

UNIT-I

Introduction to VLSI Technology

Introduction:

The invention of the transistor by William B. Shockley, Walter H. Brattain and John Bardeen of Bell Telephone Laboratories drastically changed the electronics industry and paved the way for the development of the Integrated Circuit (IC) technology. The first IC was designed by Jack Kilby at Texas Instruments at the beginning of 1960 and since that time there have already been four generations of ICs. Viz SSI (small scale integration), MSI (medium scale integration), LSI (large scale integration), and VLSI (very large scale integration). Now we are ready to see the emergence of the fifth generation, ULSI (ultra large scale integration) which is characterized by complexities in excess of 3 million devices on a single IC chip. Further miniaturization is still to come and more revolutionary advances in the application of this technology must inevitably occur.

Over the past several years, Silicon CMOS technology has become the dominant fabrication process for relatively high performance and cost effective VLSI circuits. The revolutionary nature of this development is understood by the rapid growth in which the number of transistors integrated in circuits on a single chip.

METAL-OXIDE-SEMICONDUCTOR (MOS) AND RELATED VLSI TECHNOLOGY:

The MOS technology is considered as one of the very important and promising technologies in the VLSI design process. The circuit designs are realized based on PMOS, NMOS, CMOS and BiCMOS devices.

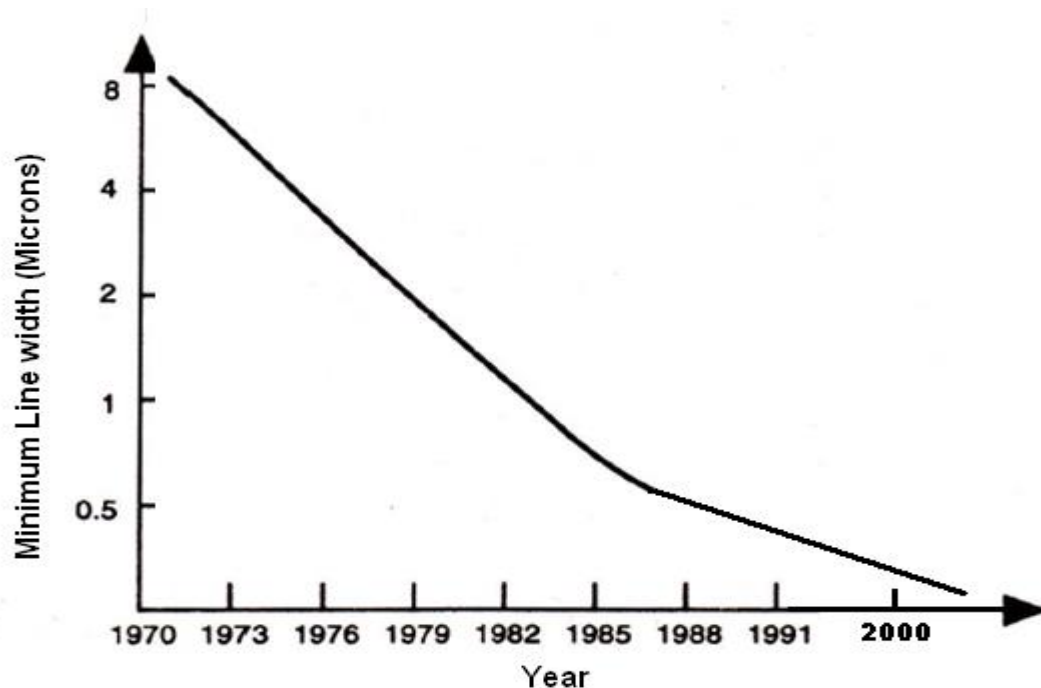
The PMOS devices are based on the p-channel MOS transistors. Specifically, the PMOS channel is part of a n-type substrate lying between two heavily doped p+ wells beneath the source and drain electrodes. Generally speaking, a PMOS transistor is only constructed in consort with an NMOS transistor.

The NMOS technology and design processes provide an excellent background for other technologies. In particular, some familiarity with NMOS allows a relatively easy transition to CMOS technology and design.

The techniques employed in NMOS technology for logic design are similar to GaAs technology.. Therefore, understanding the basics of NMOS design will help in the layout of GaAs circuits

In addition to VLSI technology, the VLSI design processes also provides a new degree of freedom for designers which helps for the significant developments. With the rapid advances in technology the size of the ICs is shrinking and the integration density is increasing.

The minimum line width of commercial products over the years is shown in the graph below.

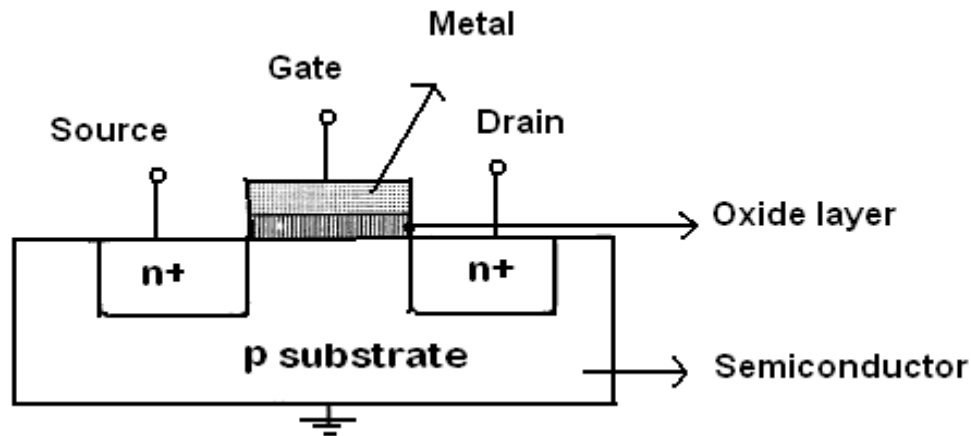


The graph shows a significant decrease in the size of the chip in recent years which implicitly indicates the advancements in the VLSI technology.

BASIC MOS TRANSISTORS:

The MOS Transistor means, Metal-Oxide-Semiconductor Field Effect Transistor which is the most basic element in the design of a large scale integrated circuits(IC).

These transistors are formed as a "sandwich" consisting of a semiconductor layer, usually a slice, or wafer, from a single crystal of silicon; a layer of silicon dioxide (the oxide) and a layer of metal. These layers are patterned in a manner which permits transistors to be formed in the semiconductor material (the "substrate"); a diagram showing a MOSFET is shown below in Figure .



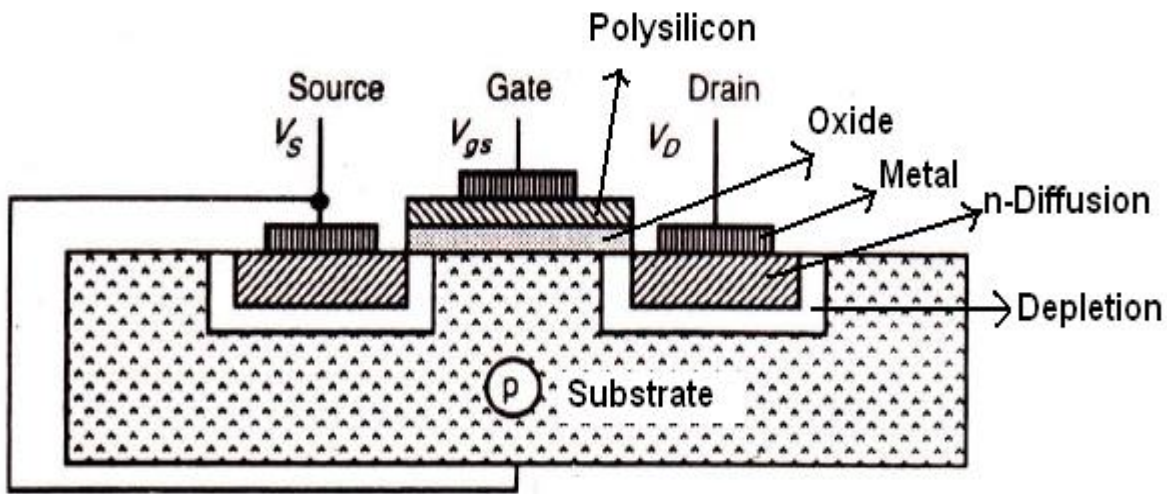
Silicon dioxide is a very good insulator, so a very thin layer, typically only a few hundred molecules thick, is used. In fact , the transistors which are used do not use metal for their gate regions, but instead use polycrystalline silicon (poly). Polysilicon gate FET's have replaced virtually all of the older devices using metal gates in large scale integrated circuits. (Both metal and polysilicon FET's are sometimes referred to as IGFET's (insulated gate field effect transistors), since the silicon dioxide under the gate is an insulator.

MOS Transistors are classified as n-MOS, p-MOS and c-MOS Transistors based on the fabrication .

NMOS devices are formed in a p-type substrate of moderate doping level. The source and drain regions are formed by diffusing n- type impurities through suitable masks into these areas to give the desired n-impurity concentration and give rise to depletion regions which extend mainly in the more lightly doped p-region . Thus, source and drain are isolated from one another by two diodes. Connections to the source and drain are made by a deposited metal layer. In order to make a useful device, there must be the capability for establishing and controlling a current between source and drain, and .this is commonly achieved in one of two ways, giving rise to the enhancement mode and depletion mode transistors.

