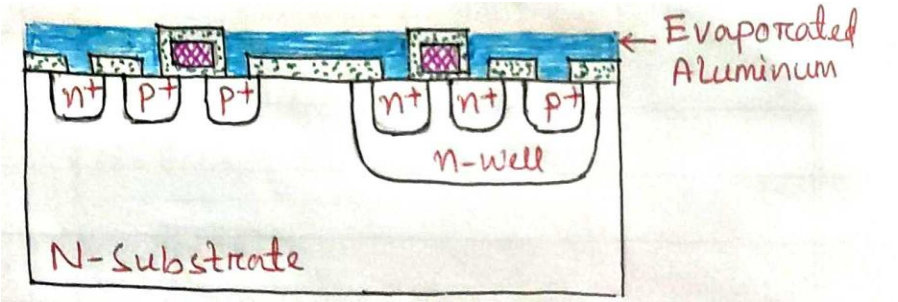
- Similar to the above process, the n type regions are formed.



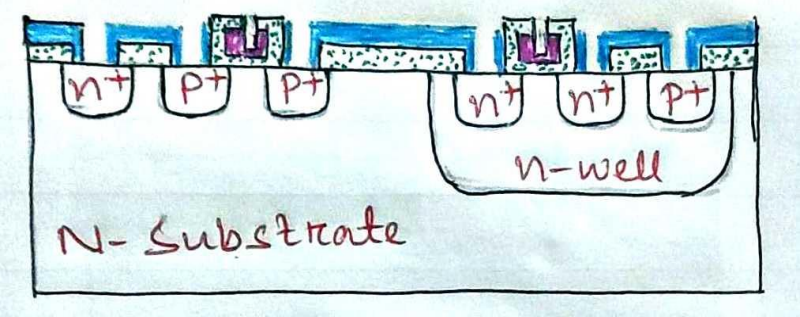
- A thick-field oxide is formed in all regions and then selectively removed.



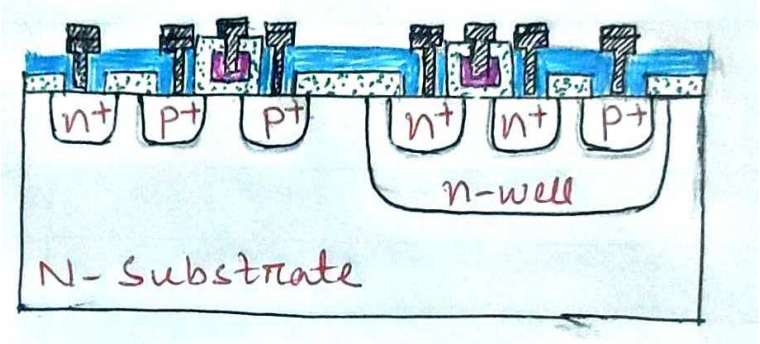
- Then the surface is covered with evaporated aluminum.



- The excess metal is removed from the surface.



- The terminals of the PMOS and NMOS are made from respective gaps.



LAYOUT DESIGN RULES-

The physical mask layout of any circuit to be manufactured using a particular process must confirm to set of geometric constraints or rules, which are generally called layout design rules.

The design rules are described in two ways-

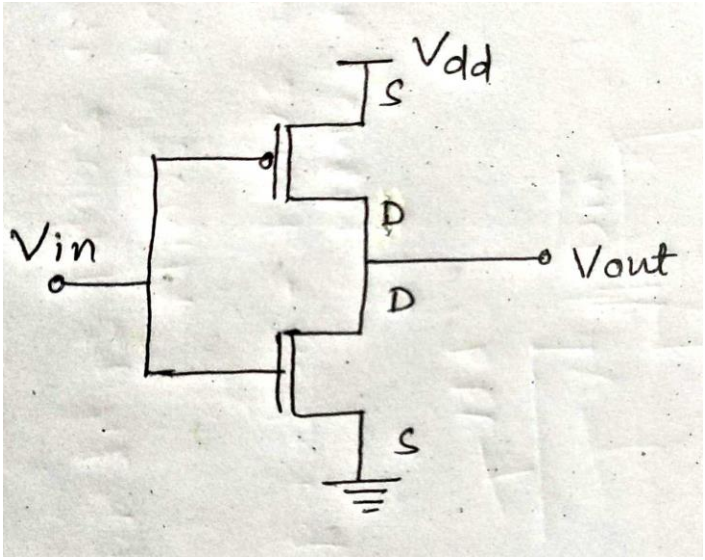
1) Micron rules-

   Micron rules, in which the layout constraints such as minimum feature sizes and minimum allowable feature separations are stated in terms of absolute dimensions in micrometers.

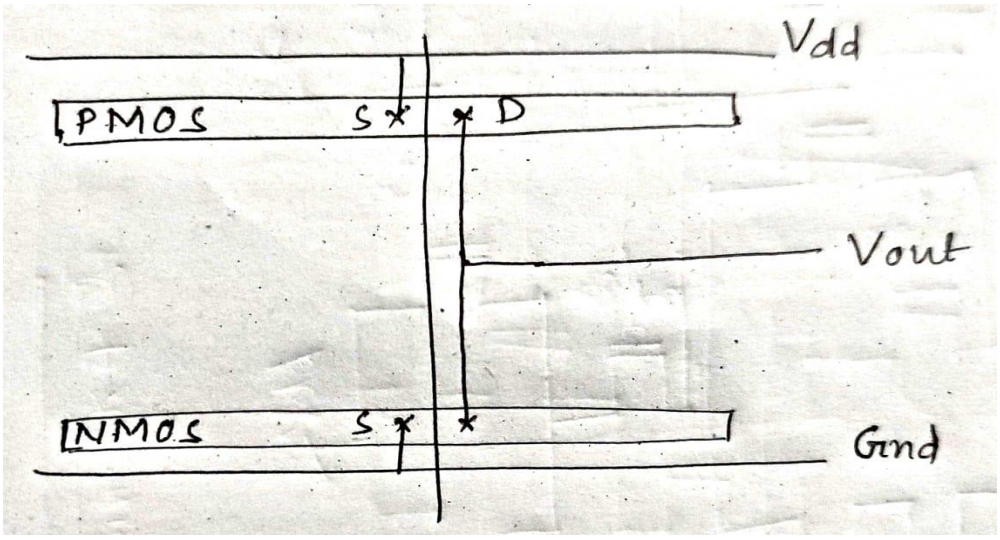
2) Lambda rules-

   Lambda rules specify the layout constraints in terms of a single parameter (λ) and thus allow linear, proportional scaling of all geometrical constraints.

STICK DIAGRAM OF CMOS INVERTER-



Stick Diagram-

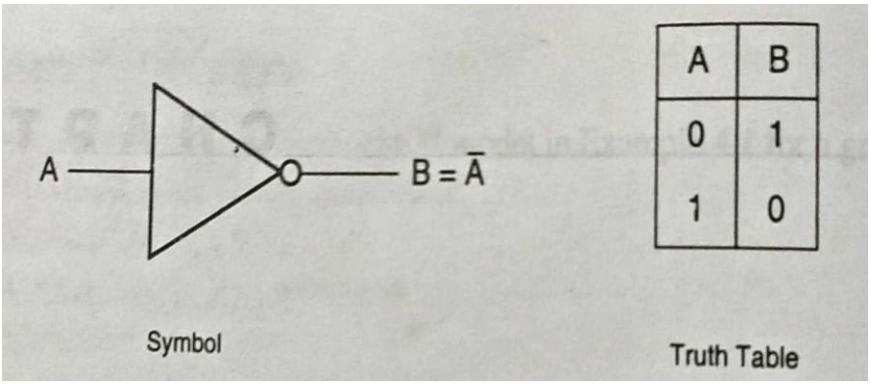


UNIT-3

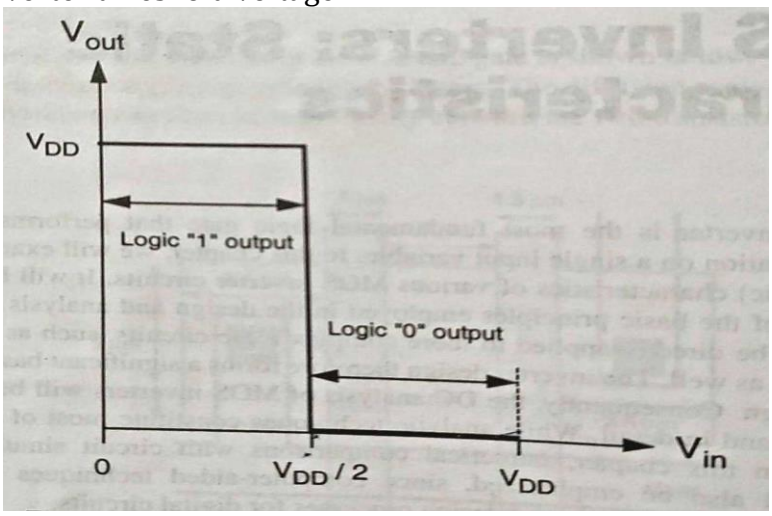
MOS INVERTER

BASIC NMOS INVERTER-

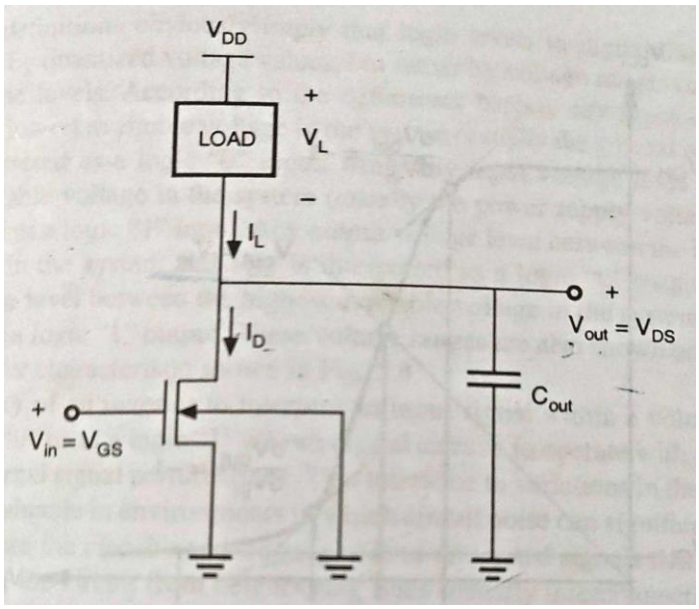
- In ideal inverter circuits, both the input variable A and the output variable B are represented by node voltages.



- Here the Boolean value of '1' means logic 1 can be represented by a high voltage of V_{DD} and the Boolean value of '0' means logic '0' can be represented by a low voltage of '0'. The voltage V_{th} is called the inverter threshold voltage.

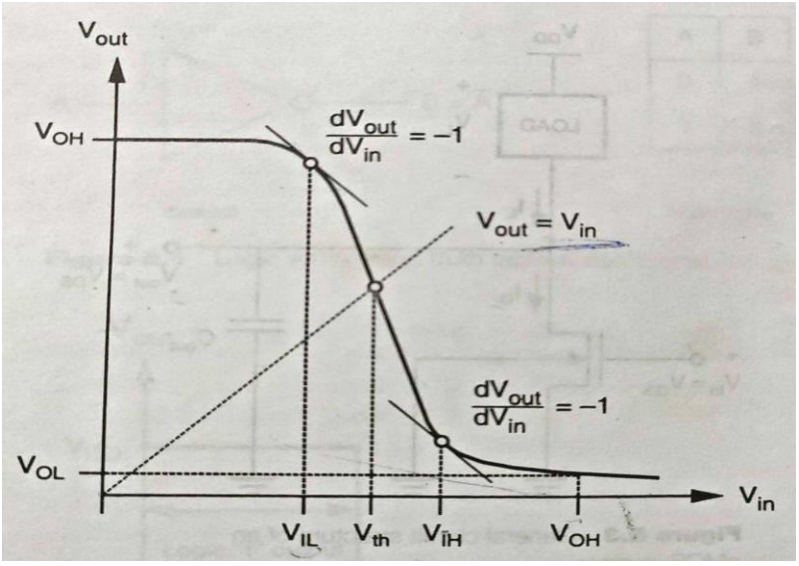


- For any input voltage between 0 to V_{th} the output voltage is equal to V_{DD}. The output switches from V_{DD} to 0 when the input is equal to V_{th}.
- For any input voltage between V_{th} andV_{DD}, the output voltage is equal to '0'. Thus an input voltage $0 \leq V_{in} < V_{th}$ is interpreted by this ideal inverter as a logic '0'.While an input voltage $V_{th} < V_{in} \leq V_{DD}$ is interpreted as a logic '1'.



- The input voltage of the inverter circuit is the gate to source voltage of the NMOS transistor. While the output voltage of the circuit is equal to the drain to source voltage.
- The source and the substrate terminals of the NMOS transistor are connected to ground potential. Hence $V_{SB} = 0$. The NMOS transistor is used as a driver transistor.
- The drain of NMOS is connected to the output terminal. The load device is represented as a two terminal circuit element with terminal currentI_L *and terminal voltage* V_L.
- One terminal of the load device is connected to the drain of the NMOS, while the other terminal is connected to V_{DD}.