Enhancement Mode Transistors:

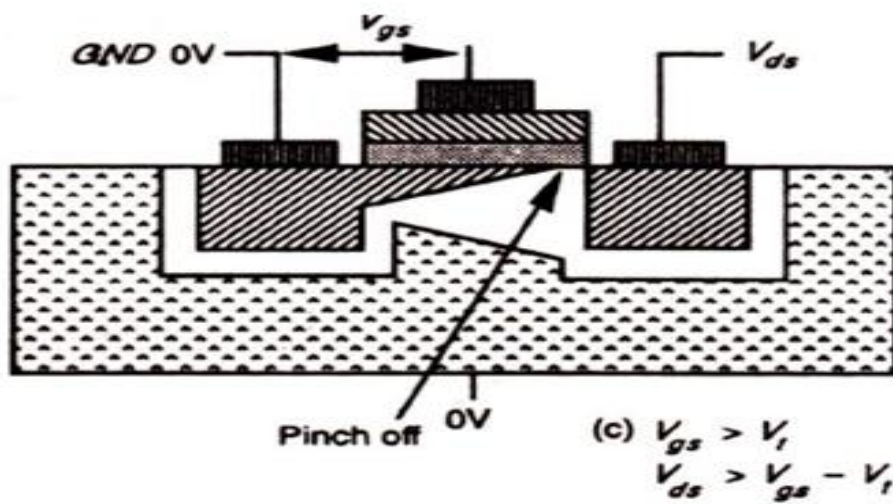
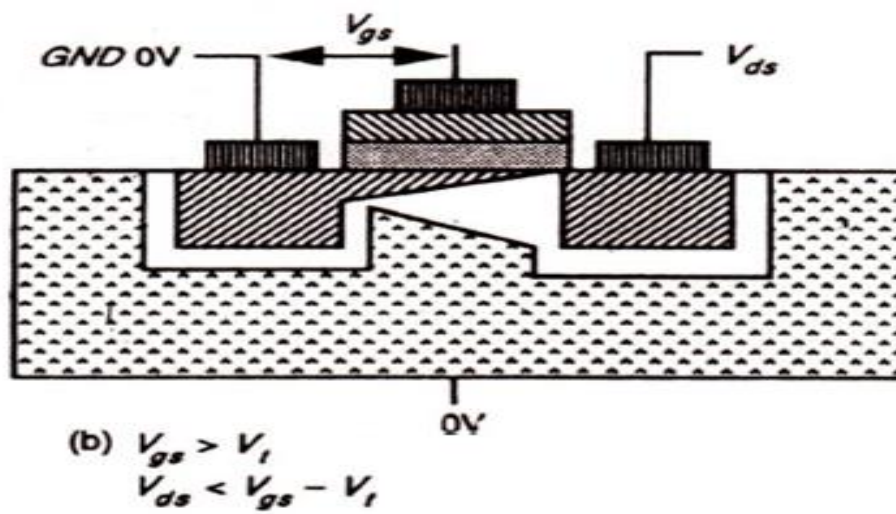
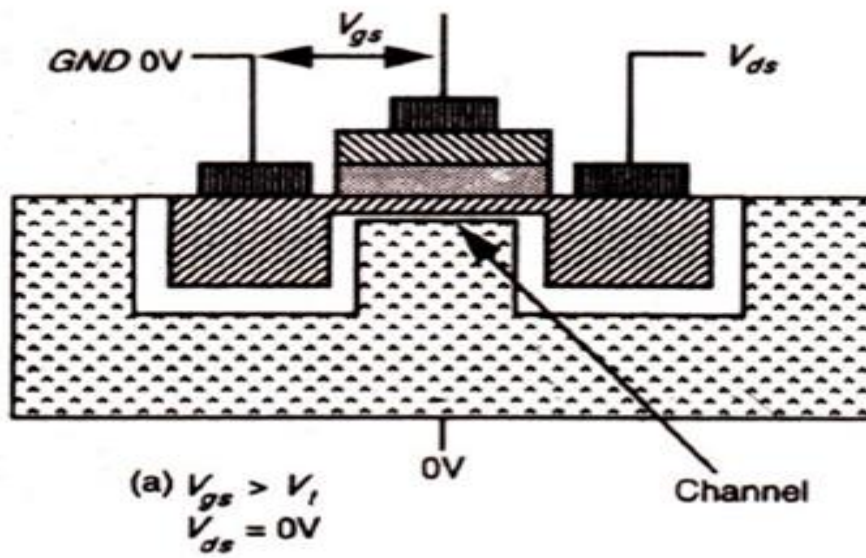
In an enhancement mode device a polysilicon gate is deposited on a layer of insulation over the region between source and drain. In the diagram below channel is not established and the device is in a non-conducting condition, i.e $V_D = V_S = V_{gs} = 0$. If this gate is connected to a suitable positive voltage with respect to the source, then the electric field established between the gate and the substrate gives rise to a charge inversion region in the substrate under the gate insulation and a conducting path or channel is formed between source and drain.



ENHANCEMENT MODE TRANSISTOR ACTION :

To understand the enhancement mechanism, let us consider the enhancement mode device. In order to establish the channel, a minimum voltage level called threshold voltage (V_t) must be established between gate and source. Fig. (a) Shows the existing situation where a channel is established but no current flowing between source and drain ($V_{ds} = 0$).

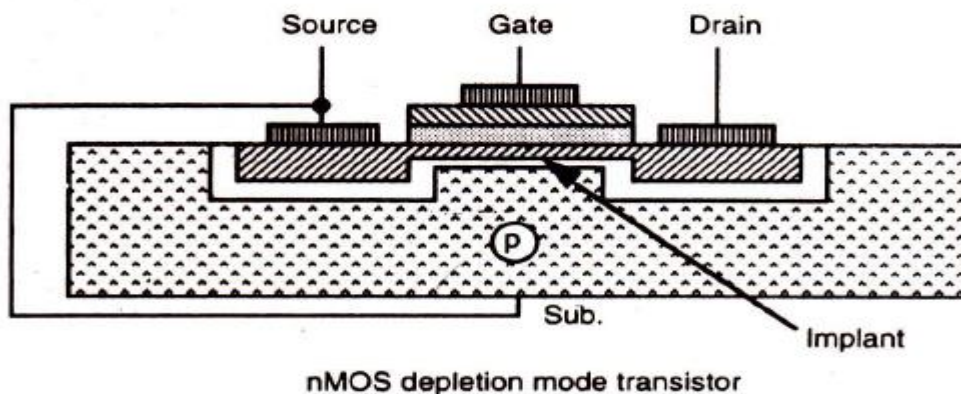
Let us now consider the conditions when current flows in the channel by applying a voltage V_{ds} between drain and source. The IR drop = V_{ds} along the channel. This develops a voltage between gate and channel varying with distance along the channel with the voltage being a maximum of V_{gs} at the source end. Since the effective gate voltage is $V_g = V_{gs} - V_t$, (no current flows when $V_{gs} < V_t$) there will be voltage available to invert the channel at the drain end so long as $V_{gs} - V_t \sim V_{ds}$. The limiting condition comes when $V_{ds} = V_{gs} - V_t$. For all voltages $V_{ds} < V_{gs} - V_t$, the device is in the non-saturated region of operation which is the condition shown in Fig. (b) below.



Let us now consider the situation when V_{ds} is increased to a level greater than $V_{gs} - V_t$. In this case, an IR drop equal to $V_{gs} - V_t$ occurs over less than the whole length of the channel such that, near the drain, there is insufficient electric field available to give rise to an inversion layer to create the channel. The channel is, therefore, 'pinched off' as shown in Fig. (c). Diffusion current completes the path from source to drain in this case, causing the channel to exhibit a high resistance and behave as a constant current source. This region, known as saturation, is characterized by almost constant current for increase of V_{ds} above $V_{ds} = V_{gs} - V_t$. In all cases, the channel will cease to exist and no current will flow when $V_{gs} < V_t$. Typically, for enhancement mode devices, $V_t = 1$ volt for $V_{DD} = 5$ V or, in general terms, $V_t = 0.2 V_{DD}$.

DEPLETION MODE TRANSISTOR ACTION

N-MOS Depletion mode mosfets are built with P-type silicon substrates, and P-channel versions are built on N-type substrates. In both cases they include a thin gate oxide formed between the source and drain regions. A conductive channel is deliberately formed below the gate oxide layer and between the source and drain by using ion-implantation. By implanting the correct ion polarity in the channel region during fabrication determines the polarity of the threshold voltage (i.e. $-V_t$ for an N channel transistor, or $+V_t$ for an P-channel transistor). The actual concentration of ions in the substrate-to-channel region is used to adjust the threshold voltage (V_t) to the desired value. Depletion-mode devices are a little more difficult to manufacture and their characteristics harder to control than enhancement types, which do not require ion implantation. In depletion mode devices the channel is established, due to the implant, even when $V_{gs} = 0$, and to cause the channel to cease a negative voltage V_{td} must be applied between gate and source.



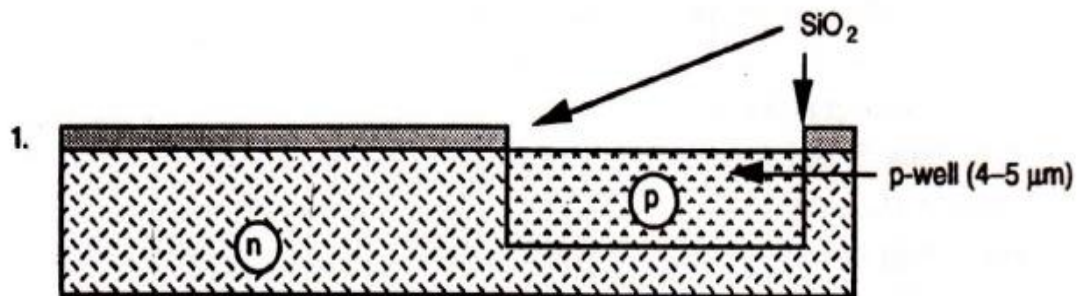
V_{td} is typically $< -0.8 V_{DD}$, depending on the implant and substrate bias, but, threshold voltage differences apart, the action is similar to that of the enhancement mode transistor.

CMOS FABRICATION :

CMOS fabrication is performed based on various methods , including the p-well, the n-well, the twin-tub, and the silicon-on-insulator processes .Among these methods the p-well process is widely used in practice and the n-well process is also popular, particularly as it is an easy retrofit to existing NMOS lines.