VTC Curve-



- VTC curve, which is a plot of input vs output voltage. The VTC indicates that for low input voltage the circuit output is high and for high input, the output decreases towards 0 volt.
- Applying Kirchhoff's current law, the load current is always equal to the NMOS drain current.

$$I_D = I_L$$

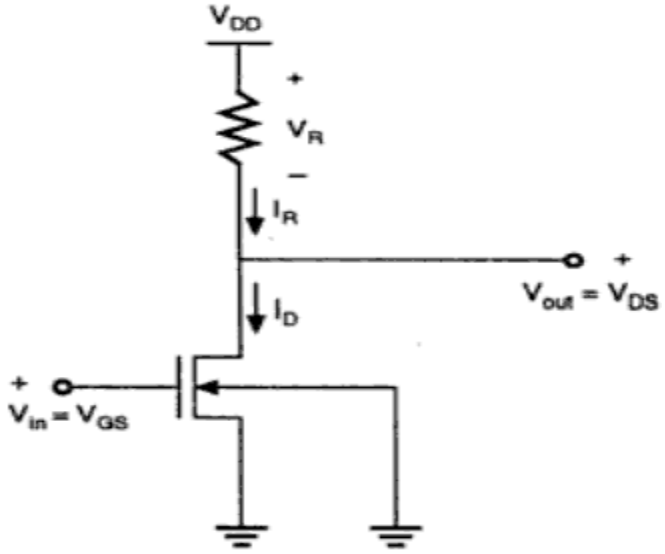
- For very low input voltage levels the output voltage V_{out} is equal to the high value of V_{OH}. The driver NMOS transistor is in cut off and hence does not conduct any current. The voltage drop across the load device is very small in magnitude and the output voltage is high.
- As the input voltage V_{in} increases, the driver transistor starts conducting a drain current and the output voltage starts to decrease. This drop in the output voltage level does not occur abruptly but in an ideal inverter it occur abruptly.
- In this curve two critical voltage points are present, where the slope becomes equal to -1.

$$\frac{dV_{out}}{dV_{in}} = -1$$

- The smaller input voltage at which first slope occur is called the input low voltage 'V_{IL}' and the larger input voltage at which second slope occur is called the input high voltage 'V_{IH}'.
- As the input voltage is further increased, the output voltage continues to drop and reaches a value of 'V_{OL}', when the input voltage is equal to 'V_{OH}'. The inverter threshold voltage V_{th} which is considered as the transition voltage is defined as the point where $V_{OUT} = V_{in}$.

RESISTIVE LOAD INVERTER-

Here, enhancement type nMOS acts as the driver transistor. The load consists of a simple linear resistor R_L. The power supply of the circuit is V_{DD} and the drain current I_D is equal to the load current I_R.



Circuit Operation

- When the input of the driver transistor is less than threshold voltage, driver transistor is in cut off region and does not conduct any current. So, the voltage drop across the load resistor is zero and output voltage is equal to the V_{DD}.
- Here $I_R = I_D$.
- So, output voltage V_{out} is

$$V_{out} = V_{DD} - I_R R$$

$$V_{out} = V_{DD} - I_D R$$

- So, Drain current equation will be

$$I_D = \frac{V_{DD} - V_{out}}{R}$$

- When the input voltage increases further, driver transistor will start conducting the non-zero current and NMOS goes in saturation region.
- if MOSFET is there in saturation region then

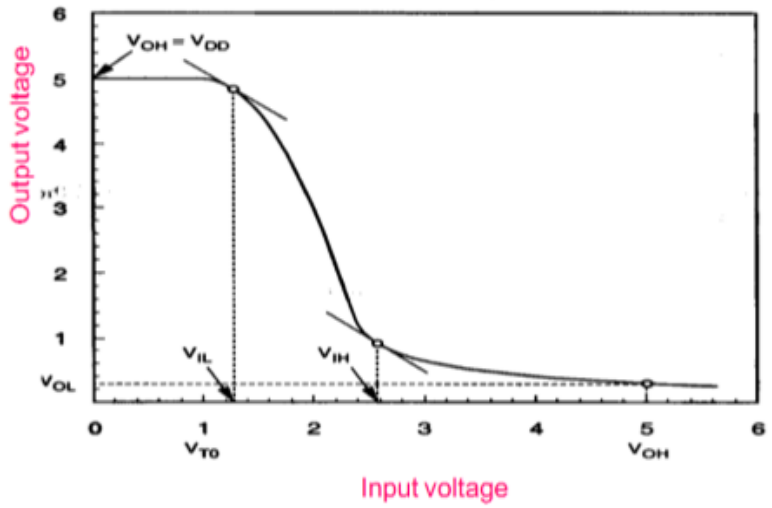
$$V_{in} - V_{To} < V_{out} \text{ and } I_D = \frac{K}{2}(V_{GS} - V_{TO})^2$$

- If MOSFET is there in linear region then

$$V_{in} - V_{To} > V_{out} \text{ and } I_D = \frac{K}{2}[2(V_{GS} - V_{TO})V_{DS} - V_{DS}^2]$$

VTC curve-

- Initially when input is at lower voltage, the NMOS is at cut off region, the V_{out} is equal to V_{DD} until the NMOS is not turned ON.
- Once the NMOS turned ON, slow decrease in output voltage starts.



V_{OH}-

- Output voltage V_{out} is

$$V_{out} = V_{DD} - I_D R$$

$$=> V_{out} = V_{DD}$$
$$=> V_{OH} = V_{DD}$$

V_{OL}-

- When $V_{in} - V_{To} > V_{out}$, MOSFET is there in linear region, so, drain current will be

$$I_D = \frac{K}{2}[2(V_{in} - V_{TO})V_{out} - {V_{out}}^2]$$

- According to kirchoff's law in the drain current is

$$I_D = \frac{V_{DD} - V_{out}}{R}$$

- If we compare these two equations

$\frac{V_{DD}-V_{out}}{R} = \frac{K}{2}[2(V_{in} - V_{TO})V_{out} - {V_{out}}^2]$

$=> \frac{V_{DD}-V_{OL}}{R} = \frac{K}{2}[2(V_{DD} - V_{TO})V_{OL} - {V_{OL}}^2]$

$=> {V_{OL}}^2 - 2(V_{DD} - V_{TO} + {}^1/_{KR})V_{OL} + {}^2/_{KR}V_{DD} = 0$

- If we solve the above equation, we get

$$V_{OL} = V_{DD} - V_{TO} + {}^1/_{KR} - \sqrt{(V_{DD} - V_{TO} + \frac{1}{KR})^2 - \frac{2V_{DD}}{KR}}$$

V_{IL}-

- When $V_{in} - V_{To} < V_{out}$ MOSFET is there in saturation region, so drain current will be

$$I_D = \frac{K}{2}(V_{in} - V_{TO})^2$$

- Again compare the equation with circuit drain current equation

$$\frac{V_{DD}-V_{out}}{R} = \frac{K}{2}(V_{in} - V_{TO})^2$$

- We have to differentiate it with respect to V_{in}

$(\frac{V_{DD}-V_{out}}{R})\frac{dV_{out}}{dV_{in}} = [\frac{K}{2}(V_{in} - V_{TO})^2]\frac{dV_{out}}{dV_{in}}$

$=> -\frac{1}{R}\frac{dV_{out}}{dV_{in}} = K(V_{in} - V_{TO})$

$=> \frac{1}{R} = K(V_{IL} - V_{TO})$

$=> V_{IL} = V_{TO} + {}^1/_{KR}$

V_{IH}-

- When $V_{in} - V_{To} > V_{out}$ MOSFET is there in linear region, so drain current will be

$$I_D = \frac{K}{2}[2(V_{in} - V_{TO})V_{out} - {V_{out}}^2]$$

Compare the equation with circuit drain current equation

$$\frac{V_{DD}-V_{out}}{R} = \frac{K}{2}[2(V_{in} - V_{TO})V_{out} - {V_{out}}^2]$$

- We have to differentiate it with respect to V_{in}

$$-\frac{1}{R}\frac{dV_{out}}{dV_{in}} = {K}/{2}\,[2(V_{in}-V_{TO})\frac{dV_{out}}{dV_{in}} + 2V_{out} - 2V_{out}\frac{dV_{out}}{dV_{in}}]$$

$$=> \frac{1}{R} = {K}/{2}\,[-2(V_{IH}-V_{TO}) + 4\,V_{out}]$$

$$=> \frac{1}{KR} = -(V_{IH}-V_{TO}) + 2\,V_{out}$$

$$=> \frac{1}{KR} = -V_{IH} + V_{TO} + 2V_{out}$$

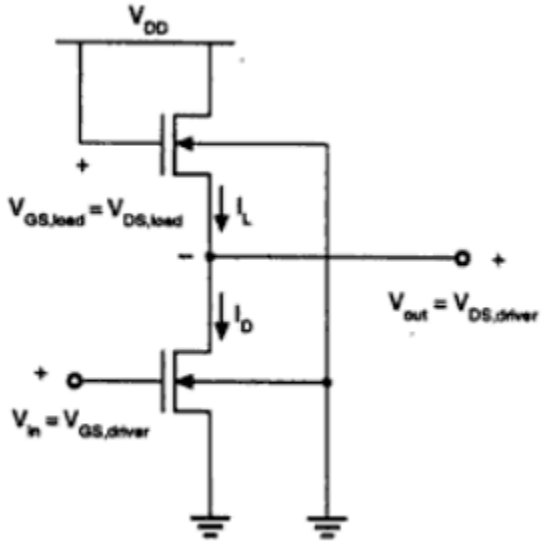
$$=> V_{IH} = V_{TO} + 2V_{out} - \frac{1}{KR}$$

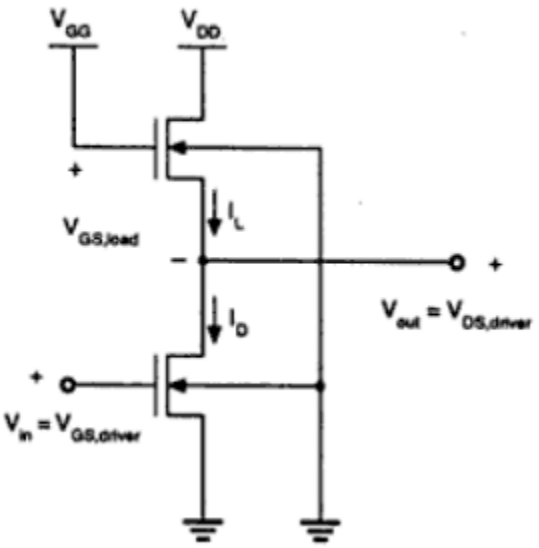
INVERTER WITH N-TYPE MOSFET LOAD-

The main advantage of using MOSFET as load device is that the silicon area occupied by the transistor is smaller than the area occupied by the resistive load. Here, MOSFET is active load and inverter with active load gives a better performance than the inverter with resistive load.

Enhancement load-

Load transistor can be operated either, in saturation region or in linear region, depending on the bias voltage applied to its gate terminal. The saturated enhancement load inverter is shown in the first figure. It requires a single voltage supply and simple fabrication process and so V_{OH} is limited to the $V_{DD} - V_T$.

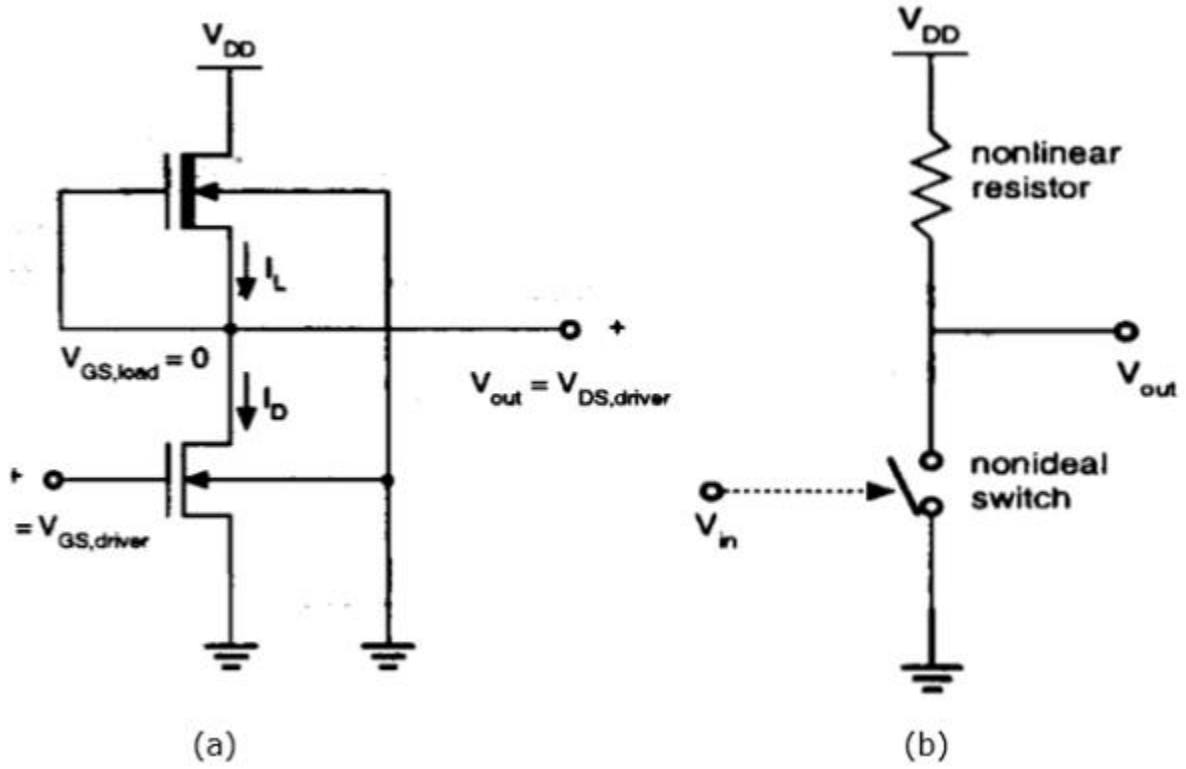




The linear enhancement load inverter is shown in the second figure. It always operates in linear region; so V_{OH} level is equal to V_{DD}.

Linear load inverter has higher noise margin compared to the saturated enhancement inverter. But, the disadvantage of linear enhancement inverter is, it requires two separate power supply and both the circuits suffer from high power dissipation. Therefore, enhancement inverters are not used in any large-scale digital applications.

Depletion load NMOS-



- Drawbacks of the enhancement load inverter can be overcome by using depletion load inverter. Compared to enhancement load inverter, depletion load inverter requires few more fabrication steps for channel implant to adjust the threshold voltage of load.
- The advantages of the depletion load inverter are - sharp VTC transition, better noise margin, single power supply and smaller overall layout area.
- The gate and source terminal of load are connected; So, $V_{GS} = 0$. Thus, the threshold voltage of the load is negative. Hence,

$$V_{GS,load} > V_{T,load}$$

Therefore, load device always has a conduction channel regardless of input and output voltage level.

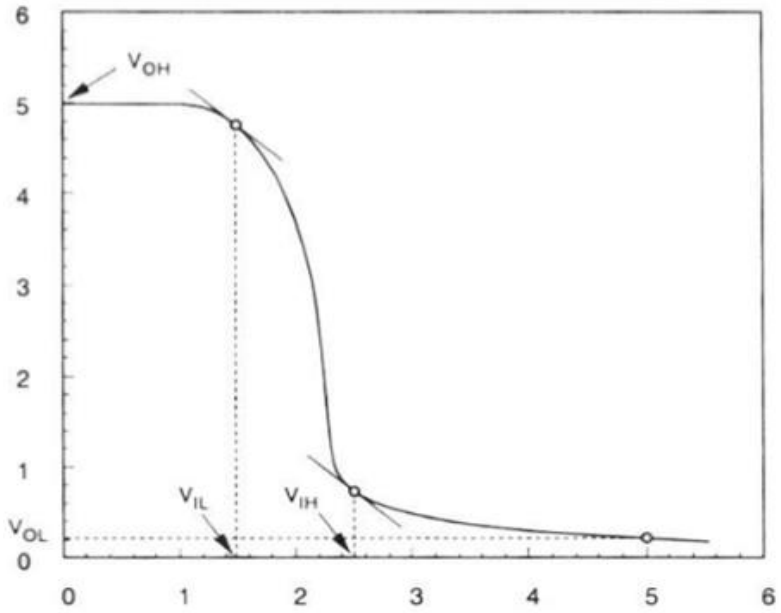
- When the load transistor is in saturation region, the load current is given by

$$I_{D,load} = \frac{K_{n,load}}{2}[-V_{T,load}(V_{out})]$$

- When the load transistor is in linear region, the load current is given by

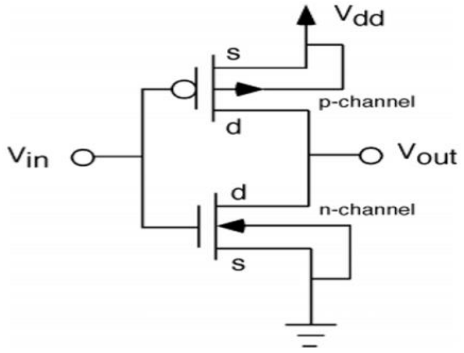
$$I_{D,load} = \frac{K_{n,load}}{2}\left[2\left|V_{T,load}(V_{out})\right|.(V_{DD} - V_{out}) - (V_{DD} - V_{out})^2\right]$$

The voltage transfer characteristics of the depletion load inverter is shown in the figure given below –



CMOS INVERTER-

In CMOS inverter NMOS work as driver and PMOS transistors work as load and always one transistor is ON, other is OFF.



This configuration is called **complementary MOS (CMOS)**. The input is connected to the gate terminal of both the transistors such that both can be driven directly with input voltages. Substrate of the NMOS is connected to the ground and substrate of the PMOS is connected to the power supply, V_{DD}.

So V_{SB} = 0 for both the transistors.

$$V_{GS,n} = V_{in}$$

$$V_{DS,n} = V_{OUT}$$

And,

$$V_{GS,n} = V_{in} - V_{DD}$$

$$V_{DS,n} = V_{out} - V_{DD}$$

When the input of nMOS is smaller than the threshold voltage ($V_{in} < V_{TO,n}$), the nMOS is cut – off and pMOS is in linear region. So, the drain current of both the transistors is zero.

$$I_{D,n} = I_{D,p} = 0$$

Therefore, the output voltage V_{OH} is equal to the supply voltage.

$$V_{out} = V_{OH} = V_{DD}$$

When the input voltage is greater than the $V_{DD} + V_{TO,p}$, the pMOS transistor is in the cutoff region and the nMOS is in the linear region, so the drain current of both the transistors is zero.

$$I_{D,n} = I_{D,p} = 0$$

Therefore, the output voltage V_{OL} is equal to zero.

$$V_{out} = V_{OL} = 0$$

The nMOS operates in the saturation region if $V_{in} > V_{TO}$ and if following conditions are satisfied.

$$V_{DS,n} \geq V_{GS,n} - V_{TO,n}$$

$$V_{out} \geq V_{in} - V_{TO,n}$$

The pMOS operates in the saturation region if $V_{in} < V_{DD} + V_{TO,p}$ and if following conditions are satisfied.

$$V_{DS,P} \leq V_{GS,P} - V_{TO,P}$$

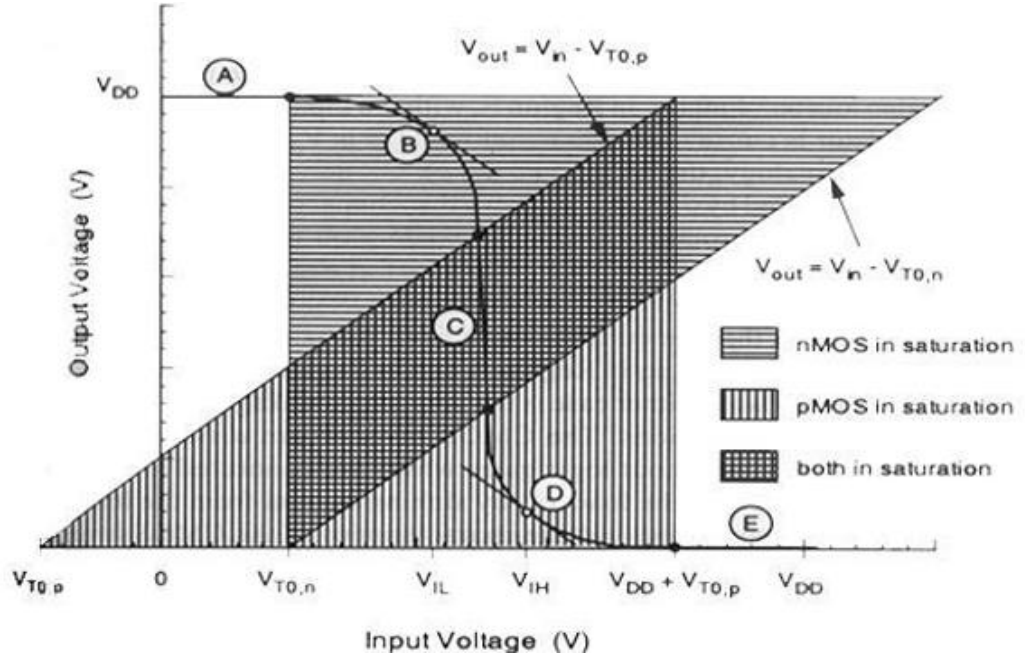
$$V_{out} \geq V_{in} - V_{TO,P}$$

For different value of input voltages, the operating regions are listed below for both transistors.

Region	**V_{in}**	**V_{out}**	**nMOS**	**pMOS**
A	$< V_{TO,n}$	V_{OH}	Cut – off	Linear
B	V_{IL}	High ≈ V_{OH}	Saturation	Linear
C	V_{th}	V_{th}	Saturation	Saturation

D	V_{IH}	Low ≈ V_{OL}	Linear	Saturation
E	> (V_{DD} + $V_{TO,p}$)	V_{OL}	Linear	Cut – off

The VTC of CMOS is shown in the figure below –



INTERCONNECT EFFECTS-

DELAY TIME DEFINATION-

The propagation delay times τ_{PHL} and τ_{PLH} determine the input to output signal delay during the high to low and low to high transitions of the output, respectively.

Definition-

τ_{PHL} is the time delay between the $V_{50\%}$ transition of the rising input voltage and the $V_{50\%}$transition of the falling output voltage.

τ_{PLH} is the time delay between the $V_{50\%}$ transition of the falling input voltage and the $V_{50\%}$transition of the rising output voltage.

τ_{PHL} becomes the time required for the output voltage to fall from V_{OH} to the $V_{50\%}$ level and

τ_{PLH} becomes the time required for the output voltage to rise from V_{OL} to the $V_{50\%}$ level.

$$V_{50\%} = V_{OL} + \frac{1}{2}(V_{OH} - V_{OL})$$

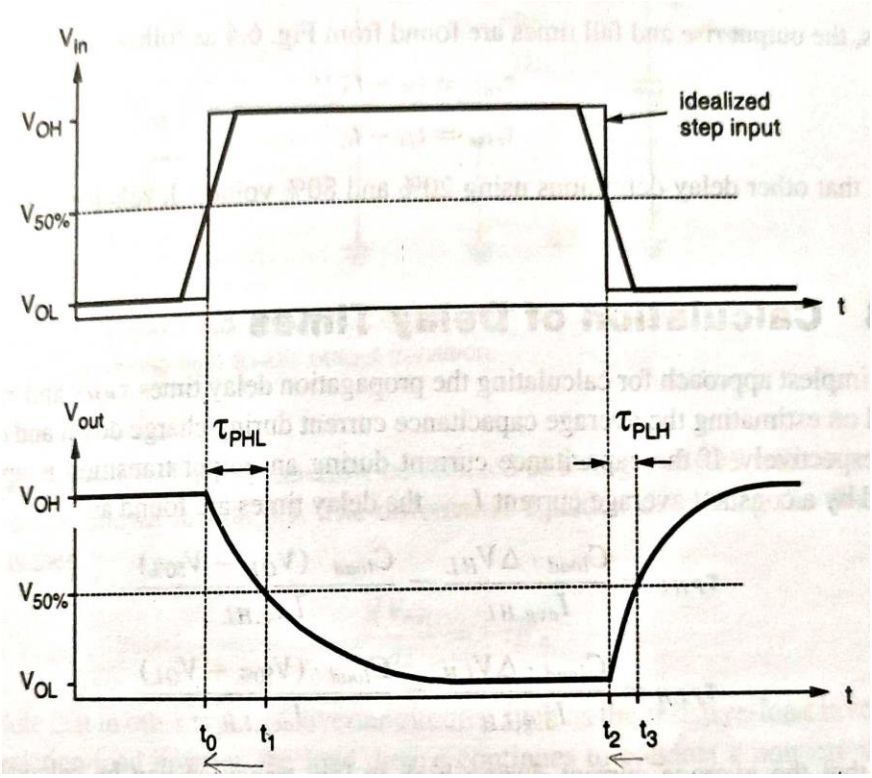
$= ½ (V_{OL} + V_{OH})$

$\tau_{PHL} = t_1 - t_0$

$\tau_{PLH} = t_3 - t_2$

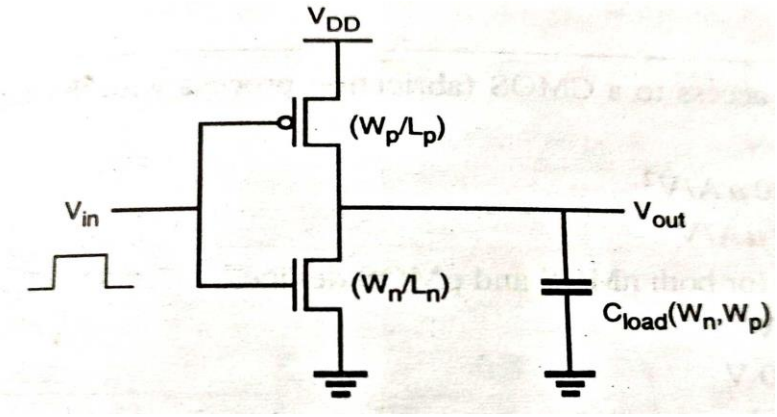
Average propagation delay is

$$\tau_p = \frac{\tau_{PHL} + \tau_{PLH}}{2}$$



INVERTER DESIGN WITH DELAY CONSTRAINTS-

The design of CMOS inverters based on timing specifications is one of the most fundamental issues in digital circuit design which ultimately determine the overall performance of complex systems.



The load capacitance C_{load} consists of intrinsic components and extrinsic components.

If C_{load} consists of extrinsic components and if this overall load capacitance can be estimated accurately and independently of the transistor dimensions, then the problem of inverter design can be reduced. Given a required delay value of τ_{PHL}, the (W/L) ratio of the NMOS transistor can be found as

$$\frac{W_n}{L_n} = \frac{C_{load}}{\tau_{PHL}\mu_n C_{ox}(V_{DD} - V_{T,n})}\left[\frac{2V_{T,n}}{V_{DD} - V_{T,n}} + ln\left(\frac{4(V_{DD} - V_{T,n})}{V_{DD}} - 1\right)\right]$$

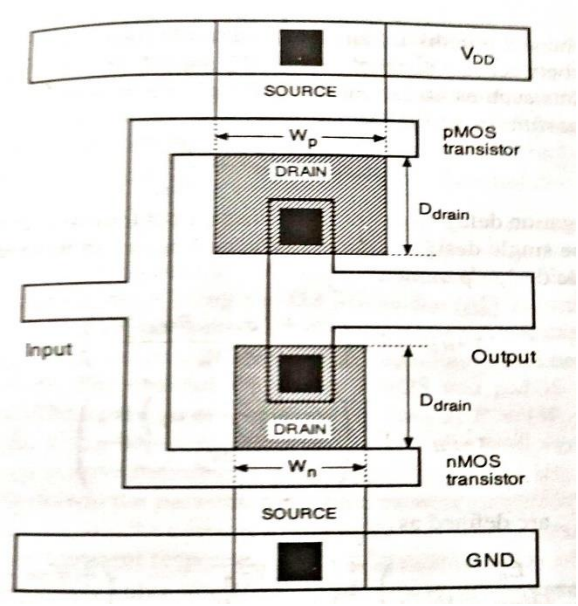
Similarly, the (W/L) ratio of the PMOS transistor to satisfy a given target value of τ_{PLH} can be calculated as

$$\frac{W_P}{L_P} = \frac{C_{load}}{\tau_{PLH}\mu_P C_{ox}(V_{DD} - |V_{T,P}|)}\left[\frac{2|V_{T,P}|}{V_{DD} - |V_{T,P}|} + ln\left(\frac{4(V_{DD} - |V_{T,P}|)}{V_{DD}} - 1\right)\right]$$

Assumed that the combined output load capacitance is mainly dominated by its extrinsic components, and hence, that is not very sensitive to device dimensions.

$$C_{load} = C_{gd,n}(W_n) + C_{gd,p}(W_p) + C_{db,n}(W_n) + C_{db,p}(W_p) + C_{int} + C_g$$
$$= \text{f}\,(W_n, W_p)$$

The fan out capacitance C_g is also a function of the device dimensions in the next stage gate.