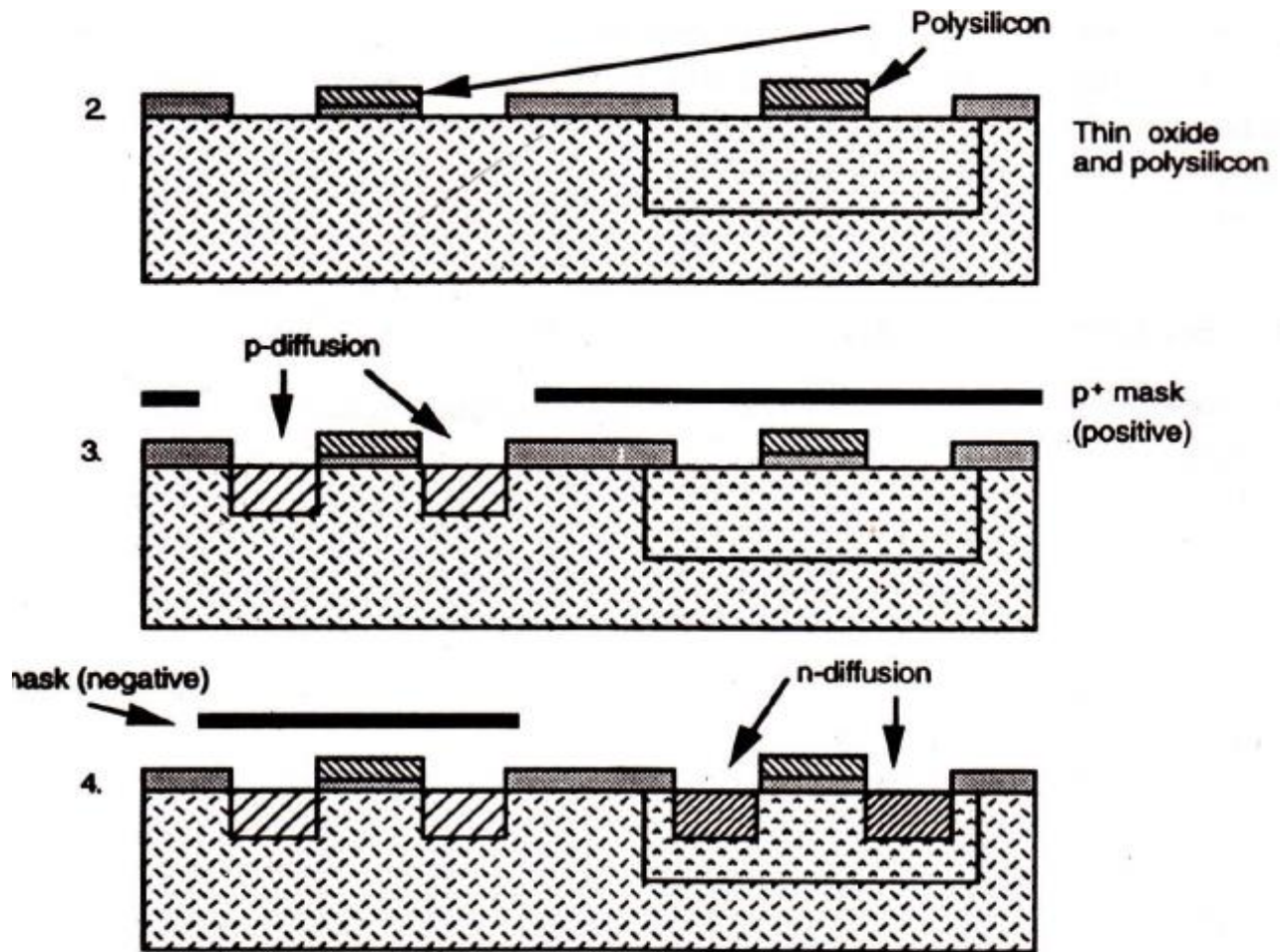
(i) The p-well Process:

The p-well structure consists of an n-type substrate in which p-devices may be formed by suitable masking and diffusion and, in order to accommodate n-type devices, a deep p-well is diffused into the n-type substrate as shown in the Fig.below.



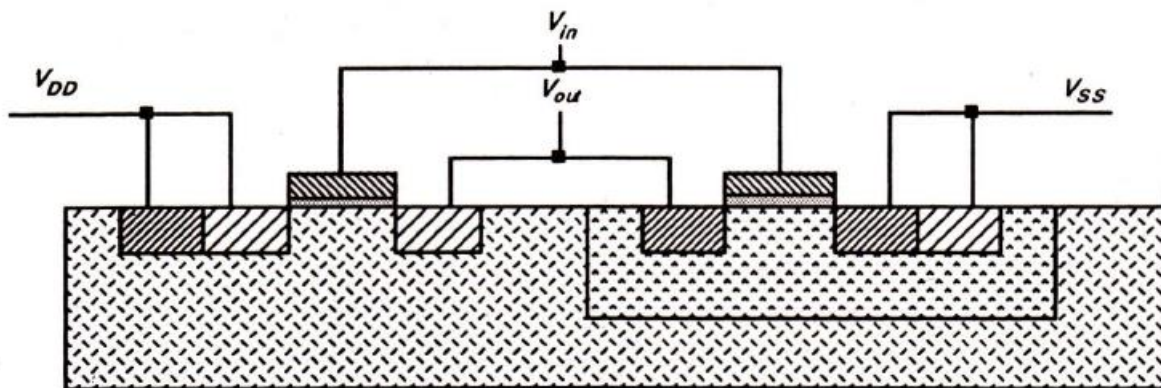
This diffusion should be carried out with special care since the p-well doping concentration and depth will affect the threshold voltages as well as the breakdown voltages of the n-transistors. To achieve low threshold voltages (0.6 to 1.0 V) either deep-well diffusion or high-well resistivity is required. However, deep wells require larger spacing between the n- and p-type transistors and wires due to lateral diffusion and therefore a larger chip area. The p-wells Act as substrates for the n-devices within the parent n-substrate, and, the two areas are electrically isolated.

Except this in all other respects- like masking, patterning, and diffusion-the process is similar to NMOS fabrication.



P-well fabrication process(Figs 1,2,3 & 4)

The diagram below shows the CMOS p-well inverter showing V_{DD} and V_{SS} substrate connections



The n-well Process : Though the p-well process is widely used in C-MOS fabrication the n-well fabrication is also very popular because of the lower substrate bias effects on transistor threshold voltage and also lower parasitic capacitances associated with source and drain regions.

The typical n-well fabrication steps are shown in the diagram below.

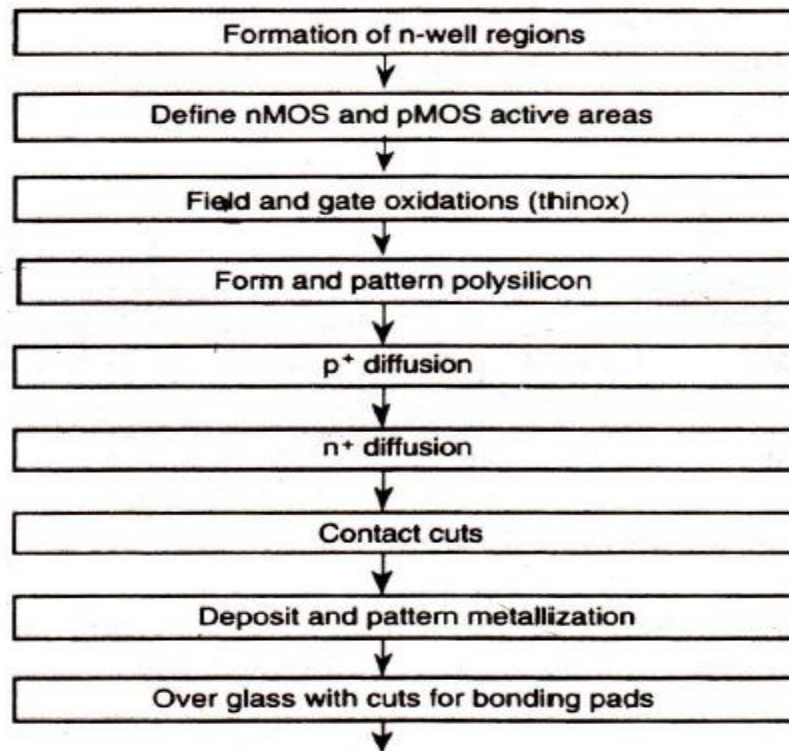
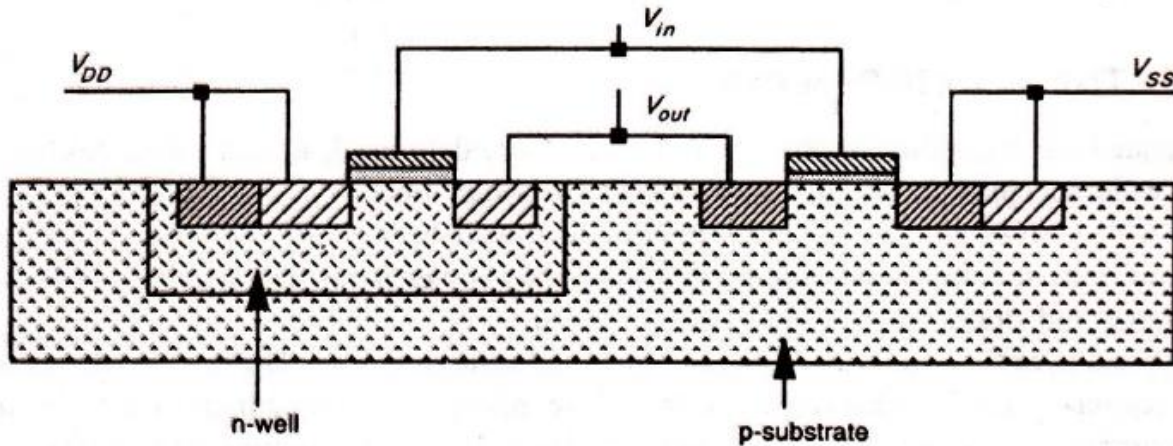


Fig. n-well fabrication steps

The first mask defines the n-well regions. This is followed by a low dose phosphorus implant driven in by a high temperature diffusion step to form the n-wells. The well depth is optimized to ensure against-substrate top⁺ diffusion breakdown without compromising then-well to n⁺ mask separation. The next steps are to define the devices and diffusion paths, grow field oxide, deposit and pattern the poly silicon, carry out the diffusions, make contact cuts, and finally metalize as before. It will be seen that an n⁺ mask and its complement may be used to define the n- and p-diffusion regions respectively. These same masks also include the V_{DD} and V_{SS} contacts (respectively). It should be noted that, alternatively, we could have used a p⁺ mask and its complement since the n⁺ and p⁺ masks are generally complementary.

The diagram below shows the Cross-sectional view of n-well CMOS Inverter.



Due to the differences in charge carrier mobilities, the n-well process creates non-optimum p-channel characteristics. However, in many CMOS designs (such as domino-logic and dynamic logic structures), this is relatively unimportant since they contain a preponderance of n-channel devices. Thus then-channel transistors are mainly those used to form logic elements, providing speed and high density of elements.

However, a factor of the n-well process is that the performance of the already poorly performing p-transistor is even further degraded. Modern process lines have come to grips with these problems, and good device performance may be achieved for both p-well and n-well fabrication.