Simplified CMOS inverter mask layout used for delay analysis

Here the diffusion areas of both NMOS and PMOS transistors have a simple rectangular geometry and the drain region length is assumed to be same for both devices. The relatively small gate to drain capacitances $C_{gd,n}$ *and* $C_{gd,p}$ will be neglected. The drain parasitic capacitances can be found as

$$C_{db,n} = W_n D_{drain} C_{j0,n} K_{eq,n} + 2(W_n + D_{drain}) C_{jsw,n} K_{eq,n}$$

$$C_{db,p} = W_p D_{drain} C_{j0,p} K_{eq,p} + 2(W_p + D_{drain}) C_{jsw,p} K_{eq,p}$$

Where $C_{j0,n}$ and $C_{j0,p}$ denote the zero bias junction capacitances for n-type and p-type diffusion regions, $C_{jsw,n}$ and $C_{jsw,p}$ denote the zero bias sidewall junction capacitances and $K_{eq,n}$and $K_{eq,p}$ denote the voltage equivalence factors. The combined output load capacitance then becomes

$$C_{load} = (W_n C_{j0,n} K_{eq,n} + W_p C_{j0,p} K_{eq,p}) D_{drain} + 2(W_n + D_{drain}) C_{jsw,n} K_{eq,n} + 2(W_p + D_{drain}) C_{jsw,p} K_{eq,p} + C_{int} + C_g$$

Thus the total capacitive load of the inverter can be expressed as

$C_{load} = \alpha_0 + \alpha_n W_n + \alpha_p W_p$

Where $\alpha_0 = 2D_{drain}(C_{jsw,n} K_{eq,n} + C_{jsw,p} K_{eq,p}) + C_{int} + C_g$

$\alpha_n = K_{eq,n}(C_{j0,n} D_{drain} + 2C_{jsw,n})$

$\alpha_p = K_{eq,p}(C_{j0,p} D_{drain} + 2C_{jsw,p})$

UNIT-4

STATIC COMBINATIONAL, SEQUENTIAL, DYNAMICS LOGIC CIRCUITS & MEMORIES

STATIC CMOS LOGIC CIRCUITS-

Static CMOS is a logic circuit design technique whereby the output is always strongly driven due to it always being connected to either VCC or GND (except when switching). This design is in contrast to Dynamic CMOS which relies on the temporary storage of signal using various load capacitances.

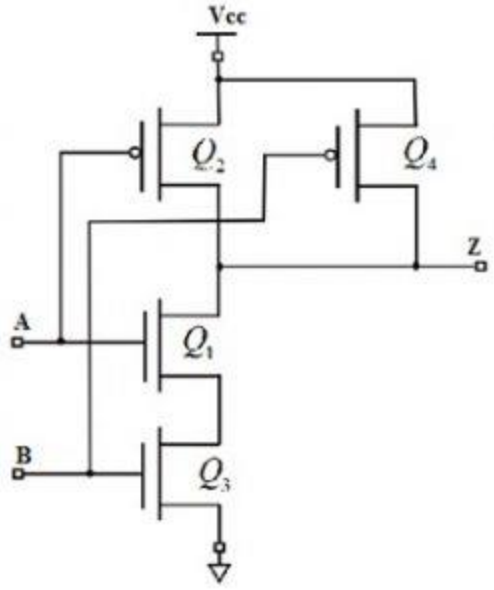
A static CMOS circuit is composed of two networks:

- pull-up network (PUN) - a set of PMOS transistors connected between V_{cc} and the output line
- pull-down network (PDN) - a set of NMOS transistors connected between GND and the output line

Components designed out pull-up and pull-down networks operate in a mutually exclusive way; in a steady state there is never a direct path between Vcc and GND. Devices that are made up of PUN/PDN are always strongly driven and therefore offers strong immunity from noise. When both the pull-up and pull-down networks are OFF, the result is high impedance. That state is important for memory elements, tristate bus drives, and various other components such as some multiplexers and buffers. When both the pull-up and pull-down networks are ON, the result is a crowbarred level. This result is typically an unwanted condition

CMOS NAND2 Gate-

- The below figure shows a 2-input Complementary MOS NAND gate. It consists of two series NMOS transistors between Y and Ground and two parallel PMOS transistors between Y and VDD.
- If either input A or B is logic 0, at least one of the NMOS transistors will be OFF, breaking the path from Y to Ground. But at least one of the PMOS transistors will be ON, creating a path from Y to VDD.



Two Input NAND Gate

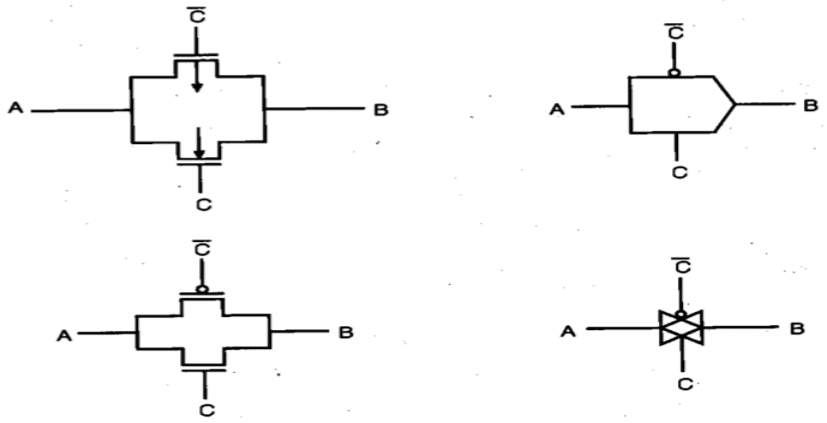
Hence, the output Y will be high. If both inputs are high, both of the nMOS transistors will be ON and both of the pMOS transistors will be OFF. Hence, the output will be logic low. The truth table of the NAND logic gate given in the below table.

A	B	Pull-Down Network	Pull-up Network	OUTPUT Y
0	0	OFF	ON	1
0	1	OFF	ON	1
1	0	OFF	ON	1
1	1	ON	OFF	0

CMOS TRANSMISSION GATES-

CMOS transmission gate consists of one NMOS and one PMOS transistor, connected in parallel. The gate voltages applied to these two transistors are also set to be complementary signals.

Symbols-

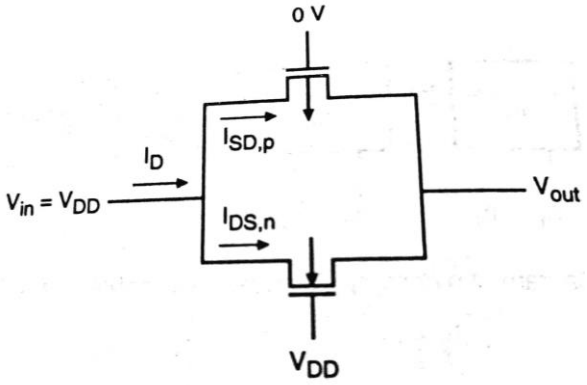


The CMOS transmission gate operates as a bidirectional switch between the nodes A and B which is controlled by signal C.

If the control signal C is

(i) Logic high i.e, equal toV_{DD}, then both transistors are turned on and provide a low resistance current path between the nodes A and B.

(ii) Logic low then both transistors will be off and the path between the nodes A and B will be an open circuit. This condition is called the high impedance state.



The substrate terminal of the NMOS transistor is connected to ground and the substrate terminal of the PMOS transistor is connected to V_{DD}.

C	A	B
0	0	High impedance State
0	1	High impedance State
1	0	0
1	1	1

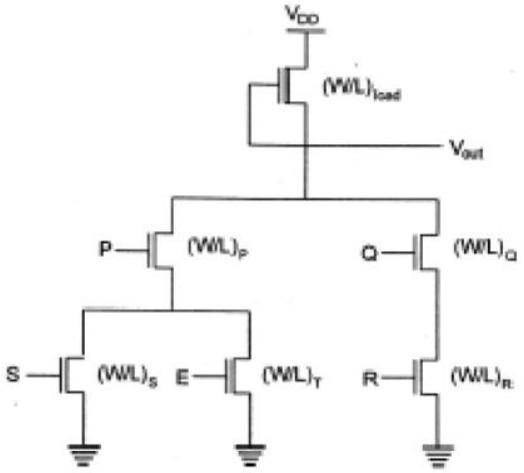
COMPLEX LOGIC CIRCUITS-

NMOS Depletion Load Complex Logic Gate

To realize complex functions of multiple input variables, the basic circuit structures and design principles developed for NOR and NAND can be extended to complex logic gates. The ability to realize complex logic functions, using a small number of transistors is one of the most attractive features of nMOS and CMOS logic circuits. Consider the following Boolean function as an example.

$$Z=(P(S+T)+QR)'$$

The nMOS depletion-load complex logic gate used to realize this function is shown in figure. In this figure, the left nMOS driver branch of three driver transistors is used to perform the logic function P (S + T), while the right-hand side branch performs the function QR. By connecting the two branches in parallel, and by placing the load transistor between the output node and the supply voltage $\mathbf{V_{DD}}$, we obtain the given complex function. Each input variable is assigned to only one driver.



Inspection of the circuit topology gives simple design principles of the pull-down network

- OR operations are performed by parallel-connected drivers.
- AND operations are performed by series-connected drivers.
- Inversion is provided by the nature of MOS circuit operation.

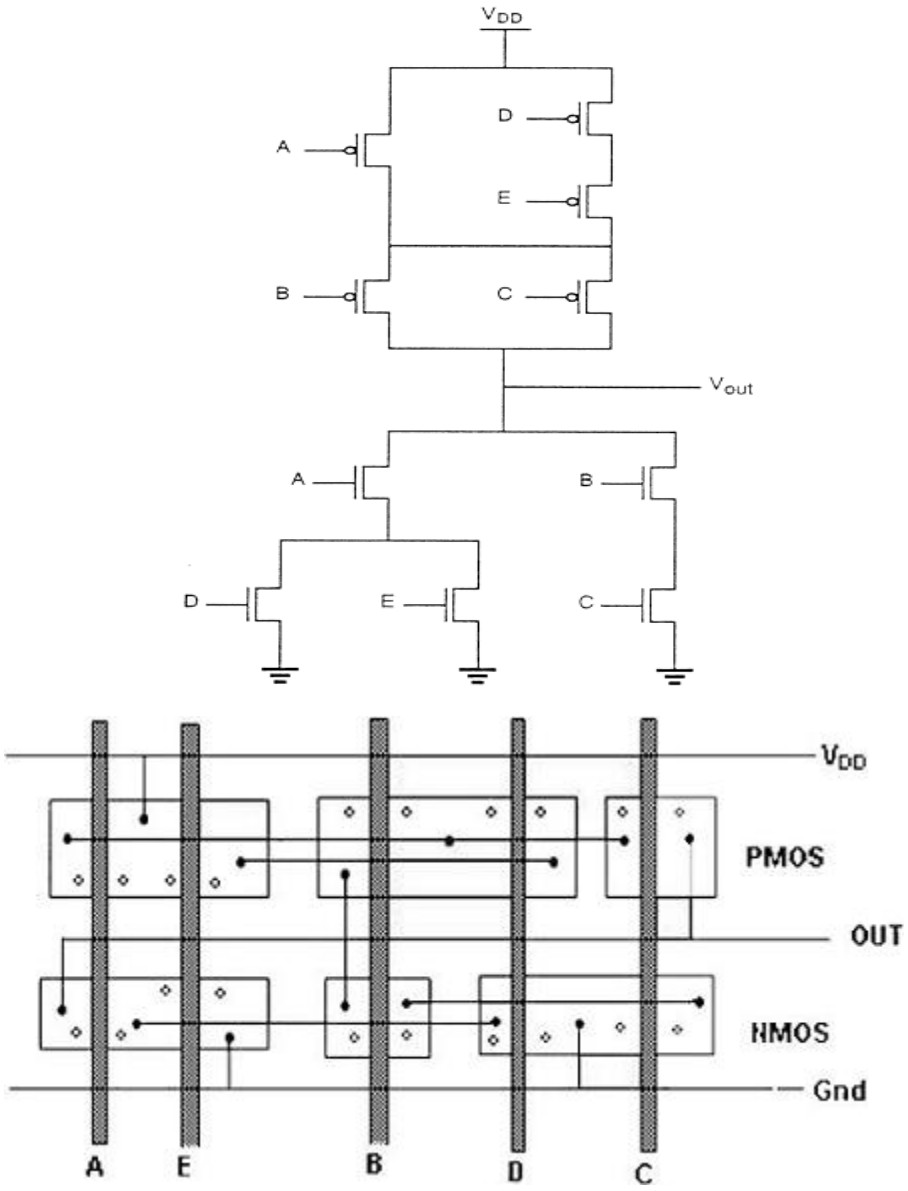
Complex CMOS Logic Gates-

The realization of the n-net, or pull-down network, is based on the same basic design principles examined for nMOS depletion-load complex logic gate. The pMOS pull-up network must be the dual network of the n-net.

It means all parallel connections in the nMOS network will correspond to a series connection in the pMOS network, and all series connection in the nMOS network correspond to a parallel connection in the pMOS network. The figure shows a simple construction of the dual p-net (pull-up) graph from the n-net (pull-down) graph.