BICMOS Technology:

A Bi-CMOS circuit of both bipolar junction transistors and MOS transistors on a single substrate. The driving capability of MOS transistors is less because of limited current sourcing and sinking capabilities of the transistors. To drive large capacitive loads Bi-CMOS technology is used. As this technology combines Bipolar and CMOS transistors in a single integrated circuit, it has the advantages of both bipolar and CMOS transistors. Bi-CMOS is able to achieve VLSI circuits with speed-power-density performance previously not possible with either technology individually. The diagram given below shows the cross section of the Bi-CMOS process which uses an NPN transistor

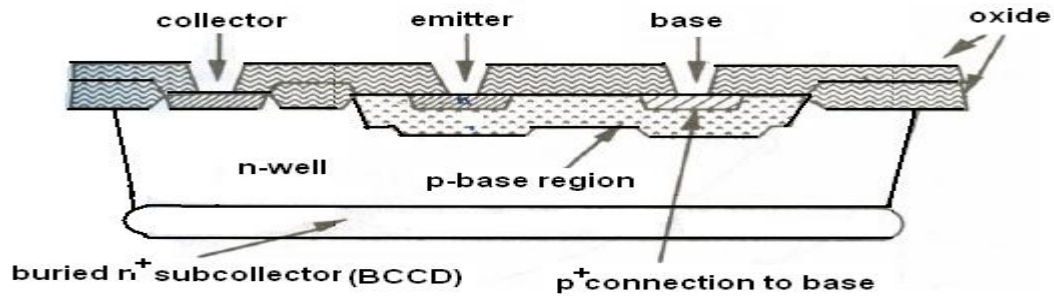
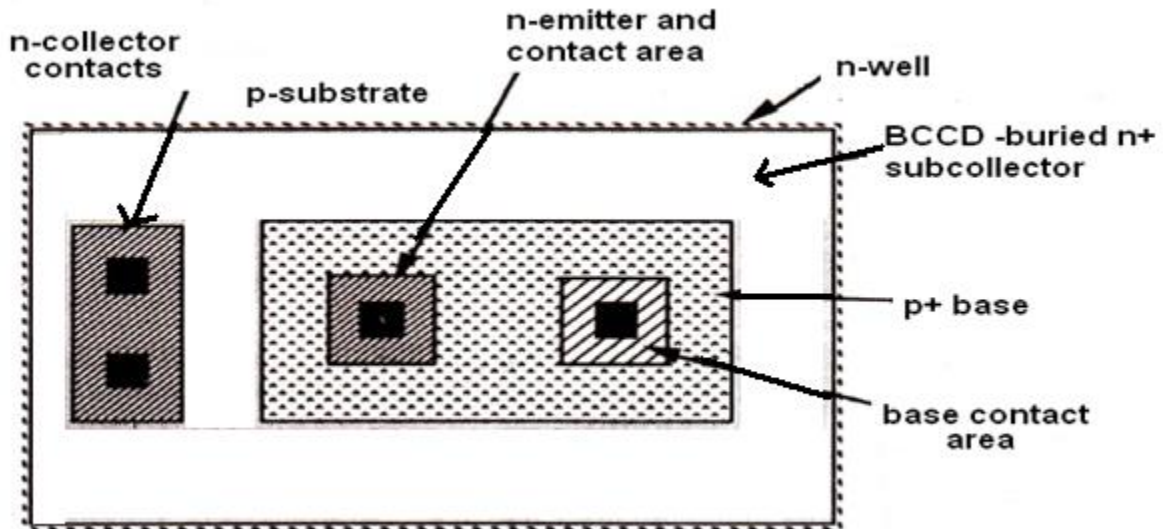


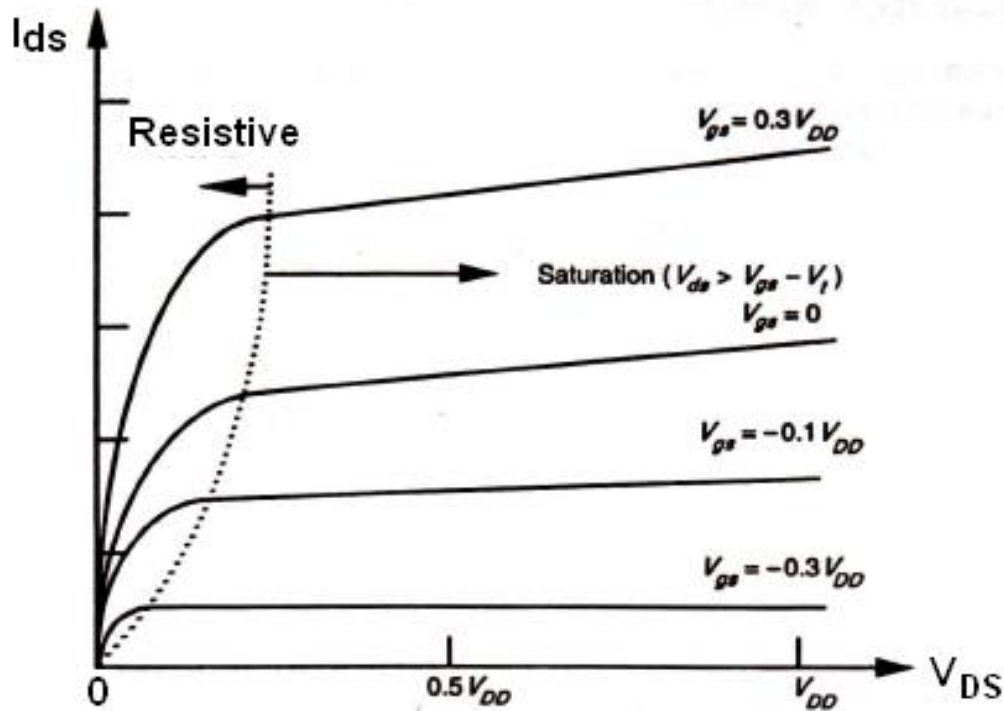
Fig. Arrangement of Bi-CMOS npn transistor

The lay-out view of Bic-MOS transistor is shown in the figure below. The fabrication of Bi-CMOS is similar to CMOS but with certain additional process steps and additional masks are considered. They are (i) the p+ base region; (ii) n+ collector area; and (iii) the buried sub collector (BCCD).



I_{DS} - V_{DS} characteristics of MOS Transistor:

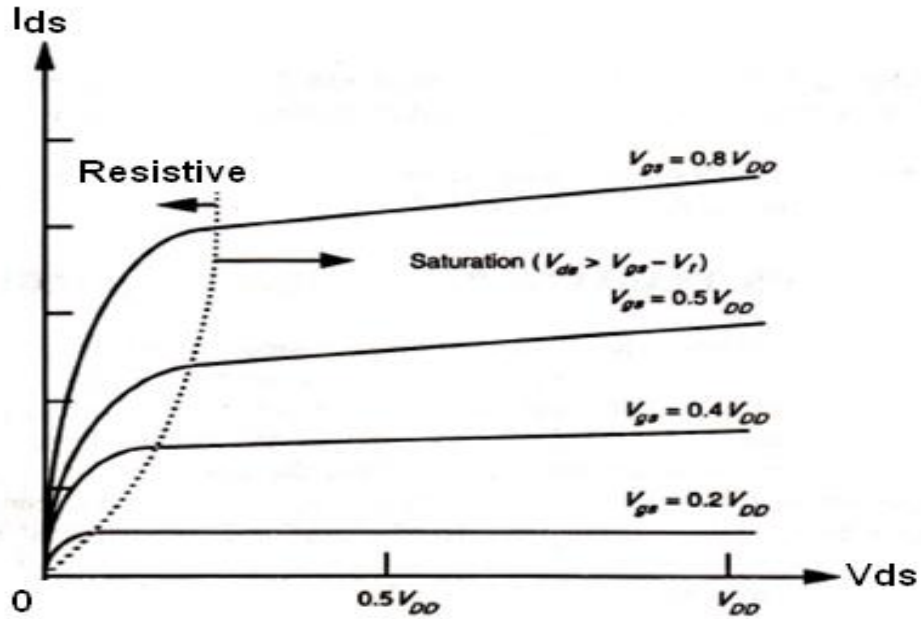
The graph below shows the I_D Vs V_{DS} characteristics of an n- MOS transistor for several values of V_{GS} . It is clear that there are two conduction states when the device is ON. The saturated state and the non-saturated state. The saturated curve is the flat portion and defines the saturation region. For $V_{gs} < V_{DS} + V_{th}$, the NMOS device is conducting and I_D is independent of V_{DS} . For $V_{gs} > V_{DS} + V_{th}$, the transistor is in the non-saturation region and the curve is a half parabola. When the transistor is OFF ($V_{gs} < V_{th}$), then I_D is zero for any V_{DS} value.



(a) Depletion mode device

The boundary of the saturation/non-saturation bias states is a point seen for each curve in the graph as the intersection of the straight line of the saturated region with the quadratic curve of the non-saturated region. This intersection point occurs at the channel pinch off voltage called V_{DSAT} . The diamond symbol marks the pinch-off voltage V_{DSAT} for each value of V_{GS} . V_{DSAT} is defined as the minimum drain-source voltage that is required to keep the transistor in saturation for a given V_{GS} . In the non-saturated state, the drain current initially increases almost linearly from the origin before bending in a parabolic response. Thus the name ohmic or linear for the non-saturated region.

The drain current in saturation is virtually independent of V_{DS} and the transistor acts as a current Source. This is because there is no carrier inversion at the drain region of the channel. Carriers are pulled into the high electric field of the drain/substrate pn junction and ejected out of the drain terminal.



(b). Enhance mode device

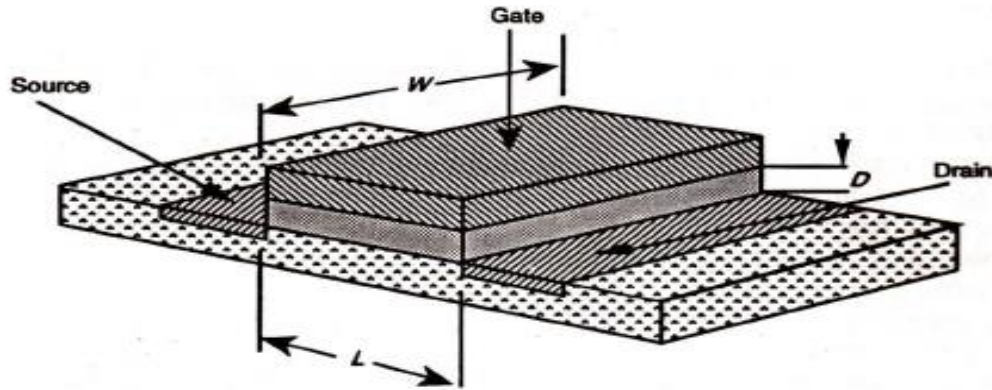
Drain-to-Source Current I_{DS} versus Voltage V_{DS} Relationships:

The working of a MOS transistor is based on the principle that the use of a voltage on the gate induce a charge in the channel between source and drain, which may then be caused to move from source to drain under the influence of an electric field created by voltage V_{ds} applied between drain and source. Since the charge induced is dependent on the gate to source voltage V_{gs} then I_{ds} is dependent on both V_{gs} and V_{ds} .