Using an arbitrary ordering of the polysilicon gate columns-

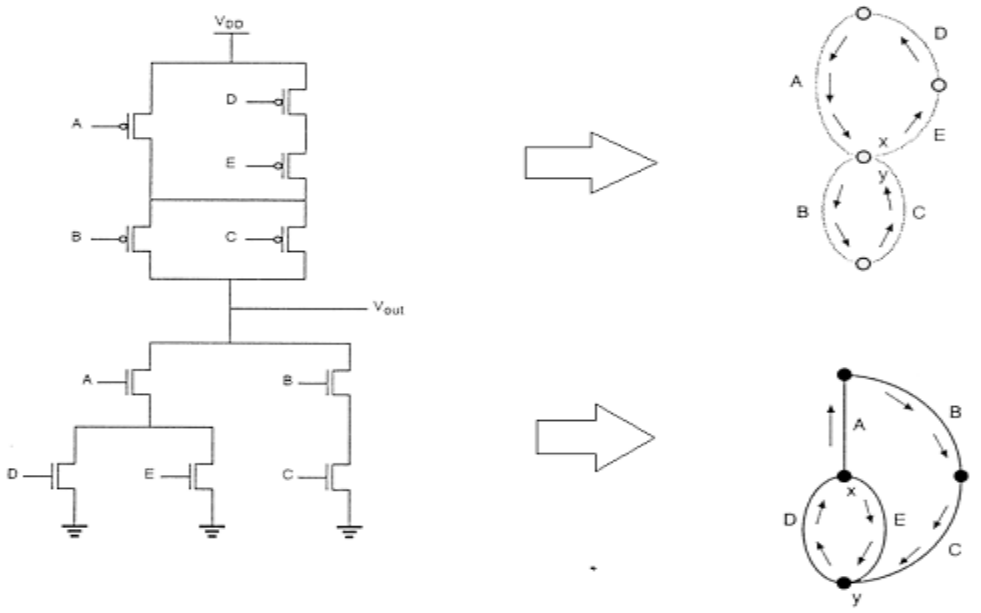
$$X= (A(D+E)+BC)'$$



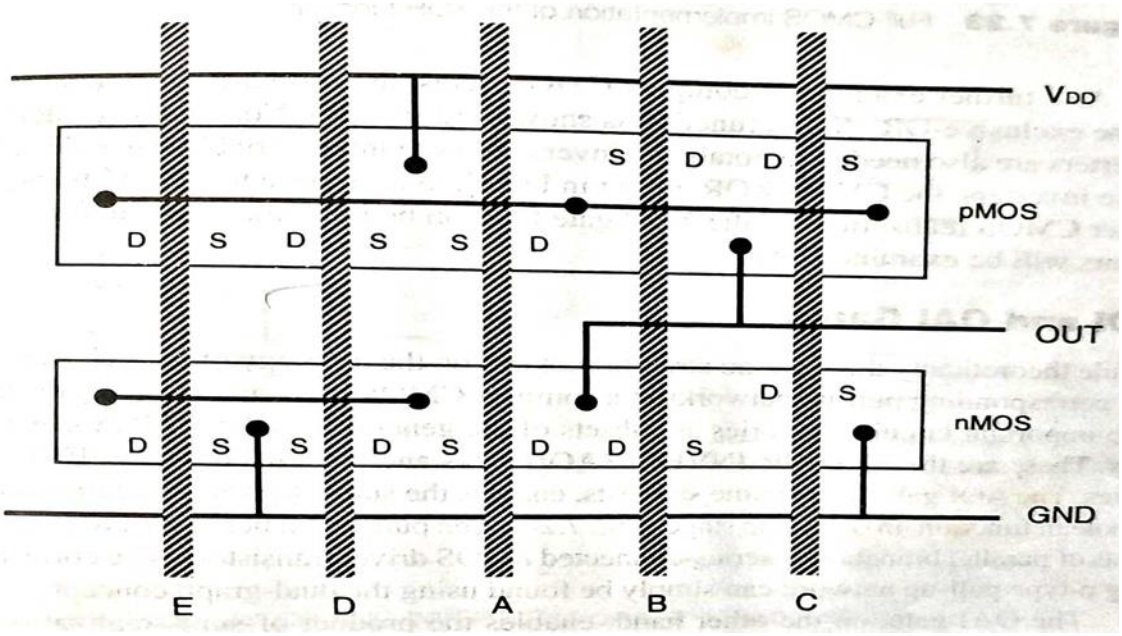
If we can minimize the number of diffusion area breaks both for NMOS and for PMOS transistors, the separation between the polysilicon gate columns can be made smaller, which will reduce the overall horizontal dimension and hence the circuit layout area. The number of diffusion breaks can be minimized by changing the ordering of the polysilicon columns.

A simple method for finding the optimum gate ordering is the Euler-path approach: find a Euler path in the pull down graph and a Euler path in the pull-up graph with identical ordering of input labels i.e, find a common Euler path for both graphs.

The Euler path is defined as an uninterrupted path that traverses each edge (branch) of the graph exactly once.

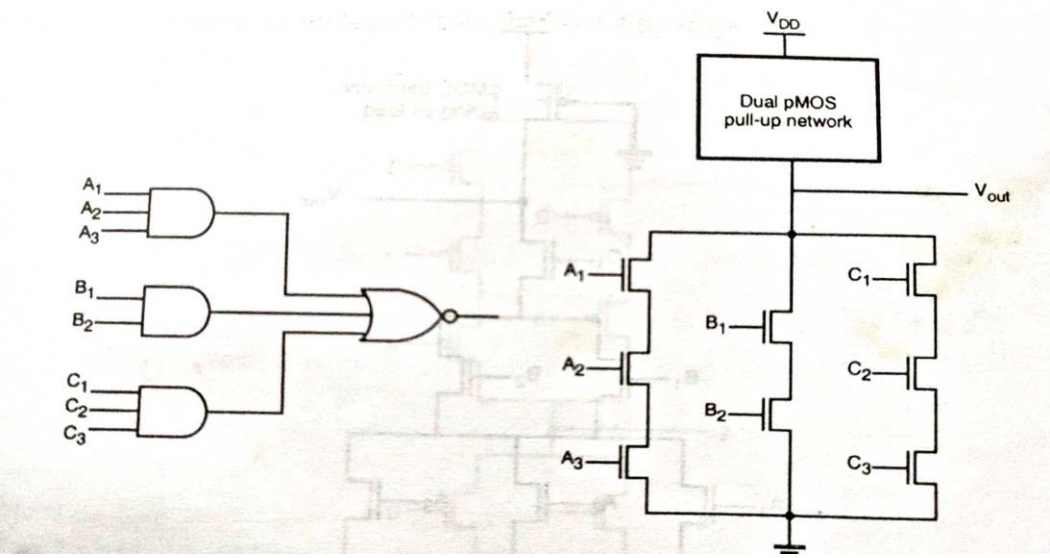


There is a common sequence (E-D-A-B-C) in both graphs i.e, a Euler path. The polysilicon gate columns can be arranged according to this sequence.

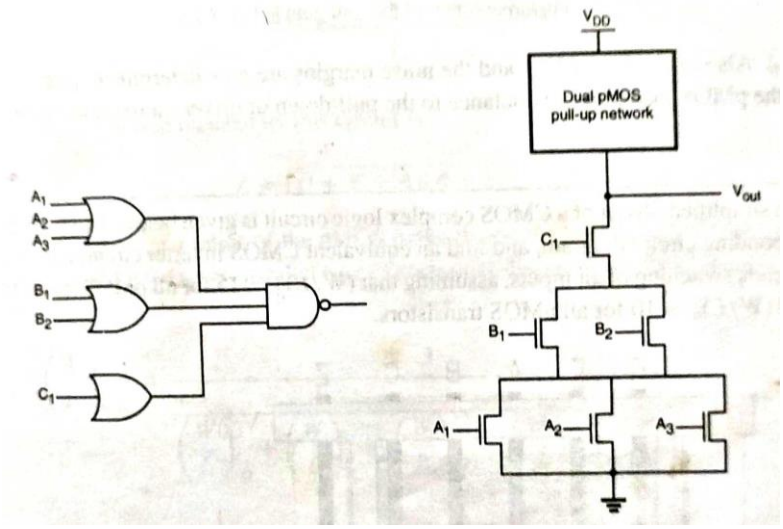


AOI and OAI gates-

- The AND-OR-INVERT (AOI) gate enables the sum of products realization of a Boolean function in one logic stage. The pull down network of the AOI gate consists of parallel branches of series connected NMOS driver transistors.

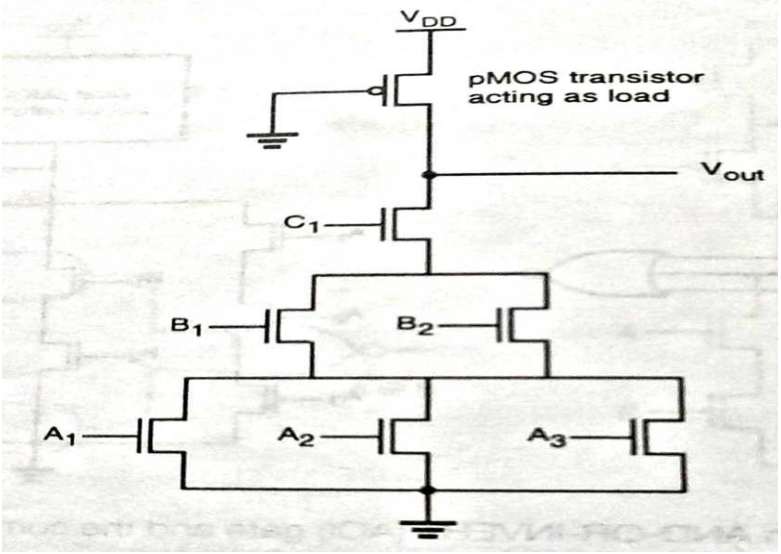


- The OAI gate, enables the product-of-sums realization of a Boolean function in one logic stage. The pull down network of the OAI gate consists of series branches of parallel connected NMOS driven transistors.



Pseudo-NMOS gates-

- The large area requirements of complex CMOS gates present a problem in high density designs, since two complementary transistors, one NMOS and one PMOS, are needed for every input.
- One possible approach to reduce the number of transistor is to use a single PMOS transistor, with its gate terminal connected to ground, as the load device.



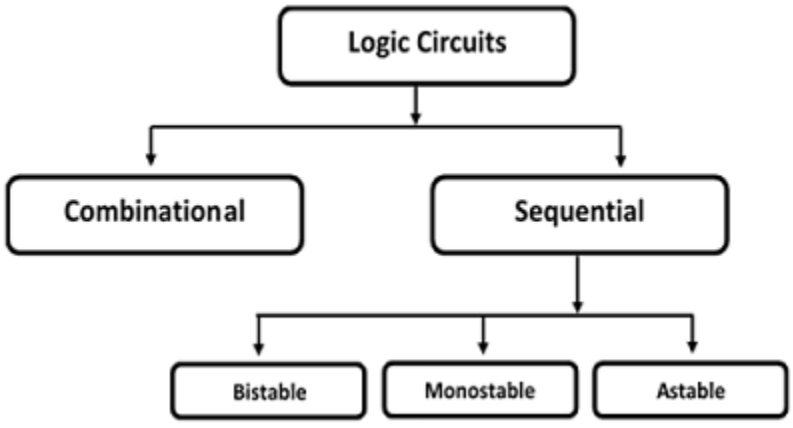
- With this simple pull up arrangement, the complex gate can be implemented with much fewer transistors.
- The disadvantages of using a pseudo NMOS gate instead of a full CMOS gate is the nonzero static power dissipation, since the always on PMOS load device conducts a steady state current when the output voltage is lower than V_{DD}.

CLASSIFICATION OF LOGIC CIRCUITS BASED ON THEIR TEMPORAL BEHAVIOUR-

Logic circuits are divided into two categories – (a) Combinational Circuits, and (b) Sequential Circuits.

In Combinational circuits, the output depends only on the condition of the latest inputs.

In Sequential circuits, the output depends not only on the latest inputs, but also on the condition of earlier inputs. Sequential circuits contain memory elements.



Sequential circuits are of three types –

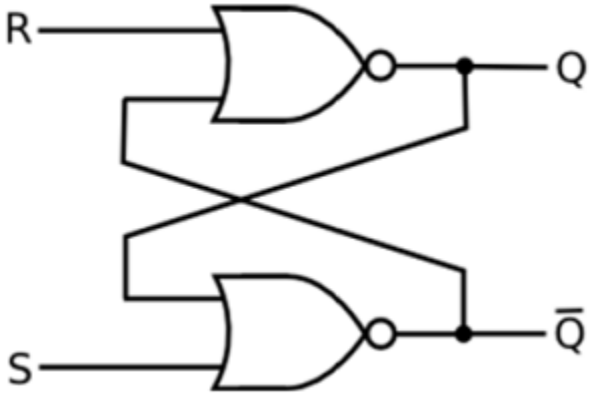
Bistable – Bistable circuits have two stable operating points and will be in either of the states. Example – Memory cells, latches, flip-flops and registers.

Monostable – Monostable circuits have only one stable operating point and even if they are temporarily perturbed to the opposite state, they will return in time to their stable operating point. Example: Timers, pulse generators.

Astable – Astable circuits have no stable operating point and oscillate between several states. Example – Ring oscillator.

SR LATCH CIRCUIT-

SR Latch based on NOR Gate-

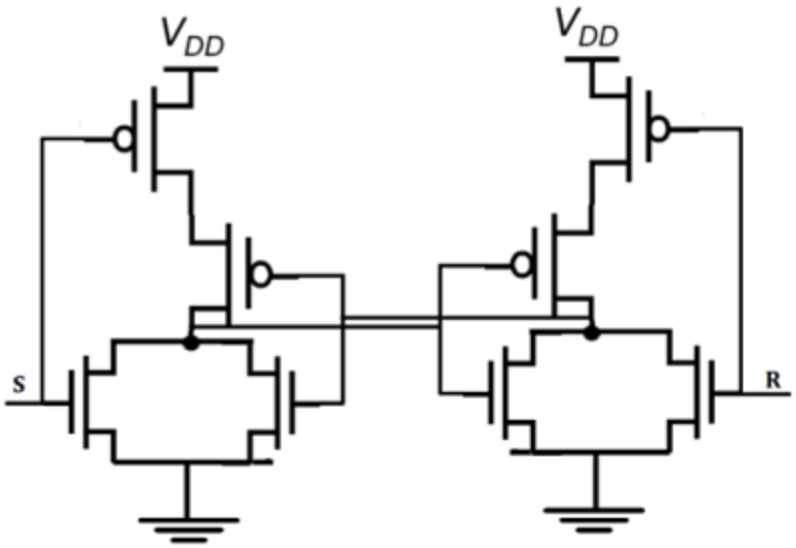


If the set input (S) is equal to logic **"1"** and the reset input is equal to logic **"0."** then the output Q will be forced to logic **"1"**. While Q' is forced to logic **"0"**. This means the SR latch will be set, irrespective of its previous state.
Similarly, if S is equal to "0" and R is equal to **"1"** then the output Q will be forced to **"0"** while Q' is forced to **"1"**. This means the latch is reset, regardless of its previously held state. Finally, if both of the inputs S and R are equal to logic **"1"** then both output will be forced to logic **"0"** which conflicts with the complementarity of Q and Q'.

Therefore, this input combination is not allowed during normal operation. Truth table of NOR based SR Latch is given in table.

S	R	Q	Q'	Operation
0	0	Q	Q'	Hold
1	0	1	0	Set
0	1	0	1	Reset
1	1	0	0	Not allowed

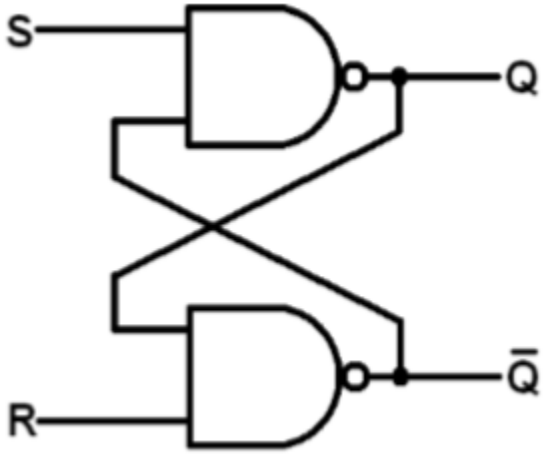
CMOS SR latch based on NOR gate is shown in the figure given below.



If the S is equal to V_{OH} and the R is equal to V_{OL}, both of the parallel-connected transistors M1 and M2 will be ON. The voltage on node Q' will assume a logic-low level of $V_{OL} = 0$.

At the same time, both M3 and M4 are turned off, which results in a logic-high voltage V_{OH} at node Q. If the R is equal to V_{OH} and the S is equal to V_{OL}, M1 and M2 turned off and M3 and M4 turned on.

SR Latch based on NAND Gate



Block diagram and gate level schematic of NAND based SR latch is shown in the figure. The small circles at the S and R input terminals represents that the circuit responds to active low input signals. The truth table of NAND based SR latch is given in table

S	**R**	**Q**	**Q′**	**OPERATION**
0	0	NC	NC	No change. Latch remained in present state.
1	0	1	0	Latch SET.

0	1	0	1	Latch RESET.
1	1	0	0	Invalid condition.

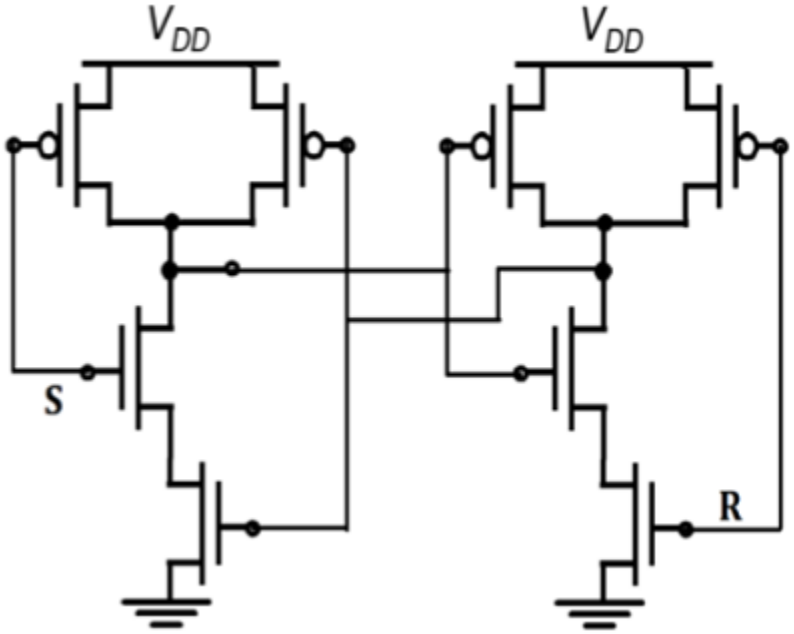
If S goes to 0 (while R = 1), Q goes high, pulling Q' low and the latch enters Set state

S = **0** then Q = **1** (if R = **1**)

If R goes to 0 (while S = 1), Q goes high, pulling Q' low and the latch is Reset

R = **0** then Q = **1** (if S = **1**)

Hold state requires both S and R to be high. If S = R = 0 then output is not allowed, as it would result in an indeterminate state. CMOS SR Latch based on NAND Gate is shown in figure.



Depletion-load nMOS SR Latch based on NAND Gate is shown in figure. The operation is similar to that of CMOS NAND SR latch. The CMOS circuit implementation has low static power dissipation and high noise margin.

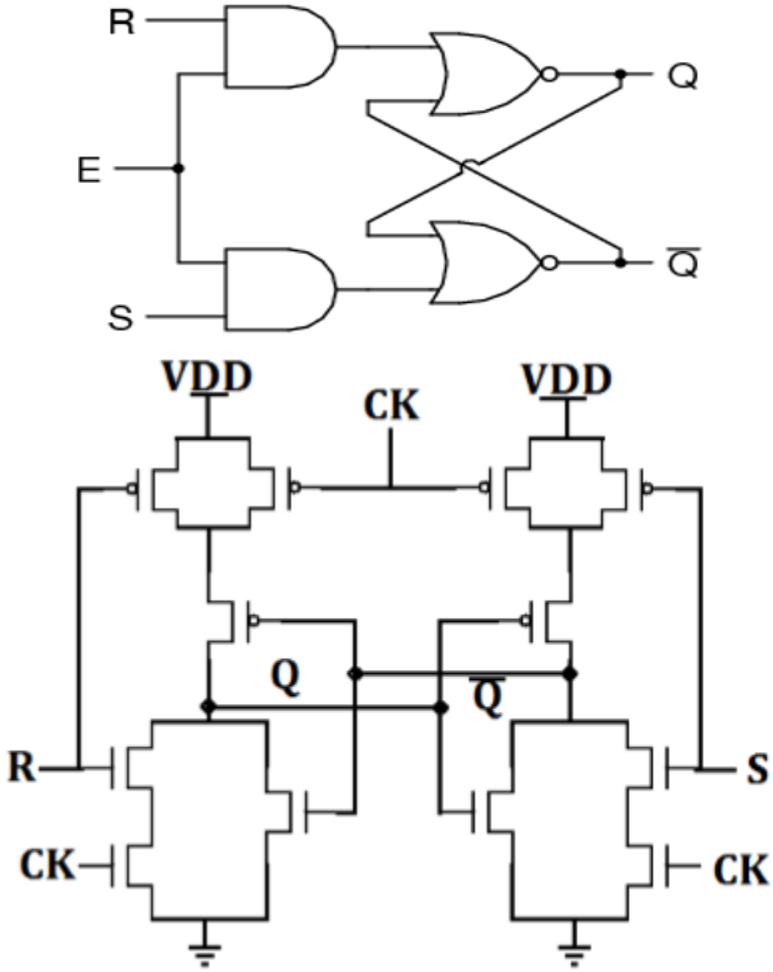
CLOCKED SR LATCH-

For synchronous operations the circuit response can be controlled by adding a gating clock signal to the circuit, so that the outputs will respond to the input levels only during the active period of a clock pulse.

If the clock (CK) is equal to logic “0”, the input signals have no influence upon the circuit response. The outputs of the two AND gates will remain at logic “0”, which forces SR latch to hold its current state regardless of the S and R input signals.

When the clock input goes to logic “1”, the logic levels applied to the S and R inputs are permitted to reach the SR latch and possibly change its state.

With both inputs S and R at logic “1”, the occurrence of clock pulse causes both outputs to go momentarily to zero. When the clock pulse is removed i.e, when it becomes “0”, the state of the latch is undermined.



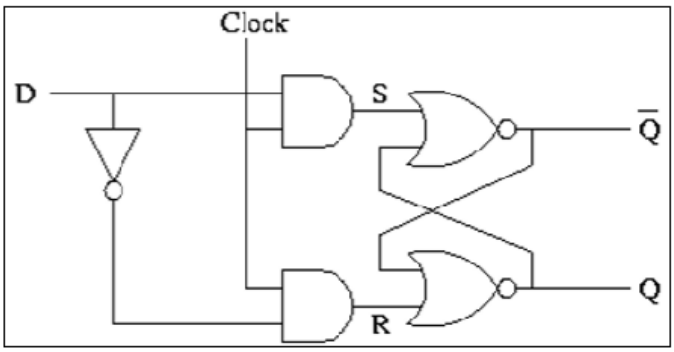
CMOS AOI implementation of clocked NOR based SR latch is shown in the figure. If this circuit is implemented with CMOS then it requires 12 transistors.

- When CLK is low, the latch retains its present state.
- When clock is high, the circuit becomes simply a NOR based CMOS latch which will respond to input S and R.

CMOS D LATCH-

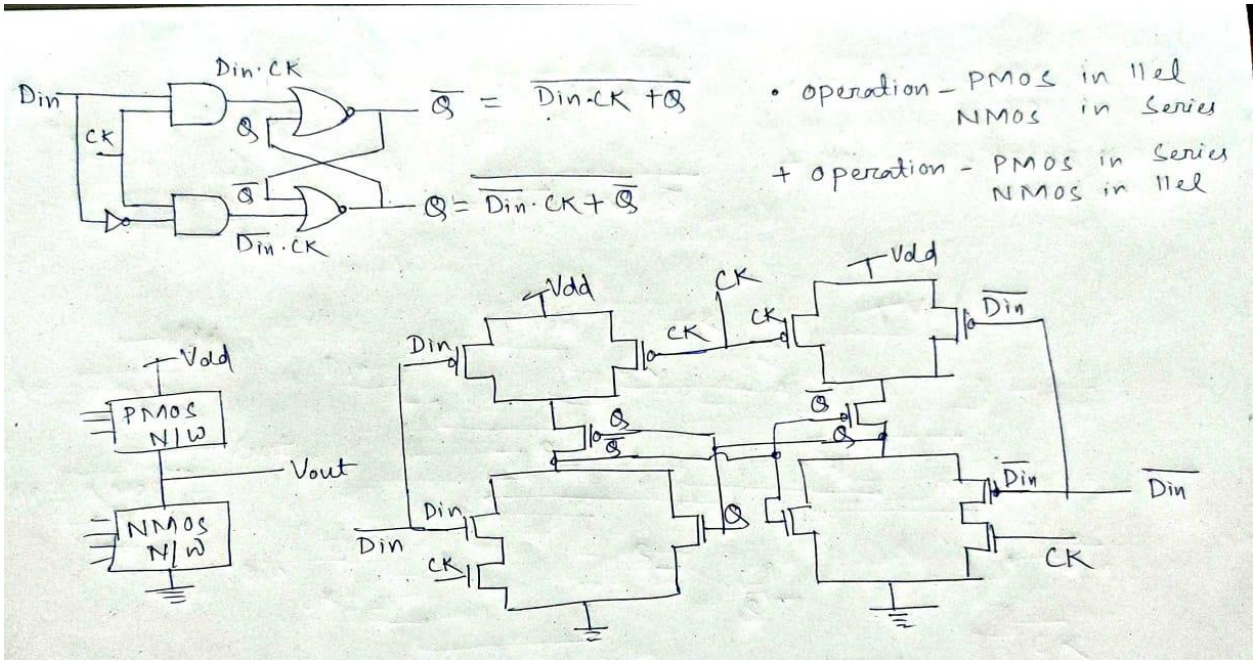
- The D latch is simply obtained by modifying the clocked NOR based SR latch circuit. Here, the circuit has a single input D, which is directly connected to the S input of the latch.
- The input variable D is also inverted and connected to the R input of the latch. The output Q assumes the value of the input D when the clock is active.

- When the clock signal goes to zero, the output will preserve its state. Thus the CK input acts as an enable signal which allows data to be accepted into the D latch.
- The D latch finds many applications mainly for temporary storage of data or as a delay element.



D latch using CMOS logic-

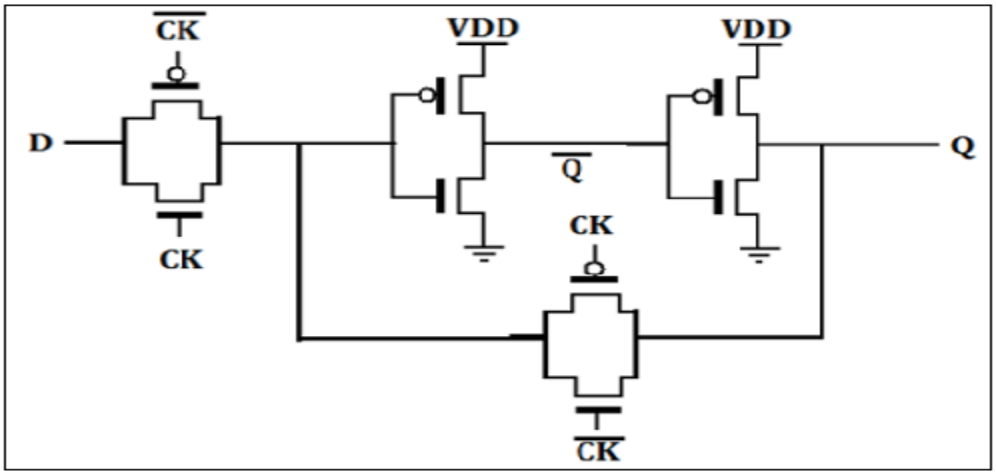
. If this circuit is implemented with CMOS then it requires 12 transistors.



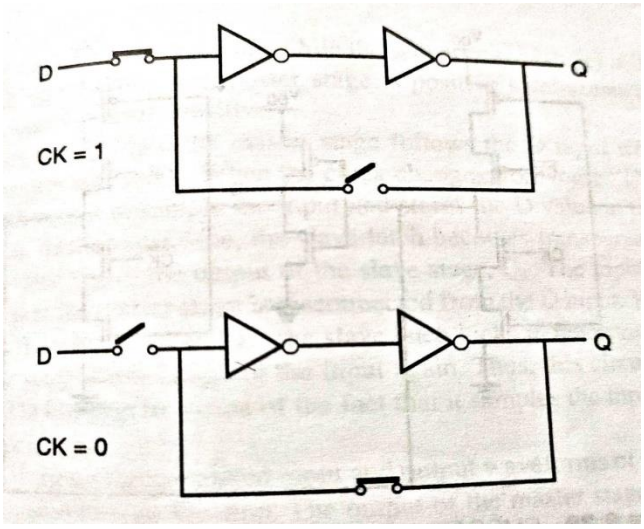
D latch using transmission gate-

The transmission gate at the input is activated by the CK signal, whereas the transmission gate in the inverter loop is activated by the inverse of the CK signal.