Let us consider the diagram below in which electrons will flow source to drain .So, the drain current is given by

$$I_{ds} = -I_{sd} = \frac{\text{Charge induced in channel (} Q_c \text{)}}{\text{Electron transit time}(\tau)}$$

$$\text{Where the transit time is given by } \tau_{sd} = \frac{\text{Length of the channel}}{\text{Velocity (} v \text{)}}$$



But velocity $v = \mu E_{ds}$

Where μ = electron or hole mobility and E_{ds} = Electric field

Also , $E_{ds} = V_{ds}/L$

So, $v = \mu \cdot V_{ds}/L$

And $\tau_{ds} = L^2 / \mu \cdot V_{ds}$

The typical values of μ at room temperature are given below.

$$\mu_n \approx 650 \text{ cm}^2/\text{V sec (surface)}$$

$$\mu_p \approx 240 \text{ cm}^2/\text{V sec (surface)}$$

The Non-saturated Region:

Let us consider the I_d vs V_d relationships in the non-saturated region. The charge induced in the channel due to the voltage difference between the gate and the channel, V_{gs} (assuming substrate connected to source). The voltage along the channel varies linearly with distance X from the source due to the IR drop in the channel. In the non-saturated state the average value is $V_{ds}/2$. Also the effective gate voltage $V_g = V_{gs} - V_t$ where V_t is the threshold voltage needed to invert the charge under the gate and establish the channel.

Hence the induced charge is

$$Q_c = E_g \epsilon_{ins} \epsilon_o W \cdot L$$

Where

E_g = average electric field gate to channel

ϵ_{ins} = relative permittivity of insulation between gate and channel

ϵ_0 = permittivity of free space.

So, we can write that

$$E_g = \frac{\left((V_{gs} - V_t) - \frac{V_{ds}}{2} \right)}{D}$$

Here D is the thickness of the oxide layer. Thus

$$Q_c = \frac{WL\epsilon_{ins}\epsilon_0}{D} \left((V_{gs} - V_t) - \frac{V_{ds}}{2} \right)$$

So, by combining the above two equations ,we get

$$I_{ds} = \frac{\epsilon_{ins}\epsilon_0\mu}{D} \frac{W}{L} \left((V_{gs} - V_t) - \frac{V_{ds}}{2} \right) V_{ds}$$

Or the above equation can be written as

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

In the non-saturated or resistive region where $V_{ds} < V_{gs} - V_t$ and

$$K = \frac{\epsilon_{ins}\epsilon_0\mu}{D}$$

Generally ,a constant β is defined as

$$\beta = K \frac{W}{L}$$

So that ,the expression for drain –source current will become

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

The gate /channel capacitance is

$$C_g = \frac{\epsilon_{ins} \epsilon_0 WL}{D} \text{ (parallel plate)}$$

Hence we can write another alternative form for the drain current as

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

Some time it is also convenient to use gate –capacitance per unit area, C_g

So, the drain current is

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

This is the relation between drain current and drain-source voltage in non-saturated region.

The Saturated Region

Saturation begins when $V_{ds} = V_{gs} - V_t$, since at this point the IR drop in the channel equals the effective gate to channel voltage at the drain and we may assume that the current remains fairly constant as V_{ds} increases further. Thus

$$I_{ds} = K \frac{W}{L} \frac{(V_{gs} - V_t)^2}{2}$$

Or we can also write that

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

Or it can also be written as

$$I_{ds} = \frac{C_g \mu}{2L^2} (V_{gs} - V_t)^2$$

Or

$$I_{ds} = C_g \mu \frac{W}{2L} (V_{gs} - V_t)^2$$

The expressions derived above for I_{ds} hold for both enhancement and depletion mode devices.

Here the threshold voltage for the NMOS depletion mode device (denoted as V_{td}) is negative.

MOS Transistor Threshold Voltage V_t :

The gate structure of a MOS transistor consists, of charges stored in the dielectric layers and in the surface to surface interfaces as well as in the substrate itself. Switching an enhancement mode MOS transistor from the off to the on state consists in applying sufficient gate voltage to neutralize these charges and enable the underlying silicon to undergo an inversion due to the electric field from the gate. Switching a depletion mode NMOS transistor from the on to the off state consists in applying enough voltage to the gate to add to the stored charge and invert the 'n' implant region to 'p'.

The threshold voltage V_t may be expressed as:

$$V_t = \phi_{ms} - \frac{Q_B - Q_{ss}}{C_o} + 2\phi_{fn}$$

Where Q_D = the charge per unit area in the depletion layer below the oxide

Q_{ss} = charge density at Si: SiO_2 interface

C_o = Capacitance per unit area.

Φ_{ns} = work function difference between gate and Si

Φ_{fn} = Fermi level potential between inverted surface and bulk Si

For polynomial gate and silicon substrate, the value of Φ_{ns} is negative but negligible and the magnitude and sign of V_t are thus determined by balancing the other terms in the equation.

To evaluate the V_t the other terms are determined as below.

$$Q_B = \sqrt{2\epsilon_0\epsilon_{Si}qN(2\phi_{fn} + V_{SB})} \text{ coulomb/m}^2$$

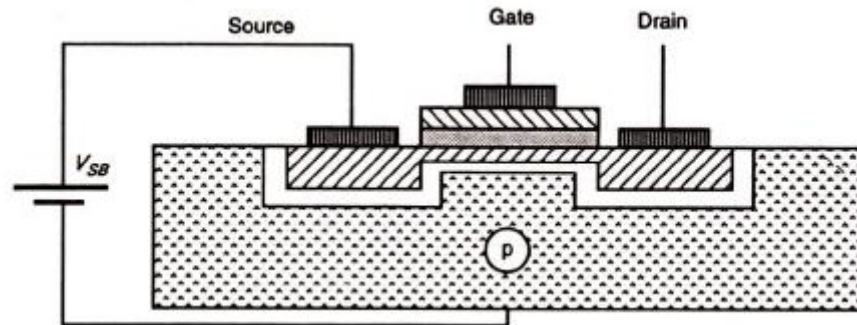
$$\phi_{fn} = \frac{kT}{q} \ln \frac{N}{n_i} \text{ volts}$$

$$Q_{ss} = (1.5 \text{ to } 8) \times 10^{-8} \text{ coulomb/m}^2$$

Body Effect :

Generally while studying the MOS transistors it is treated as a three terminal device. But ,the body of the transistor is also an implicit terminal which helps to understand the characteristics of the transistor. Considering the body of the MOS transistor as a terminal is known as the body

effect. The potential difference between the source and the body (V_{sb}) affects the threshold voltage of the transistor. In many situations, this Body Effect is relatively insignificant, so we can (unless **otherwise** stated) ignore the Body Effect. But it is not always insignificant, in some cases it can have a tremendous impact on MOSFET circuit performance.



Body effect - NMOS device

Increasing V_{sb} causes the channel to be depleted of charge carriers and thus the threshold voltage is raised. Change in V_t is given by $\delta v_t = \gamma \cdot (V_{sb})^{1/2}$ where γ is a constant which depends on substrate doping so that the more lightly doped the substrate, the smaller will be the body effect

The threshold voltage can be written as

$$V_t = V_t(0) + \left(\frac{D}{\epsilon_{ins} \epsilon_0} \right) \sqrt{2 \epsilon_0 \epsilon_{si} QN} \cdot (V_{sb})^{1/2}$$

Where $V_t(0)$ is the threshold voltage for $V_{sd} = 0$

For n-MOS depletion mode transistors, the body voltage values at different V_{DD} voltages are given below.

$$V_{SB} = 0 \text{ V} ; V_{sd} = -0.7V_{DD} (= -3.5 \text{ V for } V_{DD} = +5 \text{ V})$$

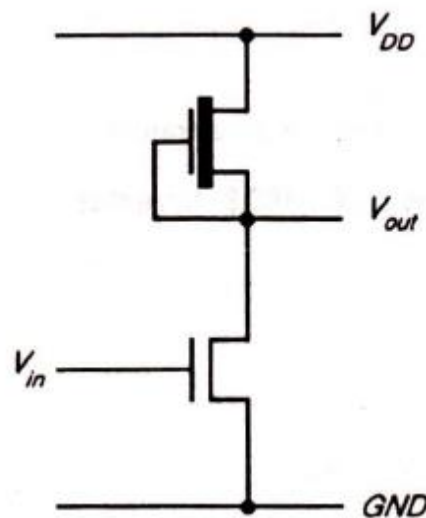
$$V_{SB} = 5 \text{ V} ; V_{sd} = -0.6V_{DD} (= -3.0 \text{ V for } V_{DD} = +5 \text{ V})$$

The NMOS INVERTER :

An inverter circuit is a very important circuit for producing a complete range of logic circuits. This is needed for restoring logic levels, for Nand and Nor gates, and for sequential and memory circuits of various forms .

A simple inverter circuit can be constructed using a transistor with source connected to ground and a load resistor connected from the drain to the positive supply rail V_{DD} . The output is taken from the drain and the input applied between gate and ground.

But, during the fabrication resistors are not conveniently produced on the silicon substrate and even small values of resistors occupy excessively large areas. Hence some other form of load resistance is used. A more convenient way to solve this problem is to use a depletion mode transistor as the load, as shown in Fig. Below.



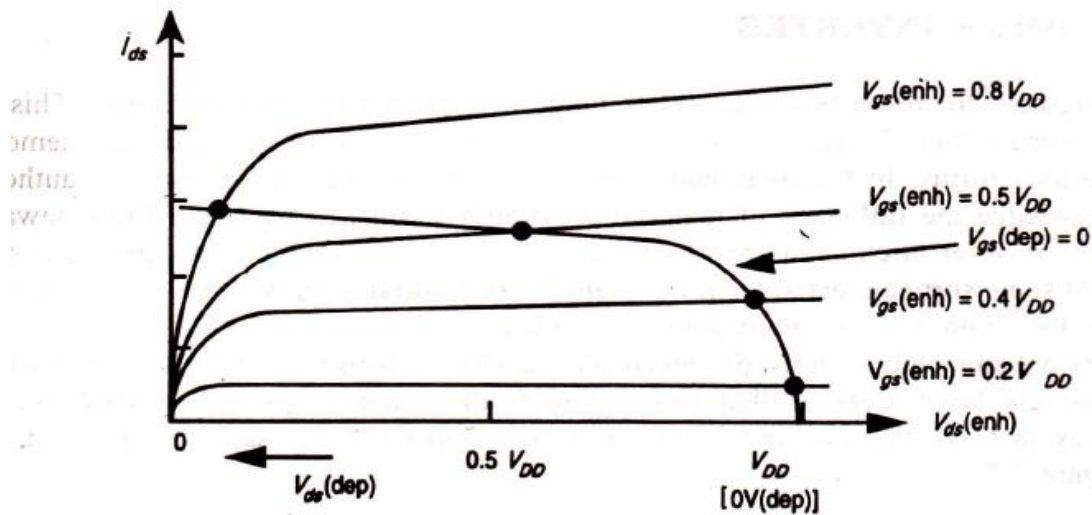
The salient features of the n-MOS inverter are

- For the depletion mode transistor, the gate is connected to the source so it is always on.
- In this configuration the depletion mode device is called the pull-up (P.U) and the enhancement mode device the pull-down (P.D) transistor.
- With no current drawn from the output, the currents I_{ds} for both transistors must be equal.