The input signal is accepted into the circuit when the clock is high and this information preserved as the state of the inverter loop when the clock is low.



The operation of the CMOS D latch circuit can be better visualized by replacing the CMOS transmission gates with simple switches.



BASIC PRINCIPLES OF DYNAMIC PASS TRANSISTOR CIRCUITS-

In static CMOS logic, logic function implemented using large no. of transistors and which may cause large time delay.

In a high performance digital implementations where reduction of circuit delay and silicon area is a major objective to achieve these dynamic logic circuits are used.

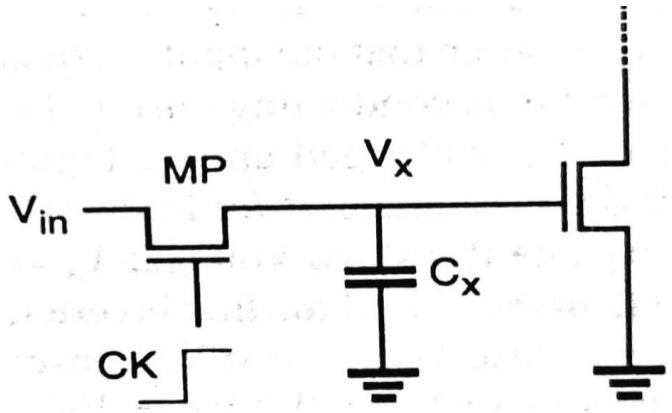
Here no. of transistor used decreases. The operation of all dynamic logic gates depends on temporary storage of charge in parasitic node capacitances.

Basic Principle

NMOS dynamic logic circuits, consisting of an NMOS pass transistor driving the gate of another NMOS transistor.

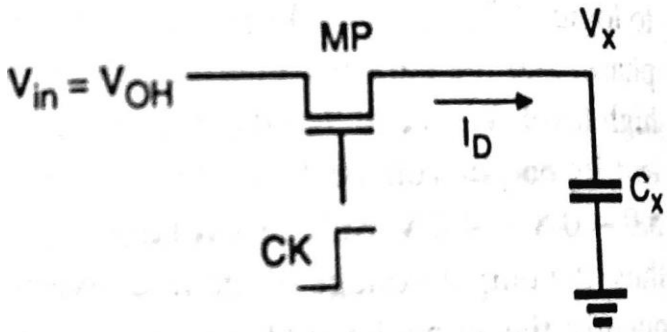
The pass transistor MP is driven by the periodic clock signal and acts as an access switch to either charge up or charge down the parasitic capacitanceC_X, depending on the input signal V_{in}.

Thus, the two possible operations when the clock signal is active (CK = 1) are the logic "1" transfer and the logic "0" transfer.



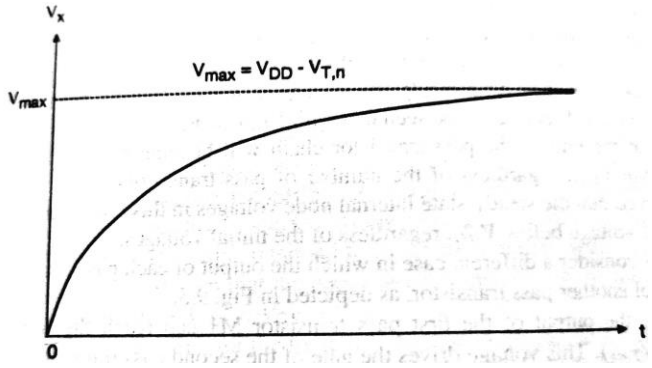
Logic "1" transfer-

Assume that the V_X node voltage is equal to 0 initially. A logic "1" level is applied to the input terminal, which corresponds to $V_{in} = V_{OH} = V_{DD}$. When the clock signal at the gate of the pass transistor becomes active, the pass transistor MP starts to conduct and that MP will operate in saturation.



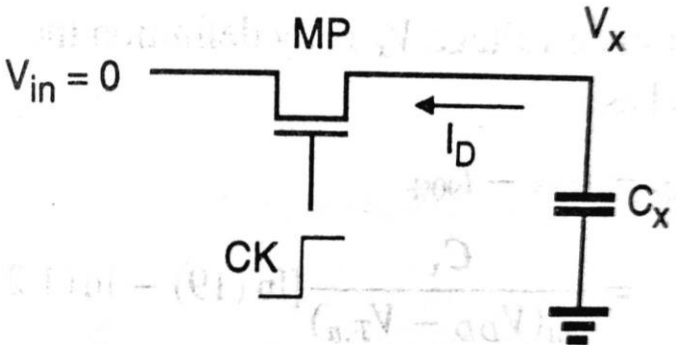
The voltage rises from its initial value of 0V and approaches a limit value for large t, but it cannot exceed its limit value of $V_{max} = V_{DD} - V_{T,n}$.

The pass transistor will turn off when $V_X = V_{max}$, since at this point , its gate to source voltage will be equal to its threshold voltage. Therefore the voltage at node X can never attain the full power supply voltage level of V_{DD} during the logic "1" transfer.

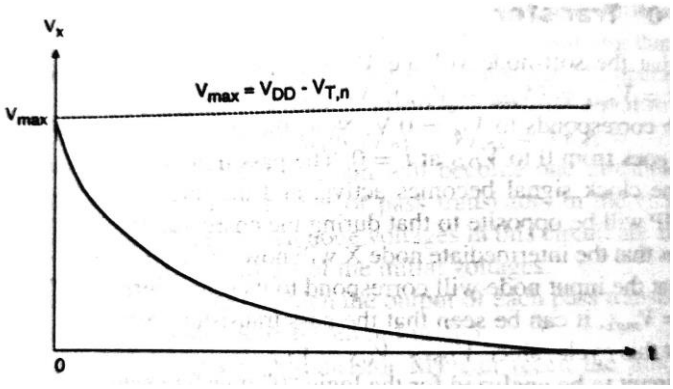


Logic "0" transfer-

Assume that the node voltage V_X is equal to a logic "1" level initially i.e, $V_X(t = 0) = V_{max} = (V_{DD} - V_{T,n})$.A logic "0" level is applied to the input terminal, which corresponds to V_{in} =0V.



The pass transistor MP starts to conduct as soon as the clock signal becomes active and the direction of drain current flow through MP will be opposite to that during the charge up (logic "1" transfer).



RAM-

The read/write memory is commonly called Random Access Memory(RAM). Here the stored data is volatile i.e, the stored data is lost when the power supply voltage is turned off.

Based on the operation type of individual data storage cells, RAMs are classified into two categories-

1) Dynamic RAMs (DRAM)

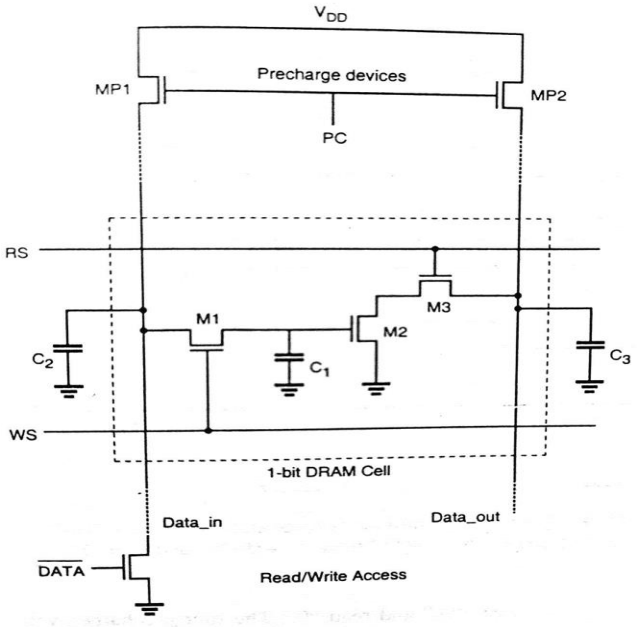
- The DRAM cell consists of a capacitor to store binary information, 1 or 0 and a transistor to access the capacitor.
- Cell information is degraded mostly due to a junction leakage current at the storage node. Therefore the cell data must be read and rewritten periodically (refresh operation) even when memory arrays are not accessed.
- Due to advantage of low cost and high density, DRAM is widely used for the main memory in personal and mainframe computers and engineering workstations

2) Static RAMs (SRAM)

- SRAM cell consists of a latch, therefore the cell data is kept as long as the power is turned on and refresh operation is not required.
- SRAM is mainly used for the cache memory in microprocessors, mainframe computers, engineering workstations due to high speed and low power consumption.

DYNAMIC RAM-

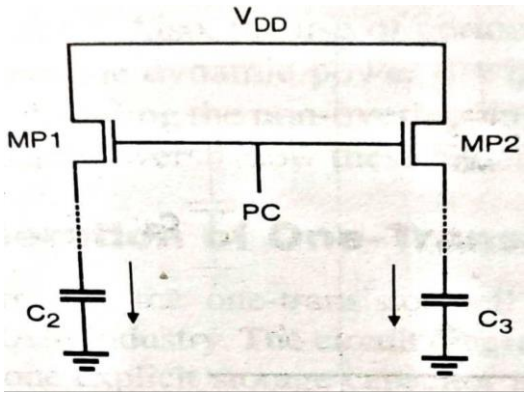
The binary information is stored in the form of charge in the parasitic node capacitance C1. The storage transistor M2 is turned on or off depending on the charge stored in C1, and the pass transistors M1 and M3 act as access switches for data read and write operations.



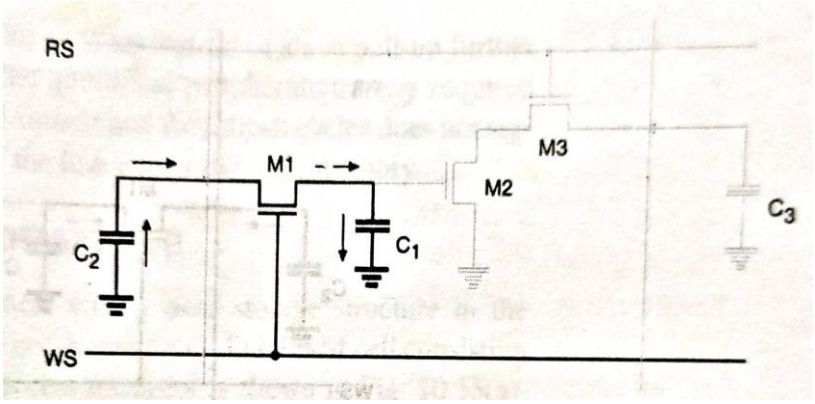
- The operation of the three transistor DRAM cell and the peripheral circuitry is based on a two phase non- overlapping clock scheme.

1) The precharge events are driven by $\emptyset_1$.
2) The "read" and "write" events are driven by $\emptyset_2$.

- Every “data read” and “data write” operation is preceded by a precharge cycle, which is initiated with the precharge signal PC going high.
- During the precharge cycle, the column pull up transistors are activated and the corresponding column capacitances C2 and C3 are charged up to logic high level.

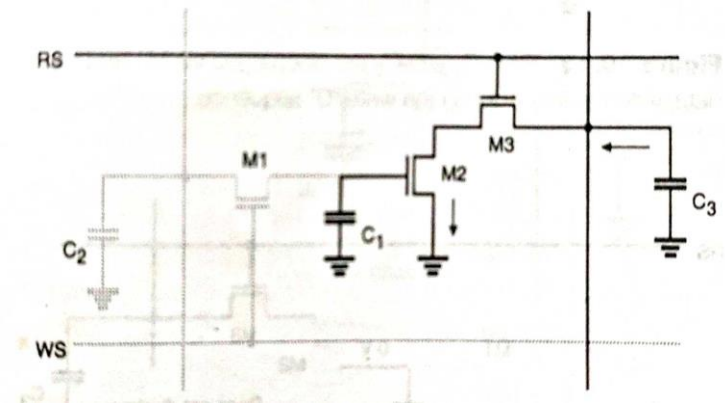


Write “1”-



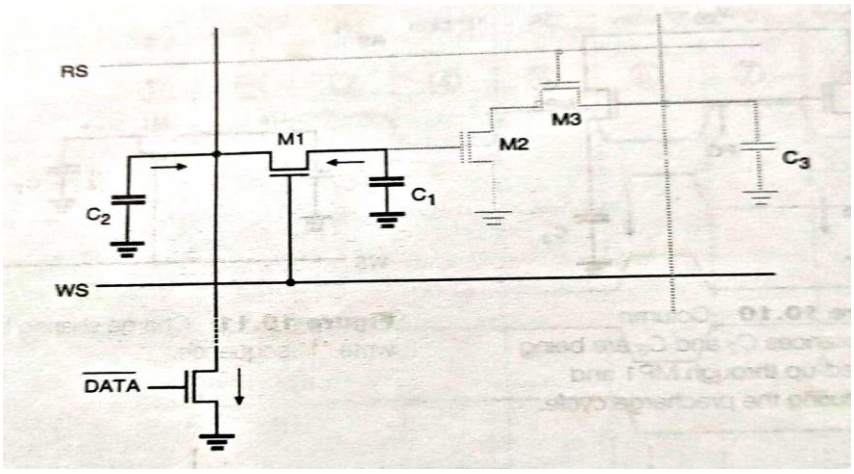
- The inverse data input is at the logic low level, because the data to be written onto the DRAM cell is logic “1”.
- The “write select” signal WS is pulled high during the active phase of $\emptyset_2$.
- The transistor M1 is turned on. With M1 conducting, the charge on C2 is shared with C1.
- Since the capacitance C2 is very large compared to C1, the storage node capacitance C1 attains approximately the same logic high level as the column capacitance C2 at the end of the charge sharing process.