NMOS Inverter transfer characteristic.

The transfer characteristic is drawn by taking V_{ds} on x-axis and I_{ds} on Y-axis for both enhancement and depletion mode transistors. So, to obtain the inverter transfer characteristic for $V_{gs} = 0$ depletion mode characteristic curve is superimposed on the family of curves for the

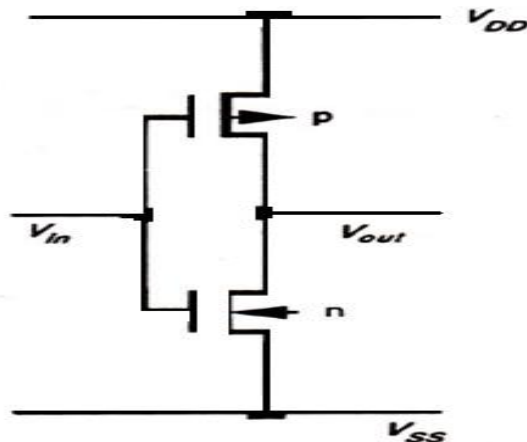
enhancement mode device and from the graph it can be seen that , maximum voltage across the enhancement mode device corresponds to minimum voltage across the depletion mode transistor.



From the graph it is clear that as $V_{in}(=V_{gs} \text{ p.d. Transistor})$ exceeds the Pulldown threshold voltage current begins to flow. The output voltage V_{out} thus decreases and the subsequent increases in V_{in} will cause the Pull down transistor to come out of saturation and become resistive.

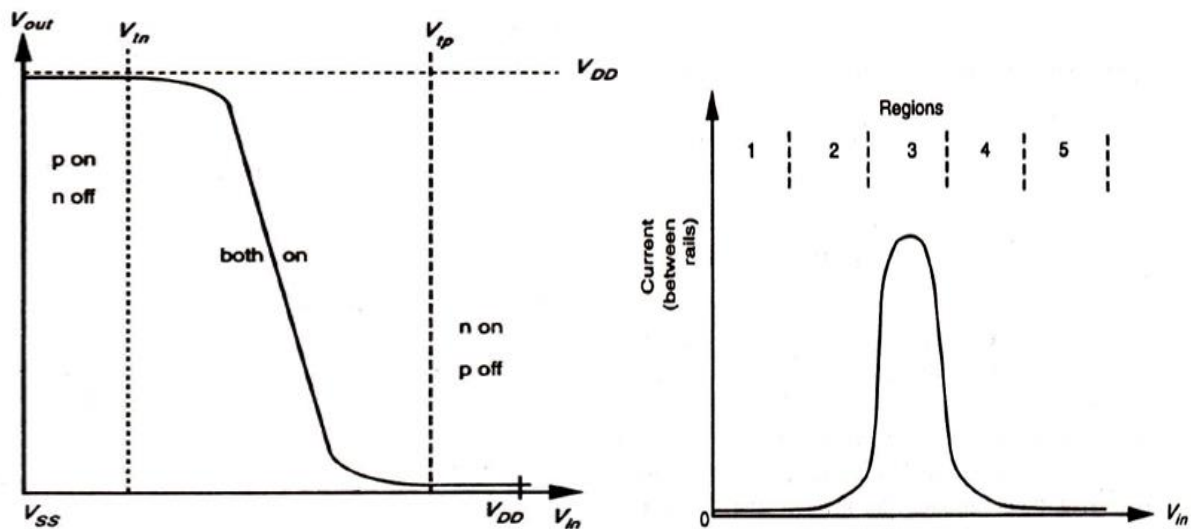
CMOS Inverter :

The inverter is the very important part of all digital designs. Once its operation and properties are clearly understood, Complex structures like NAND gates, adders, multipliers, and microprocessors can also be easily done. The electrical behavior of these complex circuits can be almost completely derived by extrapolating the results obtained for inverters. As shown in the diagram below the CMOS transistor is designed using p-MOS and n-MOS transistors.



In the inverter circuit ,if the input is high .the lower n-MOS device closes to discharge the capacitive load .Similarly ,if the input is low,the top p-MOS device is turned on to charge the capacitive load .At no time both the devices are on ,which prevents the DC current flowing from positive power supply to ground. Qualitatively this circuit acts like the switching circuit, since the p-channel transistor has exactly the opposite characteristics of the n-channel transistor. In the transition region both transistors are saturated and the circuit operates with a large voltage gain. The C-MOS transfer characteristic is shown in the below graph.

Considering the static conditions first, it may be seen that in region 1 for which $V_{in} = \text{logic } 0$, we have the p-transistor fully turned on while the n-transistor is fully turned off. Thus no current flows through the inverter and the output is directly connected to V_{DD} through the p-transistor.



Hence the output voltage is logic 1. In region 5, $V_{in} = \text{logic } 1$ and the n-transistor is fully on while the p-transistor is fully off. So, no current flows and a logic 0 appears at the output.

In region 2 the input voltage has increased to a level which just exceeds the threshold voltage of the n-transistor. The n-transistor conducts and has a large voltage between source and drain; so it is in saturation. The p-transistor is also conducting but with only a small voltage across it, it operates in the unsaturated resistive region. A small current now flows through the inverter from V_{DD} to V_{SS} . If we wish to analyze the behavior in this region, we equate the p-device resistive region current with the n-device saturation current and thus obtain the voltage and current relationships.

Region 4 is similar to region 2 but with the roles of the p- and n-transistors reversed. However, the current magnitudes in regions 2 and 4 are small and most of the energy consumed in switching from one state to the other is due to the larger current which flows in region 3.

Region 3 is the region in which the inverter exhibits gain and in which both transistors are in saturation.

The currents in each device must be the same, since the transistors are in series. So, we can write that

$$I_{dsp} = -I_{dsn}$$

where

$$I_{dsp} = \frac{\beta_p}{2} (V_{in} - V_{DD} - V_{tp})^2$$

and

$$I_{dsn} = \frac{\beta_n}{2} (V_{in} - V_{tn})^2$$

Since both transistors are in saturation, they act as current sources so that the equivalent circuit in this region is two current sources in series between V_{DD} and V_{SS} with the output voltage coming from their common point. The region is inherently unstable in consequence and the changeover from one logic level to the other is rapid.

Determination of Pull-up to Pull-Down Ratio ($Z_{p.u}/Z_{p.d}$) For an NMOS Inverter driven by another NMOS Inverter :

Let us consider the arrangement shown in Fig.(a). In which an inverter is driven from the output of another similar inverter. Consider the depletion mode transistor for which $V_{gs} = 0$ under all conditions, and also assume that in order to cascade inverters without degradation the condition

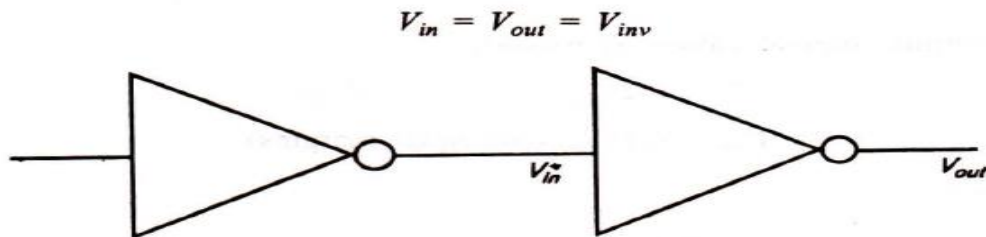


Fig.(a). Inverter driven by another inverter.

For equal margins around the inverter threshold, we set $V_{inv} = 0.5V_{DD}$. At this point both transistors are in saturation and we can write that

$$I_{ds} = K \frac{W}{L} \frac{(V_{gs} - V_t)^2}{2}$$

In the depletion mode $I_{ds} = K \frac{W_{p.u.}}{L_{p.u.}} \frac{(-V_{td})^2}{2}$ since $V_{gs} = 0$

and in the enhancement mode

$$I_{ds} = K \frac{W_{p.d.}}{L_{p.d.}} \frac{(V_{inv} - V_t)^2}{2} \text{ since } V_{gs} = V_{inv}$$

Equating (since currents are the same) we have

$$\frac{W_{p.d.}}{L_{p.d.}} (V_{inv} - V_t)^2 = \frac{W_{p.u.}}{L_{p.u.}} (-V_{td})^2$$

Where $W_{p.d.}$, $L_{p.d.}$, $W_{p.u.}$ And $L_{p.u.}$ are the widths and lengths of the pull-down and pull-up transistors respectively.

So, we can write that

$$Z_{p.d.} = \frac{L_{p.d.}}{W_{p.d.}}; Z_{p.u.} = \frac{L_{p.u.}}{W_{p.u.}}$$

we have

$$\frac{1}{Z_{p.d.}} (V_{inv} - V_t)^2 = \frac{1}{Z_{p.u.}} (-V_{td})^2$$

whence

$$V_{inv} = V_t - \frac{V_{td}}{\sqrt{Z_{p.u.}/Z_{p.d.}}}$$

The typical, values for V_t , V_{inv} and V_{td} are

$$V_t = 0.2V_{DD}; V_{td} = -0.6V_{DD}$$

$$V_{inv} = 0.5V_{DD} \text{ (for equal margins)}$$

Substituting these values in the above equation ,we get

$$0.5 = 0.2 + \frac{0.6}{\sqrt{Z_{p.u.}/Z_{p.d.}}}$$

Here

$$\sqrt{Z_{p.u.}/Z_{p.d.}} = 2$$

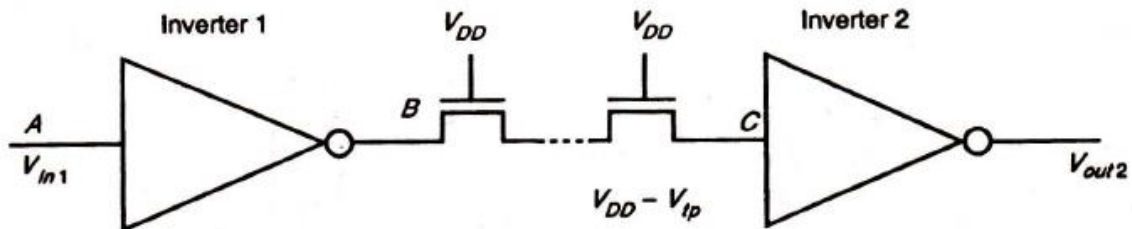
So,we get

$$Z_{p.u.}/Z_{p.d.} = 4/1$$

This is the ratio for pull-up to pull down ratio for an inverter directly driven by another inverter.