Read “1”-



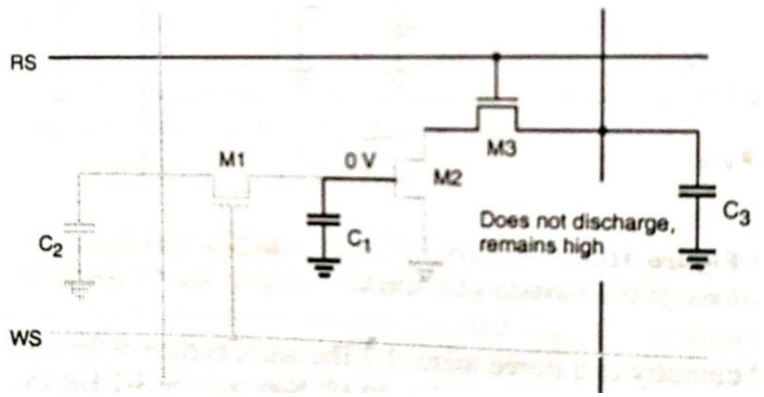
- With the storage node capacitance C1 charged up to a logic high level, transistor M2 is conducting.
- In order to read this stored "1", the "read select" signal RS pulled high during the active phase of $\emptyset_2$, following a precharge cycle.
- As the transistor M3 turns on, M2 and M3 create a conducting path between the column capacitance C3 and the ground.
- The capacitance C3 discharges through M2 and M3, and the falling column voltage is interpreted as a stored logic "1".

Write "0"-



- The inverse data input is at the logic high level, because the data to be written onto the DRAM cell is a logic "0".
- The "write select" signal WS is pulled high during the active phase of $\emptyset_2$, following a precharge cycle.
- As a result, the transistor M1 is turned on. The voltage level on C2, as well as that on the storage node C1, is pulled to logic "0" through M1.
- At the end of the write "0" sequence, the storage capacitance C1 contains a very low charge and the transistor M2 is turned off since its gate voltage is approximately equal to zero.

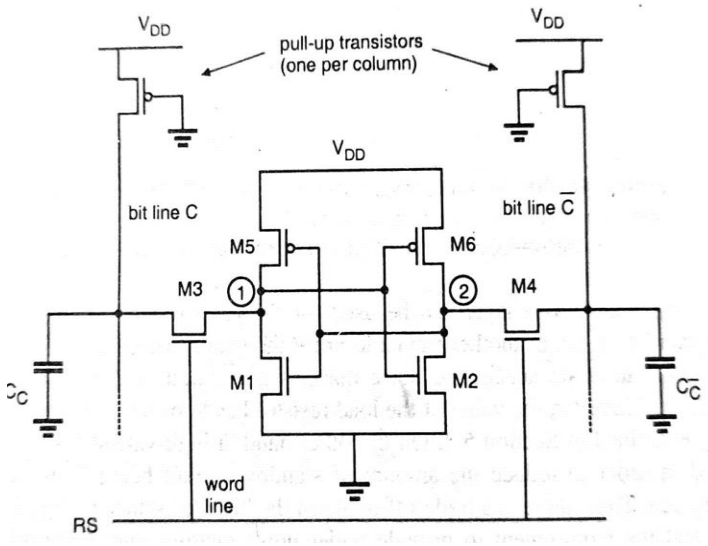
Read "0"-



- In order to read this stored "0", the read select signal RS pulled high during the active phase $\emptyset_2$, following a precharge cycle.
- The transistor M3 turns on, but since M2 is off, there is no conducting path between the C3 and ground. So, C3 does not discharge and the logic high level on the D_{out} column is interpreted as a stored "0" bit.

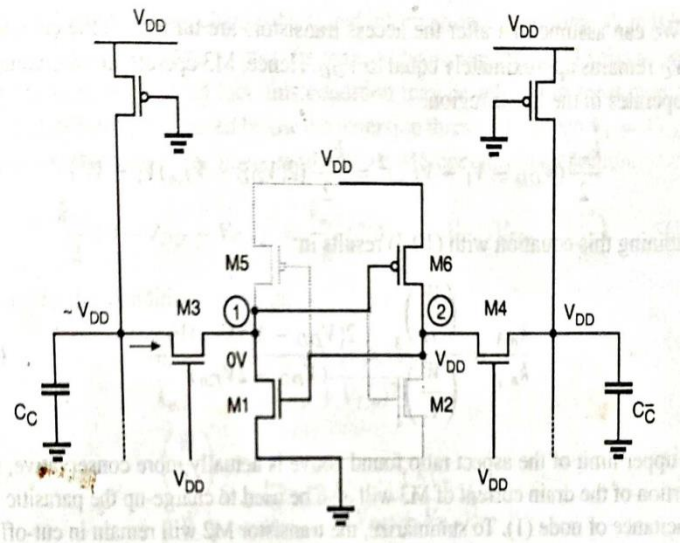
STATIC RAM (SRAM)-

A low power SRAM cell designed simply by using cross coupled CMOS inverters. The memory cell consists of a simple CMOS latch (two inverters connected back to back), and two complementary access transistors (M3 and M4).



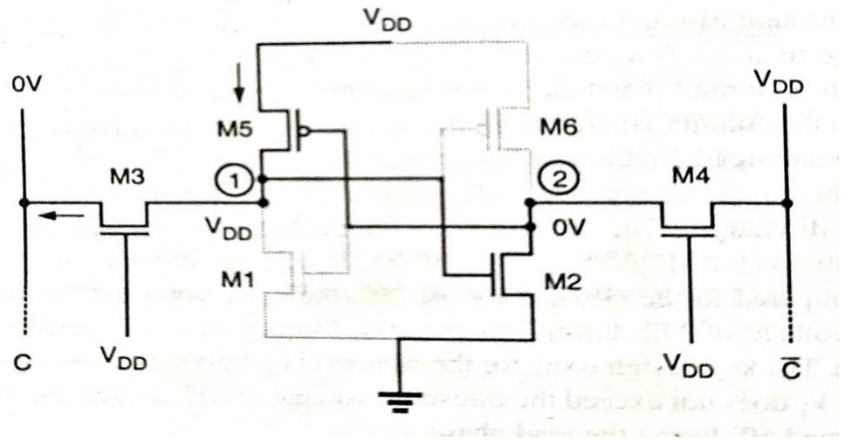
The cell will preserve one of its two possible stable states as long as the power supply is available. The access transistors are turned on whenever a word line is activated for read or write operation, connecting the cell to the complementary bit line columns.

Read "0" operation-



- Assuming that a logic "0" is stored in the cell. Here the transistors M2 and M5= off and M1 and M6= On (operate in linear mode).
- The internal node voltages are V_1 =0V and $V_2 = V_{DD}$ before M3 and M4 are turned on.
- The voltage level of column C' will not show any significant variation since no current flow through M4.
- On the other half of the cell, M3 and M1 will conduct a non-zero current and the voltage level of column C will begin to drop slightly.
- The capacitance C_c is very large, therefore the amount of decrease in the column voltage is limited to a few hundred millivolts during read phase.

Write "0" operation-



- Assuming that a logic "1" is stored in the SRAM cell initially. Here the transistors M1 and M6= off and M2 and M5= On (operate in linear mode).
- The internal node voltages are $V_1 = V_{DD}$ and $V_2 = 0V$ before M3 and M4 are turned on.

- The column voltage V_c is forced to logic "0" level by the data write circuitry. Once the pass transistors M3 and M4 are turned on, we expect that the node voltage V_2 remains below the threshold voltage of M1.
- To change the stored information i.e, to force $V_1 to\ 0V$ and $V_2 to\ V_{DD}$, the node voltage V_1 must be reduced below the threshold voltage of M2 . so that M2 turns off.
- Similarly read "1" and write "0" operation can be done.

Basic Requirements-

The two basic requirements which dictate the (W/L) ratios are-

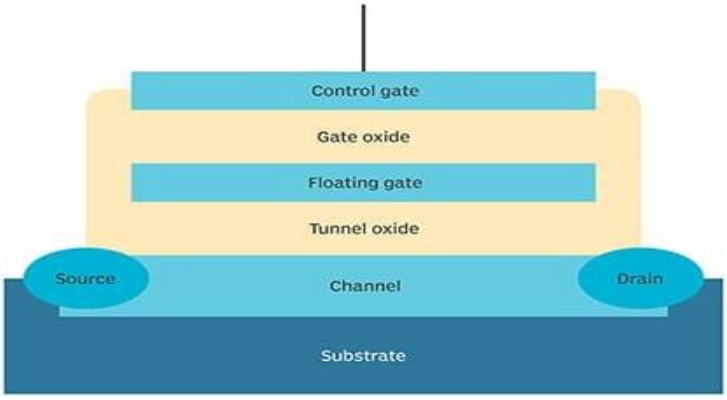
(a) The data read operation should not destroy the stored information in the SRAM cell.
(b) The cell should allow modification of the stored information during the data write phase.

Advantages-

1) The static power dissipation is very small.
2) High noise immunity due to larger noise margins and the ability to operate at lower power supply voltages.

FLASH MEMORY-

Flash memory is a non- volatile memory chip used for storage. Flash memory is a type of Electronically Erasable Programmable Read Only Memory (EEPROM).



In flash memory, each memory cell looks like standard MOSFET except that the transistor has two gates instead of one.

The cells can be seen as an electrical switch in which current flows between two terminals and is controlled by a floating gate and a control gate.

The control gate is similar to the gate in the MOS transistors, but below this there is the floating gate insulated all around by an oxide layer.

The Floating gate is electrically isolated by its insulating layer, electrons placed on it are trapped. This makes flash memory non-volatile.

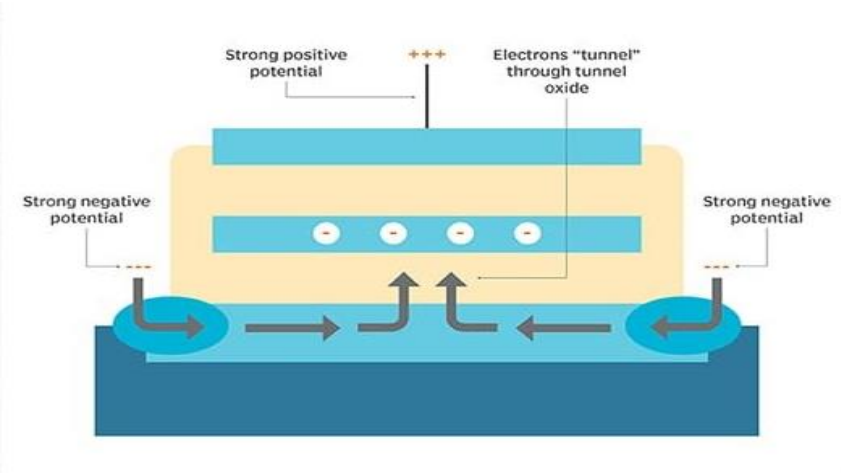
It works by adding or removing electrons to and from a floating gate. A bit “0” or “1” state depends upon whether or not the floating gate is charged or uncharged.

When electrons are present on the floating gate, current cannot flow through the transistor and the bit state is “0”.

When electrons are removed from the floating gate, current is allowed to flow and the bit state is “1”.

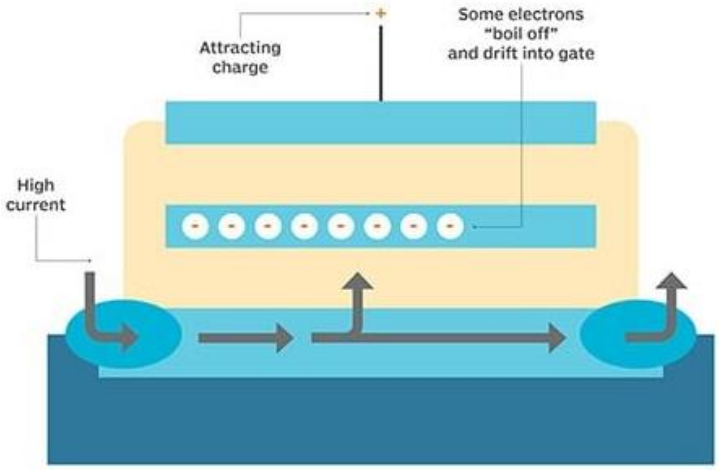
Two processes are used to add electrons in the floating gate:

1) Fowler Nordheim tunneling –



- It requires a strong electric field between negatively charged source and the positively charged control gate to draw electrons into the floating gate.
- The electrons move from the source through the thin oxide layer to the floating gate, where they are trapped between the oxide insulation layers.

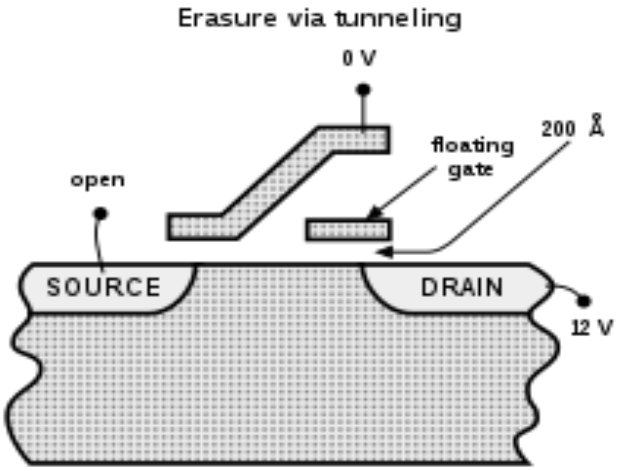
2) Hot electron injection-



- It uses a high current in the channel to give electrons sufficient energy to break through the oxide layer.

- A positive charge on the control gate attracts the electrons from the channel into the floating gate, where they become trapped.

Fowler-Nordheim tunneling is also used to remove electrons from the floating gate. A strong negative charge on the control gate forces electrons through the oxide layer into the channel, where the electrons are drawn to the strong positive charge at the source and the drain



Types of flash memory-

1) NOR flash memory
2) NAND flash memory

- In both, memory cells are arranged differently. In a NAND memory chip, all floating gate MOSFETs are organized in series. Here bit line is pulled low only if all the word lines are pulled high.
- In a NOR flash, at least one memory cell must conduct in order to pull down the bit line, because they are connected in parallel to ground.
- NOR flash uses more space than NAND to save the same amount of information, since two flash cells share the same ground potential in this configuration. Therefore NAND is cheaper.
- NOR memory needs less time to read a bit because of its direct access to individual cells.