Pull -Up to Pull-Down ratio for an NMOS Inverter driven through one or more Pass Transistors

Let us consider an arrangement in which the input to inverter 2 comes from the output of inverter 1 but passes through one or more NMOS transistors as shown in Fig. Below (These transistors are called pass transistors).



The connection of pass transistors in series will degrade the logic 1 level / into inverter 2 so that the output will not be a proper logic 0 level. The critical condition is , when point A is at 0 volts and B is thus at V_{DD} . But the voltage into inverter 2 at point C is now reduced from V_{DD} by the threshold voltage of the series pass transistor. With all pass transistor gates connected to V_{DD} there is a loss of V_{tp} , however many are connected in series, since no static current flows through them and there can be no voltage drop in the channels. Therefore, the input voltage to inverter 2 is

$$V_{in2} = V_{DD} - V_{tp}$$

Where V_{tp} = threshold voltage for a pass transistor.

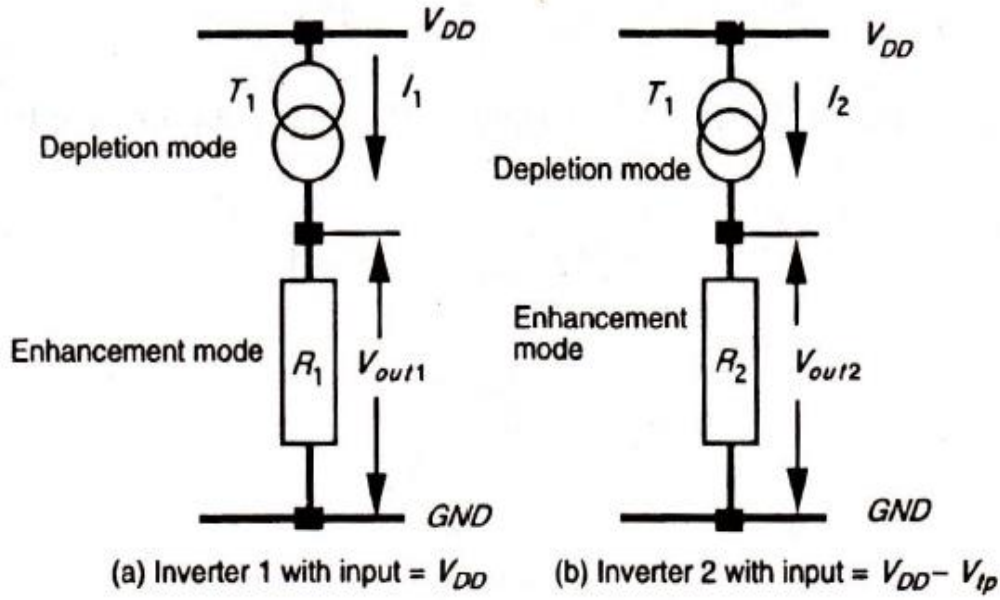
Let us consider the inverter 1 shown in Fig.(a) with input = V_{DD} . If the input is at V_{DD} , then the pull-down transistor T2 is conducting but with a low voltage across it; therefore, it is in its resistive region represented by R_1 in Fig.(a) below. Meanwhile, the pull up transistor T1 is in saturation and is represented as a current source.

For the pull down transistor

$$R_1 = \frac{V_{ds1}}{I_{ds}} = \frac{1}{K} \frac{L_{p.d.1}}{W_{p.d.1}} \left(\frac{1}{V_{DD} - V_t - \frac{V_{ds1}}{2}} \right)$$

$$I_{ds} = K \frac{W_{p.d.1}}{L_{p.d.1}} \left((V_{DD} - V_t) V_{ds1} - \frac{V_{ds1}^2}{2} \right)$$

Since V_{ds} is small, $V_{ds}/2$ can be neglected in the above expression.



So,

$$R_1 \doteq \frac{1}{K} Z_{p.d.1} \left(\frac{1}{V_{DD} - V_t} \right)$$

Now, for depletion mode pull-up transistor in saturation with $V_{gs} = 0$

$$I_1 = I_{ds} = K \frac{W_{p.u.1}}{L_{p.u.1}} \frac{(-V_{td})^2}{2}$$

The product

$$I_1 R_1 = V_{out1}$$

So,

$$V_{out1} = I_1 R_1 = \frac{Z_{p.d.1}}{Z_{p.u.1}} \left(\frac{1}{V_{DD} - V_t} \right) \frac{(V_{td})^2}{2}$$

Let us now consider the inverter 2 Fig.b .when input = $V_{DD} - V_{tp}$.

$$R_2 \doteq \frac{1}{K} Z_{p.d.2} \frac{1}{((V_{DD} - V_{tp}) - V_t)}$$

$$I_2 = K \frac{1}{Z_{p.u.2}} \frac{(-V_{td})^2}{2}$$

Whence,

$$V_{out2} = I_2 R_2 = \frac{Z_{p.d.2}}{Z_{p.u.2}} \left(\frac{1}{V_{DD} - V_{tp} - V_t} \right) \frac{(-V_{td})^2}{2}$$

If inverter 2 is to have the same output voltage under these conditions then $V_{out1} = V_{out2}$. That is

$$I_1 R_1 = I_2 R_2 \quad , \quad \text{therefore}$$

$$\frac{Z_{p.u.2}}{Z_{p.d.2}} = \frac{Z_{p.u.1}}{Z_{p.d.1}} \frac{(V_{DD} - V_t)}{(V_{DD} - V_{tp} - V_t)}$$

Considering the typical values

$$V_t = 0.2V_{DD}$$

$$V_{tp} = 0.3V_{DD}^*$$

$$\frac{Z_{p.u.2}}{Z_{p.d.2}} = \frac{Z_{p.u.1}}{Z_{p.d.1}} \frac{0.8}{0.2}$$

Therefore

$$\frac{Z_{p.u.2}}{Z_{p.d.2}} \doteq 2 \frac{Z_{p.u.1}}{Z_{p.d.1}} = \frac{8}{1}$$

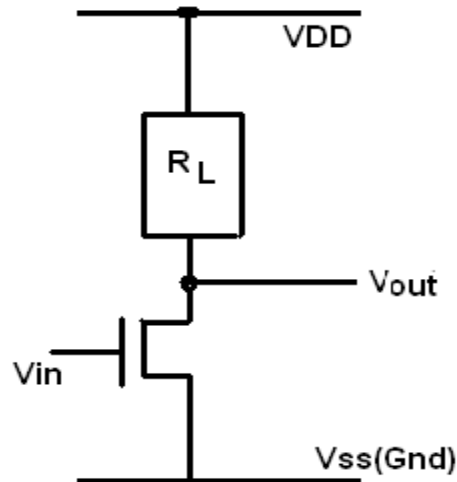
From the above theory it is clear that, for an n-MOS transistor

- (i). An inverter driven directly from the output of another should have a $Z_{p.u.}/Z_{p.d.}$ Ratio Of $\geq 4/1$.
- (ii). An inverter driven through one or more pass transistors should have a $Z_{p.u.}/Z_{p.d.}$ ratio of $\geq 8/1$

ALTERNATIVE FORMS OF PULL -UP

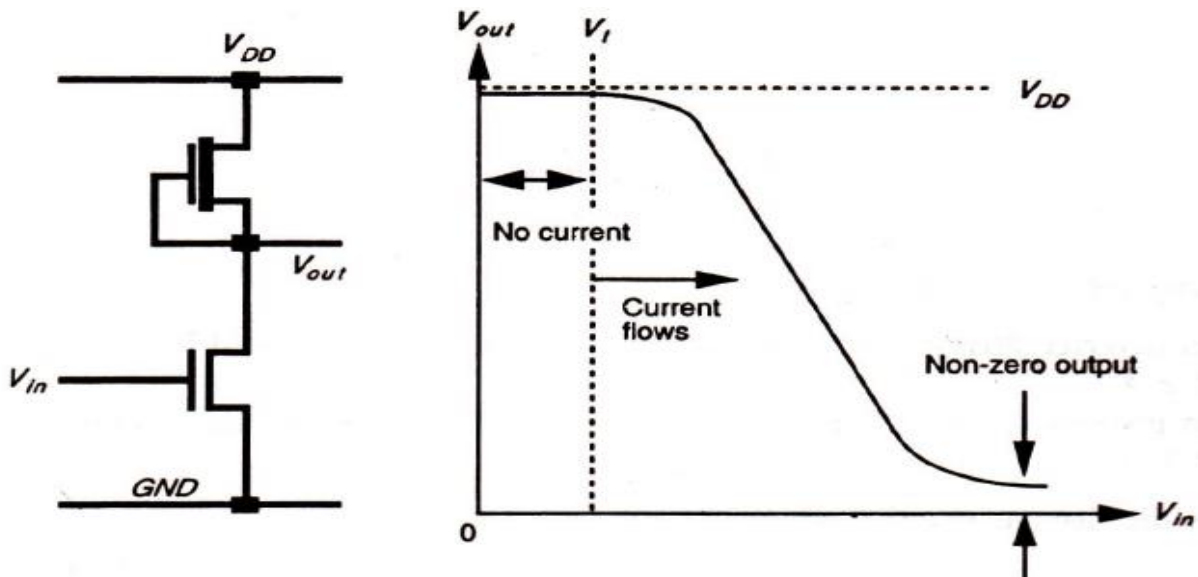
Generally the inverter circuit will have a depletion mode pull-up transistor as its load. But there are also other configurations. Let us consider four such arrangements.

(i). **Load resistance R_L** : This arrangement consists of a load resistor as a pull-up as shown in the diagram below. But it is not widely used because of the large space requirements of resistors produced in a silicon substrate.



2. NMOS depletion mode transistor pull-up : This arrangement consists of a depletion mode transistor as pull-up. The arrangement and the transfer characteristic are shown below. In this type of arrangement we observe

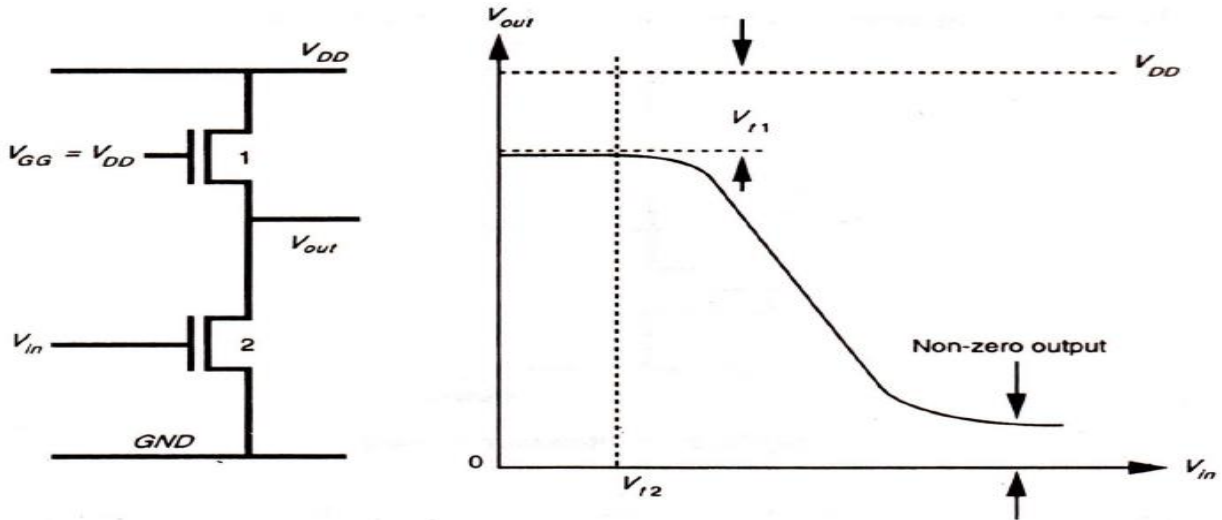
- (a) Dissipation is high, since rail to rail current flows when V_{in} = logical 1.
- (b) Switching of output from 1 to 0 begins when V_{in} exceeds V_t of pull-down device.



NMOS depletion mode transistor pull-up and transfer characteristic

(c) When switching the output from 1 to 0, the pull-up device is non-saturated initially and this presents lower resistance through which to charge capacitive loads .

3. NMOS enhancement mode pull-up : This arrangement consists of a n-MOS enhancement mode transistor as pull-up. The arrangement and the transfer characteristic are shown below.

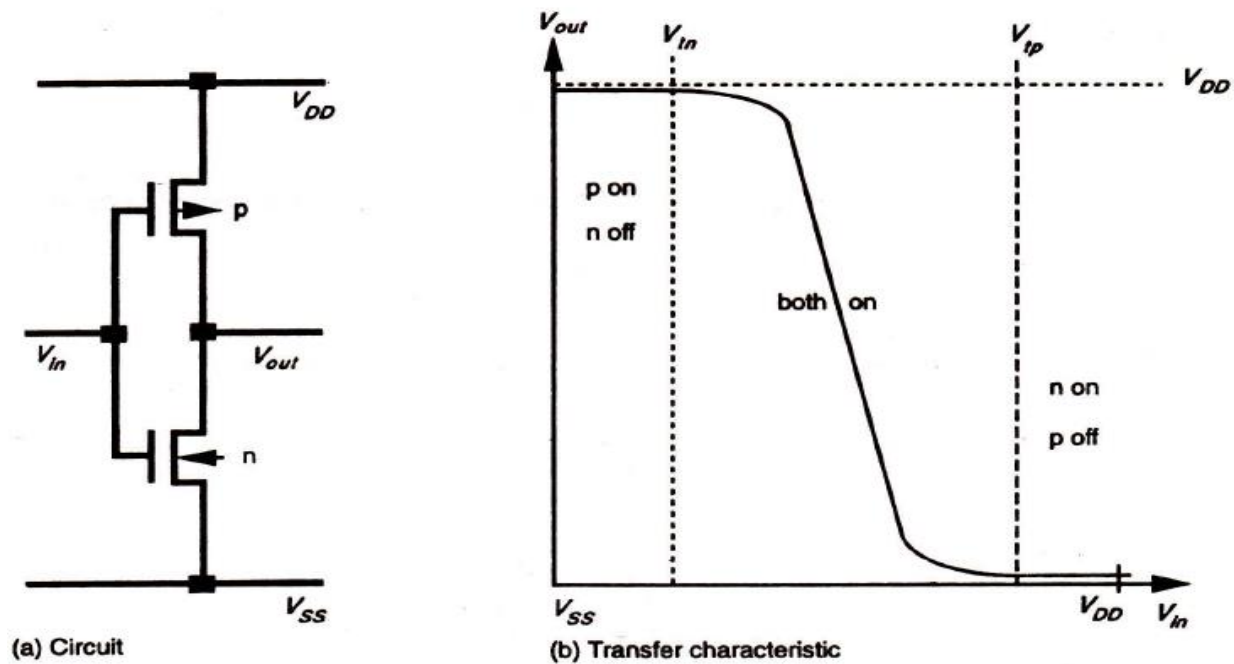


NMOS enhancement mode pull-up and transfer characteristic

The important features of this arrangement are

- (a) Dissipation is high since current flows when $V_{in} = \text{logical 1}$ (V_{GG} is returned to V_{DD}) .
- (b) V_{out} can never reach V_{DD} (logical 1) if $V_{GG} = V_{DD}$ as is normally the case.
- (c) V_{GG} may be derived from a switching source, for example, one phase of a clock, so that Dissipation can be greatly reduced.
- (d) If V_{GG} is higher than V_{DD} then an extra supply rail is required.

4. Complementary transistor pull-up (CMOS) : This arrangement consists of a C-MOS arrangement as pull-up. The arrangement and the transfer characteristic are shown below

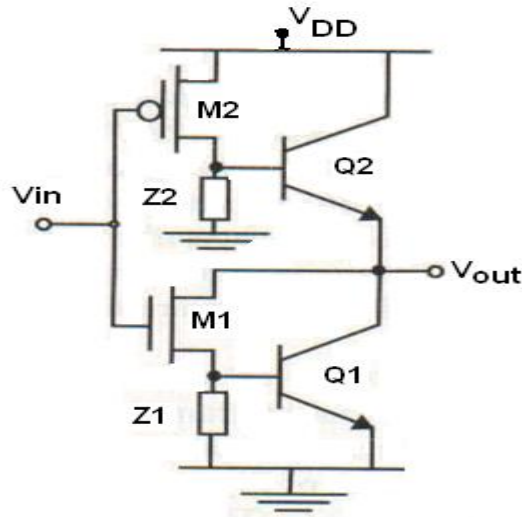


The salient features of this arrangement are

- (a) No current flows either for logical 0 or for logical 1 inputs.
- (b) Full logical 1 and 0 levels are presented at the output.
- (c) For devices of similar dimensions the p-channel is slower than the n-channel device.

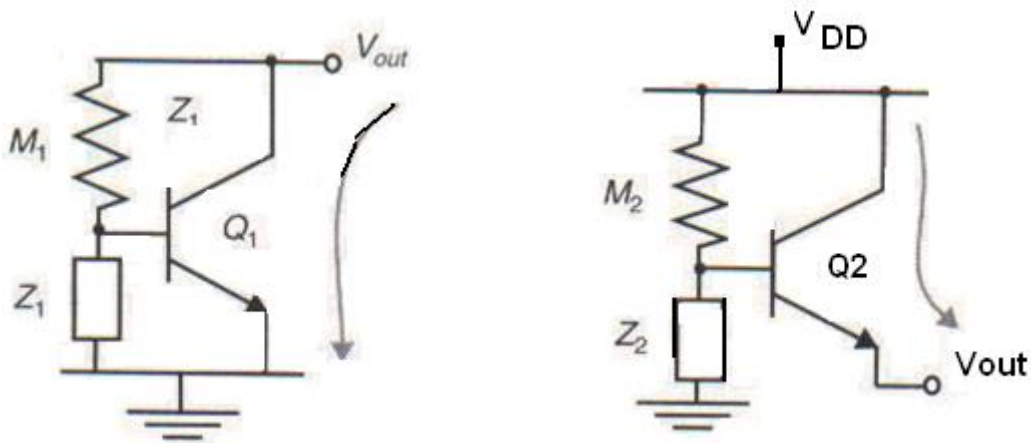
THE BiCMOS INVERTER :

A BiCMOS inverter, consists of a PMOS and NMOS transistor (M_2 and M_1), two NPN bipolar junction transistors, (Q_2 and Q_1), and two impedances which act as loads (Z_2 and Z_1) as shown in the circuit below.



When input, V_{in} , is high (V_{DD}), the NMOS transistor ($M1$), turns on, causing $Q1$ to conduct, while $M2$ and $Q2$ are off, as shown in figure (b) . Hence , a low (GND) voltage is translated to the output V_{out} . On the other hand, when the input is low, the $M2$ and $Q2$ turns on, while $m1$ and $Q1$ turns off, resulting to a high output level at the output as shown in Fig.(b).

In steady-state operation, $Q1$ and $Q2$ never turns on or off simultaneously, resulting to a lower power consumption. This leads to a push-pull bipolar output stage. Transistors $m1$ and $M2$, on the other hand, works as a phase-splitter, which results to a higher input impedance.



The impedances $Z2$ and $Z1$ are used to bias the base-emitter junction of the bipolar transistor and to ensure that base charge is removed when the transistors turn off. For example when the input voltage makes a high-to-low transition, $M1$ turns off first. To turn off $Q1$, the base charge must be removed, which can be achieved by $Z1$. With this effect, transition time reduces. However,

there exists a short time when both Q1 and Q2 are on, making a direct path from the supply (V_{DD}) to the ground. This results to a current spike that is large and has a detrimental effect on both the noise and power consumption, which makes the turning off of the bipolar transistor fast .

Comparison of BiCMOS and C-MOS technologies

The BiCMOS gates perform in the same manner as the CMOS inverter in terms of power consumption, because both gates display almost no static power consumption.

When comparing BiCMOS and CMOS in driving small capacitive loads, their performance are comparable, however, making BiCMOS consume more power than CMOS. On the other hand, driving larger capacitive loads makes BiCMOS in the advantage of consuming less power than CMOS, because the construction of CMOS inverter chains are needed to drive large capacitance loads, which is not needed in BiCMOS.

The BiCMOS inverter exhibits a substantial speed advantage over CMOS inverters, especially when driving large capacitive loads. This is due to the bipolar transistor's capability of effectively multiplying its current.

For very low capacitive loads, the CMOS gate is faster than its BiCMOS counterpart due to small values of C_{int} . This makes BiCMOS ineffective when it comes to the implementation of internal gates for logic structures such as alus, where associated load capacitances are small.

BiCMOS devices have speed degradation in the low supply voltage region and also BiCMOS is having greater manufacturing complexity than CMOS.